JMIR mHealth and uHealth

Mobile and tablet apps, ubiquitous and pervasive computing, wearable computing and domotics for health
Volume 9 (2021), Issue 10 ISSN: 2291-5222

Contents

Original Papers

Reviews

Corrigenda and Addenda

Original Paper

A Mobile Sensing App to Monitor Youth Mental Health: Observational Pilot Study

Lucy MacLeod[1*], BSc; Banuchitra Suruliraj[2*], MSc; Dominik Gall[3], MA; Kitti Bessenyei[1], MA; Sara Hamm[1], BSc; Isaac Romkey[1], BSc; Alexa Bagnell[1], MD; Manuel Mattheisen[1], MD; Viswanath Muthukumaraswamy[1], MSc; Rita Orji[2*], DPhil; Sandra Meier[1*], DPhil

[1]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada

[2]Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

[3]Department of Psychology, University of Würzburg, Würzburg, Germany

[*]these authors contributed equally

Corresponding Author:
Sandra Meier, DPhil
Department of Psychiatry
Dalhousie University
5850/5980 University Avenue, PO Box 9700
Halifax, NS, B3K 6R8
Canada
Phone: 1 782414 ext 8054
Email: sandra.m.meier@gmail.com

Abstract

Background: Internalizing disorders are the most common psychiatric problems observed among youth in Canada. Sadly, youth with internalizing disorders often avoid seeking clinical help and rarely receive adequate treatment. Current methods of assessing internalizing disorders usually rely on subjective symptom ratings, but internalizing symptoms are frequently underreported, which creates a barrier to the accurate assessment of these symptoms in youth. Therefore, novel assessment tools that use objective data need to be developed to meet the highest standards of reliability, feasibility, scalability, and affordability. Mobile sensing technologies, which unobtrusively record aspects of youth behaviors in their daily lives with the potential to make inferences about their mental health states, offer a possible method of addressing this assessment barrier.

Objective: This study aims to explore whether passively collected smartphone sensor data can be used to predict internalizing symptoms among youth in Canada.

Methods: In this study, the youth participants (N=122) completed self-report assessments of symptoms of anxiety, depression, and attention-deficit hyperactivity disorder. Next, the participants installed an app, which passively collected data about their mobility, screen time, sleep, and social interactions over 2 weeks. Then, we tested whether these passive sensor data could be used to predict internalizing symptoms among these youth participants.

Results: More severe depressive symptoms correlated with more time spent stationary (r=0.293; P=.003), less mobility (r=0.271; P=.006), higher light intensity during the night (r=0.227; P=.02), and fewer outgoing calls (r=−0.244; P=.03). In contrast, more severe anxiety symptoms correlated with less time spent stationary (r=−0.249; P=.01) and greater mobility (r=0.234; P=.02). In addition, youths with higher anxiety scores spent more time on the screen (r=0.203; P=.049). Finally, adding passively collected smartphone sensor data to the prediction models of internalizing symptoms significantly improved their fit.

Conclusions: Passively collected smartphone sensor data provide a useful way to monitor internalizing symptoms among youth. Although the results replicated findings from adult populations, to ensure clinical utility, they still need to be replicated in larger samples of youth. The work also highlights intervention opportunities via mobile technology to reduce the burden of internalizing symptoms early on.

(JMIR Mhealth Uhealth 2021;9(10):e20638) doi:10.2196/20638

KEYWORDS

mobile sensing; youth; psychiatry; feasibility; mobile phone

Introduction

Background

Internalizing disorders are the most common psychiatric problems among youth in Canada, with an estimated 6-month prevalence of 11% to 14% and more girls being affected than boys [1]. Over the last 3 decades, a steep increase in the prevalence of internalizing disorders has been observed; unfortunately, this trend is likely to continue [1]. Internalizing disorders typically have an early onset in childhood or adolescence and are characterized by high rates of relapse and chronicity, often resulting in substantial impairment across the lifespan. They are among the top 10 global causes of years lived with disability [2], and severely affect psychosocial development [1], family and social functioning, long-term health outcomes [3], and increased the risk of death by suicide [4]. Unfortunately, only 20% of youth with these disorders receive adequate mental health care [5]. If health professionals cannot predict and efficiently prevent new onset and poor outcomes of internalizing disorders in youth, then this field faces a crisis. Particularly, intervening at critical moments—that is—during mental health crises, including times of risk of suicide, self-harm, psychotic breakdown, substance use relapse, bullying, and interpersonal loss can have a major impact on improving youth mental health.

Current methods of predicting internalizing disorders usually rely on subjective symptom ratings obtained at discrete time points during routine clinical care (eg, clinical monitoring and screening). However, clinical decision-making based on such subjective information is challenging, as changes in symptoms might be subthreshold, context-dependent, or too subtle to be captured using subjective patient ratings. Symptom recall in patients might also not be accurate, potentially biased by current symptom severity [6] and environmental stressors [7]. Moreover, subjective self-reports often fail to reflect symptom variability and context reactivity. Therefore, novel prediction tools need to be developed to meet the highest standards of reliability, feasibility, scalability, and affordability. Given the large proportion of youths who experience mental health crises and do not receive treatment [8], these prediction tools must also encompass those who are not in traditional mental health care.

The recent proliferation of mobile sensing technologies offers health care professionals unprecedented access and insights into youth behavior and mood. Of profound clinical importance is the ability of these technologies to unobtrusively record aspects of youth behaviors in their daily lives with the potential to make inferences about their mental health states [9]. Crucially, these technologies make it possible to obtain detailed moment-by-moment information on youth behaviors and mood while requiring only minimal, if any, active involvement of the users [10]. Smartphones may be especially helpful prediction tools, not only because they afford a wide variety of types of behavioral data that can be automatically recorded by their built-in sensors, but also because they are integral to a youth's life and nearly 80% of youths own at least 1 smartphone. Thus, smartphones might allow long-term clinical monitoring of internalizing disorders at low cost and high scalability. Collecting high-quality, passive mobile sensing data might further enable the generation of predictive algorithms for solving previously intractable problems and identifying risk states before they manifest as full-blown internalizing disorders.

Previous Work

The utility of mobile sensing approaches to predict internalizing symptoms has been examined in multiple studies involving adult populations. Several mobile sensor features mirror internalizing symptoms in these studies [11-14]. In particular, GPS data have been shown to be of value in the prediction of internalizing symptoms. For example, the number of different locations visited was found to be negatively associated with social anxiety [11,12] and depression [15-17]. Study participants with social anxiety and depression were also found to avoid public areas and engage in less leisure activities, choosing instead to spend more time at home after school [11,12]. Studies exploring phone use data have reported the number of outgoing calls and messages to negatively correlate with internalizing symptoms [13,15,18]. In addition, ambient light data, an indicator of quality of sleep, were found to be associated with internalizing symptoms [19]. However, to date, mobile sensing studies in the context of youth have been limited [20]. We are only aware of a small pilot study exploring the utility of an Android mobile sensing app among 11 youth patients diagnosed with a major depressive disorder [21]. In that study, the daily number of steps, SMS frequency, and average call duration across a 2-week window were found to be highly correlated with the clinical symptoms of depression. However, in contrast to many studies among adults [17,22], clinical symptoms of depression correlated positively with entropy among youth [21]. This is in line with findings from Jongs et al [22], who also described strong age-related differences in entropy and other GPS-related measures of mobility in a sample aged 18-90 years. Similarly, phone use patterns have been shown to vary between youths and adults. Using a mobile sensing approach, Christensen et al [23] observed that younger age was significantly associated with more screen time. Importantly, screen time has been suggested to be differentially related to internalizing symptoms across age groups [24]. Thus, it is important to explore whether the mobile sensor features found to mirror internalizing symptoms in adults also predict such symptoms in youth.

Study Aims

Accordingly, this study aims to test the utility of passively collected smartphone sensor data (ie, incoming and outgoing calls, text messages, and GPS data), gathered over 14 days via a mobile sensing app, as the predictors of internalizing symptoms in youth. Studies have shown that internalizing symptoms are common in the general population [25], and the combination of patient and general population samples is meaningful for examining this pathology on a continuum. Therefore, this study aims to collect data from both clinical and nonclinical youth populations. Moreover, the study also uses a mobile sensing app that works on both iOS and Android operating systems, ensuring the generalizability of our results. By including iOS and Android users from clinical and nonclinical settings, we have advanced the inclusivity of current mobile sensing studies in youth [21]. In addition, we hypothesize that mobile sensing data will be associated with internalizing

symptoms among youths aged 10-21 years. As outlined, the only study using mobile sensing to predict internalizing symptoms in youth included only 11 participants, thus requiring further evaluation in a larger and more diverse sample. This is especially important as studies have suggested age-related differences in mobile sensing patterns. In contrast to previous studies, we also aim to correct for clinical comorbidities, thus enabling the identification of symptom-specific associations of mobile sensing data. Finally, given the novel nature of these data, we have some exploratory aims, namely, whether the data indicate any nuanced biomarkers that may be worth examining in future iterations of this work.

Methods

Overview

Youth participants from clinical and nonclinical populations were recruited for a 14-day observational study, in which a custom app was installed on their personal Android or iOS phone. Self-report measures of anxiety and depression were collected at the beginning of the study. Throughout the duration of the study, the smartphone app passively recorded multiple indices of the youths' daily life behaviors in an unobtrusive manner. The indices were selected based on findings that demonstrated their potential to make inferences about mental health states [26,27]. These indices included geolocation, sleep, phone use, calls, and messages. A set of features were designed and used to extract higher-level information from these data. Statistical analyses were performed to determine whether a significant relationship existed between participants' self-reported symptoms of anxiety or depression and these features. The Sciences Research Ethics Board of the Izaak Walton Killam Health Centre in Halifax, (Nova Scotia, Canada) approved this study, which also included the collection of location data.

Study Procedure

Youth were recruited via posters and social media from the outpatient psychiatry clinic at Izaak Walton Killam Health Centre and the local community in Halifax between February 2020 and July 2020. Youths interested in the study were referred over a link to the lab's screening page. The study inclusion criteria were as follows: subjects should (1) be between 10 and 21 years old, (2) reside in Canada, (3) be fluent in English, and (4) have an Android or iOS phone that would be in use as the primary mobile phone during the study period (14 days). There were no exclusion criteria for this study. Youth participants who met the inclusion criteria could read a description of the study, which included an informed consent guide. There was no deception nor were the study aims omitted for the participants. All participants provided written informed consent before participation. For youths aged 14 or younger, we required, in addition to their own assent, one of their parents to consent to their study participation. After providing consent, the participants were asked to complete a short web-based survey regarding their clinical symptomatology.

Next, the youth participants were sent instructions to install the app and provided with individualized log-in credentials for using it. Once installed, the study app guided the youths through a short setup, during which they were asked to provide the app with the necessary permissions to access their data, followed by a log-in. Immediately following the setup and log-in, the app began to continuously collect data in the background. No further actions or interactions with the study app were performed until the end of the study, and data were uploaded for 14 days. At this time, the youths received a notification, informing them that the study had ended and directed them to uninstall the app from their phone. A study ID was assigned to each participant, and documents connecting the ID and the participant's name, demographics, and clinical data were kept separate. The data on the phone were anonymous and only identifiable through the participants' study ID. To encourage youths to participate in our study, we offered them a monetary reward of Can $20.00 (US $15.82). The REDCap (Research Electronic Data Capture) platform [28] was used for web-based screening, consent, and clinical surveys. REDCap is a secured web-based app that allows researchers to collect, store, and manipulate research data.

Self-report Assessment

Youth participants completed two self-report measures of internalizing symptoms in digital form at the beginning of the 14-day study. Anxiety symptoms were assessed using the Screen for Child Anxiety-Related Emotional Disorders (SCARED) [29]. SCARED is a self-report instrument for children and their parents, including 41 items that screen for several types of 4 anxiety disorders, namely, generalized anxiety disorder, separation anxiety disorder, panic disorder, and social anxiety disorder [30]. Both the child and parent versions of SCARED in clinical samples demonstrated good internal consistency (α=.74-.93), test-retest reliability (intraclass correlation coefficients 0.70-0.90), and discriminative validity (both between anxiety and other disorders and within anxiety disorders) [29,30]. We administered a 33-item version of SCARED, excluding items related to separation anxiety disorders.

Depression symptoms were assessed using the Center for Epidemiological Studies Depression Scale for Children (CES-DC), which includes 20 items. CES-DC screens for behavioral and cognitive components of depression [31]. The sensitivity of the CES-DC is 85% and can be considered good [32]. In addition, the CES-DC showed results similar to the results of other highly validated instruments [33,34].

To be able to adjust for comorbid externalizing symptomatology, we also required the youths to complete the Adult Attention-Deficit Hyperactivity Disorder Self-Report Scale (ASRS) with adolescent prompts. On the basis of an 18-item scale, the ASRS assesses hyperactivity or impulsivity and inattentiveness symptoms. The ASRS demonstrated a good internal consistency (α=.80) and reliability (α=.79) in adolescent samples [35]. The SCARED, CES-DC, and ASRS instruments asked youth participants to evaluate their symptoms over the past 3 months.

Smartphone Data Collection

A mobile sensing app was designed and created to collect all the study data. The PROSIT (Predicting Risk and Outcomes of Social Interactions) tool captures multiple indices of a youth's

daily life behaviors via the naturalistic use of a smartphone (Figure 1). In this study, we used the following features of the PROSIT tool: first, the tool collected GPS data every time the youths moved within a 20-meter radius, preventing the identification of their home address or precise geographical location to preserve privacy. Second, the app logged call events (ie, time and date of the call, duration, and contact of both incoming and outgoing calls), and screen on or off events (ie, time and date). Finally, we sampled ambient light data as an indicator of sleep using the PROSIT tool. Calls and phone use were event-based sensor streams, whereas GPS and ambient light data were sampled as time series. Upon installation, the PROSIT tool continuously ran in the background and collected sensor and use data. The data were first stored locally and then automatically uploaded every hour to a remote server when the phone was in an idle state, and the Wi-Fi was connected. To protect user privacy, all collected sensitive personal data, such as contact details (ie, names and phone numbers), were anonymized during data collection by the app through the built-in cryptographic hash functions, thus guaranteeing that user identity was never recorded either locally or remotely. As the tool is not yet distributed through the official mobile app stores from Apple and Google, a TestFlight-based distribution for iOS and downloadable Android packages for Android was used in this study. The iOS version of the tool was built in the native Swift language using XCode 13 (Apple), whereas the Android version was built using Java and Kotlin languages using Android Studio 4.1 (Google). Importantly, the PROSIT tool also captured steps, notifications, installed apps, app use time, typed text, and music listened to. However, we decided not to include these measures in the current analyses, as these data were slightly differently sampled among Android and iOS users because of certain iOS-specific restrictions. More information on the PROSIT tool is presented in Multimedia Appendix 1 [13,36-50].

Figure 1. Sensors used to capture behavioral traits.

Participant

A total of 161 youth participants downloaded the PROSIT tool into their phone, 6.8% (11/161) withdrew from the study (only 3/161, 1.8% due to privacy concerns), 15.5% (25/161) provided less than 14 days of data, and 1.8% (3/161) did not provide data on sensors relevant to this study. Therefore, this study focused on 75.7% (122/161) of the initial youth participants.

Youth participants were, on average, 18 (SD 2.76) years old and 78.6% (96/122) were female. A total of 5.7% (7/122) participants identified themselves as indigenous, 5.7% (7/122) as black Canadian, 6.5% (8/122) as Asian, and the other 81.9% (100/122) were identified as non-Hispanic Caucasians. As a marker of the youths' socioeconomic status, we obtained information on maternal education: 4.1% (5/122) of the youths' mothers had not finished high school, 14.7% (18/122) had finished high school, 23.7% (29/122) had received further education, 54.1% (66/122) had attended university, and 3.2%

(4/122) cases lacked this information. A total of 78.6% (96/122) youths were iOS users and 21.3% (26/122) were Android users. In total, 24.5% (30/122) of the youth participants had a lifetime diagnosis of depression, 31.9% (39/122) had a lifetime diagnosis of generalized anxiety disorder, and 9.8% (12/122) had a lifetime diagnosis of social phobia. Regarding the self-report measures of internalizing symptoms, SCARED mean 33.02 (SD 14.79) and range 3-62, the CES-DC mean 32.59 (SD 15.23) and range 1-58. During the study period, 9.8% (12/122) youths were diagnosed with depression, 11.4% (14/122) were diagnosed with generalized anxiety, and 4.9% (6/122) were diagnosed with social phobia. Finally, we excluded two participants from the call analyses and one participant from the GPS analyses, as they were extreme outliers.

Feature Extraction

Raw unobtrusive mobile sensing data with respect to mobility, social interactions, phone use, and sleep were aggregated into daily summaries. As outlined above, we logged each GPS data point when the displacement from the last point was larger than 20 m. From these data, we extracted the number of locations visited per hour for at least 2 minutes. The distance between stationary states is defined by the Haversine distance function, which calculates the distance between a pair of coordinates denoted by latitude and longitude. We extracted the total distance traveled (times traveled a day), normalized entropy, and time spent at one location (ie, stationary at work or at home), 2-9 locations (ie, moving), and over 10 locations (ie, transportation) per hour for each participant on a daily basis. Normalized entropy was defined as

$$\text{Normalized Entropy} = -\sum_{i}^{N} p_i \log p_i(2) / \log N \quad \textbf{(1)}$$

where each i=1, 2,..., N represented a location cluster (in a 150-m radius), N denoted the total number of location clusters, and p_i was the percentage of time the participant spent at the location cluster i. High entropy indicated that the participant spent time more uniformly across different location clusters, whereas lower entropy indicated a greater inequality in the time spent across the clusters. Unlike entropy, the normalized entropy is invariant to the number of clusters; thus, it depends solely on the distribution of the visited location clusters. The value of normalized entropy ranges from 0 to 1, where 0 indicates that all location data points belong to the same cluster, and 1 implies that they are uniformly distributed across all clusters. For each phone call, we logged the timestamp, type (incoming, outgoing, and missing), and duration. For each user, we computed the total number of phone calls (connected, incoming, and outgoing) and total call duration on a daily basis. The raw mobile phone screen on or off events were transformed into two features: (1) the total number of times the screen was turned on per day and (2) the total amount of screen time per day (calculated as the difference between the number of screen on or off events). From the ambient light data, we logged changes in light intensity and calculated the time spent at each light intensity level. Next, we computed the total average light intensity, time spent in the lowest and highest light intensity ranges during nighttime per day (11 PM-7 AM), which was used as an indicator of sleep quality.

Statistical Analysis

After completing feature extraction and exclusion of extreme outliers, we performed the Spearman correlation analysis of sensor data over 14 days with SCARED and CES-DC scores of youth participants at a dimensional level. All analyses were adjusted for age, sex, and maternal education as the indicators of socioeconomic status, comorbidities, and the smartphone operating system. In addition, we included the month of assessment to account for potential impact of the pandemic on the results. We also fitted linear regressors by comparing models with and without passively collected smartphone sensor data as predictors. Finally, we assessed the discriminant validity of the predicted internalizing symptoms; for example, we ran correlations between the SCARED scores based on smartphone biomarkers and CES-DC and ASRS scores.

Results

Mobile Sensing Data Characteristics

The mean and SD of mobile sensing features across youth participants are presented in Table 1. To predict the assessment of internalizing symptoms (SCARED and CES-DC), we used averages across individuals as predictor variables.

Table 1. Descriptive statistics of mobile sensor data and features.

Feature	Values, mean (SD)
Mobility	
Time sedentary	18.82 (3.60)
Time moving	4.02 (3.56)
Distance traveled	6.01 (3.56)
Entropy	0.48 (0.20)
Social interaction	
Call duration	15.22 (13.63)
Incoming calls	1.81 (0.94)
Outgoing calls	2.11 (1.20)
Connected calls	2.30 (1.25)
Phone usage	
Screen use time	249.58 (109.28)
Screen use unlocks	53.66 (38.71)
Sleep	
Ambient light intensity	304,883 (95,828)
Clinical symptoms	
SCARED[a]	33.02 (14.79)
CES-DC[b]	32.59 (15.23)

[a]SCARED: Screen for Child Anxiety-Related Emotional Disorders.

[b]CES-DC: Center for Epidemiological Studies Depression Scale for Children.

Relationship Between Smartphone Data and Psychometric Scores

We used the pairwise Spearman coefficient as an indicator of the correlations between mobile sensing data and clinical self-report measures of internalizing symptoms (SCARED, CES-DC). The mobile sensing data were represented by mobility, social interaction, phone use, and sleep-related features extracted using methods described in Feature Extraction and Statistical Analysis of the Methods section. The analysis and correlations are shown in Tables 2 and 3.

Table 2 shows the correlations between mobility and internalizing symptom scores, where mobility is captured using GPS data. From the table, we observe that youths with more severe depression and higher CES-DC scores spent more time stationary (r=0.293; P=.003) and were significantly less mobile (r=−0.271; P=.006). In contrast, less time spent stationary (r=−0.249; P=.01) and higher mobility (r=0.234; P=.02) were associated with more anxiety symptoms as measured by SCARED. No correlations between normalized entropy values and internalizing symptoms were observed.

The correlations of social interactions, phone use, and sleep-related features with internalizing symptom scores are shown in Table 3. The results indicate that higher depression scores are significantly correlated with lower social interaction levels, such as fewer outgoing phone calls (r=−0.223; P=.03). Higher anxiety scores also correlated with fewer outgoing phone calls (r=−0.293; P=.003); however, this correlation became insignificant after adjusting for comorbid depression (r=−0.069; P=.50). In addition, higher ambient light intensity during the night, which is a proxy for poorer sleep quality, was correlated with higher depression scores (r=0.227; P=.02), but not anxiety scores (r=−0.15; P=.12). Conversely, higher smartphone screen use was only correlated with higher anxiety scores (r=0.203; P=.049).

Table 2. Spearman correlation coefficients between internalizing symptoms and mobility features.

Feature	Anxiety		Depression	
	Correlation coefficient (r)	P value	Correlation coefficient (r)	P value
Distance travelled	0.230	.02	−0.271	.006
Time moving	−0.248	.01	0.293	.003
Time sedentary	0.234	.02	−0.266	.007
Entropy	0.076	.46	−0.042	.69

Table 3. Spearman correlation coefficients between internalizing symptoms and social interactions, phone use, sleep-related features.

Feature	Anxiety		Depression	
	Correlation coefficient (*r*)	*P* value	Correlation coefficient (*r*)	*P* value
Call duration	0.024	.81	−0.035	.73
Incoming calls	0	.99	−0.072	.48
Outgoing calls	−0.069	.50	−0.223	.03
Connected calls	−0.106	.31	−0.110	.29
Ambient light intensity	−0.136	.12	0.227	.02
Screen use	0.203	.049	−0.113	.28

Thus, internalizing symptoms assessed by clinical instruments were significantly correlated with mobility levels and social interactions captured by passive smartphone sensor data. We also observed significant correlations between other daily living contexts (ie, ambient light intensity and smartphone screen use) and internalizing symptoms.

The inclusion of mobility, social interactions, phone use, and sleep-related features significantly improved the fit of models predicting more severe depression and higher CES-DC scores ($F_{17,62}$=2.036; *P*=.04), as well as more severe anxiety and higher SCARED scores ($F_{17,62=}$2.256; *P*=.02). Finally, predicted SCARED scores from smartphone biomarkers correlated with higher depression scores (*r*=0.258; *P*=.01), but not ASRS scores (*r*=0.062; *P*=.55). Consistently, predicted CES-DC scores correlated with higher anxiety scores (*r*=0.258; *P*=.01), but not ASRS scores (*r*=0.020; *P*=.85).

Discussion

Principal Findings

Our results support the hypothesis that passively collected smartphone sensor data can provide sufficient information for successfully predicting internalizing symptoms among youth. The proposed approach worked well for youths who were usually heavy smartphone users. Remarkably, we were able to include youths as young as 10 years of age in our completely remote study design, demonstrating the feasibility of mobile sensing among youth. The data also underlined high acceptability; even youths with severe clinical symptomology reported only minimal issues in the use of the app. Dropouts due to privacy concerns can be considered low (3/161, 1.9%).

The study indicated that there were significant correlations between internalizing symptoms and smartphone biomarkers. Lower levels of mobility, fewer social interactions, and higher ambient light intensity during the night, used as a proxy for poorer sleep quality, were correlated with higher depression symptoms. Higher levels of mobility, fewer social interactions, and higher smartphone screen use were correlated with higher anxiety symptoms among youth. The study also showed that by adding passively collected smartphone sensor data to prediction models, we were able to achieve better model fits. Finally, the results indicate the potential of smartphone biomarkers to have a discriminative nature by predicting other internalizing symptoms but not adult attention-deficit hyperactivity disorder symptoms.

Here, we expand on the only previous study among youth showing that light intensity and smartphone screen use in addition to mobility and social interactions might be valuable features for predicting internalizing symptoms among youth [21]. In addition, we confirmed the potential discriminative validity of smartphone biomarkers observed among adults in a youth sample [51]. Although our results on the number of outgoing calls negatively correlating with internalizing symptoms are in line with previous studies in adults [13,15,18], differences in mobility patterns of youths with high anxiety symptoms as compared with high depression symptoms have not been described thus far [11,14,21]. However, these findings have to be interpreted considering the recent pandemic and the behavioral changes accompanying it [52]. It is easy to argue that the time spent outside during a pandemic might increase anxiety symptoms.

Limitations and Future Directions

This study has some limitations that provide directions for future research. First, the conclusions must be considered preliminary, given the relatively small sample size of 122 youths and require further replication. Larger samples will also allow for more power to assess potentially moderating effects (eg, sex, trauma exposure), especially as the cost-effectiveness of mobile sensing by using the patients' own smartphones is fostering the scalability of the approach. However, in contrast to many previous studies [11], we were able to work with a relatively heterogeneous real-world sample of youths, including patients with severe symptomatology, which may reduce the generalizability of our findings. Second, through the use of passively collected smartphone sensor data, we may miss relevant information on social interactions; for example, it can be difficult to infer the quality of the social contact based on the metadata of call logs. Third, our study focused on identifying between-person differences; the identified predictors might not be of value in predicting within-person variability of internalizing symptoms. Fourth, it is not possible to make any claims regarding causation between mobility, screen time, sleep, social interactions, and internalizing symptoms. It cannot be determined if fewer social interactions lead to more internalizing symptoms or vice versa. Fifth, our study was conducted during the recent pandemic that led to a multitude of behavioral changes [52,53]; thus, the results might not be generalizable to a postpandemic situation. However, our study findings were largely consistent with those of previous studies conducted in

adult populations, supporting the reliability of mobile sensing results.

The next steps should focus on exploring how youth patients, their parents, and clinicians would prefer to be notified about the youths' passively collected smartphone sensor data, what inferences can be drawn about youths' mental health states from these data, and what actions they would find useful and acceptable in response to these data. For example, early warning signs can be identified from smartphone biomarkers, triggering early interventions. In addition, the enthusiasm for exciting new possibilities associated with passively collected smartphone sensor data must be carefully evaluated regarding the challenges of privacy and data security.

Conclusions

In sum, our study expands recent efforts to use passively collected smartphone sensor data for predicting internalizing symptoms among youth. Such accurate behavioral assessments might be especially important for youths with internalizing symptoms, given that the occurrence of these symptoms is more frequent than is typically reported. However, this method of harnessing naturally occurring behavioral data is certainly relevant for identifying and better understanding a range of maladaptive thoughts and behaviors that underlie psychiatric conditions more broadly. Finally, it is equally important to be sensitive and treat passively collected smartphone sensor data as sensitive health information and ensure adequate privacy and security controls are in place before wider use.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The PROSIT (Predicting Risk and Outcomes of Social Interactions) app.
[DOCX File , 1115 KB - mhealth_v9i10e20638_app1.docx]

References

1. Melkevik O, Hauge LJ, Bendtsen P, Reneflot A, Mykletun A, Aarø LE. Associations between delayed completion of high school and educational attainment and symptom levels of anxiety and depression in adulthood. BMC Psychiatry 2016 Mar 15;16:64 [FREE Full text] [doi: 10.1186/s12888-016-0765-1] [Medline: 26980092]
2. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015 Apr;72(4):334-341 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2502] [Medline: 25671328]
3. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
4. Meier SM, Mattheisen M, Mors O, Mortensen PB, Laursen TM, Penninx BW. Increased mortality among people with anxiety disorders: total population study. Br J Psychiatry 2016 Sep;209(3):216-221 [FREE Full text] [doi: 10.1192/bjp.bp.115.171975] [Medline: 27388572]
5. Hintzpeter B, Klasen F, Schön G, Voss C, Hölling H, Ravens-Sieberer U, BELLA study group. Mental health care use among children and adolescents in Germany: results of the longitudinal BELLA study. Eur Child Adolesc Psychiatry 2015 Jun;24(6):705-713. [doi: 10.1007/s00787-015-0676-6] [Medline: 25651821]
6. Wells JE, Horwood LJ. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychol Med 2004 Aug;34(6):1001-1011. [doi: 10.1017/s0033291703001843] [Medline: 15554571]
7. Goldfarb EV, Mendelevich Y, Phelps EA. Acute stress time-dependently modulates multiple memory systems. J Cogn Neurosci 2017 Nov;29(11):1877-1894. [doi: 10.1162/jocn_a_01167] [Medline: 28699809]
8. 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 Feb;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
9. 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]
10. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
11. Boukhechba M, Chow P, Fua K, Teachman BA, Barnes LE. Predicting social anxiety from global positioning system traces of college students: feasibility study. JMIR Ment Health 2018 Jul 04;5(3):e10101 [FREE Full text] [doi: 10.2196/10101] [Medline: 29973337]
12. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]

13. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]
14. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
15. Sarda A, Munuswamy S, Sarda S, Subramanian V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: cross-sectional observational study. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11041 [FREE Full text] [doi: 10.2196/11041] [Medline: 30694197]
16. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan;36(1):72-81 [FREE Full text] [doi: 10.1002/da.22822] [Medline: 30129691]
17. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
18. Wang Y, Ren X, Liu X, Zhu T. Examining the correlation between depression and social behavior on smartphones through usage metadata: empirical study. JMIR Mhealth Uhealth 2021 Jan 06;9(1):e19046 [FREE Full text] [doi: 10.2196/19046] [Medline: 33404512]
19. Narziev N, Goh H, Toshnazarov K, Lee SA, Chung K, Noh Y. STDD: Short-Term Depression Detection with passive sensing. Sensors (Basel) 2020 Mar 04;20(5):1396 [FREE Full text] [doi: 10.3390/s20051396] [Medline: 32143358]
20. Melbye S, Kessing LV, Bardram JE, Faurholt-Jepsen M. Smartphone-based self-monitoring, treatment, and automatically generated data in children, adolescents, and young adults with psychiatric disorders: systematic review. JMIR Ment Health 2020 Oct 29;7(10):e17453 [FREE Full text] [doi: 10.2196/17453] [Medline: 33118950]
21. 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]
22. Jongs N, Jagesar R, van Haren NE, Penninx BW, Reus L, Visser PJ, et al. A framework for assessing neuropsychiatric phenotypes by using smartphone-based location data. Transl Psychiatry 2020 Jul 01;10(1):211 [FREE Full text] [doi: 10.1038/s41398-020-00893-4] [Medline: 32612118]
23. Christensen MA, Bettencourt L, Kaye L, Moturu ST, Nguyen KT, Olgin JE, et al. Direct measurements of smartphone screen-time: relationships with demographics and sleep. PLoS One 2016 Nov 9;11(11):e0165331 [FREE Full text] [doi: 10.1371/journal.pone.0165331] [Medline: 27829040]
24. Sharifian N, Kraal AZ, Zaheed AB, Sol K, Morris EP, Zahodne LB. Measurement invariance of social media use in younger and older adults and links to socioemotional health. Innov Aging 2021 Mar 11;5(2):igab009 [FREE Full text] [doi: 10.1093/geroni/igab009] [Medline: 33928196]
25. Merikangas KR. Vulnerability factors for anxiety disorders in children and adolescents. Child Adolesc Psychiatr Clin N Am 2005 Oct;14(4):649-79, vii. [doi: 10.1016/j.chc.2005.06.005] [Medline: 16171697]
26. 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]
27. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
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 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
29. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry 1997 Apr;36(4):545-553. [doi: 10.1097/00004583-199704000-00018] [Medline: 9100430]
30. Monga S, Birmaher B, Chiappetta L, Brent D, Kaufman J, Bridge J, et al. Screen for Child Anxiety-Related Emotional Disorders (SCARED): convergent and divergent validity. Depress Anxiety 2000;12(2):85-91. [doi: 10.1002/1520-6394(2000)12:2<85::AID-DA4>3.0.CO;2-2] [Medline: 11091931]
31. Faulstich ME, Carey MP, Ruggiero L, Enyart P, Gresham F. Assessment of depression in childhood and adolescence: an evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC). Am J Psychiatry 1986 Aug;143(8):1024-1027. [doi: 10.1176/ajp.143.8.1024] [Medline: 3728717]
32. Garrison CZ, Addy CL, Jackson KL, McKeown RE, Waller JL. The CES-D as a screen for depression and other psychiatric disorders in adolescents. J Am Acad Child Adolesc Psychiatry 1991 Jul;30(4):636-641. [doi: 10.1097/00004583-199107000-00017] [Medline: 1890099]
33. Aebi M, Metzke CW, Steinhausen H. Prediction of major affective disorders in adolescents by self-report measures. J Affect Disord 2009 May;115(1-2):140-149. [doi: 10.1016/j.jad.2008.09.017] [Medline: 18947881]

34. Dierker LC, Albano AM, Clarke GN, Heimberg RG, Kendall PC, Merikangas KR, et al. Screening for anxiety and depression in early adolescence. J Am Acad Child Adolesc Psychiatry 2001 Aug;40(8):929-936. [doi: 10.1097/00004583-200108000-00015] [Medline: 11501693]
35. Green JG, DeYoung G, Wogan ME, Wolf EJ, Lane KL, Adler LA. Evidence for the reliability and preliminary validity of the Adult ADHD Self-Report Scale v1.1 (ASRS v1.1) Screener in an adolescent community sample. Int J Methods Psychiatr Res 2019 Mar;28(1):e1751 [FREE Full text] [doi: 10.1002/mpr.1751] [Medline: 30407687]
36. Paluska SA, Schwenk TL. Physical activity and mental health: current concepts. Sports Med 2000 Mar;29(3):167-180. [doi: 10.2165/00007256-200029030-00003] [Medline: 10739267]
37. Gomes da Silva S, Arida RM. Physical activity and brain development. Expert Rev Neurother 2015;15(9):1041-1051. [doi: 10.1586/14737175.2015.1077115] [Medline: 26289488]
38. Lederbogen F, Kirsch P, Haddad L, Streit F, Tost H, Schuch P, et al. City living and urban upbringing affect neural social stress processing in humans. Nature 2011 Jun 22;474(7352):498-501. [doi: 10.1038/nature10190] [Medline: 21697947]
39. Tost H, Reichert M, Braun U, Reinhard I, Peters R, Lautenbach S, et al. Neural correlates of individual differences in affective benefit of real-life urban green space exposure. Nat Neurosci 2019 Sep;22(9):1389-1393. [doi: 10.1038/s41593-019-0451-y] [Medline: 31358990]
40. Tarokh L, Saletin JM, Carskadon MA. Sleep in adolescence: physiology, cognition and mental health. Neurosci Biobehav Rev 2016 Nov;70:182-188 [FREE Full text] [doi: 10.1016/j.neubiorev.2016.08.008] [Medline: 27531236]
41. Walker MP, Harvey AG. Obligate symbiosis: sleep and affect. Sleep Med Rev 2010 Aug;14(4):215-217. [doi: 10.1016/j.smrv.2010.02.003] [Medline: 20427211]
42. Wehr TA. Sleep-loss as a possible mediator of diverse causes of mania. Br J Psychiatry 1991 Oct;159:576-578. [doi: 10.1192/bjp.159.4.576] [Medline: 1751874]
43. Cheng P, Kalmbach DA, Tallent G, Joseph CL, Espie CA, Drake CL. Depression prevention via digital cognitive behavioral therapy for insomnia: a randomized controlled trial. Sleep 2019 Oct 09;42(10):zsz150. [doi: 10.1093/sleep/zsz150] [Medline: 31535688]
44. Lemola S, Perkinson-Gloor N, Brand S, Dewald-Kaufmann JF, Grob A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. J Youth Adolesc 2015 Feb;44(2):405-418. [doi: 10.1007/s10964-014-0176-x] [Medline: 25204836]
45. Woods HC, Scott H. #Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. J Adolesc 2016 Aug;51:41-49. [doi: 10.1016/j.adolescence.2016.05.008] [Medline: 27294324]
46. Godsell S, White J. Adolescent perceptions of sleep and influences on sleep behaviour: a qualitative study. J Adolesc 2019 Jun;73:18-25. [doi: 10.1016/j.adolescence.2019.03.010] [Medline: 30953841]
47. Touitou Y, Touitou D, Reinberg A. Disruption of adolescents' circadian clock: the vicious circle of media use, exposure to light at night, sleep loss and risk behaviors. J Physiol Paris 2016 Nov;110(4 Pt B):467-479. [doi: 10.1016/j.jphysparis.2017.05.001] [Medline: 28487255]
48. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
49. Noel VA, Acquilano SC, Carpenter-Song E, Drake RE. Use of mobile and computer devices to support recovery in people with serious mental illness: survey study. JMIR Ment Health 2019 Feb 20;6(2):e12255 [FREE Full text] [doi: 10.2196/12255] [Medline: 30785401]
50. Low LS, Maddage NC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans Biomed Eng 2011 Mar;58(3):574-586 [FREE Full text] [doi: 10.1109/TBME.2010.2091640] [Medline: 21075715]
51. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875 [FREE Full text] [doi: 10.2196/16875] [Medline: 32348284]
52. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using smartphones and wearable devices to monitor behavioral changes during COVID-19. J Med Internet Res 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
53. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]

Abbreviations

ASRS: Adult Attention-Deficit Hyperactivity Disorder Self-Report Scale
CES-DC: Center for Epidemiological Studies Depression Scale for Children
PROSIT: Predicting Risk and Outcomes of Social Interactions
REDCap: Research Electronic Data Capture

SCARED: Screen for Child Anxiety-Related Emotional Disorders

Edited by L Buis; submitted 27.05.20; peer-reviewed by C Ralph-Nearman, A Sesagiri Raamkumar, D Hartigan, SK Mukhiya, L Castro, A Adams; comments to author 10.09.20; revised version received 02.02.21; accepted 27.07.21; published 26.10.21.

Please cite as:
MacLeod L, Suruliraj B, Gall D, Bessenyei K, Hamm S, Romkey I, Bagnell A, Mattheisen M, Muthukumaraswamy V, Orji R, Meier S
A Mobile Sensing App to Monitor Youth Mental Health: Observational Pilot Study
JMIR Mhealth Uhealth 2021;9(10):e20638
URL: https://mhealth.jmir.org/2021/10/e20638
doi:10.2196/20638
PMID:34698650

Review

Effectiveness of Telehealth Interventions for Women With Postpartum Depression: Systematic Review and Meta-analysis

Liuhong Zhao[1*], BSN; Jingfen Chen[1*], BSN; Liuying Lan[1], BSN; Ni Deng[2], BSN; Yan Liao[3], BEng; Liqun Yue[4*], BSN; Innie Chen[3,5,6], MD; Shi Wu Wen[3,5,6], PhD; Ri-hua Xie[1], PhD, FAAN

[1]Department of Nursing, The Seventh Affiliated Hospital, Southern Medical University, Foshan, China

[2]Department of Obstetrics and Gynecology, The Seventh Affiliated Hospital, Southern Medical University, Foshan, China

[3]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

[4]Department of Nursing, The Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

[5]Department of Obstetrics & Gynecology, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[6]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Ri-hua Xie, PhD, FAAN
Department of Nursing
The Seventh Affiliated Hospital
Southern Medical University
28 Liguan Road, Lishui
Foshan, 528244
China
Phone: 86 189 2869 7126
Email: xierihua928@hotmail.com

Abstract

Background: Postpartum depression (PPD) is a prevalent mental health problem with serious adverse consequences for affected women and their infants. Clinical trials have found that telehealth interventions for women with PPD result in increased accessibility and improved treatment effectiveness. However, no comprehensive synthesis of evidence from clinical trials by systematic review has been conducted.

Objective: The aim of this study is to evaluate the effectiveness of telehealth interventions in reducing depressive symptoms and anxiety in women with PPD. To enhance the homogeneity and interpretability of the findings, this systematic review focuses on PPD measured by the Edinburgh Postnatal Depression Scale (EPDS).

Methods: PubMed, The Cochrane Library, CINAHL, PsycINFO, CNKI, and Wanfang were electronically searched to identify randomized controlled trials (RCTs) on the effectiveness of telehealth interventions for women with PPD from inception to February 28, 2021. Data extraction and quality assessment were performed independently by two researchers. The quality of included studies was assessed using the Cochrane risk-of-bias tool, and meta-analysis was performed using RevMan 5.4 software.

Results: Following the search, 9 RCTs with a total of 1958 women with PPD were included. The EPDS (mean difference=–2.99, 95% CI –4.52 to –1.46; *P*<.001) and anxiety (standardized mean difference=–0.39, 95% CI –0.67 to –0.12; *P*=.005) scores were significantly lower in the telehealth group compared with the control group. Significant subgroup differences were found in depressive symptoms according to the severity of PPD, telehealth technology, specific therapy, and follow-up time (*P*<.001).

Conclusions: Telehealth interventions could effectively reduce the symptoms of depression and anxiety in women with PPD. However, better designed and more rigorous large-scale RCTs targeting specific therapies are needed to further explore the potential of telehealth interventions for PPD.

Trial Registration: PROSPERO CRD42021258541; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=258541

(JMIR Mhealth Uhealth 2021;9(10):e32544) doi:10.2196/32544

KEYWORDS

telehealth; postpartum depression; anxiety; meta-analysis

Introduction

Postpartum depression (PPD) is one of the most common mental health disorders in women after giving birth. A systematic review comprised of 58 articles with a total of 37,294 women reported that the overall prevalence of PPD was 17% among healthy mothers [1]. PPD symptoms in women manifest as sleep disorders, mood swings, sadness and crying, loss of appetite, a lack of interest in daily activities, or even more serious adverse outcomes such as suicide and infanticide [2]. PPD may also be associated with an increased risk of cognitive and behavioral problems in infants [3]. Timely access to effective interventions such as psychotherapy and pharmacotherapy is important for women affected by PPD [4,5]. However, breastfeeding mothers may have concerns regarding their infant's exposure to medications because of reported side effects of antidepressant exposure in infants (eg, excessive crying, colic, irritability, sedation, poor feeding, and sleep problems) [6]. Therefore, antidepressant medications are recommended only for women with severe depression, while psychotherapy is the first-line method for prevention and treatment of mild to moderate PPD [7]. The clinical effectiveness of common psychotherapies for PPD such as peer support therapy [8], interpersonal therapy [9], cognitive behavioral therapy [10], and mindfulness therapy [11] has been demonstrated. However, psychotherapy conducted in a traditional face-to-face manner may not be accessible for many women due to time and financial hurdles, childcare concerns, and fear of social stigma [12].

Women who have challenges accessing face-to-face psychotherapy may benefit from telehealth interventions, through which health care and health education could be provided to them at home [13]. Telehealth technologies include telephones, websites, videoconferences, and apps that allow consultation, assessment, and intervention services to be provided remotely by health professionals or peer support [14], which have been widely used to help manage diseases in various domains, including diabetes self-management [15], pulmonary rehabilitation [16], and palliative home care [17]. In addition, telehealth has been gaining momentum in continuous obstetrical care [18]. Telehealth facilitates interactions and communication between specialists and patients in the field of maternal-fetal medicine, especially in the postpartum period for breastfeeding and lactation assistance in rural communities [19]. Telehealth care has many benefits, including increased access and convenience and decreased social stigma and costs [20].

A previous study [21] showed that there are 18 scales for screening depression symptoms, including the Patient Health Questionnaire-9 Item [22], the Beck Depression Inventory II scale [23], the Postpartum Depression Screening Scale [24], and the Edinburgh Postnatal Depression Scale (EPDS) [25,26]. These scales have different sensitivity, specificity, and disease predictivity. Of them, EPDS is the most reliable scale in terms of disease predictivity and adaptivity for differences in population profiles, and it has therefore been the most frequently used scale in clinical and research settings to screen for PPD [25,26]. To enhance the homogeneity and interpretability of the findings, this systematic review targets adult women with PPD measured by EPDS [25,26].

Methods

Overview

This systematic review was conducted following the guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [27]. The study protocol was registered in PROSPERO (the International Prospective Register of Systematic Reviews) as CRD42021258541.

Search Strategy

For this study, 4 English databases (PubMed, the Cochrane Library, CINAHL, and PsycINFO) and 2 Chinese databases (CNKI and Wanfang) were electronically searched to identify randomized controlled trials (RCTs) on the effectiveness of telehealth interventions for women with PPD from inception to February 28, 2021. Advanced searches were performed using a combination of two groups of terms according to the syntax rules of each database: (1) telehealth-related terms, including telehealth, telemedicine, telecommunication, telephone, remote consultation, information technology, mobile health, m-Health, e-Health, internet, web-based, social media, application, and software, and (2) PPD-related terms, including postpartum depression, PPD, post-partum depression, postnatal depression, post-natal depression, maternal depression, postpartum mental disorder, and puerperal disorder. In addition, ClinicalTrials.gov, the World Health Organization International Clinical Trials Registry Platform, and the Chinese Clinical Trial Register were searched for unpublished trials relevant to this review.

Inclusion and Exclusion Criteria

To be eligible, RCTs had to meet the following criteria: (1) target adult women with EPDS scores ≥9 points; (2) use telehealth interventions including mobile phones, apps, websites, or other remote technologies compared with routine care in the control group (participants in the control group were inaccessible to any telehealth technologies, but were free to receive routine care including any offline treatment at public hospitals or maternal and child health care centers); (3) assess the primary outcome of depression symptoms using EPDS, and secondary outcomes including any improvement of social support, loneliness, and anxiety using any scale; and (4) be published in English or in Chinese.

Studies were excluded for the following reasons: (1) study included women with severe physical illnesses, a history of mental illnesses and medication treatments, drug and/or alcohol abuse, or suicidal tendency; (2) study included women whose infants had adverse neonatal outcomes such as 5-minute Apgar score <5 points, assisted ventilation for more than 6 hours, neonatal seizure, birth injury, or neonatal death; and (3) study was an RCT protocol or duplicate.

Study Selection

The software Endnote X9 (Clarivate Analytics) was used to import all the references and remove duplicates. The remaining studies were assessed against the inclusion and exclusion criteria by two independent reviewers (LZ and JC). Study selection was conducted in a stepwise manner. First, titles and abstracts of all studies were independently screened for potential eligibility.

Any disagreements were discussed until consensus was reached. Second, the full papers of all included abstracts were independently assessed. Any discrepancies that arose during the assessment were resolved by a third reviewer (RHX).

Quality Assessment

The risk of bias was assessed according to the guidelines provided by the Cochrane risk-of-bias tool for randomized trials (version 2.0) [28]. Risk ratings of "low risk," "unclear," and "high risk" were assigned to each type of bias based on the presence of selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias. Any disagreements with respect to study quality were resolved by a third reviewer (RHX).

Data Extraction

The data were extracted independently by two reviewers (LZ and JC) at the same time and disagreements were resolved by consensus. Data extracted from each relevant trial included author, the year of publication, country, participant, inclusion criterion (EPDS scores), sample size, telehealth technology, specific therapy, follow-up time, and outcomes (primary and secondary outcomes) with measure scales. Any disagreements between the two reviewers were resolved by a third reviewer (RHX).

Data Synthesis and Analysis

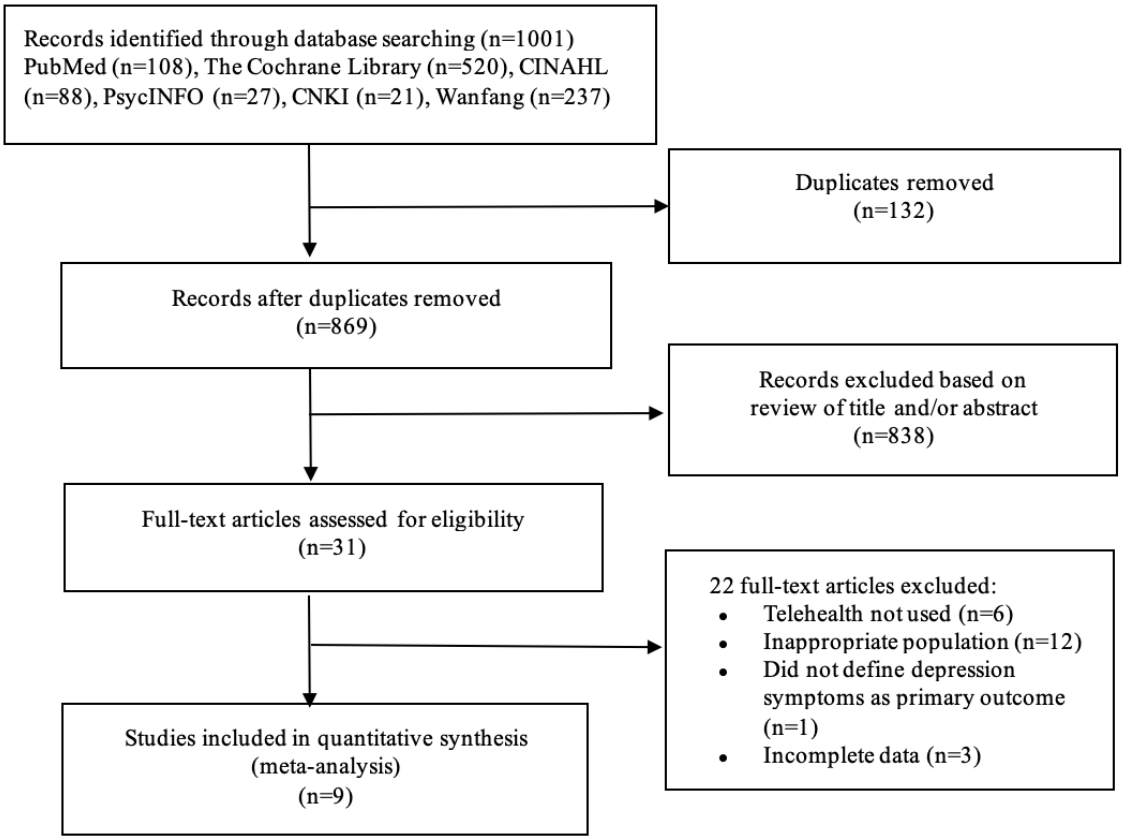
RevMan 5.4 software (The Cochrane Collaboration) was applied in the meta-analysis of the data. The effect estimate was expressed as means and standard deviations (SD) for continuous data. The standardized mean differences (SMD) with their corresponding 95% CI were applied to combine studies that measured the same outcome with different scales. If the same scales were used to evaluate one outcome, mean difference (MD) with its 95% CI could be employed. MD was derived from inverse variance methods. If the SD was not reported, it was computed from standard error (SE) following the Cochrane Handbook [28]: . Heterogeneity was assessed by I^2, with significant statistical heterogeneity as I^2>50%. The fixed-effect model was used to estimate one true effect in cases without significant heterogeneity (I^2<50%), whereas the random-effect model was employed to estimate the effects in cases with significant heterogeneity between studies (I^2>50%). When there was significant heterogeneity, data would be double-checked and then subgroup analysis or sensitivity analysis was performed to explore the sources of heterogeneity. A P value <.05 was considered statistically significant.

Results

Search Results

The PRISMA flow diagram for this study is shown in Figure 1. The search strategies yielded 1001 potentially relevant citations from the 6 electronic databases searched. After excluding duplications and screening titles and abstracts for eligibility, 31 studies were retained for full-text evaluation. Of them, 22 studies were excluded according to the inclusion and exclusion criteria. A total of 9 studies [29-37] were included in this systematic review.

Figure 1. Flow diagram of study selection.



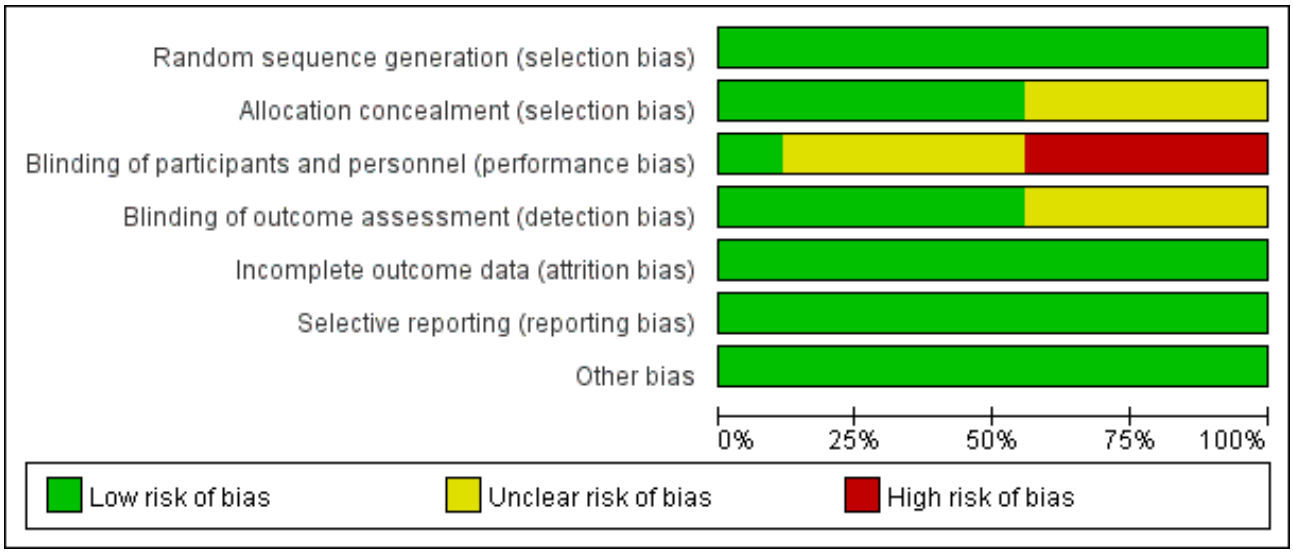
Characteristics of Included Studies

This systematic review included 9 RCTs [29-37] with a total of 1958 participants. The 9 studies included were conducted in Singapore, Canada, Portugal, Iran, China, and the United Kingdom and were published between 2003 and 2020. Participants were postpartum women aged ≥18 years. All recruited participants had baseline EPDS scores ≥9 points. The sample sizes ranged from 42 to 910. The 9 included studies applied telehealth technologies including telephones [29-31,34,36,37], apps [33,37], and websites [32,35,36]. Of them, 3 studies [29,30,37] used peer support therapy, 5 studies [32-36] used cognitive behavioral therapy and behavioral activation therapy, and 1 study [31] used interpersonal therapy. The follow-up period of these 9 studies [29-37] ranged from 4 weeks to 36 weeks after completion of telehealth interventions. Characteristics of the 9 studies included are presented in Multimedia Appendix 1.

Risk of Bias

We summarized the findings for risk of bias in Figure 2 and Multimedia Appendix 2. Among the 9 RCTs [29-37], 5 studies [30-33,37] used a computer-generated random sequence, 1 study [34] used a random number table, 2 studies [35,36] used a minimization algorithm including a stochastic element, and 1 study [29] only mentioned the word “random,” with no details of the randomization method used. Allocation concealment was done in 5 studies [29,34-37]. Furthermore, 1 study [37] implemented the blinding of participants and personnel and outcome assessment, 4 studies [32,33,35,36] were unclear about the blinding of outcome assessment, and 4 studies [30,31,33,34] had high risk of bias on implementing the blinding of participants and personnel. All 9 studies included had low risk of bias on incomplete outcome data, selective reporting, or other bias.

Figure 2. Risk of bias graph.

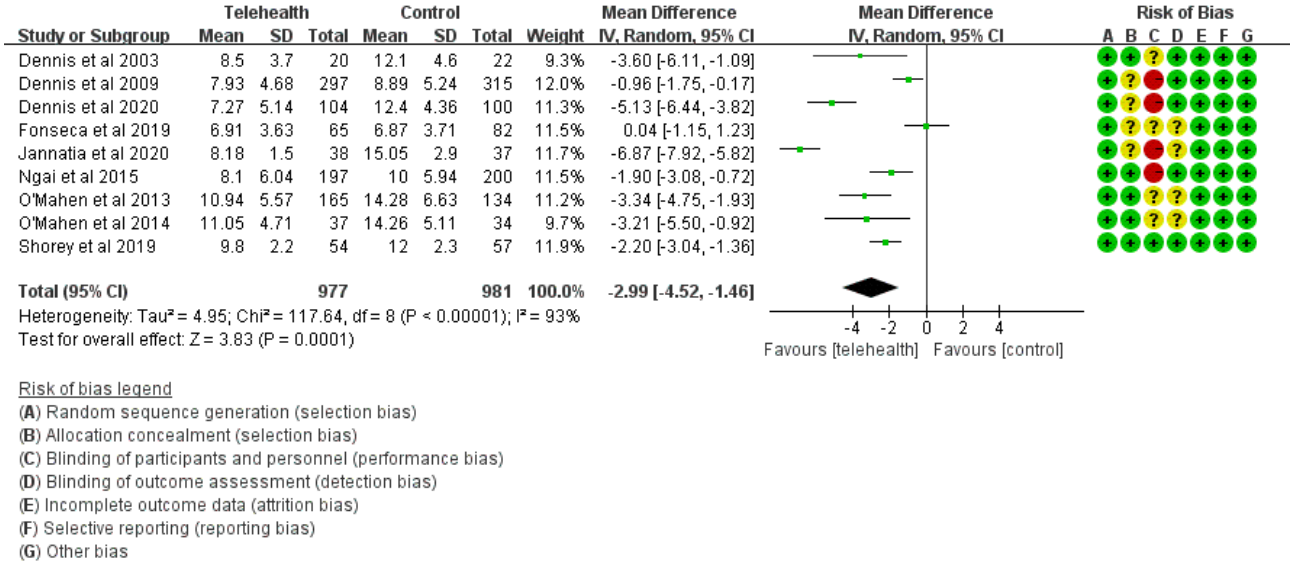


Meta-Analysis

Depressive Symptoms in the Telehealth Group and the Control Group

The 9 included studies [29-37] assessed the effectiveness of telehealth interventions on depressive symptoms among 1958 women by comparing the telehealth group and the control group. Depressive symptoms were all measured using EPDS. Considerable heterogeneity among studies was detected (I^2>50%); thus, a randomized model was used. The overall pooled analysis demonstrated that total EPDS scores of women with PPD who received telehealth interventions were significantly lower than the control group (MD=–2.99, 95% CI –4.52 to –1.46; P<.001; I^2=93%; Figure 3).

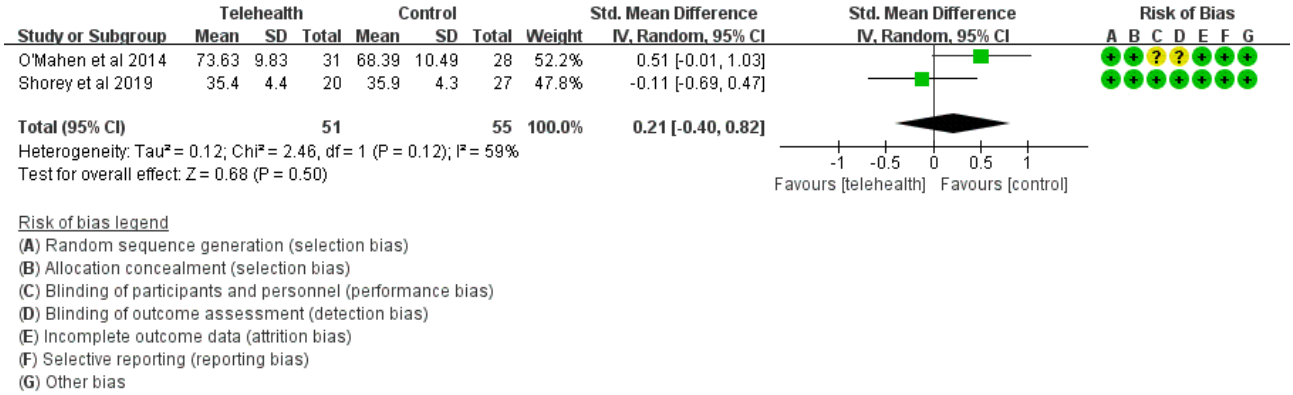
Figure 3. Depressive symptom scores in the telehealth group and the control group.



Social Support in the Telehealth Group and the Control Group

A total of 2 studies [36,37] reported social support improvement after interventions by the Perceived Social Support for Parenting instrument [37] and the Social Provision Scale [36]. The efficacy of telehealth was evaluated among 101 women by comparing the telehealth and control groups. There was no statistically significant difference between the two groups (SMD=–0.21, 95% CI –0.40 to 0.82; P=.50, I^2=59%; Figure 4).

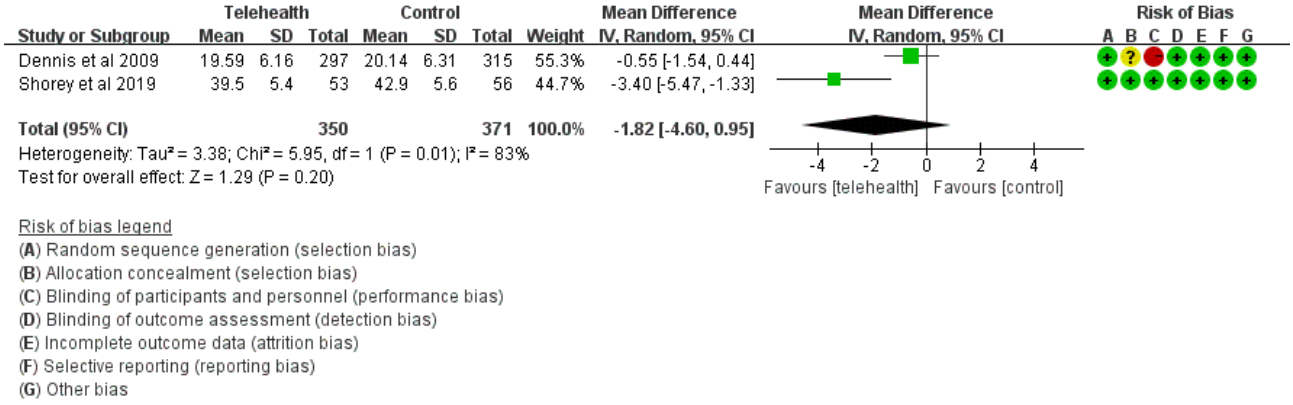
Figure 4. Social support scores in the telehealth group and the control group.



Loneliness in the Telehealth Group and the Control Group

A total of 2 studies [30,37] reported loneliness scores after interventions using the University of California, Los Angeles Loneliness Scale. The efficacy of telehealth was evaluated among 721 participants by comparing the telehealth and control groups. Meta-analysis showed no statistically significant difference between the two groups (MD=–1.82, 95% CI –4.60 to 0.95; P=.20, I^2 = 83%; Figure 5).

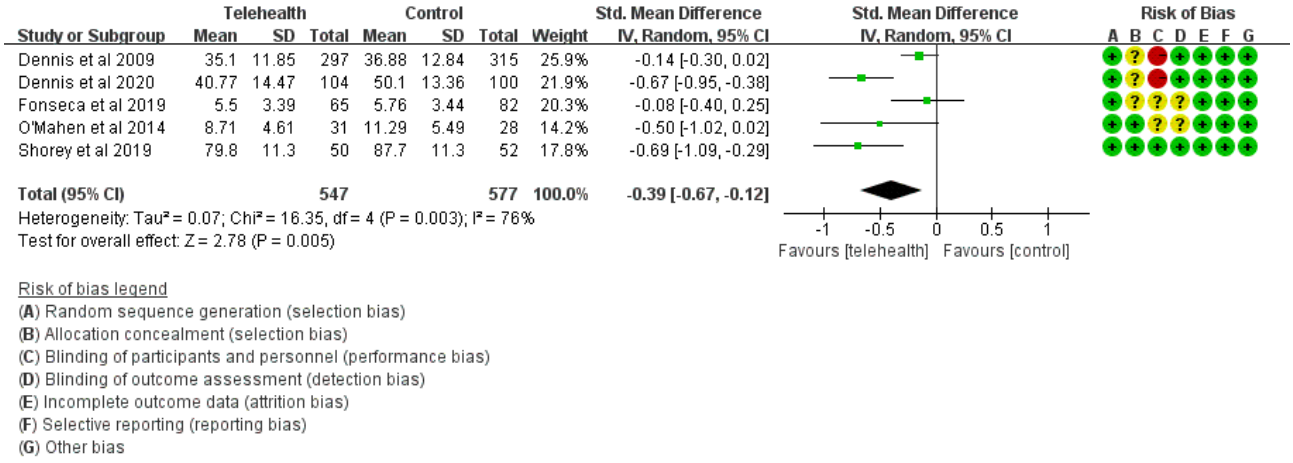
Figure 5. Loneliness scores in the telehealth group and the control group.



Anxiety in the Telehealth Group and the Control Group

A total of 5 studies [30-32,36,37] presented anxiety symptoms after interventions. Anxiety symptoms were measured through the State-Trait Anxiety Inventory [30,31,37], the Hospital Anxiety and Depression Scale [32], and the Generalized Anxiety Disorder 7-item scale [36]. Scores of anxiety symptoms in the telehealth group were lower than in the control group (SMD=–0.39, 95% CI –0.67 to –0.12; *P*=.005; I^2=76%; Figure 6).

Figure 6. Anxiety symptom scores in the telehealth group and the control group.



Sensitivity and Subgroup Analysis

In relation to the primary outcome (ie, depressive symptoms improvement), the pooled estimates were consistent when excluding one study at a time, indicating that the meta-analysis results of this study were stable and reliable. A series of subgroup analyses were performed to uncover more information about the heterogeneity (as shown in Table 1). Subgroup analyses were performed according to the severity of PPD, telehealth technology, specific therapy, and follow-up time. However, considerable heterogeneity (I^2≥71%) existed in each subgroup analysis.

Table 1. Subgroup analyses of the effect of telehealth interventions on depressive symptoms.

Subgroup analyses	Number of randomized controlled trials	Mean difference (95% CI)	*Z* statistic	*P* value	I^2, %
Severity of postpartum depression				<.001	
EPDS[a] scores ≥9	3 [29,30,37]	–1.64 (–2.20 to –1.08)	5.76	<.001	71
EPDS scores ≥10	2 [32,34]	–0.94 (–1.78 to –0.10)	2.2	.03	81
EPDS scores ≥12	4 [31,33,35,36]	–5.27 (–5.95 to –4.60)	15.29	<.001	84
Telehealth technology				<.001	
Telephone	6 [29-31,34,36,37]	–2.18 (–2.64 to –1.72)	9.26	<.001	84
App	2 [33,37]	–4.02 (–4.67 to –3.36)	12.03	<.001	98
Website	3 [32,35,36]	–1.62 (–2.47 to –0.78)	3.76	<.001	87
Specific therapy				<.001	
Peer support	3 [29,30,37]	–1.64 (–2.20 to –1.08)	5.76	<.001	71
Interpersonal psychotherapy	1 [31]	–5.13 (–6.44 to –3.82)	7.7	<.001	N/A[b]
Cognitive behavioral therapy	3 [32-34]	–3.25 (–3.91 to –2.60)	9.73	<.001	97
Behavioral activation therapy	2 [35,36]	–3.30 (–4.50 to –2.10)	5.4	<.001	0
Follow-up time (weeks)				<.001	
4	2 [29,37]	–1.21 (–1.93 to –0.49)	3.3	.001	73
6	1 [34]	–1.90 (–3.08 to –0.72)	3.16	.002	N/A
8	3 [29,32,33]	–3.86 (–4.64 to –3.07)	9.6	<.001	99
12	3 [30,31,37]	–2.12 (–2.65 to –1.60)	7.92	<.001	93
15	1 [35]	–3.34 (–4.75 to –1.93)	4.65	<.001	N/A
17	1 [36]	–3.21 (–5.50 to –0.92)	2.74	.006	N/A
24	3 [30,31,34]	–1.58 (–2.14 to –1.03)	5.55	<.001	94
36	1 [31]	–3.34 (–4.75 to –1.93)	4.65	<.001	N/A

[a]EPDS: Edinburgh Postnatal Depression Scale.

[b]N/A: not applicable.

Discussion

Overview

Our systematic review summarized the effectiveness of telehealth interventions on PPD and associated maternal mental health problems in the postpartum period (defined as ≤12 months after childbirth). We found that telehealth interventions could significantly improve depression and anxiety symptoms, although their effectiveness in improving social support and reducing loneliness was less certain.

Effectiveness of Telehealth Interventions on Postpartum Depression

In this study, we demonstrated that maternal depression scores were significantly lower in the telehealth group compared to the control group. Previous studies [30,34,38] were inconsistent regarding whether telehealth could improve maternal depression in the long term. One study [38] reported that the EPDS scores of women in the telehealth group increased at 24 weeks after interventions. Another study [30] discovered no significant difference in EPDS scores at 24 weeks after interventions between the two groups. However, one study [34] reported that EPDS scores at 24 weeks after interventions were significantly lower in the telehealth group. The synthesized result of this systematic review suggested that telehealth was effective in reducing EPDS scores at 24 weeks after the interventions, which was consistent with an earlier systematic review [39] comprised of 7 RCTs with a total of 1106 participants.

Effectiveness of Telehealth Interventions on Social Support, Loneliness, and Anxiety

In this study, we also summarized the effectiveness of telehealth interventions on mental health issues associated with PPD including social support, loneliness, and anxiety. However, telehealth interventions were not significantly effective at improving social support and loneliness in women affected by PPD. This may be related to the use of different assessment tools and the small number of included studies in the literature reporting on these outcomes. In an internet-based peer therapy project in Singapore [37], the authors found that the degree of loneliness in depressed women was reduced while social support was increased after an online peer support intervention that used telephones and apps. On the other hand, the symptoms of anxiety were significantly reduced in the telehealth group, updating the results in the study which found no evidence on the effectiveness

of telehealth for anxiety [40]. Coexistence of PPD and anxiety was common [41] and was mainly related to negative life events experienced by mothers with inadequate social support and increased childcare burden [42]. Results of this systematic review suggested that the potential for telehealth to improve mental health care for either PPD or anxiety is being increasingly recognized by affected women and health care providers [43]. Many women have already used publicly available online apps to access informational support, to consult a team of specialists, or to seek and find resources to alleviate PPD and anxiety [44].

Strengths and Limitations

This review has multiple strengths. The results of subgroup and sensitivity analyses suggested the findings are robust. More than half of the included trials were of high quality, with a relatively high degree of evidence that telehealth interventions could be effective in PPD treatment. Furthermore, the studies included were conducted in both developed and developing countries, expanding generalizability. Finally, this review analyzed not only the effectiveness of telehealth interventions on PPD, but also on social support, loneliness, and anxiety.

However, there are several limitations in this review. First, there could be selection bias in the original studies as most women participated in the studies on a voluntary basis and were recruited online. Second, most of the included studies relied on self-report measure scales at either recruitment or follow-up, which may lead to inflated estimates of effect sizes. Third, the findings of this meta-analysis were limited by major heterogeneity. There was methodological heterogeneity among the studies included in terms of the severity of PPD, telehealth technology, specific therapy, and follow-up time. In addition, the secondary outcomes in the 9 RCTs included were assessed using different scales.

Implications for Practice

Through telehealth services, women could have access to the relevant knowledge of psychological interventions anytime and anywhere. The anonymity of chat rooms in telehealth services could help protect women's privacy, providing a new treatment option for women who do not want to receive a face-to-face treatment due to social stigma.

Implications for Future Research

This review highlights some directions for future research, including increasing research attention on antenatal and peripartum depression and determining the applicability of telehealth interventions for adolescent mothers who may be more comfortable with novel technologies. Consideration of an intervention including the mother's partner may be an important approach in future research, especially in cases where triggering events (such as adverse infant outcomes, major negative life events) may affect both the woman and her partner. Future studies should also aim to enhance methodological quality through consistent design and execution on such aspects of the study including telehealth technology, follow-up time, and severity of PPD.

Conclusion

Telehealth interventions could effectively reduce the symptoms of depression and anxiety in women with PPD. However, better designed and more rigorous large-scale RCTs targeting specific therapies are needed to further explore the potential of telehealth interventions for PPD.

Acknowledgments

This work has been partly supported by the Canadian Institute of Health Research (FDN-148438), Clinical Research Startup Program of Southern Medical University by High-level University Construction Funding of Guangdong Provincial Department of Education (LC2019ZD019), The Seventh Affiliated Hospital, Southern Medical University (YNKT201802), and Key Project of Nursing Scientific Research Plan of Southern Medical University (Z2019001).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the 9 included randomized controlled trials.
[DOCX File , 21 KB - mhealth_v9i10e32544_app1.docx]

Multimedia Appendix 2
Risk of bias chart.
[DOCX File , 21 KB - mhealth_v9i10e32544_app2.docx]

References

1. Shorey S, Chee CYI, Ng ED, Chan YH, Tam WWS, Chong YS. Prevalence and incidence of postpartum depression among healthy mothers: A systematic review and meta-analysis. J Psychiatr Res 2018 Sep;104:235-248. [doi: 10.1016/j.jpsychires.2018.08.001] [Medline: 30114665]
2. Norhayati M, Hazlina NHN, Asrenee A, Emilin WMAW. Magnitude and risk factors for postpartum symptoms: a literature review. J Affect Disord 2015 Apr 01;175:34-52 [FREE Full text] [doi: 10.1016/j.jad.2014.12.041] [Medline: 25590764]

3. 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]
4. Branquinho M, Rodriguez-Muñoz MF, Maia BR, Marques M, Matos M, Osma J, et al. Effectiveness of psychological interventions in the treatment of perinatal depression: A systematic review of systematic reviews and meta-analyses. J Affect Disord 2021 Aug 01;291:294-306 [FREE Full text] [doi: 10.1016/j.jad.2021.05.010] [Medline: 34062397]
5. Stewart DE, Vigod SN. Postpartum Depression: Pathophysiology, Treatment, and Emerging Therapeutics. Annu Rev Med 2019 Jan 27;70:183-196. [doi: 10.1146/annurev-med-041217-011106] [Medline: 30691372]
6. Cuomo A, Maina G, Neal SM, De Montis G, Rosso G, Scheggi S, et al. Using sertraline in postpartum and breastfeeding: balancing risks and benefits. Expert Opin Drug Saf 2018 Jul;17(7):719-725. [doi: 10.1080/14740338.2018.1491546] [Medline: 29927667]
7. Guille C, Newman R, Fryml LD, Lifton CK, Epperson CN. Management of postpartum depression. J Midwifery Womens Health 2013;58(6):643-653 [FREE Full text] [doi: 10.1111/jmwh.12104] [Medline: 24131708]
8. Prevatt B, Lowder EM, Desmarais SL. Peer-support intervention for postpartum depression: Participant satisfaction and program effectiveness. Midwifery 2018 Sep;64:38-47. [doi: 10.1016/j.midw.2018.05.009] [Medline: 29908406]
9. Bina R, Barak A, Posmontier B, Glasser S, Cinamon T. Social workers' perceptions of barriers to interpersonal therapy implementation for treating postpartum depression in a primary care setting in Israel. Health Soc Care Community 2018 Jan;26(1):e75-e84. [doi: 10.1111/hsc.12479] [Medline: 28726342]
10. Li Z, Liu Y, Wang J, Liu J, Zhang C, Liu Y. Effectiveness of cognitive behavioural therapy for perinatal depression: A systematic review and meta-analysis. J Clin Nurs 2020 Sep;29(17-18):3170-3182. [doi: 10.1111/jocn.15378] [Medline: 32563204]
11. Guo L, Zhang J, Mu L, Ye Z. Preventing Postpartum Depression With Mindful Self-Compassion Intervention: A Randomized Control Study. J Nerv Ment Dis 2020 Feb;208(2):101-107. [doi: 10.1097/NMD.0000000000001096] [Medline: 31868776]
12. Cacciola E, Psouni E. Insecure Attachment and Other Help-Seeking Barriers among Women Depressed Postpartum. Int J Environ Res Public Health 2020 May 30;17(11):3887 [FREE Full text] [doi: 10.3390/ijerph17113887] [Medline: 32486285]
13. Badowski ME, Walker S, Bacchus S, Bartlett S, Chan J, Cochran KA, et al. Providing Comprehensive Medication Management in Telehealth. Pharmacotherapy 2018 Feb;38(2):e7-e16. [doi: 10.1002/phar.2071] [Medline: 29239004]
14. Speyer R, Denman D, Wilkes-Gillan S, Chen Y, Bogaardt H, Kim J, et al. Effects of telehealth by allied health professionals and nurses in rural and remote areas: A systematic review and meta-analysis. J Rehabil Med 2018 Feb 28;50(3):225-235 [FREE Full text] [doi: 10.2340/16501977-2297] [Medline: 29257195]
15. So CF, Chung JW. Telehealth for diabetes self-management in primary healthcare: A systematic review and meta-analysis. J Telemed Telecare 2018 Jun;24(5):356-364. [doi: 10.1177/1357633X17700552] [Medline: 28463033]
16. Selzler A, Wald J, Sedeno M, Jourdain T, Janaudis-Ferreira T, Goldstein R, et al. Telehealth pulmonary rehabilitation: A review of the literature and an example of a nationwide initiative to improve the accessibility of pulmonary rehabilitation. Chron Respir Dis 2018 Feb;15(1):41-47 [FREE Full text] [doi: 10.1177/1479972317724570] [Medline: 28786297]
17. Steindal SA, Nes AAG, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' Experiences of Telehealth in Palliative Home Care: Scoping Review. J Med Internet Res 2020 May 05;22(5):e16218 [FREE Full text] [doi: 10.2196/16218] [Medline: 32369037]
18. 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]
19. Magann EF, Bronstein J, McKelvey SS, Wendel P, Smith DM, Lowery CL. Evolving trends in maternal fetal medicine referrals in a rural state using telemedicine. Arch Gynecol Obstet 2012 Dec;286(6):1383-1392 [FREE Full text] [doi: 10.1007/s00404-012-2465-5] [Medline: 22821508]
20. Brearly TW, Shura RD, Martindale SL, Lazowski RA, Luxton DD, Shenal BV, et al. Neuropsychological Test Administration by Videoconference: A Systematic Review and Meta-Analysis. Neuropsychol Rev 2017 Jun;27(2):174-186. [doi: 10.1007/s11065-017-9349-1] [Medline: 28623461]
21. Cheng PGF, Ramos RM, Bitsch JA, Jonas SM, Ix T, See PLQ, et al. Psychologist in a Pocket: Lexicon Development and Content Validation of a Mobile-Based App for Depression Screening. JMIR mHealth uHealth 2016 Jul 20;4(3):e88 [FREE Full text] [doi: 10.2196/mhealth.5284] [Medline: 27439444]
22. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
23. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio: Psychological Corporation; 1996.
24. Beck CT, Gable RK. Postpartum Depression Screening Scale: development and psychometric testing. Nurs Res 2000;49(5):272-282. [doi: 10.1097/00006199-200009000-00006] [Medline: 11009122]
25. 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]

26. Lee DT, Yip SK, Chiu HF, Leung TY, Chan KP, Chau IO, et al. Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry 1998 May;172:433-437. [doi: 10.1192/bjp.172.5.433] [Medline: 9747407]
27. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
28. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Cochrane Handbook for Systematic Reviews of Interventions, version 6.2.: Cochrane; 2021. URL: http://www.training.cochrane.org/handbook [accessed 2021-08-30]
29. Dennis C. The effect of peer support on postpartum depression: a pilot randomized controlled trial. Can J Psychiatry 2003 Mar;48(2):115-124. [doi: 10.1177/070674370304800209] [Medline: 12655910]
30. Dennis C, Hodnett E, Kenton L, Weston J, Zupancic J, Stewart DE, et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. BMJ 2009 Jan 15;338:a3064 [FREE Full text] [doi: 10.1136/bmj.a3064] [Medline: 19147637]
31. Dennis C, Grigoriadis S, Zupancic J, Kiss A, Ravitz P. Telephone-based nurse-delivered interpersonal psychotherapy for postpartum depression: nationwide randomised controlled trial. Br J Psychiatry 2020 Apr;216(4):189-196. [doi: 10.1192/bjp.2019.275] [Medline: 32029010]
32. Fonseca A, Alves S, Monteiro F, Gorayeb R, Canavarro MC. Be a Mom, a Web-Based Intervention to Prevent Postpartum Depression: Results From a Pilot Randomized Controlled Trial. Behav Ther 2020 Jul;51(4):616-633. [doi: 10.1016/j.beth.2019.09.007] [Medline: 32586434]
33. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: A randomized controlled trial. Int J Med Inform 2020 Sep;141:104145. [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
34. Ngai F, Wong PW, Leung K, Chau P, Chung K. The Effect of Telephone-Based Cognitive-Behavioral Therapy on Postnatal Depression: A Randomized Controlled Trial. Psychother Psychosom 2015;84(5):294-303. [doi: 10.1159/000430449] [Medline: 26278623]
35. O'Mahen HA, Woodford J, McGinley J, Warren FC, Richards DA, Lynch TR, et al. Internet-based behavioral activation--treatment for postnatal depression (Netmums): a randomized controlled trial. J Affect Disord 2013 Sep 25;150(3):814-822. [doi: 10.1016/j.jad.2013.03.005] [Medline: 23602514]
36. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]
37. Shorey S, Chee CYI, Ng ED, Lau Y, Dennis C, Chan YH. Evaluation of a Technology-Based Peer-Support Intervention Program for Preventing Postnatal Depression (Part 1): Randomized Controlled Trial. J Med Internet Res 2019 Aug 29;21(8):e12410 [FREE Full text] [doi: 10.2196/12410] [Medline: 31469084]
38. Jiao N, Zhu L, Chong YS, Chan WS, Luo N, Wang W, et al. Web-based versus home-based postnatal psychoeducational interventions for first-time mothers: A randomised controlled trial. Int J Nurs Stud 2019 Nov;99:103385. [doi: 10.1016/j.ijnurstu.2019.07.002] [Medline: 31442783]
39. Hanach N, de Vries N, Radwan H, Bissani N. The effectiveness of telemedicine interventions, delivered exclusively during the postnatal period, on postpartum depression in mothers without history or existing mental disorders: A systematic review and meta-analysis. Midwifery 2021 Mar;94:102906 [FREE Full text] [doi: 10.1016/j.midw.2020.102906] [Medline: 33360589]
40. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
41. Lee EW, Denison FC, Hor K, Reynolds RM. Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review. BMC Pregnancy Childbirth 2016;16:38 [FREE Full text] [doi: 10.1186/s12884-016-0831-1] [Medline: 26928898]
42. Dol J, Richardson B, Grant A, Aston M, McMillan D, Tomblin Murphy G, et al. Influence of parity and infant age on maternal self-efficacy, social support, postpartum anxiety, and postpartum depression in the first six months in the Maritime Provinces, Canada. Birth 2021 Sep;48(3):438-447. [doi: 10.1111/birt.12553] [Medline: 34008241]
43. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile Health for Perinatal Depression and Anxiety: Scoping Review. J Med Internet Res 2020 Apr 13;22(4):e17011 [FREE Full text] [doi: 10.2196/17011] [Medline: 32281939]
44. Doherty K, Barry M, Marcano-Belisario J, Arnaud B, Morrison C, Car J, et al. A Mobile App for the Self-Report of Psychological Well-Being During Pregnancy (BrightSelf): Qualitative Design Study. JMIR Ment Health 2018 Nov 27;5(4):e10007 [FREE Full text] [doi: 10.2196/10007] [Medline: 30482742]

Abbreviations

EPDS: Edinburgh Postnatal Depression Scale

MD: mean difference
PPD: postpartum depression
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by G Eysenbach; submitted 01.08.21; peer-reviewed by M Ahmed Kamal; comments to author 29.08.21; revised version received 05.09.21; accepted 17.09.21; published 07.10.21.

Please cite as:
Zhao L, Chen J, Lan L, Deng N, Liao Y, Yue L, Chen I, Wen SW, Xie RH
Effectiveness of Telehealth Interventions for Women With Postpartum Depression: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2021;9(10):e32544
URL: https://mhealth.jmir.org/2021/10/e32544
doi:10.2196/32544
PMID:34617909

Review

Appropriation of mHealth Interventions for Maternal Health Care in Sub-Saharan Africa: Hermeneutic Review

Priscilla Maliwichi[1,2], BSc, MSc; Wallace Chigona[1], BSocSci, MSc, PhD; Karen Sowon[1], BSc, MSc, PhD

[1]Department of Information Systems, Faculty of Commerce, University of Cape Town, Cape Town, South Africa

[2]Department of Computer Science and Information Technology, Malawi Institute of Technology, Malawi University of Science and Technology, Thyolo, Malawi

Corresponding Author:
Priscilla Maliwichi, BSc, MSc
Department of Information Systems
Faculty of Commerce
University of Cape Town
Private Bag X1
Rondebosch
Cape Town, 7701
South Africa
Phone: 27 21 650 2261
Email: pmaliwichi@must.ac.mw

Abstract

Background: Many maternal clients from poorly resourced communities die from preventable pregnancy-related complications. The situation is especially grave in sub-Saharan Africa. Mobile health (mHealth) interventions have the potential to improve maternal health outcomes. mHealth interventions are used to encourage behavioral change for health care–seeking by maternal clients. However, the appropriation of such interventions among maternal health clients is not always guaranteed.

Objective: This study aims to understand how maternal clients appropriate mHealth interventions and the factors that affect this appropriation.

Methods: This study used a hermeneutic literature review informed by the model of technology appropriation. We used data from three mHealth case studies in sub-Saharan Africa: Mobile Technology for Community Health, MomConnect, and Chipatala Cha Pa Foni. We used the search and acquisition hermeneutic circle to identify and retrieve peer-reviewed and gray literature from the Web of Science, Google Scholar, Google, and PubMed. We selected 17 papers for analysis. We organized the findings using three levels of the appropriation process: adoption, adaptation, and integration.

Results: This study found that several factors affected how maternal clients appropriated mHealth interventions. The study noted that it is paramount that mHealth designers and implementers should consider the context of mHealth interventions when designing and implementing interventions. However, the usefulness of an mHealth intervention may enhance how maternal health clients appropriate it. Furthermore, a community of purpose around the maternal client may be vital to the success of the mHealth intervention.

Conclusions: The design and implementation of interventions have the potential to exacerbate inequalities within communities. To mitigate against inequalities during appropriation, it is recommended that communities of purpose be included in the design and implementation of maternal mHealth interventions.

(JMIR Mhealth Uhealth 2021;9(10):e22653) doi:10.2196/22653

KEYWORDS

mHealth; appropriation; mobile phones; model of technology appropriation; maternal health; community of purpose; hermeneutic literature review

Introduction

Background

Approximately 295,000 women die globally from pregnancy- and childbirth-related complications [1]. Most of these deaths are preventable [1]. The numbers are particularly high in transitional countries. For instance, sub-Saharan Africa recorded approximately 196,000 of these maternal deaths, and most of these deaths occurred in poorly resourced settings [1]. This translates to 533 deaths per 100,000 live births in sub-Saharan Africa [1]. Sustainable Development Goal 3 seeks to reduce the maternal mortality ratio to <70 deaths per 100,000 live births [1]. To contribute toward Sustainable Development Goal 3, information and communication technologies are used to improve maternal health care–seeking behavior. For example, mobile phones have been used to send health tips and reminders to visit antenatal care clinics and health facilities for delivery. The use of mobile phones in health care is known as mobile health (mHealth) [2].

Previous studies have pointed to the low uptake and low efficacy of mHealth interventions, especially in transitional countries [3,4]. For mHealth interventions to meet maternal health care needs, maternal health clients must not only adopt [5] but also appropriate interventions. Appropriation is the way technologies are adopted, adapted, and incorporated into everyday life [6]. The appropriation of technology goes beyond mere adoption. Appropriation also deals with how users engage with the technology, and this might differ from how the designers of the technology had intended. Information systems researchers have explored the phenomenon of technology appropriation [7-9]. Carroll et al [8] focused on the appropriation of technologies over time. Others have argued that the focus should be on how technologies are appropriated to get the job done or the intended outcome achieved [7]. Furthermore, technology appropriation may influence users for social changes [10]. Once technologies become a routine part of daily life, they often generate particular forms of habituated practice and a specific form of sociality.

Objectives

There is a need to investigate the appropriation of maternal mHealth interventions by maternal clients in transitional countries [11,12]. Most studies on the appropriation of mobile technologies have been conducted in resource-rich countries where mobile phone ownership is high and infrastructure is developed [7,11]. In contrast, in transitional countries, the adoption and appropriation of mobile phones to support health care are affected by demographic factors, such as low levels of literacy and low mobile phone ownership, and structural challenges, such as low connectivity [13]. To understand mobile technology use, it is necessary to understand technology appropriation in different contexts [14]. Therefore, this study seeks to investigate how maternal health clients in transitional countries appropriate mHealth interventions. The following research questions guided this study: (1) How do maternal clients appropriate maternal mHealth interventions? (2) What factors affect the appropriation of maternal mHealth interventions?

Methods

Study Design

This study used a hermeneutic literature review. Data were collected and analyzed using a hermeneutic framework for reviews. This study used the model of technology appropriation (MTA) as a theoretical lens.

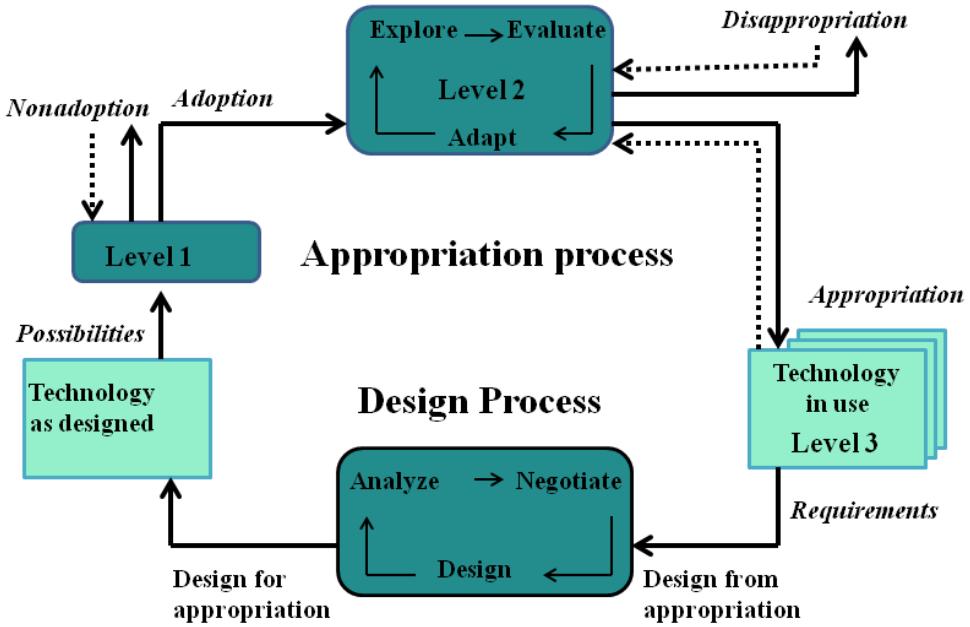
Theoretical Framework: MTA

MTA was developed by Carroll et al [6] to explain how young people adopt and use technologies. MTA has been used in mHealth for decades. Imperatore and Dunlop [15] used MTA to assess how people with aphasia (lack of language abilities) appropriate smartphones. Humans interact with mobile technologies in diverse and dispersed contexts. In maternal health, maternal clients may opt to not, or fail to, exploit the capabilities of an mHealth intervention, which may result in nonappropriation of the mHealth intervention. However, deciding to register for maternal mHealth interventions initiates the process of appropriation. The process of appropriation may result in either integrating the technology in their everyday life (appropriation) or disappropriation, that is, stopping using a technology.

Technology not only shapes users' behaviors, but users, in turn, shape how systems are created through use [9,16]. The design of systems is completed through the process of appropriation, whereby the use and performance of design change over time. Therefore, the focus of appropriation is twofold: (1) it draws attention to the context of use and the need to use evaluations that are situated in the context of the phenomenon and (2) the unfolding of use over time associated with appropriation suggests that evaluations conducted to support the design of technologies should continue after completion of the initial design process [17].

According to MTA, the process of appropriating a technology has three stages: adoption, adaptation, and integration (Figure 1). At the *adoption stage*, the user interacts with the technology as intended by the designers [18]. Designers develop the technology to address specific needs in an organization or society. In this study, the interventions were designed to reduce maternal mortality and to assist in maternal home-based care by creating a link between the maternal client and the health facility. During the initial interaction with the technology, users evaluate the intervention and decide whether to adopt it [18]. For mHealth interventions, clients might be motivated to continue using the intervention if they find it valuable. However, a maternal client may not adopt the intervention because of other factors such as failing to register or not finding value in the use of the intervention.

Figure 1. Model of technology appropriation, adapted from the study by Carroll [18].



At the *adaptation stage*, users evaluate the technology more by exploring and using it [18]. Users not only familiarize themselves with the technology but also learn how the technology can support their practices or needs. Carroll [18] argued that at this stage, mutual adaptation occurs, with people adapting practices associated with the use of the technology and also adapting the technology itself. During this stage, users may come across influences that can encourage or discourage them from using the technology [18]. For example, the maternal client may realize that the information they received via the intervention was helpful. However, maternal clients may *disappropriate* when the mobile phone malfunctions or encounters system failures multiple times.

At the *integration stage*, the user incorporates the technology into their everyday lives [18]. For example, a maternal client may call the intervention call center when she feels something is wrong to get advice or get referred to the clinic. At this stage, the technology is in use and is working as expected. However, maternal clients may *disappropriate* the intervention when they have a miscarriage or stillbirth.

As illustrated in Figure 1, the appropriation process was not linear. Users may move forward and backward during these stages. A user at the appropriation stage may move back to the adaptation stage or may decide to *disappropriate.* Subsequently, the user may adopt the technology again. However, over time, technologies can be evaluated and redesigned for appropriation to meet new user requirements [18]. For example, after the pilot phase, mHealth interventions can be evaluated to determine their performance. This may inform the modifications to the design of mHealth interventions.

During appropriation, users evaluate technology in use [17]. Evaluation of the performance of a product is crucial to human experience [17]; individuals evaluate the things they come across. The evaluations inform user attitudes and behaviors as well as future actions, such as recommendations to friends. These evaluations are usually informal; however, frameworks, methods, and techniques have been developed to formalize the evaluation process [17]. An example of a formal evaluation method is the mHealth Evaluation, Reporting and Assessment checklist [19].

Hermeneutic Literature Review

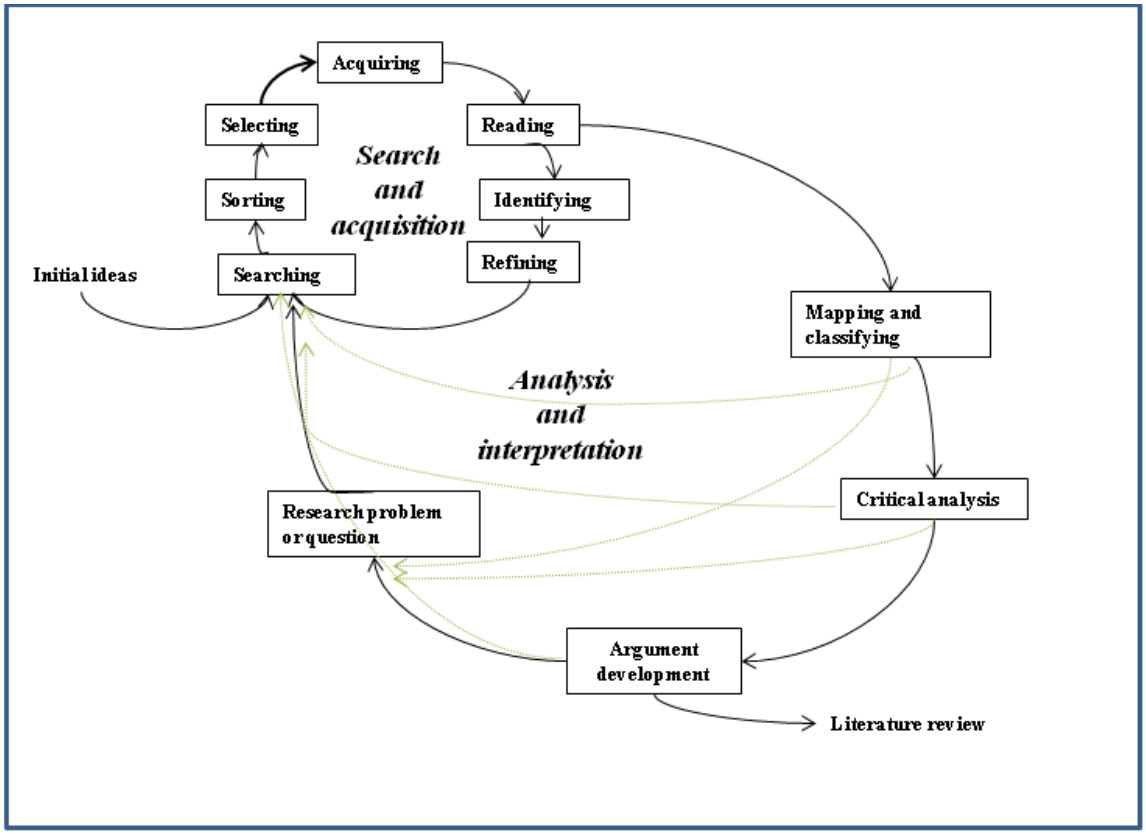
Overview

The hermeneutic literature review was deemed appropriate for this study because of its ability to create a contextual interpretive understanding of a phenomenon under investigation. The unstructured and flexible nature of the hermeneutic literature review made a hermeneutic literature review suitable for this study [20]. The search for relevant papers when using a hermeneutic literature review extends beyond database searches, as it allows the identification of evidence through snowballing and citation tracking [21]. Furthermore, a hermeneutic literature review allows the researcher to move from a general to a more specific search to identify relevant literature [21]. This is in contrast to a systematic literature review that encourages the use of a predefined set of keywords. A systematic literature review has the limitation that it may miss publications using different wording.

When using a hermeneutics circle, understanding the meaning and importance of individual texts depends on the understanding of the whole corpus of relevant literature. In turn, an understanding of the corpus of literature is built up through the understanding of individual articles [22]. This is an iterative process. A hermeneutic literature review uses the interpretive process, whereby a researcher expands and increases their understanding of the relevant literature [22].

Specifically, Figure 2 illustrates two circles: (1) search and acquisition and (2) analysis and interpretation. Textboxes 1 and 2 summarize the hermeneutic search and acquisition circle and the hermeneutic analysis and interpretation circle, respectively.

Figure 2. A hermeneutic framework for the literature review process [21].



Textbox 1. Overview of the hermeneutic search and acquisition circle.

Activity and Description

Searching

- When identifying publications using the hermeneutic framework, small sets of highly relevant publications are preferred over huge sets of documents whose relevance cannot be ascertained.

Sorting

- The results can be sorted based on the determined criteria, such as relevance rankings or publication dates.

Selecting

- Individual publications are selected for acquisition and reading.

Acquiring

- Full texts are acquired.

Reading

- Reading of acquired publications is initially orientational, leading to further selection of publications. Through orientational reading, the researcher gains a general understanding of the wider literature.

Identifying

- On the basis of the reading, researchers identify further search terms, additional publications (through citation tracking), authors, journals, conferences, and other sources.

Refining

- Search strategies can be used to refine searches. In particular, “citation pearl grow,” “successive fractions,” or “building blocks” can help in locating additional literature.

Textbox 2. Overview of the hermeneutic analysis and interpretation circle.

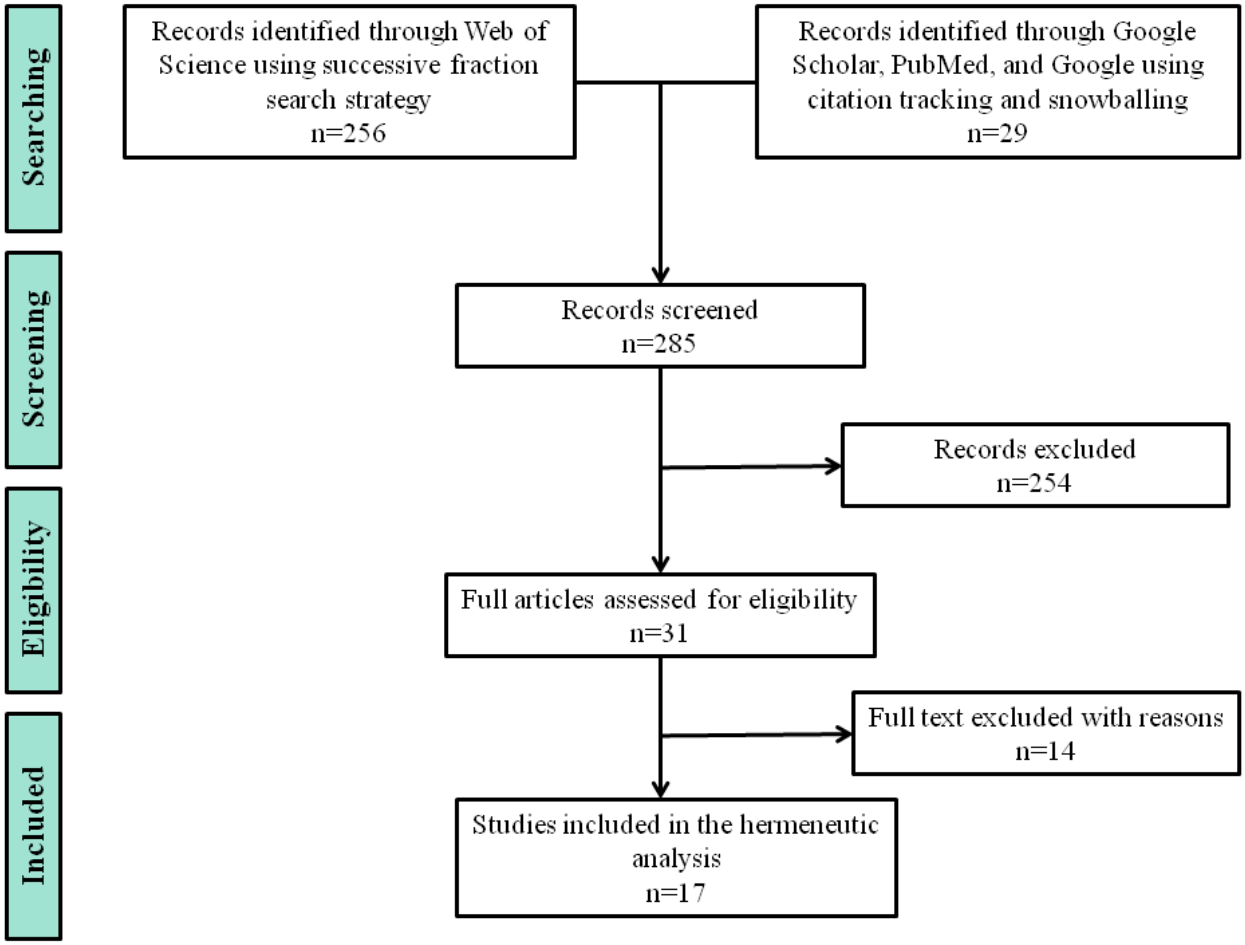
Activity and Description
Reading • Through analytic reading, the researcher identifies key concepts, findings, and theories and their interpretations. They also infer assumptions and a methodological approach; these may not be explicitly stated.
Mapping and classifying • Mapping and classifying provide a systematic analysis and classification of relevant ideas, findings, and contributions to knowledge within a body of literature.
Critical assessment • Critical assessment examines the body of literature on the basis of what is known and how knowledge is produced and acquired. The researcher also assesses how useful different types of knowledge are in understanding and explaining the problem of interest and where the boundaries and weaknesses of existing knowledge are.
Argument development • The argument development builds from the mapping and classification and also critical assessment, leading to the construction of a gap or problematization, which provides the motivation for further research. Through argumentation, future directions of research and the rationale for specific research questions are developed.
Research problem or question • Research questions can be formulated at a general, abstract level and at a more specific, empirical level. The former will logically follow from the gap in the literature or problematization of existing knowledge. The latter is typically transformed into one or more specific questions that can be empirically explored.
Searching • Searching leads to the identification of additional literature for further reading.

Search and Acquisition Circle

Owing to the nascency of mHealth, we opted to use both peer-reviewed and gray literature to obtain holistic descriptions of mHealth interventions. We searched the Web of Science, Google Scholar, Google, and PubMed and did not impose any year restrictions. The databases were selected for their coverage of mHealth literature. We used a combination of the following search terms: *Maternal*, *mHealth*, *mobile phone*, *appropriation*, *developing countries*, *Africa*.

The details of the search and selection strategies are presented in Figure 3. The search was conducted from December 2019 to March 2020.

Figure 3. Steps involved in a hermeneutic literature review.



We read the abstracts of the identified papers. While reading, we made notes of specific ideas from the text to refine the search. This prompted a second round of search, sort, and selection, where we also used citation tracking. On the basis of this reading, we compiled a list of mHealth interventions that were implemented in sub-Saharan Africa. We were interested in the history of the interventions, the technologies used, experiences of the maternal clients, and evaluations of those technologies. Finally, we selected interventions that met the following criteria: interventions that (1) were piloted and then scaled up (this allowed us to observe the progression of the intervention), (2) had evaluated both how maternal clients used the system and the technical aspects of the intervention, (3) worked on a basic phone, and (4) had run for a minimum of 3 years.

As our unit of analysis was the maternal client, we excluded interventions where the community health workers were primary beneficiaries.

Following this search and selection process, we identified five mHealth interventions: Wired Mothers (Tanzania), Rapid SMS (Rwanda), Mobile Technology for Community Health (MOTECH; Ghana), MomConnect (South Africa), and Chipatala Cha Pa Foni (CCPF; Malawi). Only MOTECH, MomConnect, and CCPF met the inclusion criteria. Textbox 3 summarizes the interventions and publications that qualified for analysis.

Textbox 3. Summary of papers used for analysis in the study.

Project Name and Publications

Mobile Technology for Community Health, Ghana

- Grameen Foundation [23]
- Lefevre et al [24]
- Macleod et al [25]
- Willcox et al [14]

MomConnect, South Africa

- Skinner et al [26]
- Seebregts et al [13]
- Coleman and Xiong [27]
- Lefevre et al [28]
- Barron et al [29]
- Seebregts et al [30]

Chipatala Cha Pa Foni, Malawi

- Nyemba-Mudenda and Chigona [31]
- Crawford et al [32]
- Larsen-Cooper et al [33]
- Larsen-Cooper et al [34]
- Blauvelt et al [35]
- Fotso et al [36]
- VillageReach [37]

Mapping and Classifying

During mapping and classifying, different factors such as the unit of analysis, major concepts, theoretical lens, and conceptual framework are considered [21]. In this study, we used MTA as the theoretical lens to map and classify our findings.

The synthesis of the selected articles involved repeated reading, looking at how different mobile technology functions have been used, and the experience of maternal clients as they appropriate these technologies for maternal health. Excel (Microsoft Inc) was used to tabulate the findings.

Descriptions of the Interventions

Overview

All 3 interventions implemented mHealth interventions that could work on a basic phone. These interventions used push SMS text messaging, push voice messages, and retrieved voice messages, that is, basic functionalities of a mobile phone. A hotline service is integrated into the system to advise maternal clients in real time, and in some cases, the helpdesk is used to report queries encountered when appropriating the intervention.

All 3 interventions in this study evaluated the technological performance of their mHealth intervention after the pilot phase and after operating for a few years after scaling up. This enabled the implementer to modify the system to optimize its performance.

Mobile Technology for Community Health

MOTECH was launched in rural Ghana in 2010 in the Upper East region and later scaled up to seven districts across four regions [24]. The project aimed to leverage mHealth to increase the quantity and quality of prenatal and neonatal care in the Upper East region and create a replication in the Awatu Senya district and to improve health outcomes for mothers and their newborn babies [23]. MOTECH was scaled in clusters over a 3-year period to reach 78.7% (170/216) of Ghana's districts [14].

The system has a component for maternal clients called Mobile Midwife app as well as the nurses' apps [23]. The Mobile Midwife service provides pregnant women and their families with SMS text messages or voice messages that provide time-specific information about their pregnancy each week. These messages include alerts and reminders for care-seeking, actionable information and advice, and educational information. The messages were written in local languages.

Maternal clients can register for Mobile Midwife through either a community health worker who captures their details on a MOTECH registration form on the phone or by calling the MOTECH call center [23]. Users who do not have a personal or household phone may access their messages by calling a toll-free number from a phone on any telecommunications provider in the country. Once connected to MOTECH, the user

interacts with the Mobile Midwife interactive voice response (IVR) system.

MomConnect

The MomConnect initiative is run by the country's department of health [29]. The initiative sends SMS messages to maternal clients and new mothers in South Africa. "In three years, MomConnect has been taken to scale to reach over 95% of public health facilities and has reached 63% of all pregnant women attending their first antenatal appointment" [29]. MomConnect provides maternal clients with maternal health information and encourages them to register at an antenatal care clinic. It is expected that the intervention would provide a valuable service to new mothers, complementing the current set of health care services by informing mothers about maternal health and childcare [26]. Maternal clients subscribe to MomConnect via Unstructured Supplementary Service Data (USSD). To register on the system, a nurse must first confirm that the woman is pregnant [29].

SMS text messages sent to the maternal clients include antenatal care and access to care during labor, diet and nutrition, nonpregnancy-related infections, hypertension, newborn care, breastfeeding, and immunization. The system sends between 1 and 3 messages per week, depending on the stage of the pregnancy. The messages continue until the child is 1 year old [26]. The registration and the sending and receiving of messages are free of charge to the user. If a mother does not own a phone, she can opt to receive the messages via a phone owned by an acquaintance [26]. Maternal clients can register for the MomConnect service at any public health clinic in the country. MomConnect also has a help desk where mothers send messages. The messages are forwarded to the management of the concerned health facilities [29].

Chipatala Cha Pa Foni

CCPF (translates to *Health Center by Phone*) is a health hotline that was started in one district in Malawi in 2011. The initiative was later scaled up to the entire country, available 24 hours every day [35]. It was started as a pilot in the Balaka district, which was experiencing a high maternal mortality rate [38]. During the pilot phase, the intervention provided only maternal and child health services. The topics of calls ranged from danger signs needing emergency care to maternal clients calling to inquire about their expected due date [38]. Callers were provided with one-on-one health counseling with a care provider and were encouraged to provide home-based care and to seek appropriate care for themselves or their children when appropriate.

Furthermore, maternal clients were registered for the tips and reminders service during their first call. This service provides women with the opportunity to receive text messages or listen to recorded messages through the IVR system about how to care for themselves and their infants [38]. Messages were targeted to provide relevant and timely health information and reminders based on the stage of pregnancy or age of the child, such as reminders for antenatal care visits; birth planning; immunization timing; and the promotion of positive health behaviors, such as mosquito net use and exclusively breastfeeding [38].

The intervention evolved to become a general hotline, and the IVR system expanded to include different topics (in addition to the pregnancy topic), such as nutrition and hygiene [35]. Anyone could access the IVR system to speak with a hotline care provider or listen to specific messages [35]. From June 2019, the CCPF has been fully owned by the Government of Malawi, Ministry of Health [35].

Results

Overview

Our findings suggest that maternal clients appropriate mHealth interventions regardless of their mobile ownership status. Using MTA, the findings of this review were synthesized using the stages of the appropriation process, namely, adoption, adaptation, and integration (appropriation). We identified a number of factors as enablers and hindrances at different stages of appropriation. Table 1 summarizes the findings.

This section discusses the factors that influenced the different phases of appropriation of maternal mHealth interventions.

Table 1. Summary of findings.

Stages of appropriation	Enablers	Hindrances
Level 1: adoption	Easy to use [23,26,27,31,35]; content in local languages [23,24,35]; able to access the intervention on any mobile phone [13,30,31,35]; use of methods familiar to users (eg, SMS) [13,23,26,29,33,36]; and clear messages [14,23,27,31-33]	Inconsistent network connection [23,24,28,31]; user timeouts [26,28]; mobile phone skills [31-33]; and low literacy levels [23,31-33,35]
Level 2: adaptation	New information learned [23,27,29,31-33,35]; trusting of the message [23,26,27,31,35,36]; convenience of the service [23,27,31-33,36]; able to share information with husbands and friends [23,31,33,34]; and able to get situation-specific advice [23,31,33]	Mobile numbers cannot be changed [29]; messages not delivered [23,29,32,33]; malfunction of the keypad or mobile phone [31,33]; call congestion [23,35]; and bottlenecks in voice messages [14,23]
Level 3: integration	Empowered in decision-making [27,31,33,35,36]; improved number of antenatal visits [13,23,27,31,35]; improved food and medicine consumption [27,31,36]; place of delivery (health facility) [14,23,27,31-33,35]; exclusive breastfeeding [23,27,29,31]; improved number of vaccines [23,26,27,31]; and improved number of postnatal visits [23,26,27,31]	Messages not useful [27-29]; miscarriage [23,28]; stillbirth [23,28]; and baby loss [23,28]

Adoption Stage

The design of all the 3 interventions took the context of a transitional country into account. This helped to increase the chances of adoption for a wide range of clients. The adoption of the 3 mHealth interventions was influenced by (1) the low cost of accessing the intervention, (2) the frugality of the design of the interventions, and (3) the inclusion of clients with no mobile phones. The services on all the 3 interventions were provided free of charge to the user. This reduced the chances of others being excluded from benefiting from the intervention based on their economic status. In transitional countries, women are more severely disadvantaged than men; hence, this is particularly useful because the interventions primarily targeted underserved communities that are burdened by economic hardships [39].

The interventions were frugal in that they were based on technologies that work on basic phones (eg, SMS text messaging, USSD, and voice) [23,26,27,31,35]. Although there is a growing number of mobile phones in Africa, most poor and rural women do not own smartphones [13]. Furthermore, in rural areas, mobile phone networks may not always support internet-based apps [33]. In such a context, technologically sophisticated interventions based on smartphones would serve little purpose. In addition to the ubiquity of the functionalities across phone types, the use of basic phone functionalities also ensured that the users were already familiar with such functionalities from their normal mobile phone use [13,23,26,29,33,36].

All the 3 interventions were designed to cater to both clients who owned and those who did not own a mobile phone. The interventions allowed those who did not own phones to use third-party phones [13,30,31,35]. The CCPF used community volunteers to provide maternal clients access to mobile phones. However, MomConnect and MOTECH allowed women to use mobile phones of husbands and friends. Hence, maternal clients could adopt the interventions regardless of their mobile phone ownership status. However, for CCPF, the use of community volunteers faced a number of challenges, such as sustaining volunteer motivation, challenges in accessing volunteers, phone maintenance, and mobile phone charging [33].

CCPF and MOTECH had the option for the clients to call a hotline or to interact with the IVR system to retrieve voice messages [23,35,40]. Most maternal clients used the pushed (voice messages sent to the client's mobile phone) or retrieved voice messages (voice messages that are listened on demand through the IVR system). Maternal clients interacted with the IVR system to access voice messages [23]. The preference for voice messages could be because of low literacy levels in rural areas, especially among women [31]. Furthermore, this could be due to the fact that some African communities are oral societies and, therefore, prefer voice messages over written text [31].

Adaptation Stage

Adaptation occurred when a maternal client had registered for the intervention and had familiarized herself with the intervention. Adaptation was influenced by (1) the need to learn new information and practices, (2) convenience of the service, and (3) trustworthiness of the information. The new information that the maternal clients learned about maternal health and nutrition influenced appropriation. The CCPF baseline survey showed that clients could list the information that was new to them [32]. The clients may have valued the intervention as a source of new information because clients struggle to obtain information from the clinics, as the clinics are too busy and have long queues. Furthermore, because of the culture that limits women from talking to strangers about pregnancy-related matters, women might have shied away from seeking the information from face-to-face consultations with clinicians [40].

The maternal clients felt that the interventions were convenient for them [23,27,31-33,36]. When the client did not feel well during pregnancy, they called the call center to determine whether their condition required medical attention. They saved time and money by not traveling long distances to the health facility, only to be told that they did not require medical attention. In rural areas of transitional countries, maternal clients travel long distances to the nearest health facility, and raising transport costs are a challenge [31].

Maternal clients trusted the information they received from the interventions and trusted the call center workers [23,26,27,31,35,36]. All the interventions were part of the health services provided by the department of health of their respective countries, which could be the reason why the maternal clients trusted the information [13,35].

Furthermore, there is evidence that the clients used the interventions and the information provided by the interventions [23,27,29,31-33,35]. On the basis of the information obtained from the interventions, the clients could make decisions about seeking care [27,31,33,35,36]. The messages helped the maternal clients make better maternal and infant health decisions. The maternal clients felt empowered and felt they could manage their pregnancy [26].

Integration Stage

Integration is reached when using mHealth interventions becomes routine in the maternal client's everyday life. The integration of the intervention in the clients' lives was influenced by (1) attitudes and behaviors of the user and (2) performance of the technology [32,41]. At this stage, the use of the mHealth intervention influenced the maternal clients to attend all antenatal care clinics, take medication and have a balanced diet, deliver at the health facility, take the child to the clinic, and receive all the vaccines. An independent evaluation of CCPF linked the intervention with improved knowledge of maternal and child health as well as certain behaviors, such as increased use of antenatal care clinics within the first trimester [13,23,27,31,35], increased use of a mosquito net during pregnancy and also for children under the age of 5 years [32,34], increased rates of early initialization of breastfeeding, and increased knowledge of health behaviors in pregnancy and the postnatal period [35]. However, the evaluation showed reduced use during the postnatal period [23,26,27,31]. This could be because of the fact that some clients found that the messages were not useful [40].

Factors That Affect Appropriation

Appropriation of the intervention was affected in different ways at all stages. The factors may be categorized as personal and technological. Personal factors such as low levels of literacy [23,31-33,35] and low mobile phone skills [31-33] influence the likelihood of clients not adopting the intervention. For CCPF, nonadoption occurred because the majority of community volunteers and users were not familiar with the IVR system. One of the challenges that maternal clients encountered when using the IVR system was that the messages could not play [32]. This may have been caused by low mobile phone skills or malfunction of the system itself. This is similar to other findings, such as barriers to IVR use are related to lack of familiarity with the technology and social barriers, including lack of mobile phone use skills and infrastructure challenges [42]. The implementers of CCPF overcame this challenge by training community volunteers or community health workers who, in turn, trained the maternal clients in their communities [33]. Hence, when interventions are being introduced, there should be a provision of bespoke training to improve familiarity of the intervention among the communities [43].

The technical challenges were related to the actual phone [31,33] as well as the network [23,24,28,31]. One challenge was related to instances such as when a client loses a mobile number [29]. The client could no longer receive the messages because the system did not allow change of the mobile number. Messages sent to these numbers were recorded in the system as dropped. In some circumstances, because of the low quality of mobile phones, the keypad could not function properly for the clients to interact with the mHealth system or the mobile phone stopped working during the period in which the client was supposed to be using the mHealth intervention.

The challenge of unreliable networks and user timeout [26,28] hindered maternal clients from registering with the interventions. MomConnect clients used USSD to register. Although this function is ubiquitous across different mobile phone types, it is prone to both network and user timeouts. Mobile network providers place a high priority on voice calls; therefore, in areas where the service is limited, USSD sessions are dropped and replaced by voice calls [28]. This challenge during the registration into the interventions might demotivate potential clients from adopting the intervention. Furthermore, call congestion influenced the maternal client to not appropriate properly.

At the integration stage, unexpected circumstances forced some maternal clients to withdraw from the intervention. The most common reasons for withdrawing were miscarriages, stillborn babies, and baby deaths [23,28].

Discussion

Principal Findings

This study suggests that several enablers influence maternal clients appropriate maternal mHealth interventions. The interventions were available free of charge to the clients, were implemented on technologies that were familiar to the potential clients, and were enabled to use regardless of mobile phone ownership status. Furthermore, the study noted a myriad of factors that hinder maternal clients' appropriation of technological interventions.

Considerations of the mHealth Intervention Context

mHealth technologies are enablers in the provision of intervention services. The use of SMS text messaging ensured that mHealth implementers could reach the most vulnerable maternal clients in hard-to-reach areas. However, the same SMS technology has raised several challenges. In all 3 interventions, some pushed SMS messages (SMS sent by the intervention to the maternal client mobile phone) sent to maternal clients were dropped [28]. Several factors contributed to the dropped SMSs. Some SMSs dropped because the recipients' mobile phones were off or unavailable. Users in rural areas with limited electricity infrastructure typically switch off their mobile phones to preserve battery power. However, unavailability was because of the poor coverage of mobile networks in rural areas. Furthermore, the delivery rate of the pushed SMS messages depended on the mobile service provider. The high SMS drop rate could also be explained by some policy about changing phone numbers. MomConnect did not allow their clients to change their mobile phone numbers, and the clients had to register their new numbers. As such, pushed SMS messages for clients who had lost their mobile phones were recorded as dropped [29]. Furthermore, the delivery rate of pushed messages was observed to be dependent on the infrastructure and network coverage of mobile service providers.

Owing to the oral culture and low levels of literacy among women in rural areas, voice messages could have been a more appropriate option for message delivery than SMSs. However, the findings show that the delivery rate for pushed voice messages for MOTECH and CCPF was lower than that for the pushed SMS text messages [24]. This points to the role of infrastructural limitations in the design of mHealth interventions. Although some technologies may be more appropriate than others based on context, the limitations in infrastructure do not always allow designers to adopt user-centric designs. These challenges allude to the trade-offs between the design goals for low-resource and underprivileged settings. For example, the goal of implementing frugal innovations may not be congruent with the goals of technical reliability. Although the use of USSD addressed the goal of providing a low-cost option that was ubiquitous across all types of phones, this option did not offer technical reliability. Similarly, the goal of the need for voice messages was incongruent with the goal of ease of use.

Potential candidates for exclusion were those who did not own mobile phones. All the 3 interventions sought to include maternal clients who did not have a mobile phone. All interventions included the option of using a third-party phone [23,33]. The provision of asynchronous messages afforded the clients who did not own phones the flexibility to negotiate mobile phone use with the phone owners. CCPF reported that approximately 20% of maternal clients who accessed the service used third-party mobile phones [35].

The Influence of Usefulness on Appropriation

The usefulness of an mHealth intervention may enhance how maternal health clients appropriate it. In this study, the maternal health clients used the messages from the interventions to improve their knowledge on how to take care of themselves during pregnancy, how to prepare for birth, and how to care for the baby after birth. Our finding is similar to that of a study in Bangladesh, which noted that maternal clients found maternal health care information received from an mHealth intervention valuable [44].

The hotline for the CCPF afforded women an opportunity to ask questions and obtain advice from the hotline workers. This, to an extent, was a shift from the cultural practices of avoiding talking about pregnancy-related matters too early and with people outside one's own family. The hotline consultation afforded the women a sense of anonymity; they could talk to a person who could not see them and, therefore, had no power to harm their pregnancy. Here, it can be argued that the intervention mediated the interaction between clients and health care providers. The literature also claims that this interaction has improved women's freedom to talk about pregnancy with health care workers [45]. Furthermore, the hotline consultation allowed the women to talk about their pregnancy to a health care provider who was not from their community and who could not see them. Here, the women sought medical care while maintaining what was socially required of them.

The Role of Community of Purpose in the Appropriation of Maternal mHealth Interventions

The findings showed that a *community of purpose* around the maternal client may be vital to the success of the mHealth intervention. A community of purpose is the voluntary coming together of individuals with commitments and an organization with a mission [12]. The community of purpose has different members who may have different roles but are working together toward a shared purpose. The main purpose of the maternal health community of purpose is to promote the well-being of maternal clients. The mHealth intervention was one of the tools that the community could use to achieve its goals. In all the interventions, a variety of stakeholders, such as community leaders, community health volunteers, nurses, traditional healers, and other key community members, were engaged in the design of the programs [23,35]. The involvement of these stakeholders in the design process ensured that the implemented interventions were contextually relevant and sensitive. Involving actors in the health sector and people within the communities helps to legitimize the information being disseminated by the intervention [23].

CCPF and MOTECH train communities to know how to support maternal clients at home and in their communities. For example, a community could arrange for the transport of maternal clients to the health facility on the onset of labor [46]. Communities of purpose support maternal clients by ensuring that the clients have access to the intervention, even in cases where they do not own a mobile phone [35]. Leaving out key stakeholders could have negative consequences on the appropriation of the intervention.

Conclusions

This study analyzed how maternal clients appropriate mHealth interventions for maternal health. The study used the cases of three maternal mHealth interventions in sub-Saharan Africa. The study noted that a myriad of factors play a role in the way clients appropriate technological interventions at different stages of the appropriation process. The study also noted that the socioeconomic status of the intended clients may affect their appropriation. If the designers fail to take into account the context in which the intervention is deployed, the intervention may perpetuate and even exacerbate existing inequalities. Although mHealth interventions may serve to include maternal clients in the information society, there is always a risk that some people could be left behind if the mediating factors in the context are not considered. To reduce inequalities during the appropriation process, it is also recommended that the interventions seek to create and leverage on communities of purpose around the use of the intervention.

Future Work

This study used secondary data to understand how maternal clients appropriate mHealth interventions. Future studies should consider using primary data. This study did not distinguish the appropriation based on mobile phone ownership. It is likely that maternal clients who do not own a mobile phone and use third-party access experience the appropriation differently. It would be interesting to explore how maternal clients who do not own mobile phones appropriate maternal mHealth interventions.

Conflicts of Interest

None declared.

References

1. World Health Organization. Trends in maternal mortality 2000 to 2017: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division: executive summary. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/327596 [accessed 2020-03-20]
2. Noordam AC, Kuepper BM, Stekelenburg J, Milen A. Improvement of maternal health services through the use of mobile phones. Trop Med Int Health 2011 May;16(5):622-626 [FREE Full text] [doi: 10.1111/j.1365-3156.2011.02747.x] [Medline: 21342374]
3. Cho Y, Lee S, Islam SM, Kim S. Theories applied to m-Health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018:727-741. [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]

4. 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]
5. Rossmann C, Riesmeyer C, Brew-Sam N, Karnowski V, Joeckel S, Chib A, et al. Appropriation of mobile health for diabetes self-management: lessons from two qualitative studies. JMIR Diabetes 2019;4(1):e10271 [FREE Full text] [doi: 10.2196/10271] [Medline: 30924786]
6. Carroll J, Howard S, Vetere F. Identity, power and fragmentation in cyberspace: technology appropriation by young people. Working paper 01/IDG/2001, Department of Information Systems, University of Melbourne. 2001. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.6847&rep=rep1&type=pdf [accessed 2020-03-02]
7. Gaskin J, Lyytinen K. Psychological ownership and the individual appropriation of technology. In: Proceedings of the 43rd Hawaii International Conference on System Sciences. 2010 Presented at: 43rd Hawaii International Conference on System Sciences; Jan. 5-8, 2010; Honolulu, HI, USA p. 1-8. [doi: 10.1109/hicss.2010.472]
8. Carroll J, Howard S, Peck J, Murphy J. A field study of perceptions and use of mobile telephones by 16 to 22 year olds. J Inf Technol Theory Appl 2002;4(2):Article 6 [FREE Full text]
9. DeSanctis G, Poole MS. Capturing the complexity in advanced technology use: adaptive structuration theory. Organ Sci 1994;5(2):121-147. [doi: 10.1287/orsc.5.2.121]
10. Bar F, Weber MS, Pisani F. Mobile technology appropriation in a distant mirror: baroquization, creolization, and cannibalism. New Media Soc 2016;18(4):617-636. [doi: 10.1177/1461444816629474]
11. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
12. Nathan M. White-Collar Blues: Management loyalties in an age of corporate restructuring. Acad Manag Perspect 1995;9(4):97-99. [doi: 10.5465/ame.1995.9512032214]
13. Seebregts C, Dane P, Parsons AN, Fogwill T, Rogers D, Bekker M, et al. Designing for scale: optimising the health information system architecture for mobile maternal health messaging in South Africa (MomConnect). BMJ Glob Health 2018;3(Suppl 2):e000563 [FREE Full text] [doi: 10.1136/bmjgh-2017-000563] [Medline: 29713506]
14. Willcox M, Moorthy A, Mohan D, Romano K, Hutchful D, Mehl G, et al. Mobile Technology for Community Health in Ghana: Is maternal messaging and provider use of technology cost-effective in improving maternal and child health outcomes at scale? J Med Internet Res 2019;21(2):e11268 [FREE Full text] [doi: 10.2196/11268] [Medline: 30758296]
15. Imperatore G, Dunlop M. An investigation into appropriation of portable smart devices by users with aphasia. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility. 2015 Presented at: ASSETS '15: The 17th International ACM SIGACCESS Conference on Computers and Accessibility; October 26 - 28, 2015; Lisbon Portugal p. 323-324. [doi: 10.1145/2700648.2811360]
16. Orlikowski WJ. The duality of technology: rethinking the concept of technology in organizations. Organ Sci 1992;3(3):398-427. [doi: 10.1287/orsc.3.3.398]
17. Fidock J, Rynne A, Carroll J. Evaluating information systems: an appropriation perspective. In: Hart DN, Gregor SD, editors. Information Systems Foundations: The Role of Design Science. Australian National University, Canberra, Australia: ANU E Press; 2010:1-318.
18. Carroll J. Completing design in use: closing the appropriation cycle. In: Proceedings of the European Conference on Information Systems. 2004 Presented at: ECIS 2004; June 14-16, 2004; Turku, Finland URL: https://aisel.aisnet.org/ecis2004/44/
19. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Br Med J 2016;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
20. Geeling S, Brown I, Weimann P. Information systems and culture-a systematic hermeneutic literature review. CONF-IRM 2016 Proceedings. Paper 40. 2016. URL: https://aisel.aisnet.org/confirm2016/40 [accessed 2020-03-02]
21. Boell SK, Cecez-Kecmanovic D. A hermeneutic approach for conducting literature reviews and literature searches. Commun Assoc Inf Syst 2014;34:Article 12. [doi: 10.17705/1cais.03412]
22. Boell SK, Cecez-Kecmanovic D. Literature reviews and the hermeneutic circle. Aust Acad Res Libr 2010;41(2):129-144. [doi: 10.1080/00048623.2010.10721450]
23. Mobile Technology for Community Health in Ghana: What it is what Grameen Foundation has learned so far. Second edition. Grameen Foundation. 2012. URL: https://grameenfoundation.org/documents/5ozoc8ste2hjjkxekff8.pdf [accessed 2020-02-20]
24. LeFevre AE, Mohan D, Hutchful D, Jennings L, Mehl G, Labrique A, et al. Mobile Technology for Community Health in Ghana: what happens when technical functionality threatens the effectiveness of digital health programs? BMC Med Inform Decis Mak 2017;17(1):Article no 27 [FREE Full text] [doi: 10.1186/s12911-017-0421-9] [Medline: 28292288]
25. Macleod B, Phillips J, Stone AE, Walji A, Awoonor-Williams JK. The architecture of a software system for supporting community-based primary health care with mobile technology: the Mobile Technology for Community Health (MoTeCH) Initiative in Ghana. Online J Public Health Inform 2012;4(1):1-17 [FREE Full text] [doi: 10.5210/ojphi.v4i1.3910] [Medline: 23569631]

26. Skinner D, Delobelle P, Pappin M, Pieterse D, Esterhuizen TM, Barron P, et al. User assessments and the use of information from MomConnect, a mobile phone text-based information service, by pregnant women and new mothers in South Africa. BMJ Glob Health 2018;3(Suppl 2):e000561 [FREE Full text] [doi: 10.1136/bmjgh-2017-000561] [Medline: 29713504]
27. Coleman J, Xiong K. Gauging the impact of MomConnect on maternal health service utilisation by women and their infants in Johannesburg, South Africa. MomConnect 2017:1-26. [doi: 10.13140/RG.2.2.21122.79048]
28. LeFevre AE, Dane P, Copley CJ, Pienaar C, Parsons AN, Engelhard M, et al. Unpacking the performance of a mobile health information messaging program for mothers (MomConnect) in South Africa: evidence on program reach and messaging exposure. BMJ Glob Health 2018;3(Suppl 2):e000583 [FREE Full text] [doi: 10.1136/bmjgh-2017-000583] [Medline: 29713510]
29. Barron P, Peter J, LeFevre AE, Sebidi J, Bekker M, Allen R, et al. Mobile health messaging service and helpdesk for South African mothers (MomConnect): history, successes and challenges. BMJ Glob Health 2018;3(Suppl 2):e000559 [FREE Full text] [doi: 10.1136/bmjgh-2017-000559] [Medline: 29713503]
30. Seebregts C, Tanna G, Fogwill T, Barron P, Benjamin P. MomConnect: an exemplar implementation of the Health Normative Standards Framework in South Africa. CSIR ResearchSpace 2016:125-136 [FREE Full text]
31. Nyemba-Mudenda M, Chigona W. mHealth outcomes for pregnant mothers in Malawi: a capability perspective. Inf Technol Dev 2017;24(2):245-278. [doi: 10.1080/02681102.2017.1397594]
32. Crawford J, Larsen-Cooper E, Jezman Z, Cunningham SC, Bancroft E. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. Glob Health Sci Pract 2014;2(1):35-46. [doi: 10.9745/ghsp-d-13-00155]
33. Larsen-Cooper E, Bancroft E, O'Toole M, Jezman Z. Where there is no phone: The benefits and limitations of using intermediaries to extend the reach of mHealth to individuals without personal phones in Malawi. African Popul Stud Spec Ed 2015;29(1):1628-1642. [doi: 10.11564/29-1-714]
34. Larsen-Cooper E, Bancroft E, Rajagopal S, O'Toole M, Levin A. Scale Matters: A cost-outcome analysis of an m-Health intervention in Malawi. Telemed J E Health 2016;22(4):317-324 [FREE Full text] [doi: 10.1089/tmj.2015.0060] [Medline: 26348994]
35. Blauvelt C, West M, Maxim L, Kasiya A, Dambula I, Kachila U, et al. Scaling up a health and nutrition hotline in Malawi: the benefits of multisectoral collaboration. Br Med J 2018;363:k4590 [FREE Full text] [doi: 10.1136/bmj.k4590] [Medline: 30530659]
36. Fotso JC, Bellhouse L, Vesel L, Jezman Z. Strengthening the home-to-facility continuum of newborn and child health care through mHealth: evidence from an intervention in rural Malawi. African Popul Stud Spec Ed 2015 Jun 04;29(1):1663-1682. [doi: 10.11564/29-1-717]
37. VillageReach. Evaluation of the information and communications technology for maternal newborn and child health project. VillageReach. 2014. URL: https://www.villagereach.org/wp-content/uploads/2009/08/VR-Exec-Summary-of-CCPF-Eval_Feb-21_FINAL.pdf [accessed 2021-07-29]
38. Kachila U. A national mobile health hotline in Malawi: one step closer. VillageReach. 2017. URL: https://www.villagereach.org/2017/07/22/digital-health-innovation-last-mile-sustainability-partnership/#:~:text=I [accessed 2020-03-20]
39. Rowntree O. The Mobile Gender Gap Report 2019. GSM Association. 2019. URL: https://www.gsma.com/mobilefordevelopment/resources/mobile-gender-gap-report-2019/ [accessed 2020-02-20]
40. Nyemba-Mudenda M. A pathway through which mHealth outcomes are produced for maternal healthcare consumers in a developing country context. PhD / Doctoral Theses, University of Cape Town. 2015. URL: http://hdl.handle.net/11427/16688 [accessed 2020-01-20]
41. Baghizadeh Z, Cecez-Kecmanovic D, Schlagwein D. Review and critique of the information systems development project failure literature: an argument for exploring information systems development project distress. J Inf Technol 2019;35(2):123-142. [doi: 10.1177/0268396219832010]
42. Serçekuş P, Başkale H. Effects of antenatal education on fear of childbirth, maternal self-efficacy and parental attachment. Midwifery 2016;34:166-172. [doi: 10.1016/j.midw.2015.11.016] [Medline: 26656473]
43. Brinkel J, Dako-Gyeke P, Krämer A, May J, Fobil J. An investigation of users' attitudes, requirements and willingness to use mobile phone-based interactive voice response systems for seeking healthcare in Ghana: a qualitative study. Public Health 2017;144:125-133. [doi: 10.1016/j.puhe.2016.11.017] [Medline: 28274374]
44. 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):Article 101 [FREE Full text] [doi: 10.1186/s12911-017-0502-9] [Medline: 28683742]
45. Mildon A, Sellen D. Use of mobile phones for behavior change communication to improve maternal, newborn and child health: a scoping review. J Glob Health 2019;9(2):020425 [FREE Full text] [doi: 10.7189/jogh.09.020425] [Medline: 31893032]
46. Portela A, Santarelli C. Empowerment of women, men, families and communities: true partners for improving maternal and newborn health. Br Med Bull 2003;67(1):59-72. [doi: 10.1093/bmb/ldg013] [Medline: 14711754]

Abbreviations

CCPF: Chipatala Cha Pa Foni
IVR: interactive voice response
MOTECH: mobile technology for community health
MTA: model of technology appropriation
USSD: Unstructured Supplementary Service Data

Edited by L Buis; submitted 20.07.20; peer-reviewed by M McKinley, H Pratomo; comments to author 07.12.20; revised version received 14.04.21; accepted 24.06.21; published 06.10.21.

Please cite as:
Maliwichi P, Chigona W, Sowon K
Appropriation of mHealth Interventions for Maternal Health Care in Sub-Saharan Africa: Hermeneutic Review
JMIR Mhealth Uhealth 2021;9(10):e22653
URL: https://mhealth.jmir.org/2021/10/e22653
doi:10.2196/22653
PMID:34612835

Original Paper

Adaptation and Assessment of a Text Messaging Smoking Cessation Intervention in Vietnam: Pilot Randomized Controlled Trial

Nan Jiang[1], PhD; Nam Nguyen[2], PhD; Nina Siman[3], MSc; Charles M Cleland[1], PhD; Trang Nguyen[2], MA; Hue Thi Doan[4], MS; Lorien C Abroms[5], MA, ScD; Donna R Shelley[6], MPH, MD

[1]Department of Population Health, Grossman School of Medicine, New York University, New York, NY, United States

[2]Institute of Social and Medical Studies, Hanoi, Vietnam

[3]Ronald O Perelman Department of Emergency Medicine, Grossman School of Medicine, New York University, New York, NY, United States

[4]Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam

[5]Milken Institute School of Public Health, George Washington University, Washington, DC, United States

[6]School of Global Public Health, New York University, New York, NY, United States

Corresponding Author:
Nan Jiang, PhD
Department of Population Health, Grossman School of Medicine, New York University
180 Madison Ave
Room #17-54
New York, NY, 10016
United States
Phone: 1 646 501 3553
Email: Nan.Jiang@nyulangone.org

Abstract

Background: Text message (ie, short message service, SMS) smoking cessation interventions have demonstrated efficacy in high-income countries but are less well studied in low- and middle-income countries, including Vietnam.

Objective: The goal of the research is to assess the feasibility, acceptability, and preliminary efficacy of a fully automated bidirectional SMS cessation intervention adapted for Vietnamese smokers.

Methods: The study was conducted in 3 phases. In phase 1, we adapted the SMS library from US-based SMS cessation programs (ie, SmokefreeTXT and Text2Quit). The adaptation process consisted of 7 focus groups with 58 smokers to provide data on culturally relevant patterns of tobacco use and assess message preferences. In phase 2, we conducted a single-arm pilot test of the SMS intervention with 40 smokers followed by in-depth interviews with 10 participants to inform additional changes to the SMS library. In phase 3, we conducted a 2-arm pilot randomized controlled trial (RCT) with 100 smokers. Participants received either the SMS program (intervention; n=50) or weekly text assessment on smoking status (control; n=50). The 6-week SMS program consisted of a 2-week prequit period and a 4-week postquit period. Participants received 2 to 4 automated messages per day. The main outcomes were engagement and acceptability which were assessed at 6 weeks (end of intervention). We assessed biochemically confirmed smoking abstinence at 6 weeks and 12 weeks. Postintervention in-depth interviews explored user experiences among a random sample of 16 participants in the intervention arm.

Results: Participants in both arms reported high levels of engagement and acceptability. Participants reported using the program for an average of 36.4 (SD 3.4) days for the intervention arm and 36.0 (SD 3.9) days for the control arm. Four of the 50 participants in the intervention arm (8%) reset the quit date and 19 (38%) texted the keyword TIPS. The majority of participants in both arms reported that they always or usually read the text messages. Compared to the control arm, a higher proportion of participants in the intervention arm reported being satisfied with the program (98% [49/50] vs 82% [41/50]). Biochemically verified abstinence was higher in the intervention arm at 6 weeks (20% [10/50] vs 2% [1/50]; *P*=.01), but the effect was not significant at 12 weeks (12% [6/50] vs 6% [3/50]; *P*=.49). In-depth interviews conducted after the RCT suggested additional modifications to enhance the program including tailoring the timing of messages, adding more opportunities to interact with the program, and placing a greater emphasis on messages that described the harms of smoking.

Conclusions: The study supported the feasibility and acceptability of an SMS program adapted for Vietnamese smokers. Future studies need to assess whether, with additional modifications, the program is associated with prolonged abstinence.

Trial Registration: ClinicalTrials.gov NCT03219541; https://clinicaltrials.gov/ct2/show/NCT03219541

(JMIR Mhealth Uhealth 2021;9(10):e27478) doi:10.2196/27478

KEYWORDS

smoking cessation; text messaging; mHealth; mobile health; low- and middle-income country; smoking; developing countries; SMS; Vietnam

Introduction

Of the world's 1.1 billion smokers, 80% live in low- and middle-income countries (LMICs) [1]. As a result, tobacco use is a major contributor to the high burden of noncommunicable disease and premature death in LMICs [2]. Promoting cessation is the key to reversing current global trends in tobacco-related morbidity and mortality over the next few decades [3].

Vietnam, an LMIC, has one of the highest smoking rates in the world [4]. According to the 2015 Global Adult Tobacco Survey, 45.3% of Vietnamese men were current smokers [5]. The country has implemented a range of evidence-based tobacco control policies as defined by the World Health Organization's (WHO) Framework Convention on Tobacco Control including a national toll-free Quitline, launched in 2015 [6]. However, most smokers who attempt quitting do not call the Quitline or use cessation treatment [7]. In 2015, only 2.3% of recent quitters (who quit for less than 12 months) and current smokers who made past-year quit attempts received in-person or telephone treatment for smoking cessation [7].

To continue to meet goals for decreasing smoking prevalence globally, effective cessation interventions must be easily accessible, adapted to local languages and cultural contexts, and scalable. Mobile technology (mHealth) that uses text messaging or short message service (SMS) meets these criteria by creating a relatively low-cost platform for wide dissemination of tailored tobacco cessation interventions. A growing literature indicates that automated, bidirectional SMS cessation programs can be effective in increasing smoking cessation compared to minimal or no smoking cessation support [8-12]. However, this research has largely been conducted in upper middle-income countries [13]. The WHO Tobacco Free Initiative has emphasized the importance of developing mHealth solutions for increasing access to evidence-based tobacco cessation interventions in LMICs [14].

This study was conducted to address the gap in the literature by assessing the feasibility, acceptability, and preliminary efficacy of an SMS intervention for tobacco users in Vietnam. The study also provided an important opportunity to describe methods for adapting text message interventions found efficacious in high-income countries to different sociocultural contexts and forms of tobacco use.

Despite the large number of SMS studies conducted in high-income countries, few studies have compared the efficacy of combining SMS programs with additional cessation support to SMS alone, and findings have been mixed [12]. For example, Kruse et al [15] found that SMS combined with nicotine replacement therapy did not result in higher smoking abstinence compared with SMS alone. In contrast, the pilot study by White et al [16] found that combining SMS with personalized text message support from peer mentors increased cessation rates compared to SMS alone. Although there may be advantages to enhancing SMS interventions with additional support, this approach risks diminishing the potential cost advantage and scalability of automated SMS interventions [17]. Therefore, the goal of this study was to first adapt an SMS intervention to the sociocultural context of Vietnamese smokers and then compare the 6-week intervention to a control group that received a single text assessment on smoking status per week.

Methods

Study Design

The study was conducted in 3 phases. In the first phase, we conducted 7 focus groups (n=58 participants) to adapt text messages from SmokefreeTXT, a freely available public resource [18]. We supplemented that library with messages from Text2Quit to add topics not included in SmokefreeTXT like refusal skills and additional messages on harmful effects of tobacco use [19]. The second phase included a single-arm pilot test of the adapted SMS library with 40 participants. In the final phase, we conducted a 2-arm pilot randomized controlled trial (RCT) with 100 participants (98 males and 2 females). For all phases, eligible participants were (1) aged 21 to 55 years, (2) smoked ≥5 cigarettes per day (including dual users who used both cigarettes and waterpipe), (3) planned to quit smoking within the next 30 days, (4) had a mobile phone, (5) used text messaging in the past 6 months, and (6) lived in Hanoi, Vietnam. Exclusion criteria included current participation in other smoking cessation treatment and waterpipe-only users.

Recruitment and Enrollment

We partnered with a community health center in Hanoi to recruit participants for each phase of the study. Community health collaborators, who are similar to community health workers and are assigned to work with the community health centers, were trained to disseminate study information through community outreach activities. During their routine outreach, they assessed smoking status of community members and shared study information with current tobacco users. If interested these individuals were asked for permission to share their contact information with research staff. Research staff then contacted potential participants to provide additional details, obtain consent, and enroll them in the study. This study was approved by the institutional review boards of New York University

Grossman School of Medicine and Institute of Social and Medical Studies in Vietnam.

SMS Adaptation Procedures

Conceptual Framework

The content of efficacious SMS interventions is based on a combination of several theoretical frameworks including social cognitive theory, transtheoretical model, and cognitive behavioral theory [8,19-23]. This includes the SMS libraries that we adapted for Vietnamese smokers [24,25]. These theories guided the design of the focus group and interview guides used in the formative data collection and message modifications. For example, in the prequit phase of the intervention, messages were designed to promote readiness to quit and increase motivation (eg, reasons for quitting), address outcome expectancies (eg, harms of smoking, benefits of quitting), reinforce self-efficacy, and offer advice for how to prepare for the quit date. In the postquit phase of the intervention, messages continued to offer motivational messages similar to those in the prequit phase but added an emphasis on the importance of obtaining social support and offered cognitive and behavioral strategies for dealing with social, emotional, and environmental triggers; coping adaptively with cravings; and resuming quit attempts after a slip or relapse.

SMS Intervention Adaptation

The final message library was developed through an iterative process that included first translating messages from the English language SMS libraries into Vietnamese with some initial changes to align the content to the Vietnamese context. For example, strategies for coping with nicotine cravings were edited to include practices that were relevant to Vietnamese smokers. We then conducted focus groups to assess message preferences; elicit suggestions to guide further adaptations; and assess reasons for and barriers to quitting, smoking triggers, and participants' social networks and their influence on smoking behavior.

Focus groups included quantitative assessment of message preference followed by group discussions. Participants were asked to rate 46 text messages on a 1 to 4 scale (1=strongly dislike, 2=dislike, 3=like, 4=strongly like). Ratings were summarized while the focus groups elicited more details about smoking patterns and past quit attempts, reasons for quitting, and barriers to quitting. We then discussed a sample of the messages that were rated across the response scale options to gain additional insights about what types of messages were preferred and elicit suggestions for improving the messages.

Focus groups were moderated by two researchers and were audiorecorded, transcribed, and translated into English. Qualitative data analyses were conducted using NVivo 12 (QSR International). Using an inductive analytic approach, two research team members independently read a subset of transcripts (2-3) to identify preliminary themes, relevant patterns, and clustered concepts and generate questions [26,27]. Using an iterative process, the team continued to review transcripts until they reached consensus on a final codebook. One team member then coded the remaining transcripts.

Based on the findings from focus groups, the messages were further adapted. For example, compared to the original SMS programs, greater emphasis was placed on the impact of smoking on the family's health and the dangers of smoking. Other content that was added to align with Vietnamese culture and specific tobacco use patterns included the health hazards of waterpipe use, which is still commonly used in Vietnam. Findings from focus groups also pointed to a need to develop additional messages that encouraged smokers to identify people in their network who could support their quit attempt and offered suggestions on how to refuse an offer of cigarettes or decline to smoke when others are smoking, a common scenario in a country with high male smoking rates (eg, "Think of your children when someone offers you to smoke... Tell them 'I promised my children I wouldn't smoke'").

After finalizing the first draft of the SMS library, we enrolled 40 participants in a single-arm pilot test. The participants received automated bidirectional text messages for 6 weeks. At the end of the pilot test, two researchers conducted in-depth interviews with a random sample of 10 participants. The interviews explored 5 areas:

- Overall perceptions about the program (eg, "What did you think of the program?")
- Perceptions about specific program features such as the opportunity to type in the keyword TIPS to obtain additional advice on how to deal with cravings
- Perceptions about specific message themes (eg, "What types of messages were most helpful? Which were least helpful?"). We read some text messages that participants received during the intervention and asked what they liked and didn't like about the messages, as well as their suggestions (eg, "How can we improve the messages to make them more helpful for you?")
- Perceptions about other program characteristics including the number and timing of text messages and length of the program
- Suggestions for improving the SMS program.

Similar to the focus groups, the in-depth interview was guided by our integrated conceptual framework. Two researchers moderated and audiorecorded the interviews. Interviews were transcribed and translated into English. Two researchers used the same approach used to analyze the focus group data. Findings from the single-arm pilot test demonstrated the feasibility of retaining participants in the 6-week SMS cessation intervention and informed additional modifications to the content of the message library. These included a greater emphasis on harms of smoking versus benefits of quitting and adding more messages that offered concrete advice about coping with cravings rather than vague messages meant to motivate smokers (eg, "Stay strong, you can do it").

RCT Procedures

We conducted the pilot RCT (Multimedia Appendix 1) between November 2018 and March 2019 with 100 participants including 98 men and 2 women. Participants provided written consent at the time of enrollment and were randomized to the intervention (n=50) or control arm (n=50) using block randomization stratified by cigarette consumption per day (CPD; 5-10 vs >10 CPD). Participants completed a baseline survey at enrollment and follow-up surveys at 6 weeks and 12 weeks postenrollment.

All surveys were administered in person by a research assistant. At the end of the intervention period, we conducted in-depth interviews with a random sample of 16 participants from the intervention arm to obtain more in-depth information about their experiences with the program. Participants were compensated for text messaging charges that occurred during the intervention and received VND 50,000 (US $2.23) for each survey and VND 100,000 (US $4.46) for the in-depth interview.

SMS Intervention

The final message library consisted of 188 text messages. Messages were designed to increase knowledge about smoking and motivation to quit (elicit reasons for quitting, describe harms of smoking and secondhand smoke exposure and harms of waterpipe use, provide information about the Quitline); change outcome expectancies (benefits of quitting); and offer cognitive and behavioral strategies such as refusal skills to assist smokers in maintaining the quit attempt. Behavioral strategies encouraged self-efficacy for quitting and encouraged smokers to obtain social support from family and friends.

The intervention consisted of a 2-week prequit period and a 4-week postquit period. Participants received 2 or 3 messages per day during the 2-week prequit period, 4 on the quit date, 3 or 4 per day during the first 2 weeks after the quit date, and 2 or 3 messages per day during the subsequent postquit period. Prequit messages encouraged smokers to track their smoking behavior and identify triggers, reinforced reasons for quitting, elicited smokers' reasons for quitting, and provided advice on obtaining social support as they neared their quit date. Postquit messages were oriented toward relapse prevention and maintaining motivation and included the themes described above. In addition to programmed outgoing messages, participants could send the keyword TIPS to the program to trigger on-demand messages for additional support.

Starting from the quit date, participants received a weekly bidirectional text message to assess smoking status as follows: "Are you smoke-free? Reply: Yes or No." Those who responded yes continued to receive postquit messages. Those who answered no received a message asking if they preferred to set a new quit date. Those who responded no continued to receive postquit messages, and those answering yes received a call from the research assistant to obtain a new quit date and reset the quit date in the SMS program which returned to the prequit protocol. In addition, participants received bidirectional text messages that assessed their level of craving (hi, med, low) on days 1, 3, 5, 8, 15, and 25. A high or medium craving response triggered an automated message offering TIPS from the SMS program. Participants could opt out of the SMS program at any time by texting STOP. At the start of the program, participants were made aware that they had the option to text STOP at any time during the trial to discontinue receiving messages.

Control Arm

Participants in the control arm received one text assessment message per week at a fixed time in the evening during the 6-week intervention period: "Are you smoke-free? Reply: Yes or No." The control condition was consistent with previous SMS cessation trials that included minimum exposure for participants in the control group [15,28].

Measures

We conducted surveys at baseline and at 6 weeks (end of treatment) and 12 weeks. The baseline survey captured sociodemographic information such as gender, age, education, and household income level; text messaging habits; and smoking behavior, including CPD and waterpipe sessions per day. Participants were dichotomized as dual users if they reported waterpipe use on some day or every day or cigarette-only smokers if they responded not at all to the waterpipe use question.

Measures of feasibility included reach (ie, the proportion of individuals approached who enrolled) and survey assessment response rates. We also tracked if participants experienced technical problems.

Two measures of program engagement were assessed using the 6-week survey: (1) the number of weeks that participants reported using the program (calculated as the mean number of days using the program based on that response) and (2) self-reported frequency of reading the messages (always/usually/sometimes/never). Two additional measures included the proportion of participants who responded to the bidirectional text message assessments with mean number of times they responded and the proportion who texted the keywords (eg, TIPS) with mean number of times they texted the keyword.

Program acceptability was assessed at 6 weeks by asking participants to rate their overall satisfaction with the SMS program (for intervention arm) or the weekly text assessment (for control arm; very satisfied/satisfied/unsatisfied/very unsatisfied), perceived number of messages (too many/just right/too few), and their agreement with statements such as "The text messages helped me quit smoking" using a 4-point Likert scale from "strongly disagree" to "strongly agree."

Smoking abstinence was assessed at 6 weeks and 12 weeks. Abstinence was defined as self-reported no smoking in the past 7 days confirmed with a carbon monoxide of 10 ppm or less [29]. Quit attempts were assessed by asking participants if they had ever stopped smoking cigarettes for a day or more during the intervention period because they were trying to quit (yes/no). Reductions in cigarettes smoked per day was calculated as the difference in CPDs at baseline compared with 6 weeks and 12 weeks.

In-Depth Interview Procedures

We conducted in-depth interviews with 16 randomly selected participants from the intervention arm. Using a semistructured guide similar to the single-arm pilot test, two researchers moderated and audiorecorded the interviews to obtain more in-depth information about what they liked and didn't like about the program and elicit recommendations for improving the SMS program. Interviews were transcribed and translated into English.

Data Analysis

We analyzed quantitative data using the R statistical computing environment (R Foundation for Statistical Computing) [30]. Sociodemographic characteristics of the sample, the intention-to-treat abstinence rates, and other cessation outcomes were compared by study arm using Pearson chi-square and *t* tests. We used descriptive statistics to summarize program acceptability and engagement results. All tests of statistical significance were 2-tailed, and $P<.05$ was considered significant. The process for qualitative analyses was the same across the 3 aims and described above.

Results

Participant Demographics and Smoking Behavior

On average, participants were aged 38.9 (SD 8.2) years (Table 1). A total of 77.0% of participants (77/100) graduated from high school or had attended vocational school or college, and 70.0% (70/100) had a household income level of more than VND 100,000,000 (US $4,455.5). Our sample included more cigarette-only smokers (58/100, 58.0%) than dual users (42/100, 42.0%). Participants smoked an average of 15.4 (SD 8.2) CPD. Dual users reported a mean of 11.8 (SD 10.4) waterpipe sessions per day.

Table 1. Sociodemographic and tobacco use characteristics of participants at baseline by study arm.

Characteristic	Total (n=100)	Intervention arm (n=50)	Control arm (n=50)	*P* value
Age (years), mean (SD)	38.9 (8.2)	40.0 (7.5)	37.7 (8.7)	.17
Educational attainment, n (%)	—[a]	—	—	.63
Primary school or less	2 (2.0)	2 (4.0)	0 (0)	—
Middle school	21 (21.0)	10 (20.0)	11 (22.0)	—
High school	38 (38.0)	20 (40.0)	18 (36.0)	—
Vocational school or college	39 (39.0)	18 (36.0)	21 (42.0)	—
Household income level, n (%)	—	—	—	.12
<50,000,000 VND	4 (4.0)	3 (6.0)	1 (2.0)	—
50,000,000-100,000,000 VND	24 (24.0)	15 (30.0)	9 (18.0)	—
>100,000,000 VND	70 (70.0)	30 (60.0)	40 (80.0)	—
Unreported	2 (2.0)	2 (4.0)	0 (0)	—
Type of smoker, n (%)	—	—	—	.31
Cigarette-only smoker	58 (58.0)	32 (64.0)	26 (52.0)	—
Dual user	42 (42.0)	18 (36.0)	24 (48.0)	—
Cigarette consumption per day, mean (SD)	15.4 (8.2)	15.4 (8.0)	15.4 (8.6)	.98
Number of waterpipe sessions per day[b], mean (SD)	11.8 (10.4)	14.3 (12.9)	9.9 (7.7)	.18
Cigarette quit attempt in the past 12 months, n (%)	—	—	—	.41
Yes	36 (36.0)	34 (68.0)	30 (60.0)	—
No	64 (64.0)	16 (32.0)	20 (40.0)	—

[a]Not applicable.

[b]Among dual users only (n=42).

Feasibility

Almost all of those screened were eligible and 99.0% (100/101 eligible participants) enrolled in the study. All participants completed the 6-week and 12-week follow-up surveys. There were no technical issues reported by participants or the SMS vendor.

Engagement

The mean number of days that participants reported using the program was 36.4 (SD 3.4), out of a total of 42 days, in the intervention arm and 36.0 (SD 3.9) for the control arm (Table 2). None of the participants texted the keyword STOP to unsubscribe from the program. Among participants in the intervention arm, 8% (4/50) reset the quit date and 38% (19/50) texted the keyword TIPS to trigger on-demand messages at least once (only the intervention arm has these options). The majority of participants in both arms reported that they always or usually read the text messages.

Table 2. Participant engagement and program acceptability.

Measure	Intervention arm (n=50)	Control arm (n=50)
Engagement		
Number of days used program[a], mean (SD)	36.4 (3.4)	36.0 (3.9)
Ever texted TIPS to the SMS program, n (%)	19 (38)	—[b]
Mean number of times texted TIPS[c], mean (SD)	5.1 (8.1)	—
Ever responded to text assessment, n (%)	36 (72)	—
Mean number of times responded to text assessment[d], mean (SD)	5.3 (4.1)	—
Frequency of reading messages, n (%)		
Always	27 (54)	18 (36)
Usually	14 (28)	22 (44)
Sometimes	9 (18)	10 (20)
Never	0 (0)	0 (0)
Acceptability		
Overall satisfaction with the program[a], n (%)		
Very satisfied	14 (28)	0 (0)
Satisfied	35 (70)	41 (82)
Unsatisfied	1 (2)	0 (0)
Very unsatisfied	0 (0)	9 (18)
Number of messages received from the program[a], n (%)		
Too many	11 (22)	3 (6)
Just right	39 (78)	38 (76)
Too few	0 (0)	9 (18)
Agreed or strongly agreed with the statements, n (%)		
“The text messages helped me quit smoking”	47 (94)	40 (80)
“I learned a lot from using the text program”	48 (96)	36 (72)
“The text program gave me confidence to quit”	43 (86)	40 (80)
“The text messages motivated me to quit smoking”	47 (94)	41 (82)
“Using the text program helped with cravings and triggers”	41 (92)	35 (70)
“Using the text program motivated me to try to quit again if I quit and then started to smoke again”	45 (90)	37 (74)
“I trusted the information in the messages”	49 (98)	—
“The text messages gave me ideas about how to refuse cigarettes offered by others”	41 (92)	—

[a]The program refers to the SMS cessation program for the intervention arm and weekly text assessment for the control arm.

[b]Not applicable.

[c]Among participants who had ever texted keywords to trigger on-demand messages (n=19).

[d]Among participants who had ever responded to text assessment (n=36).

Acceptability

All but one participant in the intervention arm reported being satisfied or very satisfied with the program overall and none reported being very unsatisfied (Table 2). In contrast, none of the participants in the control arm reported being very satisfied and 18% (9/50) reported being very unsatisfied. The majority of participants in both arms perceived the number of messages as just right, however 18% (9/50) of participants in the control arm responded that there were too few compared with none in the intervention arm. Although for both arms there was a high level of agreement that the program increased confidence, was helpful, and increased motivation, participants in the intervention arm consistently expressed higher levels of acceptability across these measures than those in the control arm.

Smoking Abstinence, Quit Attempts, and Reduction in CPDs

The biochemically verified abstinence was higher in the intervention arm than the control arm (6 week: 20% [10/50] vs 2% [1/50]; *P*=.01; 12 week: 12% [6/50] vs 6% [3/50]; *P*=.49), although the difference was not significant at 12 weeks (Table 3). The proportion of participants who reported quit attempts increased from 6 weeks to 12 weeks in both arms, but we observed no difference between the two arms. Similarly, the reduction in CPD increased over time but there was no difference by arm.

Table 3. Abstinence outcomes at 6-week and 12-week follow-up.

Measure	Intervention (n=50)	Control arm (n=50)	*P* value
6-week follow-up			
Biochemically verified abstinence, n (%)	10 (20)	1 (2)	.01
Quit attempt, n (%)	21 (68)	28 (68)	>.99
Reduction in CPD[a] as compared to baseline[b], mean (SD)	9.3 (7.8)	6.8 (5.8)	.13
12-week follow-up			
Biochemically verified abstinence, n (%)	6 (12)	3 (6)	.49
Quit attempt, n (%)	27 (79)	32 (87)	.53
Reduction in CPD as compared to baseline[b], mean (SD)	11.1 (8.3)	9.9 (7.3)	.52

[a]CPD: cigarette consumption per day.

[b]Among participants who reported not quit yet.

Qualitative Findings

The qualitative data supported and expanded on the survey findings for the intervention arm. The main themes that emerged included the overall value of the SMS program, message preferences, perceptions about specific features (eg, timing, bidirectional messaging), and recommendations for enhancing the program. Almost all of the participants liked the SMS program and described it as helpful, primarily because it offered encouragement and enhanced motivation and served as a reminder to stay on track.

> *The messages motivated me and reminded me not to smoke.* [#15, 39 years, male]
>
> *I felt like they [text messages] made me determined to quit.* [#16, 43 years, female]

Messages that included craving management strategies and addressed the harmful effects of smoking were described as particularly useful. A participant noted that he "learned many things…about the ways to overcome cravings" [#1, 54 years, male]. Another explained that messages about the negative consequences of tobacco were "like a warning, helping us understand the danger of smoking and benefits of quitting. So we became conscious and then decided to quit" [#8, 50 years, male].

Reactions to messages that suggested strategies for refusing cigarettes in social situations were mostly positive. One participant noted that those messages "provided the most simple and effective way to refuse invitations to smoke" [#10, 50 years, male]. However, a few participants suggested rewording these in ways that were more consistent with how they communicate with friends and family.

Many of the participants preferred a more tailored approach in terms of message timing.

> *You should send text messages to the relevant time frame of each individual.* [#6, 34 years, male]

One participant wanted the messages to be sent when he had the cravings.

> *You could send the messages at those time.* [#6, 34 years, male]

Participants did have the option to proactively text the keyword TIPS to generate messages during those difficult times, but few routinely used the option.

Participants suggested several additional modifications to enhance the program. Additional content changes including adding more text messages about the health consequences of smoking.

> *I want to receive more text messages on the risks of smoking so my determination in quitting may be stronger.* [#8, 50 years, male]

Participants expressed a strong interest in a more interactive approach.

> *Sometimes I wanted to interact with the person who sent the messages, but I could not do that. The interaction was limited to the responses of yes or no.* [#6, 34 years, male]

Similarly, another suggested allowing users to "ask [text] my own questions" (#4, 51 *years*, male) rather than using the keyword. A few participants suggested adding telephone interactions with a counselor. Last, the majority of smokers suggested extending the intervention duration.

> *I want to receive text messages for a longer time.* [#6, 34 years, male]
>
> *[I]t may be more effective if the duration is longer.* [#11, 33 years, male]

Discussion

Principal Findings

We found support for the feasibility and acceptability of a culturally and linguistically adapted SMS program developed for tobacco users in Vietnam. Smokers overwhelmingly agreed that the messages were helpful, motivated them to quit, and that they would continue to use the program if it was available. A majority of those in the intervention arm also reported that they usually or always read the messages. However, few participants took advantage of the interactive feature that offered them the opportunity to elicit additional support by texting the keyword. Qualitative data suggested that Vietnamese smokers preferred to receive tips as part of the main program and to interact on an as needed basis in a way that would allow them to ask questions and receive tailored responses.

Engagement, defined as the mean number of days that participants used the program, was relatively high compared to previous studies. In the intervention arm, 60% remained in the program for at least 5 weeks. This is in contrast to studies in high income countries that have reported challenges retaining SMS participants [22]. This may be related to the novel nature of this type of intervention in Vietnam. Additional modifications to the program design, as suggested by the participants, may further increase engagement.

The need for significant changes in content and tone of the original message libraries demonstrated the importance of local adaptation in LMICs. As an example, the original message libraries included very few messages about the dangers of tobacco use and instead focused on the benefits of quitting. In contrast, at each stage of development, smokers expressed a preference for more messages that used negative framing of health risks (eg, “If you continue to smoke, your risk of dying from cancer is 25% higher than nonsmokers”). A review of studies that analyzed the impact of emphasizing benefits (gain framed) versus costs (loss framed) found a small advantage of gain versus loss framed messages, but findings were mixed, and these data are based on studies in high income countries [31]. In the process of adapting programs to LMIC contexts, there is an opportunity to continue to explore how messages can be more effectively designed to motivate long-term abstinence. Whether gain or loss framed, an emphasis on health risks seems important in LMIC contexts.

The control arm’s high level of engagement and satisfaction with the program that only included weekly text assessments was an unexpected finding. This may reflect the lack of prior experiences with any smoking cessation services among Vietnamese smokers. Vietnam has not widely disseminated WHO guidelines for integrating routine tobacco use screening and brief advice into the primary care health system [6], and although there is a Quitline, smokers were largely unaware of this resource. Hence, receiving even a weekly text question about their smoking status may have generated the perception that they were receiving tobacco cessation support. Despite their overall satisfaction, compared to the intervention arm, the control arm was less likely to achieve biochemically verified abstinence at 6 weeks. This provides some support for the SMS program’s specific content.

These results should be interpreted as preliminary given the small sample size of this pilot study, as well as the relative brevity of our program as compared to some of the existing SMS cessation interventions [19,20,32,33]. Participants were interested in engaging for longer durations, which could increase abstinence rates over time. However, the results are consistent with one of the few studies conducted in an LMIC. For example, in China, Liao et al [34] conducted a 12-week SMS cessation intervention and reported higher biochemically verified abstinence in intervention groups (high frequency message group: 6.5%; low-frequency message group: 6.0%) than the control group (1.9%). A recent review of mHealth cessation interventions in LMICs concluded that more rigorous studies, with longer follow up and biochemical verification, are needed to further study efficacy in LMIC [13].

Finally, the review by Krishnan et al [13] also suggested the need to compare different characteristics of mHealth cessation interventions. For example, participants requested more message tailoring. Tailoring messages to readiness may increase program effectiveness but findings from studies using this approach are not definitive [15,35]. A few studies outside of LMIC settings have also tested and reported promising findings for strategies that combine SMS programs with interpersonal supports such as peer mentoring from former smokers [16], individual counseling led by professional [35], and counselors’ responses to user composed questions [32]. Adding interpersonal supports may improve user experience and engagement [16,32,35,36] and was requested by our participants. However, the costs of integrating this format into automated SMS programs may reduce the feasibility of scaling programs nationally [37]. New approaches such as the use of automated chatbots may address one of the recommendations from participants to create opportunities for more human interaction with marginal costs. Chatbots offer a conversation interface that can both answer questions posed by the user in a natural language and ask them questions creating a virtual coach experience [38].

Limitations

Our study has limitations. First, participants were recruited from two urban wards in one city. Thus, the findings may not be generalizable to all smokers in Vietnam. In addition, we excluded waterpipe-only smokers. Additional research is needed to explore the value of tailoring mHealth programs to specific types of tobacco users. Second, we did not conduct qualitative interviews with the control arm, which could have provided additional insights into their experience and reasons for the high levels of engagement and acceptability of weekly text assessment. Third, while the study sample was small, a strength of the study was the high retention rates, with all participants completing both follow-up assessments. Last, our study sample included only 2 women, which is consistent with the low smoking rate among Vietnamese women (1.1%) as smoking is less acceptable among women [39,40]. This may limit our ability to generalize to this population of smokers.

Conclusion

Despite these limitations, this study contributes to the small body of literature on mobile phone smoking cessation treatment carried out in LMICs. The data supported the feasibility and acceptability of a culturally adapted SMS cessation treatment program and demonstrated short-term efficacy in promoting abstinence among Vietnamese smokers. While ongoing and future research continues to grow the evidence for effective mHealth cessation programs in a given context, LMICs are beginning to adopt and scale these programs [41]. Therefore, it is equally important to support the design and integration of low-cost monitoring and evaluation systems to guide program modifications that respond to user feedback and sustain and enhance program impact [42].

Acknowledgments

We are grateful for Dr. Paul Krebs' advice and assistance in qualitative data analysis. This study was supported by grant R21CA225852 from the National Cancer Institute (NCI) of the National Institutes of Health. NCI played no role in study design, data collection and analysis, manuscript writing, or the decision to submit the paper for publication.

Authors' Contributions

DRS, NN, and LCA contributed to the conception and design of the study. NN, TN, and HTD collected the data. NJ analyzed qualitative data. NS, CMC, and TN analyzed quantitative data. NJ wrote the first draft of the manuscript. All authors revised the manuscript and approved the final manuscript.

Conflicts of Interest

LCA receives royalties for the sale of Text2Quit and has stock in Welltok Inc. Others declared that there are no conflicts of interests.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2198 KB - mhealth_v9i10e27478_app1.pdf]

References

1. WHO Report on the Global Tobacco Epidemic. Geneva: World Health Organization; 2019. URL: https://apps.who.int/iris/rest/bitstreams/1239531/retrieve [accessed 2021-09-18]
2. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov 28;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
3. Jha P. Avoidable global cancer deaths and total deaths from smoking. Nat Rev Cancer 2009 Sep 20;9(9):655-664. [doi: 10.1038/nrc2703] [Medline: 19693096]
4. Drope J, Schluger N, Cahn Z, Drope J, Hamill S, Islami F. The Tobacco Atlas. Atlanta: American Cancer Society and Vital Strategies; 2018. URL: https://files.tobaccoatlas.org/wp-content/uploads/2018/03/TobaccoAtlas_6thEdition_LoRes.pdf [accessed 2021-09-18]
5. Van Minh H, Giang KB, Ngoc NB, Hai PT, Huyen DTT, Khue LN, et al. Prevalence of tobacco smoking in Vietnam: findings from the Global Adult Tobacco Survey 2015. Int J Public Health 2017 Feb 22;62(Suppl 1):121-129. [doi: 10.1007/s00038-017-0955-8] [Medline: 28229183]
6. WHO Framework Convention on Tobacco Control. Geneva: World Health Organization; 2003. URL: https://apps.who.int/iris/rest/bitstreams/50793/retrieve [accessed 2021-09-18]
7. Global Adult Tobacco Survey (GATS) Viet Nam 2015. Geneva: World Health Organization; 2016. URL: https://www.who.int/tobacco/surveillance/survey/gats/VN-2015_FactSheet_Standalone_E_Oct2016.pdf?ua=1 [accessed 2021-09-18]
8. Kong G, Ells DM, Camenga DR, Krishnan-Sarin S. Text messaging-based smoking cessation intervention: a narrative review. Addict Behav 2014 May;39(5):907-917 [FREE Full text] [doi: 10.1016/j.addbeh.2013.11.024] [Medline: 24462528]
9. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
10. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
11. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]

12. 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 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
13. Krishnan N, Gu J, Abroms LC. Mobile phone-based messaging for tobacco cessation in low and middle-income countries: A systematic review. Addict Behav 2021 Feb;113:106676. [doi: 10.1016/j.addbeh.2020.106676] [Medline: 33038676]
14. Mobile health (mHealth) for tobacco control. Geneva: World Health Organization URL: https://www.who.int/tobacco/mhealth/en/ [accessed 2021-04-06]
15. Kruse G, Park E, Chang Y, Haberer J, Abroms L, Shahid N, et al. Proactively offered text messages and mailed nicotine replacement therapy for smokers in primary care practices: a pilot randomized trial. Nicotine Tob Res 2020 Aug 24;22(9):1509-1514 [FREE Full text] [doi: 10.1093/ntr/ntaa050] [Medline: 32198520]
16. White J, Toussaert S, Thrul J, Bontemps-Jones J, Abroms L, Westmaas J. Peer mentoring and automated text messages for smoking cessation: a randomized pilot trial. Nicotine Tob Res 2020 Mar 16;22(3):371-380. [doi: 10.1093/ntr/ntz047] [Medline: 30892616]
17. Guerriero C, Cairns J, Roberts I, Rodgers A, Whittaker R, Free C. The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop. Eur J Health Econ 2013 Oct 9;14(5):789-797 [FREE Full text] [doi: 10.1007/s10198-012-0424-5] [Medline: 22961230]
18. Cole-Lewis H, Augustson E, 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 Nov;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]
19. 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 Sep;47(3):242-250. [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
20. 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;378(9785):49-55. [doi: 10.1016/S0140-6736(11)60701-0]
21. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 24;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
22. Abroms LC, Ahuja M, Kodl Y, Thaweethai L, Sims J, Winickoff JP, et al. Text2Quit: results from a pilot test of a personalized, interactive mobile health smoking cessation program. J Health Commun 2012;17 Suppl 1:44-53 [FREE Full text] [doi: 10.1080/10810730.2011.649159] [Medline: 22548598]
23. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214 [FREE Full text] [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
24. Shelley D, Nguyen N, Peng C, Chin M, Chang M, Fahs M. Increasing access to evidence-based smoking cessation treatment: effectiveness of a free nicotine patch program among Chinese immigrants. J Immigr Minor Health 2010 Apr 30;12(2):198-205 [FREE Full text] [doi: 10.1007/s10903-008-9194-7] [Medline: 18825498]
25. Muench F, Weiss RA, Kuerbis A, Morgenstern J. Developing a theory driven text messaging intervention for addiction care with user driven content. Psychol Addict Behav 2013 Mar;27(1):315-321 [FREE Full text] [doi: 10.1037/a0029963] [Medline: 22963375]
26. Lewins A, Silver C. Using Software in Qualitative Research: A Step-by-Step Guide. London: SAGE Publications; 2007.
27. Saldaña J. Coding Manual for Qualitative Researchers. London: SAGE Publications; 2009.
28. Ybarra M, Bağci Bosi AT, Korchmaros J, Emri S. A text messaging-based smoking cessation program for adult smokers: randomized controlled trial. J Med Internet Res 2012 Dec 27;14(6):e172 [FREE Full text] [doi: 10.2196/jmir.2231] [Medline: 23271159]
29. Bailey SR, Bryson SW, Killen JD. Predicting successful 24-hr quit attempt in a smoking cessation intervention. Nicotine Tob Res 2011 Nov 10;13(11):1092-1097 [FREE Full text] [doi: 10.1093/ntr/ntr151] [Medline: 21832272]
30. The R project for statistical computing. URL: https://www.r-project.org/ [accessed 2021-09-18]
31. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [doi: 10.1037/0033-2909.121.1.3] [Medline: 9000890]
32. Cupertino AP, Cartujano-Barrera F, Ramírez M, Rodríguez-Bolaños R, Thrasher JF, Pérez-Rubio G, et al. A mobile smoking cessation intervention for Mexico (Vive sin Tabaco... ¡Decídete!): single-arm pilot study. JMIR Mhealth Uhealth 2019 Apr 25;7(4):e12482 [FREE Full text] [doi: 10.2196/12482] [Medline: 31021326]
33. Liao Y, Wu Q, Tang J, Zhang F, Wang X, Qi C, et al. The efficacy of mobile phone-based text message interventions ('Happy Quit') for smoking cessation in China. BMC Public Health 2016 Aug 19;16(1):833 [FREE Full text] [doi: 10.1186/s12889-016-3528-5] [Medline: 27543164]
34. Liao Y, Wu Q, Kelly BC, Zhang F, Tang Y, Wang Q, et al. Effectiveness of a text-messaging-based smoking cessation intervention ("Happy Quit") for smoking cessation in China: A randomized controlled trial. PLoS Med 2018 Dec 18;15(12):e1002713 [FREE Full text] [doi: 10.1371/journal.pmed.1002713] [Medline: 30562352]

35. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul;1(2):e17 [FREE Full text] [doi: 10.2196/mhealth.2522] [Medline: 25098502]
36. Wang MP, Luk TT, Wu Y, Li WH, Cheung DY, Kwong AC, et al. Chat-based instant messaging support integrated with brief interventions for smoking cessation: a community-based, pragmatic, cluster-randomised controlled trial. Lancet Digital Health 2019 Aug;1(4):e183-e192. [doi: 10.1016/s2589-7500(19)30082-2]
37. Ybarra ML, Holtrop JS, Bağci BAT, Emri S. Design considerations in developing a text messaging program aimed at smoking cessation. J Med Internet Res 2012;14(4):e103 [FREE Full text] [doi: 10.2196/jmir.2061] [Medline: 22832182]
38. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
39. Morrow M, Ngoc DH, Hoang TT, Trinh TH. Smoking and young women in Vietnam: the influence of normative gender roles. Soc Sci Med 2002 Aug;55(4):681-690. [doi: 10.1016/s0277-9536(01)00310-0]
40. Global Adult Tobacco Survey: fact sheet Viet Nam 2015. 2016. URL: http://www.who.int/tobacco/surveillance/survey/gats/VN-2015_FactSheet_Standalone_E_Oct2016.pdf?ua=1 [accessed 2017-01-13]
41. Gopinathan P, Kaur J, Joshi S, Prasad VM, Pujari S, Panda P, et al. Self-reported quit rates and quit attempts among subscribers of a mobile text messaging-based tobacco cessation programme in India. BMJ Innov 2018 Jul 25;4(4):147-154. [doi: 10.1136/bmjinnov-2018-000285]
42. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. Geneva: World Health Organization; 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-09-18]

Abbreviations

CPD: cigarette consumption per day
LMIC: low- and middle-income country
mHealth: mobile health
NCI: National Cancer Institute
RCT: randomized controlled trial
SMS: short message service
WHO: World Health Organization

Edited by L Buis; submitted 26.01.21; peer-reviewed by G Kruse, R Rosen, R Jayasuriya; comments to author 08.03.21; revised version received 03.05.21; accepted 03.09.21; published 08.10.21.

Please cite as:
Jiang N, Nguyen N, Siman N, Cleland CM, Nguyen T, Doan HT, Abroms LC, Shelley DR
Adaptation and Assessment of a Text Messaging Smoking Cessation Intervention in Vietnam: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(10):e27478
URL: https://mhealth.jmir.org/2021/10/e27478
doi:10.2196/27478
PMID:34623318

Original Paper

Integration of mHealth Information and Communication Technologies Into the Clinical Settings of Hospitals in Sub-Saharan Africa: Qualitative Study

Oluwamayowa Oaikhena Ogundaini[1*], BSc, MTech, PhD; Retha de la Harpe[2], DTech; Nyx McLean[1*], PhD

[1]Department of Information Technology, Faculty of Informatics and Design, Cape Peninsula University of Technology, Cape Town, South Africa
[2]Graduate Centre for Management, Faculty of Business and Management Services, Cape Peninsula University of Technology, Cape Town, South Africa
[*]these authors contributed equally

Corresponding Author:
Oluwamayowa Oaikhena Ogundaini, BSc, MTech, PhD
Department of Information Technology
Faculty of Informatics and Design
Cape Peninsula University of Technology
District Six Campus
Hanover Street, 7925
Cape Town, 8000
South Africa
Phone: 27 735989341
Email: ogundainio@cput.ac.za

Abstract

Background: There is a rapid uptake of mobile-enabled technologies in lower- and upper-middle–income countries because of its portability, ability to reduce mobility, and facilitation of communication. However, there is limited empirical evidence on the usefulness of mobile health (mHealth) information and communication technologies (ICTs) to address constraints associated with the work activities of health care professionals at points of care in hospital settings.

Objective: This study aims to explore opportunities for integrating mHealth ICTs into the work activities of health care professionals at points of care in clinical settings of hospitals in Sub-Saharan Africa. Thus, the research question is, “How can mHealth ICTs be integrated into the work activities of health care professionals at points of care in hospital settings?”

Methods: A qualitative approach was adopted to understand the work activities and points at which mHealth ICTs could be integrated to support health care professionals. The techniques of inquiry were semistructured interviews and co-design activities. These techniques were used to ensure the participation of frontline end users and determine how mHealth ICTs could be integrated into the point of care in hospital settings. Purposive and snowball sampling techniques were used to select tertiary hospitals and participants for this study from South Africa and Nigeria. A total of 19 participants, including physicians, nurses, and hospital managers, were engaged in the study. Ethical clearance was granted by the University research committee and the respective hospitals. The data collected were sorted and interpreted using thematic analysis and Activity Analysis and Development model.

Results: The findings show that mHealth ICTs are suitable at points where health care professionals consult with patients in the hospital clinics, remote communication is needed, and management of referrals and report writing are required. It was inferred that mHealth ICTs could be negatively disruptive, and some participants perceived the use of mobile devices while engaging with patients as unprofessional. These findings were informed by the outcomes of the interplay between human attributes and technology capabilities during the transformation of the motives of work activity into the intended goal, which is enhanced service delivery.

Conclusions: The opportunities to integrate mHealth ICTs into clinical settings depend on the inefficiencies of interaction moments experienced by health care professionals at points of care during patient consultation, remote communication, referrals, and report writing. Thus, the timeliness of mHealth ICTs to address constraints experienced by health care professionals during work activities should take into consideration the type of work activity and the contextual factors that may result in contradictions in relation to technology features. This study contributes toward the design of mHealth ICTs by industry vendors and its usability evaluation for the work activity outcomes of health care professionals.

(JMIR Mhealth Uhealth 2021;9(10):e26358) doi:10.2196/26358

KEYWORDS

mHealth; health care professionals; co-design; hospitals; ActAD model; work activity; Sub-Saharan Africa; referrals; VULA mobile app; WhatsApp; mobile phone

Introduction

Background

The emergence of mobile health (mHealth) technologies originated from a need to provide support for health care professionals, workers, and patients to deliver and receive services efficiently at any time and anywhere. The concept of mHealth transcends beyond delocalized delivery of health care services. It includes the use of portable technologies (and sometimes enabled wireless communication) as a means of facilitating the processing and provision of timely information on the management of health care [1,2]. Thus, mHealth is referred to as the use of mobile technologies such as smartphones, tablets, and sensor-enabled hand-held devices for health-related purposes [3]. However, similar to other digital systems, there are challenges that hinder the effective use of mHealth information and communication technologies (ICTs) to support tasks within the work activities of health care professionals at points of care [4,5]. Health care professionals represent a skilled workforce that includes specialist physicians and nurses who carry out clinical work activities to provide care services at the point of care. Point of care refers to locations within clinical settings where health care professionals perform tasks that encompass the management of patients' health and service delivery [6]. Thus, this paper explores how mHealth ICTs can be integrated into the work activities of health care professionals at the point of care in hospital settings to support the delivery of safe and quality care.

The existing literature has alluded to the claims that the outcome of technology-enabled work activities is influenced by the ability of health care professionals to make timely and informed decisions that ensure the quality of service delivery and, ultimately, improve patient well-being [7,8]. However, when adopted and designed without adequate consideration of the work activities executed by health care professionals, implemented health ICTs, including mHealth apps, are used partially or eventually discarded [9]. Although there is an increasing global penetration of mobile communication and devices in every other sector, such as commerce and banking, the application of mHealth ICTs by health care professionals at point of care in hospital settings is rarely explored, and the impact on service delivery is unclear [5,10,11].

The use of mHealth ICTs by health care professionals at the point of care has potential benefits. Owing to the information-intensive nature of health care services and the dependence of timely decision-making on up-to-date records, desktop computers may restrict the access, retrieval, and exchange of information to specific locations in hospital settings. Conversely, a major advantage of mHealth ICTs is that it enables real-time access to patient information and communication at different points of care and does not restrict the mobility of health care professionals [5]. It reduces time- and location-related constraints that may hinder health care professionals from quickly retrieving patient information for timely decision-making, especially where access to computer-based records is limited [12]. In case of emergencies, the use of mHealth ICTs provides a platform where health care professionals can easily verify clinical guidelines and related prescriptions in academic journals to make informed decisions and monitor patient vitals [6,13,14]. Ideally, mHealth ICTs are beneficial for addressing the shortcomings associated with timely retrieval of information and communication at point of care to ensure job and patient satisfaction after health care service delivery.

There are examples of initiatives in which mHealth ICTs have been developed to improve the care delivered by health care professionals to patients. Most of the initiatives are focused on specific public health–related health challenges for data collection by community- and home-based health care workers [15,16]. For example, MomConnect is built for affordable mobile handsets, with support for unstructured supplementary service data, SMS text messages, and voice communication. It offers messaging services that enable pregnant women to interact with health care providers through SMS text messages, asking questions, and submitting complaints to help improve their health as well as the health of their babies [17]. The Clinical Patient Administration Kit is a mobile-enabled electronic medical record tablet that assists clinicians in facilitating the tracking and reporting of treatment and outcomes in maternal and infant care services [18]. Clinical Patient Administration Kit improved the tasks performed by clinicians by eliminating repetition of tasks and records and providing timely care to patients at clinics.

It was drawn from the literature that treatments and well-being of patients have been extended out of the hospital to homes as a result of the access to and ownership of mobile technologies; however, there are limited studies on how tasks within work activities inform the use of mHealth ICTs by health care professionals at point of care in hospital settings. Other studies have investigated the attitudes and behaviors of health care professionals toward the acceptance and use of mHealth ICTs for its benefits [2,19-22]. In contrast to its benefits, there are concerns that mHealth ICTs could inhibit health care professionals when the nature of tasks performed within work activities is not adequately understood or considered.

Concerns have been raised by some authors regarding the unintended consequences associated with the use of mHealth ICTs and related applications within health care settings. In this paper, the authors refer to unintended consequences as contradictions that result from the tensions between sociotechnical interactions [23]. Unintended consequences can be either desirable or undesirable. For example, Watson et al [24] shared their views that despite its usefulness, WhatsApp (Facebook Inc) is not a secure communication channel for the

exchange of patient clinical data because of weak encryption and the vulnerability of information during transmission. Therefore, it does not satisfy regulatory standards for use in health care [11]. In addition, Wallis et al [11] stated that mHealth ICTs are mostly not designed to interoperate with third-party apps or legacy systems. As a result, this causes fragmentation of data within the health care system. Other concerns highlighted by authors include the inhibition of workflow [7,25], time inefficiency [26], and influence on the patient-physician relationship [27].

Activity Analysis and Development Model

This paper explores opportunities for integrating mHealth ICTs into the work activities of health care professionals at point of care using the Activity Analysis and Development (ActAD) model. The argument of this paper is that understanding the interaction between the elements of a work activity presents an opportunity for the integration of mHealth ICTs to enable the work activities of health care professionals at point of care and identify the contradictions that may influence unintended consequences associated with technology-enabled work activities.

The ActAD model is a theoretical approach used to describe the elements of a work activity and to gain insights on how their interactions inform the development of information systems [28]. A work activity comprises actors (individuals or groups) that use tools (means) to perform actions on an object that could be shared as informed by the motive of the activity and guided by a set of rules [29]. Ultimately, the motive of the activity is transformed into a goal or an intended outcome.

In the context of this study, work activities are defined as the set of actions performed by health care professionals according to their medical practice toward delivering health care services. Actions may include collection, access, and retrieval of information; sharing and exchange of information; communication between health care professionals; and use of the information to make informed decisions about patients' health and well-being. The actions can be enabled by means of action that would be mHealth ICTs in the context of this study. The object of activity is the patient information or records, whereas the motive of the activity is to make an informed decision. Patient information could be a *shared object of networking* or *shared object of action*, depending on how it serves a group of health care professionals. Patient information enables health care professionals to make informed decisions on the state of well-being of patients to improve their health conditions. Ultimately, the motive of the activity is transformed into the intended goal or outcome of the activity.

Next, the research strategy and methods used to engage participants in exploring opportunities to integrate mHealth ICTs into the work activities of health care professionals at point of care are discussed, followed by data analysis results and interpretation of the research findings.

Methods

Qualitative Approach

The study was exploratory, as informed by the research objectives. The researchers adopted an interpretivist standpoint because of the subjective nature of socially constructed realities of different health care professionals who use (or do not use) mHealth ICTs for their work activities at the point of care. A qualitative strategy, including multiple techniques, specifically semistructured interviews and co-design activities, was used to engage the participants.

According to Myers and Newman [30], semistructured interviews enable the use of open-ended questions to gain insights and to afford opportunities for new areas to emerge for further exploration of a research problem. The use of co-design activities enables the users or potential users of a product or service to express their lived experiences and expectations visually, such that a researcher or designer can gain further insights into the use contexts [31].

The use of multiple techniques enabled researchers to draw inferences from the thoughts, experiences, and expectations of participants to understand the work activities of the health care service delivery process. More importantly, the techniques helped to understand the service interactions between health care professionals and technology-enabled work activities in a bid to identify opportunities to integrate mHealth ICTs into point of care hospital settings.

Participants were selected using purposive and snowball techniques. The 2 are nonrandom sampling techniques applied to select a sample size from a research population based on the biased judgment of the researcher in line with a research objective [32]. These techniques were applied to identify health care professionals who could provide a detailed account of daily clinical work activities and the tools used to enable specific actions within the process of care delivery. The snowball technique was adopted because of the difficulty in reaching available physicians who were willing to participate in the study because of their busy schedule. Thus, participants suggested some of their colleagues that could describe how technologies are currently being used and opportunities to integrate mHealth ICTs at point of care.

Recruitment of Participants

Overview

A study by Abyaomi et al [22] on the factors that enable or inhibit the behavioral intention of physicians to use clinical informatics selected Nigeria and South Africa as leading economic powerhouses in Africa. According to a study on e-readiness of African countries conducted by Ifinedo [33], South Africa and Nigeria ranked within the top 5 countries in Sub-Saharan Africa with emerging technological innovation for a networked economy.

Health and the quality of health care service delivery contribute to the skills and well-being of human resources required for developing and sustaining economies. In this paper, the researchers considered Nigeria and South Africa to be 2

countries with comparable economic growth and are ideal for exploring the opportunities to integrate mHealth ICTs into the work activities of health care professionals to address the challenges that inhibit job satisfaction and quality health care service delivery.

Clinical Settings

The study was conducted in 2 tertiary hospitals, one in the Western Cape Province of South Africa and in the South Western geopolitical region of Nigeria because they provide specialized clinical care. The researchers purposively chose to investigate the workflow of health care professionals in clinical departments where activities are technology-enabled to ensure that the participants selected align with the aim of the study. For example, the authors considered clinical departments that deal with medical imaging and reporting where executing work activities involves the use of electronic systems and digitized records.

The 2 contexts comprised contrasting situations. In one of the tertiary hospitals, the physicians were using their smartphones and health ICTs, including a mobile app called VULA created by Dr William Mapham, whereas health care professionals in the other hospital largely used a paper-based system at points of care. In both tertiary hospitals, WhatsApp is used as a means of facilitating communication and sharing patient information.

WhatsApp is a consumer-oriented mobile app that can be installed to facilitate instant messaging in the form of text, pictures, audio, or video between two or multiple internet-enabled devices, including smartphones [34,35]. It offers a platform for health care professionals to exchange health-related information and communicate with each other or to follow up with patients on the status of their well-being [36]. WhatsApp could be interpreted as an improved version of the two-way pager used by health care professionals, which saves time, requires no computer, and limits communication barriers between colleagues.

Similar to WhatsApp, the VULA mobile app was designed to facilitate remote communication between health care professionals and workers as well as to manage referrals at points of care [37]. The VULA app allows health workers to seek advice and receive training from their more experienced colleagues or specialist physicians. The VULA app is an established referral platform used in the Western Cape Province of South Africa by clinical units, including orthopedics and ophthalmology, as a point of entry to facilitate the exchange of medical images and instant messaging related to real-time diagnostic and treatment consultation [38].

As described in Table 1, a total of 19 participants were engaged in this study and identified through purposive and snowball sampling techniques. In the tertiary hospital selected in South Africa, participants included 2 ophthalmology specialist physicians; 6 orthopedic specialist physicians and 3 nurses who manage trauma care wards, intensive care units and theater wards; and the deputy nursing manager of the hospital.

Table 1. Sampled participants and techniques of engagement.

Pseudocode for participants	Clinical department	Technique of engagement
oph1_H1[a], oph2_H1	Ophthalmology	Face-to-face interviews at hospital H1 in the physicians' library
ort1_H1, ort2_H1, ort3_H1, ort4_H1, ort5_H1, ort6_H1	Orthopedic	Face-to-face interviews and co-design activity at the physicians' library
Deputy nursing manager (dnm_H1), trauma unit operational manager (tn_H1), theater operational manager (ton_H1), intensive care unit operational manager (icn_H1)	Nursing	Face-to-face interviews and co-design activities at the nursing board room
RR1_H2[b], RR2_H2	Radiology	Face-to-face interviews at hospital H2 in the physicians' consulting room and co-design activities
ortC_H2	Orthopedic	Face-to-face interviews and co-design activity at hospital H2 in the physicians' consulting room
General surgeon (gs_H2); neurosurgeon (ns_H2)	Surgery	Face-to-face interviews and co-design activity at hospital H2 in the physicians' consulting room
Assistant director of nursing services (adn_H2); Orthopedic nurse (on_H2)	Nursing	Face-to-face interviews and co-design activity at hospital H2 in the physicians' consulting room

[a]H1: South African hospital.

[b]H2: Nigerian hospital.

In the Nigerian tertiary hospital, participants were 2 radiologist physicians, 1 orthopedic nurse, 1 assistant nursing director in radiology, 1 orthopedic consultant, 1 neurosurgery consultant, and 1 general surgery consultant. A common drawback in the 2 contexts is that physicians were mostly unable to honor appointments because of their busy schedules. This is a common phenomenon reported in a study by Kabanda and Rother [39].

Data Collection Process

Data were collected from both tertiary hospitals from September 2018-January 2019. Hospital managers were the first point of contact at the hospitals. The hospital managers directed the heads of clinical departments that deal with medical imaging and reporting to assist experienced physicians and nurses who were interested and willing to participate in the study. The

interested participants were contacted by email, and appointments were set up based on their convenience and subject to their availability.

Two sets of data were collected from the participants. The interviews and co-design activities took place within the hospitals' premises as convenient for each participant. Each interview lasted an average of 40 minutes, whereas each of the co-design activities lasted an average of 60 minutes. The participants were allowed time to express themselves as they could, whereas the sessions were recorded with a portable voice recorder.

During each co-design activity, one of the researchers assumed the role of a facilitator to incorporate their knowledge of the issues of investigation and the research design procedure to encourage participants' active participation. The motivation for using a co-design activity was to comply with ethical concerns and avoid any disruptions that are associated with directly observing health care professionals' workflow in real time.

In preparation for the co-design activities, a set of paper cut-out graphic representations of physicians, nurses, and the tools that portray health care work activities, as indicated in the literature, were provided to participants. Cut-out representations of actors and tools provide an opportunity for participants to visually express and map their actions and experiences to researchers [40]. In addition, large pieces of paper, pencils, erasers, and stickers were provided to health care professionals as writing materials to illustrate the directions of their workflow.

The participants performed three tasks during the coactivity sessions. The first task was for participants to visually illustrate the start of their daily work activities, how they perform their tasks, and the tools used, if any. Subsequently, the participants used the cut-outs to represent themselves, as actors, on a large piece of paper using the stickers, thereby giving practical descriptions of the actions undertaken at the point of care. The outcome of the first task was a visual illustration of the workflow or user journeys of health care professionals from the first encounter with a patient until the patient is discharged, transferred, or deceased.

For the second task, the researcher used visual illustrations generated by participants to identify the touchpoints within the user journeys. In this paper, touchpoints are described as contact points where a network of human actors and objects interact to produce or influence an outcome [41]. By identifying the touchpoints, the researcher was able to determine the interaction moments of tasks performed by health care professionals during work activities. The interaction moments could be human-to-technology, human-to-human, or machine-to-human actions.

The researcher then probed using open-ended questions to determine whether there were any challenges experienced by health care professionals at any point or while using health ICTs within the visual illustration generated. The outcome of this task produced the challenges, unintended consequences, and the resulting effects attributed to the use of health ICTs by health care professionals.

For the third task, the outcomes of the second task were used to facilitate a discussion between the researcher and participants on how mHealth ICTs could be integrated into their work activities at the point of care. The outcome of the third task presented an opportunity to identify the characteristics of the expected features that a fit-for-purpose technology could have to enable health care professionals' work activities. This marked the end of the co-design activities.

Ultimately, the data collected from the semistructured interviews were used to inform and complement the coactivity sessions used to engage with physicians and nurses. By doing so, the researcher was able to validate what health care professionals said against their actions. Both techniques enabled the researcher to identify opportunities to integrate mHealth ICTs into the work activities of health care professionals at the point of care in a hospital setting.

Ethical Considerations

To ensure that the research was executed in a manner that guarantees safety and harm-free procedures to the participants, the researchers, and the environment, an ethical clearance was obtained from all relevant authorities. The authorities include, first, the University's research ethics committee, then the researcher applied to the Western Cape Provincial department of health, South Africa, and the tertiary hospital in Nigeria. Permission was granted on the condition that the researchers adhered to the delineation and safety measures highlighted in the ethics clearance application.

An information sheet was provided to each of the individual participants explaining that participation was voluntary, and consent could be withdrawn if they felt that the questions being asked were uncomfortable. In other words, participation was not incentivized. The participants signed a consent form that allowed interview and co-design sessions to be recorded, and they were aware that the information collected would be used in a confidential manner and responses would not be misconstrued. For security purposes, the data obtained from the participants were stowed away in a password-protected folder on a secured desktop computer at all times.

Data Analysis Process

Thematic analysis was used to sort and organize qualitative data. The process of thematic analysis enabled the researcher to determine the frequency of attributes to present the findings [42]. First, the recorded data collected from participants were transcribed from audio to verbatim text to allow the researcher to easily categorize the data into themes. To group the data into themes, the researcher used descriptive codes to tag the emerging attributes, which are words or phrases that imply the issues being investigated, and then categorize the codes into themes [43]. The processes through which attributes emerge are identified were conceptualization and operationalization.

Conceptualization refers to defining the keywords of a research question or objective, whereas operationalization is used to identify the attributes that characterize the keywords through a process of coding [32]. The coding process involved assigning a descriptive word to identify points where mHealth ICTs were being used and could be integrated into the work activities of

health care professionals at the point of care. Ultimately, this enabled the researcher to sort and organize the data collected from the interviews and the co-design activity transcripts. The themes were informed by the elements of the ActAD model, as shown in Textbox 1.

Textbox 1. Themes and findings generated from the categorizing of transcript codes.

Research objective

- Explore opportunities for integrating mobile health (mHealth) information and communication technologies into the work activities of health care professionals at point of care

Themes

- Actions: Work activities are patient consultations, remote referrals, and formulating treatment plans
- Tools: mobile apps—VULA and WhatsApp
- Outcomes of the work activities: this attributed to the use of mobile apps for work activities
- Unintended consequences associated with the use of mobile apps for work activities
- Transformation process: alignment of mHealth apps to work activities of health care professionals

Findings

- The nature of work activities is routinized and require health care professionals to have timely access to updated patient information and to reduce their mobility between point of care.
- VULA and WhatsApp are mobile apps used by physicians to facilitate communication, exchange patient information, and enable remote consultations.
- Other hospital information systems are used to facilitate the retrieval of patient records (objects of activity) and booking of clinical examinations to diagnose and make informed decisions (motive of work activity).
- VULA has simplified the process of referrals and reduced unnecessary referrals.
- VULA enables accountability of the referral process.
- VULA and WhatsApp have improved how physicians communicate and collaborate with each other remotely.
- VULA referral notifications interrupt physicians during patient consultation multiple times.
- Responding to VULA referral notifications is time consuming.
- Use of mobile apps by physicians at point of care might be perceived by patients as unprofessional.
- Use of mobile apps at point of care could enable distractions from the work activities.
- VULA mobile app is suitable for referrals, retrieval, and the exchange of health-related information.
- VULA mobile app requires a push notification of referrals to health care professionals that are available while on call.
- mHealth app suitable for patient consultation may include a voice recorder with speech recognition to capture and transcribe verbal communication.
- Other features may include high-resolution cameras and web-enabled and instant messaging to reduce interaction time at point of care.

Results

Overview

The results were presented using the elements of the ActAD model to deductively describe how mHealth ICTs could be integrated into the work activities of health care professionals, as informed by the interaction moments of tasks and outcomes. Insights were drawn from the interviews and co-design activities used to engage with all participants. The analysis showed that patient consultation, referral management, treatment planning, and report writing were the main work activities executed by health care professionals.

Consultation With Patients

Overview

The physicians indicated that their first point of contact with patients in the hospital was by the outpatient clinics where they had to access and retrieve patient records or engage with their patients to collect a history record.

During this activity, physicians use hospital information systems on computer desktops, mobile phones, and paper. The interaction moments during patient consultation include verbal communication, taking notes, retrieving patient records, and requesting clinical examination:

> *In our clinics, for all patients that are seen notes are also made by hand and those notes as well as all referrals goes into a patient's folder. All those notes are sent to the scanning department. And all notes*

> *get scanned into our ECM. All those notes eventually do become available on a computer.* [oph1_H1]

On the basis of the accounts of specialist physicians, it is evident that work activities at point of care are supported using technology. The electronic content manager is used to store patients' records electronically. Another physician explained as follows:

> *VULA is a smartphone based app where doctors and healthcare professionals, that includes more than just doctors, can have direct communication with on-call doctor or Orthopaedic person, to ask for advice and or refer patients by a means of a list of questions as well as photos of x-rays that can be sent through to us.* [ort2_H1]

In hospital H2, physicians and nurses stated that they used their mobile phones at point of care but did not use any forms of hospital information systems or mHealth ICTs. However, they were aware of the benefits associated with technology-enabled work activities at point of care:

> *During call hours we make use of our mobile phones and some social media apps: WhatsApp, Snapchat. Now, this we use mostly to send images, radiological images and even some laboratory reports. So, a lot of times if a resident is reviewing a patient somewhere and is conversing over the phone, a lot of times I tell them to send images on my phone.* [ortC_H2]

> *At the clinic, doctors and the nurses make use of paper to take the record of the patients and when they are through...to assess the patient, then all the records are documented.* [on_H2]

It can be inferred that when mobile phone apps are used at point of care, it serves as a quicker means to facilitate communication and easier information exchange between health care professionals. The use of mobile phones provides an opportunity for physicians and nurses to capture information digitally, thereby reducing the amount of time and papers used manually. Thus, there is an opportunity to improve the outcome of tasks performed during consultation with patients or when seeking medical advice remotely.

Outcome of Consultation Enabled by mHealth ICT

Health care professionals are able to communicate directly with each other to view necessary medical images remotely and to save time on patient diagnosis and treatment decision-making. Ultimately, VULA mobile apps have reduced both the time and cost implications associated with clinical decision-making by health care professionals. Physicians explained as follows:

> *When referring a patient with VULA application the referring doctor must give a lot of important information about the patient to us that include a photo of the eye and then we can have a conversation with them to get a better idea of the problem. This way of referring is much more comprehensive that a telephonic referral, and we can get a better idea of the problem.* [oph2_H1]

> *VULA makes it easier to make clinical decisions without seeing patients personally. It's often difficult to determine if a patient can be seen as an outpatient or do they need urgent referral, and using the VULA app with clinical pictures helps with this decision making.* [ort4_H1]

Technology-enabled referrals afford physicians the opportunity to request detailed information from their colleagues about patients before making informed decisions, without the need for a face-to-face consultation. One of the physicians in hospital H2 alluded to the importance of technology-enabled work activities:

> *Digitising would really help a lot, then I will not have to be going on all around. That will help a lot to reduce unnecessary waiting time. Many times, we are at the clinic, we have to be waiting for folders to arrive before we start.* [ns_H2]

The responses show that VULA app as an example of mHealth ICTs is useful for the work of health care professionals. The usefulness is evident in how mobile devices can reduce the time taken to access and retrieve patient records remotely at any time or at different point of care. mHealth ICTs provide a platform for conducting extended investigations through the use of instant messaging. Despite the benefits associated with mHealth apps, there is a concern about VULA notifications being disruptive to work activities while physicians are attending to patients at point of care during consultations.

Disruption by Mobile Referral Notifications

Physicians in hospital H1 revealed that during patient consultations, they received VULA referral notifications on their mobile phones. The referral notifications are received multiple times and from multiple sources, and they cause distractions to their work activity at point of care:

> *During the day it actually interferes and it slows you down massively. Definitely because you have a lot of patients that you need to see, you need to answer the phone; you need to answer your bleeps and then you also get VULA referrals. And some of them—I would say half of them, aren't emergencies and those people want an answer now because they've got a patient sitting in front of them but I've got 50 patients sitting outside.* [oph1_H1]

It is evident that there are contradictions that result from a lack of understanding of how the tasks within work activities of health care professionals are carried out at the point of care. Thus, the design of VULA mobile app influences how mHealth ICTs are experienced at the point of care:

> *I feel that it seems unprofessional to constantly be looking at your phone screen whilst consulting patients.* [ort5_H1]

According to ortC_H2, mobile devices are used to access academic journals while teaching:

> *These days we tend to use our mobile device in clinic, but really we should have a laptop...if you see patients most times, you have residents and students. You can*

> *check up journals' abstracts to compare and to show them some images and pictures.* [ortC_H2]

The physician further mentioned that mobile devices are seldom used in wards to facilitate teaching of medical students "because they believe it tends to be a distraction." The response indicates that the nature of work activities should be adequately considered before designing mHealth ICTs relevant to tasks performed at points of care.

Referral Management

The use of mHealth ICTs enables accountability by tracking the referral source and receiving comprehensive feedback on consultations. As established earlier, the VULA app enables health care professionals to remotely communicate and access real-time information (including text and images). The motive of activity in the instance of patient referrals enabled by VULA app as a means of action is transformed into the intended goal of activity by health care professionals:

> *There is a massive advantage. So, firstly is that I know who is sending me the referrals, I know where the referral is coming from. One thing we have picked up with regards to where referral is coming from is sometimes the name of the referring healthcare worker comes up but also their location. But that location is what they put in. It's not a location that is picked up by the smartphone...so it doesn't give you a location at that time but it's a location is filled in.* [ort2_H1]

At hospital H2, one of the specialist physicians indicated challenges with the use of paper folders:

> *Many times, you don't find many folders when you see patients. For example, that shows that a lot of them are misplaced somewhere or maybe didn't find them properly at the record...some claim they are taken by one doctor or another...* [ns_H2]

Accountability is attributed to the use of the VULA appl. However, there are instances of inconsistency with geotagged locations and locations filled by the referring health care professional. This raises concerns when health care providers attempt to locate the exact institution of the referring health care personnel and the particular area where a patient is likely to be referred from. In addition, where paper folders are being used at point of care, there are no adequate accountability measures to track location or movement.

Collaborative Treatment Planning

Health care professionals, including specialist physicians and nurses, use their personal phones to make calls and use WhatsApp as a means of communication and coordination to plan care delivery. Furthermore, having a personal mobile phone helps health care professionals locate their colleagues, especially when the official communication medium (switchboard bleep) takes a longer time or stops being functional:

> *Within the department, everyone's got a work group on WhatsApp, so that's to communicate about meetings. And then sometimes you just pick up your own phone...if switchboard takes too long to speak to someone...unless you don't have the doctor's number then you have to wait for switchboard. I use my phone actually; we do use our phones a lot.* [oph1_H1]

In hospital H2, one physician explained as follows:

> *There is disjointed management of patients, where a doctor comes and does his job and the nurse is not there; and then a physiotherapist comes and everyone sees the patient disjointedly. There is no team approach to patient care.* [ortC_H2]

The convenience of having a means of action such as mobile technology and an mHealth ICTs app facilitates quicker communication between health care professionals and eliminates any restrictions associated with mobility between point of care in a hospital setting. Thus, the communication gap during collaborative care coordination is bridged.

Report Writing and Administering of Care

The responses indicate that nursing activities can be categorized into administering of care to patients and report writing while assisting physicians in facilitating clinical examination:

> *Some documentation gets lost during movement of patients. It is time consuming because whatever procedure is completed, the nurse got to come down and do the right thing afterwards. The paper, it's illegible many times due to the fact that peoples' handwriting differ.* [dnm_H1]

The nursing staff indicated that report writing was a major part of their nursing activities and required a lot of paper, as they were required to accompany patients to and from the different point of care. A shared sentiment among nursing staff who participated in the interview and co-design activity is that continuous use of paper can be addressed by introducing mHealth ICTs designed to facilitate report writing and documentation of patient care on-the-move:

> *From the points of assessment—casualty around the ward, from the point of assessment we need a gadget that can take patients' parameters. It will be better if it's mobile, since this would allow easier movement and should be interconnected to all wards, clinics so that at a click, it sends the information concerning the patient to other nurses at the next points-of-care.* [ons_H2]

The findings drawn from the *Results* section indicate that the opportunities to use a means of action such as VULA and WhatsApp by health care professionals at point of care depend on how the technology enables the motive for performing an activity is eventually transformed into the intended goal of an activity. Otherwise, contradictions as a result of a lack of mHealth ICTs *fit* could mediate how the motive of activity results into an unintended outcome.

Discussion

Principal Findings

On the basis of the accounts of most health care professionals, the use of mHealth ICTs for work-related activities has mostly

yielded and would ensure positive outcomes at the organizational and individual levels. For example, physicians in particular use hospital information systems to facilitate easier access to, and retrieval of, patient records electronically during patient visits. In this study, the use of VULA mobile app for referrals simplified the referral process and reduced the unnecessary referrals to the hospital. The findings support the arguments by O'Connor et al [5] that the use of mHealth ICTs can enhance the task performance of health care professionals at point of care. In terms of communication and the writing of clinical notes, physicians can use mobile smart devices to record their conversations to obtain a comprehensive history from patients [44]. Ultimately, the relevance of mHealth ICTs in clinical settings is evident in information access, exchange, and communication.

WhatsApp was indicated as a means used by most physicians to communicate and exchange information. According to mHealth studies by Wallis et al [11] and Ganasegeran et al [45], health care professionals tend to communicate easily through WhatsApp by sharing patient information, including pictures, and seeking advice. However, the authors highlighted that the use of WhatsApp has potential risks to patient confidentiality because there are no built-in security measures and consent is a requirement for the exchange of patient information.

WhatsApp features could enable physicians to give adequate attention to patients while engaging in verbal communication during consultation. In this study, the voice-to-text feature on WhatsApp seemed to be ideal to assist health care professionals in digitally capturing patient history because it is readily available on smartphones. However, there are limitations to the use of the WhatsApp voice-to-text feature that makes the app inadequate to enable interaction moments during patient consultations.

One of the limitations of WhatsApp voice-to-text feature is that it only coherently transcribes verbal communication in English. This puts a patient who speaks an indigenous language at a disadvantage. In addition, there is no time stamp distinction between the voices of the physician and a patient, which would result in a cluttered clinical history note. Therefore, an mHealth voice-to-text and speech recognition app is essential for consultation but should ensure that there is a clear distinction between voices and during transcription from audio to text when used in the clinical context of health care service delivery.

In the wards, where the availability of desktop computers is limited and fixed, smartphones can be used to access electronically stored patient records. This requires health data records to be integrated across all digital platforms. In addition, physicians tend to use their smartphones to organize their work activities. An investigation by Greer et al [46] established that health care professionals use smartphones as a means to access or offer answers to medical questions and cross-reference evidence-based treatment medication from web-based academic sources in clinical settings. Similarly, O'Connor et al [5] investigated the impact of mHealth on the provision of care as perceived by physicians. The authors suggested that mHealth ICTs should be easily adaptable to perform the tasks within the work routine of health care professionals in hospital settings. Thus, the integration of health ICTs plays an important role in enhancing the decision-making of health care professionals in diagnosis and treatment processes to ensure the delivery of quality and safe care in clinical settings.

According to Cresswell et al [47], when systems inadequately meet the preferences of the intended end users to achieve an anticipated outcome, they are likely to adopt alternate means or discard the technology perceived as inadequate. Despite benefits attributed to the perceived usefulness and ease of use of health ICTs, there are unintended consequences experienced by health care professionals. The physicians expressed that the frustration caused by referral notifications received from the VULA mobile app can be disruptive multiple times while consulting with patients and during off-peak hours away from the hospital.

The unintended consequences could be attributed to social and technical mediators that inhibit the quality of care provided at point of care. This causes a delay in the time it takes to complete a consultation to the extent that some physicians feel uncomfortable with attending to VULA referrals. Gagnon et al [7] identified the compatibility of mHealth technologies with tasks as one of the several factors perceived to enable or inhibit the workflow of health care professionals. According to Yahya [26], time constraints and perceived reactions of patients to the use of mHealth ICTs by physicians could inhibit the optimal use of mHealth ICTs. The use of mHealth ICTs in clinical settings is more efficient when location is an inhibiting factor and tasks require minimal to no distractions during work activities, provided there are sufficient wireless bandwidth connections.

The nursing staff in tertiary hospitals is heavily saddled with the responsibility of patient administration, including admission, discharge, and transfer. As the amount of paper becomes cumbersome depending on the extent of a patient's journey, a mobile device is ideal to support report writing and other information management tasks when nurses move from one point of care to the other. The task benefits or unintended consequences experienced by health care professionals while using mHealth ICTs to enable interaction moments influence the degree of its suitability, job, and patient satisfaction [48]. Otherwise, the mHealth ICTs would be partially used or even discarded when health care professionals are not satisfied with the values attributed to use, and ultimately, health care service delivery continues to be impaired by a lack of adequate technology integration [20,47,48].

These findings contribute to pertinent considerations when making informed decisions to design, develop, or purchase mHealth ICTs. This study encourages ICT and health care professionals to work in a transdisciplinary team during the design phase of health technologies. This saves time and costs involved in facilitating training for health care professionals in the pilot and postimplementation phases.

Limitations and Future Research

There were difficulties and delays in setting up interviews with health care professionals in tertiary hospitals. The difficulties were because of the busy schedules of health care professionals

in public hospitals. Hence, this restricted the data collection process to clinical departments that provide medical imaging and reporting care only. In the South African context, health care professionals used health ICTs, including mHealth, whereas in the Nigerian context, there were no health ICTs used. The authors adopted a pragmatic approach to include unstructured open-ended questions.

The results of this study are not generalizable because the investigation considered only 2 tertiary hospitals in 2 countries in Sub-Saharan Africa, and the data are subjective. However, the methodology can be replicated, and the lessons learned in these contexts can be compared with similar hospital settings. Future studies should consider conducting extensive usability tests of mHealth ICTs applicable to the work environment of health care professionals and standardization of information sharing across different information systems in hospitals. For example, the time efficiency and usability of mHealth ICTs during patient consultation require speech recognition, medical semantics (context sensitivity), and transcription app. Additional features such as a time stamp are also needed to distinguish between voices during transcription from audio to text. These mHealth ICT mechanisms would reduce the amount of time expended by health care professionals at point of care, given the number of patients and referral consultations attended to on a daily basis.

Conclusions and Contribution

The findings of this study show that mHealth ICTs could be used during clinic consultations where physicians need to capture verbal communication with patients and observe their body language. The VULA mobile app is useful for managing referrals and information exchange remotely in a timely manner, except for its interruptions when physicians are busy with face-to-face consultations and other urgent activities. Owing to the mobile nature of nurses, mHealth ICTs could ease information capture, retrieval, and report writing between different point of care, particularly where computer access is limited and to reduce the amount of paper used for patient care administration in hospital settings. WhatsApp enables easier communication through instant messaging and offers a means to collaborate between professionals; however, in health care contexts, it does not guarantee the privacy of patients' health information.

The opportunities identified can be considered by local health policy makers and ICT vendors when designing mHealth ICTs to enable the work activities of health care professionals at point of care in hospital settings, especially in Sub-Saharan Africa. This paper provides a possible direction for stakeholders in the public health system of Sub-Saharan Africa that have recognized and adopted the use of technology as a driver to address time and location constraints that impede service delivery.

This study applied the ActAD model as a lens to understand opportunities to integrate mHealth ICTs into the work activities of health care professionals in hospital settings. The authors established that interaction moments of tasks performed by health care professionals during complex work activities are essential to identify relevant mHealth ICTs for different point of care. In addition, the transformation of the purpose of a work activity into the intended outcome is mediated by the interplay between attributes of human agency and contradictions that emerge from the contextual conditions and the technical characteristics.

Authors' Contributions

All authors contributed equally to ensure data collection, securing ethics clearance, and writing up the interpretation of the results.

Conflicts of Interest

None declared.

References

1. Junglas I, Abraham C, Ives B. Mobile technology at the frontlines of patient care: Understanding fit and human drives in utilization decisions and performance. Decision Support Systems 2009 Feb;46(3):634-647. [doi: 10.1016/j.dss.2008.11.012]
2. Liu F, Ngai E, Ju X. Understanding mobile health service use: An investigation of routine and emergency use intentions. Int J Inf Manag 2019 Apr;45:107-117. [doi: 10.1016/j.ijinfomgt.2018.09.004]
3. Bloom G, Waldman L, Labrique A, Hampshire K. Making mHealth work for all. Impact Initiative / Inst Dev Stud 2017 May:1-3. [doi: 10.35648/20.500.12413/11781/ii286]
4. Ojo AI. mHealth interventions in South Africa: A review. SAGE Open 2018 Mar 28;8(1):23-30. [doi: 10.1177/2158244018767223]
5. O'Connor Y, Andreev P, O'Reilly P. MHealth and perceived quality of care delivery: A conceptual model and validation. BMC Med Inform Decis Mak 2020 Feb 27;20(1):1-13 [FREE Full text] [doi: 10.1186/s12911-020-1049-8] [Medline: 32103746]
6. 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]
7. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
8. Kaium A, Bao Y, Alam MZ, Hasan N, Hoque R. Understanding the insight of factors affecting mHealth adoption. Int J Res buiness Soc Sci 2019 Oct 26;8(6):181-200. [doi: 10.20525/ijrbs.v8i6.522]

9. O' Connor Y, O' Reilly P. Examining the infusion of mobile technology by healthcare practitioners in a hospital setting. Inf Syst Front 2018 Dec 6;20(6):1297-1317. [doi: 10.1007/s10796-016-9728-9]
10. Fortuin J, Salie F, Abdullahi LH, Douglas TS. The impact of mHealth interventions on health systems: A systematic review protocol. Syst Rev 2016 Nov 25;5(1):200-206 [FREE Full text] [doi: 10.1186/s13643-016-0387-1] [Medline: 27884180]
11. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Glob Health Action 2017 Jun 25;10(sup3):29-36 [FREE Full text] [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
12. Martin G, Khajuria A, Arora S, King D, Ashrafian H, Darzi A. The impact of mobile technology on teamwork and communication in hospitals: A systematic review. J Am Med Inform Assoc 2019 Jan 25;26(4):339-355 [FREE Full text] [doi: 10.1093/jamia/ocy175] [Medline: 30689893]
13. Mutula SM. Factors influencing perceptions and attitudes of nurses towards the use of ICT in Patient Care in KwaZulu Natal Province, South Africa. African J Inf Syst 2015;8(1):1-14 [FREE Full text]
14. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth and Uhealth 2016 Jun 09;4(2):1-18 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
15. Chaiyachati KH, Loveday M, Lorenz S, Lesh N, Larkan L, Cinti S, et al. A pilot study of an mHealth application for healthcare workers: Poor uptake despite high reported acceptability at a rural South African community-based MDR-TB treatment program. PLoS One 2013 May 28;8(5):e64662 [FREE Full text] [doi: 10.1371/journal.pone.0064662] [Medline: 23724075]
16. Botha A, Booi V. mHealth implementation in South Africa. In: Proceedings of the IST-Africa Week Conference. 2016 May 11 Presented at: IST-Africa Week Conference; May 11-13, 2016; Durban, South Africa p. 1-13. [doi: 10.1109/istafrica.2016.7530667]
17. Wolff-Piggott B, Coleman J, Rivett U. The clinic-level perspective on mHealth implementation: A South African case study. Inf Technol Dev 2017 Jan 06;24(3):532-553. [doi: 10.1080/02681102.2016.1233858]
18. West DM. Using mobile technology to improve maternal health and fight Ebola: A case study of mobile innovation in Nigeria. Brookings Instit 2015 Mar:1-19 [FREE Full text]
19. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: A systematic review of studies based on social cognitive theories. Implement Sci 2008 Jul 16;3(1):36 [FREE Full text] [doi: 10.1186/1748-5908-3-36] [Medline: 18631386]
20. Irinoye OO, Ayandiran EO, Fakunle I, Mtshali N. Nurses' perception and barriers to use of information communication technology in a teaching hospital in Nigeria. Comput Inform Nurs 2013 Aug;31(8):394-400. [doi: 10.1097/NXN.0b013e3182997a6e] [Medline: 23851710]
21. Puckree T, Maharaj S, Mshunquane N. Status of usage of information communications technology by academic staff at a selected nursing training institution in South Africa. Int J Edu Sci 2017 Sep 25;8(3):511-519. [doi: 10.1080/09751122.2015.11890273]
22. Abyaomi OK, Evans ND, Ocholla DN. Factors that influence medical doctors' behavioural intention to use clinical informatics. Mousaion 2017 Dec 11;35(1):130-154. [doi: 10.25159/0027-2639/2321]
23. Behrens A, Noteboom CB, Bishop D. How can health technology project communications be improved in a hospital. In: Proceedings of the Faculty Research & Publications, College of Business and Information Systems Conference. United States: Association of Information Systems Electronic Library; 2019 May 21 Presented at: MWAIS 2019; May 2019; Wisconsin p. 1-6 URL: https://scholar.dsu.edu/cgi/viewcontent.cgi?article=1065&context=bispapers
24. Watson L, Pathiraja F, Depala A, O'Brien B, Beyzade S. Ensuring safe communication in health care: A response to Johnston et al on their paper "Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams". Am J Surg 2016 Jan 01;211(1):302-303. [doi: 10.1016/j.amjsurg.2015.04.017] [Medline: 26184352]
25. de Keizer NF, Ammenwerth E. The quality of evidence in health informatics: How did the quality of healthcare IT evaluation publications develop from 1982 to 2005? Int J Med Inform 2008 Jan;77(1):41-49. [doi: 10.1016/j.ijmedinf.2006.11.009] [Medline: 17208040]
26. Yahya H. Healthcare-related smartphone use among doctors in hospitals in Kaduna, Nigeria - A survey. Niger J Clin Pract 2019 Jul;22(7):897-905 [FREE Full text] [doi: 10.4103/njcp.njcp_454_18] [Medline: 31293252]
27. Katurura MC, Cilliers L. Electronic health record system in the public health care sector of South Africa: A systematic literature review. Afr J Prim Health Care Fam Med 2018 Nov 20;10(1):1-8 [FREE Full text] [doi: 10.4102/phcfm.v10i1.1746] [Medline: 30456963]
28. Korpela M, Mursu A, Soriyan A, Eerola A, Häkkinen H, Toivanen M. Information systems research and development by Activity Analysis and Development: Dead horse or the next wave? In: Inf Syst Res. Boston: Springer; 2004:453-471.
29. Mursu Á, Luukkonen I, Toivanen M, Korpela M. Activity Theory in information systems research and practice: Theoretical underpinnings for an information systems development model. Inf Res Int Eelectron J 2007 Apr;12(3):1-21 [FREE Full text]
30. Myers MD, Newman M. The qualitative interview in IS research: Examining the craft. Info Organ 2007 Jan;17(1):2-26. [doi: 10.1016/j.infoandorg.2006.11.001]

31. Steen M, Manschot M, De Koning N. Benefits of co-design in service design projects. Int J Design 2011;5(2):53-60 [FREE Full text]
32. Creswell J. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, California, United States: SAGE Publications; 2002:1-272.
33. Ifinedo P. Measuring Africa's e-readiness in the global networked economy: A nine-country data analysis. Int J Educ Dev 2005 Apr;1(1):53-71 [FREE Full text]
34. Johnston MJ, King D, Arora S, Behar N, Athanasiou T, Sevdalis N, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. Am J Surg 2015 Jan;209(1):45-51. [doi: 10.1016/j.amjsurg.2014.08.030] [Medline: 25454952]
35. Wallis L, Hasselberg M, Barkman C, Bogoch I, Broomhead S, Dumont G, et al. A roadmap for the implementation of mHealth innovations for image-based diagnostic support in clinical and public-health settings: A focus on front-line health workers and health-system organizations. Global Health Action 2017 Jun 25;10(sup3):59-68 [FREE Full text] [doi: 10.1080/16549716.2017.1340254] [Medline: 28838310]
36. De Benedictis A, Lettieri E, Masella C, Gastaldi L, Macchini G, Santu C, et al. WhatsApp in hospital? An empirical investigation of individual and organizational determinants to use. PLoS One 2019 Jan 11;14(1):e0209873 [FREE Full text] [doi: 10.1371/journal.pone.0209873] [Medline: 30633754]
37. Watkins JO, Goudge J, Gómez-Olivé FX, Griffiths F. Mobile phone use among patients and health workers to enhance primary healthcare: A qualitative study in rural South Africa. Soc Sci Med 2018 Feb;198:139-147 [FREE Full text] [doi: 10.1016/j.socscimed.2018.01.011] [Medline: 29335160]
38. Hasselberg M, Wallis L, Blessing P, Laflamme L. A smartphone-based consultation system for acute burns - methodological challenges related to follow-up of the system. Global Health Action 2017 Jun 25;10(sup3):1328168 [FREE Full text] [doi: 10.1080/16549716.2017.1328168] [Medline: 28838311]
39. Kabanda S, Rother H. Evaluating a South African mobile application for healthcare professionals to improve diagnosis and notification of pesticide poisonings. BMC Med Inform Decis Mak 2019 Mar 11;19(1):40 [FREE Full text] [doi: 10.1186/s12911-019-0791-2] [Medline: 30857525]
40. Debrah RD, de la Harpe R, M'Rithaa MK. Design probes and toolkits for healthcare: Identifying information needs in African communities through service design. The Design Journal 2017 Sep 06;20(sup1):2120-2134. [doi: 10.1080/14606925.2017.1352730]
41. Clatworthy S. Service innovation through touch-points: Development of an innovation toolkit for the first stages of new service development. Int J Des 2011;5(2):15-28 [FREE Full text]
42. Vainikainen V, Soriyan HA, Korpela M, Saranto K. Good practices to enhance the perceived usefulness of computerized hospital information systems?case study in Nigeria. J Heal Informatics in Africa 2014;2(2):55-70. [doi: 10.12856/JHIA-2014-v2-i2-109]
43. Saldana J. The Coding Manual for Qualitative Researchers. Thousand Oaks, CA: SAGE Publications Ltd; 2009:1-224.
44. Adebara O, Adebara I, Olaide R, Emmanuel G, Olanrewaju O. Knowledge, attitude and willingness to use mhealth technology among doctors at a semi urban tertiary hospital in Nigeria. J Adv Med Med Res 2017 Jan 10;22(8):1-10. [doi: 10.9734/jammr/2017/33232]
45. Ganasegeran K, Renganathan P, Rashid A, Al-Dubai S. The m-Health revolution: Exploring perceived benefits of WhatsApp use in clinical practice. Int J Med Inform 2017 Jan;97:145-151. [doi: 10.1016/j.ijmedinf.2016.10.013] [Medline: 27919374]
46. Greer DB. Exploring nursing students' smartphone use in the clinical setting. MEDSURG Nurs 2019 May;28(2):163-182 [FREE Full text]
47. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun 01;20(e1):9-13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
48. Ladan M, Wharrad H, Windle R. eHealth adoption and use among healthcare professionals in a tertiary hospital in Sub-Saharan Africa: A Qmethodology study. PeerJ 2019 Apr;7:e6326 [FREE Full text] [doi: 10.7717/peerj.6326] [Medline: 31041146]

Abbreviations

ActAD: Activity Analysis and Development
ICT: information and communication technology
mHealth: mobile health

Edited by L Buis; submitted 08.12.20; peer-reviewed by L Cilliers, A Ojo; comments to author 19.01.21; revised version received 25.01.21; accepted 18.05.21; published 13.10.21.

Please cite as:
Ogundaini OO, de la Harpe R, McLean N
Integration of mHealth Information and Communication Technologies Into the Clinical Settings of Hospitals in Sub-Saharan Africa: Qualitative Study
JMIR Mhealth Uhealth 2021;9(10):e26358
URL: https://mhealth.jmir.org/2021/10/e26358
doi:10.2196/26358
PMID:34643540

Original Paper

Impact of a Mobile App on Paramedics' Perceived and Physiologic Stress Response During Simulated Prehospital Pediatric Cardiopulmonary Resuscitation: Study Nested Within a Multicenter Randomized Controlled Trial

Matthieu Lacour[1], MMed; Laurie Bloudeau[2], EMT-P; Christophe Combescure[1,3], PhD; Kevin Haddad[4], RNCS; Florence Hugon[4], RNCS; Laurent Suppan[1,5], MD; Frédérique Rodieux[1,6], MD; Christian Lovis[1,7], MPH, MD; Alain Gervaix[1,4], MD; Frédéric Ehrler[1,7], PhD; Sergio Manzano[1,4], MD; Johan N Siebert[1,4], MD; PedAMINES Prehospital Group[8]

[1]University of Geneva, Faculty of Medicine, Geneva, Switzerland

[2]A.C.E. Geneva Ambulances SA, Geneva, Switzerland

[3]Division of Clinical Epidemiology, Department of Health and Community Medicine, Geneva University Hospitals, Geneva, Switzerland

[4]Department of Pediatric Emergency Medicine, Geneva Children's Hospital, Geneva University Hospitals, Geneva, Switzerland

[5]Department of Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland

[6]Division of Clinical Pharmacology and Toxicology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Geneva University Hospitals, Geneva, Switzerland

[7]Division of Medical Information Sciences, Department of Radiology and Medical Informatics, Geneva University Hospitals, Geneva, Switzerland

[8]See Authors' Contributions

Corresponding Author:
Johan N Siebert, MD
Department of Pediatric Emergency Medicine, Geneva Children's Hospital
Geneva University Hospitals
Avenue de la Roseraie, 47
Geneva, 1205
Switzerland
Phone: 41 (0)795534072
Fax: 41 (0)223725405
Email: Johan.Siebert@hcuge.ch

Abstract

Background: Out-of-hospital cardiac arrests (OHCAs) are stressful, high-stake events that are associated with low survival rates. Acute stress experienced in this situation is associated with lower cardiopulmonary resuscitation performance in calculating drug dosages by emergency medical services. Children are particularly vulnerable to such errors. To date, no app has been validated to specifically support emergency drug preparation by paramedics through reducing the stress level of this procedure and medication errors.

Objective: This study aims to determine the effectiveness of an evidence-based mobile app compared with that of the conventional preparation methods in reducing acute stress in paramedics at the psychological and physiological levels while safely preparing emergency drugs during simulated pediatric OHCA scenarios.

Methods: In a parent, multicenter, randomized controlled trial of 14 emergency medical services, perceived and physiologic stress of advanced paramedics with drug preparation autonomy was assessed during a 20-minute, standardized, fully video-recorded, and highly realistic pediatric OHCA scenario in an 18-month-old child. The primary outcome was participants' self-reported psychological stress perceived during sequential preparations of 4 intravenous emergency drugs (epinephrine, midazolam, 10% dextrose, and sodium bicarbonate) with the support of the PedAMINES (Pediatric Accurate Medication in Emergency Situations) app designed to help pediatric drug preparation (intervention) or conventional methods (control). The State-Trait Anxiety Inventory and Visual Analog Scale questionnaires were used to measure perceived stress. The secondary outcome was physiologic stress, measured by a single continuous measurement of the participants' heart rate with optical photoplethysmography.

Results: From September 3, 2019, to January 21, 2020, 150 advanced paramedics underwent randomization. A total of 74 participants were assigned to the mobile app (intervention group), and 76 did not use the app (control group). A total of 600 drug doses were prepared. Higher State-Trait Anxiety Inventory–perceived stress increase from baseline was observed during the scenario using the conventional methods (mean 35.4, SD 8.2 to mean 49.8, SD 13.2; a 41.3%, 35.0 increase) than when using the app (mean 36.1, SD 8.1 to mean 39.0, SD 8.4; a 12.3%, 29.0 increase). This revealed a 30.1% (95% CI 20.5%-39.8%; *P*<.001) lower relative change in stress response in participants who used the app. On the Visual Analog Scale questionnaire, participants in the control group reported a higher increase in stress at the peak of the scenario (mean 7.1, SD 1.8 vs mean 6.4, SD 1.9; difference: −0.8, 95% CI −1.3 to −0.2; *P*=.005). Increase in heart rate during the scenario and over the 4 drugs was not different between the 2 groups.

Conclusions: Compared with the conventional method, dedicated mobile apps can reduce acute perceived stress during the preparation of emergency drugs in the prehospital setting during critical situations. These findings can help advance the development and evaluation of mobile apps for OHCA management and should be encouraged.

Trial Registration: ClinicalTrials.gov NCT03921346; https://clinicaltrials.gov/ct2/show/NCT03921346

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-019-3726-4

(JMIR Mhealth Uhealth 2021;9(10):e31748) doi:10.2196/31748

KEYWORDS

cardiopulmonary resuscitation; drugs; emergency medical services; medication errors; mobile health; mobile apps; out-of-hospital cardiac arrest; paramedics; pediatrics; State-Trait Anxiety Inventory; stress

Introduction

Background

Out-of-hospital cardiac arrest (OHCA) is a major concern for health care systems worldwide, affecting millions of people each year [1]. Despite advances in resuscitation science and improvement of cardiac arrest survival over the past decades, the survival rates following adult and pediatric OHCA are reportedly low, evaluated at 10.4% and 11.4%, respectively [1]. High-quality cardiopulmonary resuscitation (CPR) for OHCA patients is the primary determinant of survival and favorable neurological outcome [2,3]. Evaluations and decisions must be made quickly and accurately. However, acute mental stress experienced by rescuers during CPR may impair decision making and optimal performance [4-12], independent of professional experience [4]. This can, in turn, adversely affect patient safety [5]. OHCA-induced acute stress response relies mainly on an interplay between the individual's cognitive perception and appraisal made about the perceived demand and ability to compensate through both internal and environmental resources [6,13]. The degree to which this compensation occurs determines the nature and magnitude of one's stress response [12]. Some individuals show stress responses with associated active coping. Others perceive stress as excessive and outweighing their coping abilities, thus hindering their ability to adapt quickly and perform under pressure.

The OHCA setting is a stressful and high-stakes environment where safeguards and resources, both human and material, are limited [6]. In many countries, paramedics have the autonomy to prepare and administer emergency drugs. However, the impact of acute stress experienced by paramedics during OHCA on emergency drug preparation has rarely been studied. LeBlanc et al [14] observed that paramedics under simulated high-stress conditions performed worse on drug dosage calculations than those under calm, relaxed conditions. These findings are particularly concerning in pediatric CPR, where the accurate and safe preparation and administration of intravenous drugs is mandatory [15-19]. Most drugs administered intravenously to children are provided in vials that were originally prepared for the adult population. This leads to the need for an initial onsite complicated, individual, weight-based dose calculation, and drug preparation for each child, which varies widely across age groups [20]. Combined with other risk factors such as excessive extraneous cognitive load due to onsite emotional stress and time pressure [14,21-23], and pediatric-specific, age-related variations in pharmacokinetics, onsite administration of emergency drugs by paramedics is particularly challenging. Furthermore, pediatric situations only account for approximately 7% of emergency medical services (EMS) calls, and paramedics have little exposure to critically ill children and occasions to prepare emergency drugs at pediatric dosages [24-26]. Relying solely on their expertise and knowledge to take decisions during care provision, a single paramedic is often in charge of determining the child's weight, choosing the most suitable drug, calculating the drug dose and appropriate volume to inject, and administering it without delay. For this purpose, paramedics are still dependent on conventional paper-based support, empirical calculators, or spreadsheets to ensure correct drug delivery. This places children at higher risk for life-threatening prehospital medication errors than adults [17,20,27-30], with a reported error rate of more than 60% [31,32].

Some authors have advocated replacing tasks inducing stress and cognitive load during resuscitation as much as possible by automated actions to optimize patient care and diminish medication errors [22,33]. The US National Highway Traffic Safety Administration advocated, in its recent vision for the future of pediatric prehospital care to be achieved by 2050, to develop processes that do not require providers to calculate dosing of medications [34]. Supported by the rapid spread of mobile devices and their innovative features (eg, connectivity, embedded computing capabilities, small size, and versatility), mobile health (mHealth) apps have great potential as tools to support out-of-hospital emergency drug preparation at the point

of care. However, a recent systematic review showed that few mHealth apps are available for prehospital settings [35]. Among these mHealth apps, none has been validated to specifically support emergency drug preparation by EMS personnel with the aim of reducing medication dosing errors and the stress hassle of this procedure.

Previous Work

In previous randomized trials, we reported fewer medication errors and shorter times to drug preparation and delivery during in-hospital pediatric CPR when using a mobile app—the PedAMINES (Pediatric Accurate Medication in Emergency Situations) app compared with conventional preparation methods [36,37]. Although similarities exist, the prehospital environment is distinctly different in many regards. Recent findings of a multicenter randomized trial showed that this app was also able to reduce medication error rates during pediatrics OHCA in a simulated model [38]. However, its impact on situational stress experienced by paramedics during CPR remains to be determined.

Aim

This study aims to determine the effectiveness of the PedAMINES app in reducing acute stress while safely preparing emergency drugs during CPR for pediatric OHCA patients compared with the conventional preparation methods.

Methods

Study Design

This study was registered at ClinicalTrials.gov (NCT03921346) as a nested study within the context of an open-label, multicenter, randomized controlled trial. The parent trial had the broader and primary aim of assessing rates of medication dosing errors during simulation-based pediatric OHCA scenarios using a high-fidelity manikin [38]. The trial protocol, containing details about the scenario, has been previously published [39]. No changes were made to the app or intervention during the study.

Trial Participants

The trial was conducted at 14 EMS covering a population of more than 2.3 million people in Switzerland. Eligible participants were registered paramedics working in these EMS. They had undergone a 3-year education program and were trained in advanced life support procedures, including defibrillation, airway management, peripheral intravenous line cannulation, and the administration of medications to ensure advanced independent emergency prehospital care. Similar to the Anglo-American model [40,41], paramedics in Switzerland constitute the initial response team and are qualified to independently administer a range of medications. To achieve adequate participant enrolment to reach the target sample size (ie, 120 paramedics [39]), shift-working paramedics were randomly recruited weeks before the start of the study by a blinded noninvestigator in each EMS center. In several EMS centers, additional paramedics were recruited to ensure that the target sample size would ultimately be achieved. At the end of the trial, an additional number of paramedics was included in the study and the analysis was refined. To prevent preparation bias, all paramedics were informed of the upcoming simulation study, but not of its purpose and outcomes. The study excluded emergency medical technicians because they had no drug preparation autonomy. In our study, resuscitation was led by a physician (JNS) to standardize the choice of drugs prescribed across the EMS and to avoid any deviation from the study protocol. Written informed consent was obtained from all the participants before their voluntary involvement. Blinding to the purpose of the trial during recruitment was maintained to minimize preparation bias. Participants were unblinded during the intervention when they were asked to prepare the drugs either with the support of the app or conventional methods. Study participants were neither involved in the design of the app, study design, choice of outcome measures, nor study conduct. No participant was asked to advice on the interpretation or write the results.

Setting

The trial took place in a simulated, regular child's bedroom environment at each EMS center. This was intended to mimic as closely as possible a stressful prehospital environment where paramedics could actually intervene to increase realism. The intervention was standardized across all sites to follow the same chronological progression and range of difficulty in order to ensure that each participant was exposed to exactly the same situation, with similar challenges in decision making and treatment preparation provided on the same manikin (Laerdal New SimBaby, Tetherless, Laerdal Medical, Stavanger, Norway). The study team members maintained a stressful resuscitation atmosphere. Importantly, we did not organize pretests to minimize priming and preparation biases to avoid influencing the perceived stress during the scenario.

Intervention

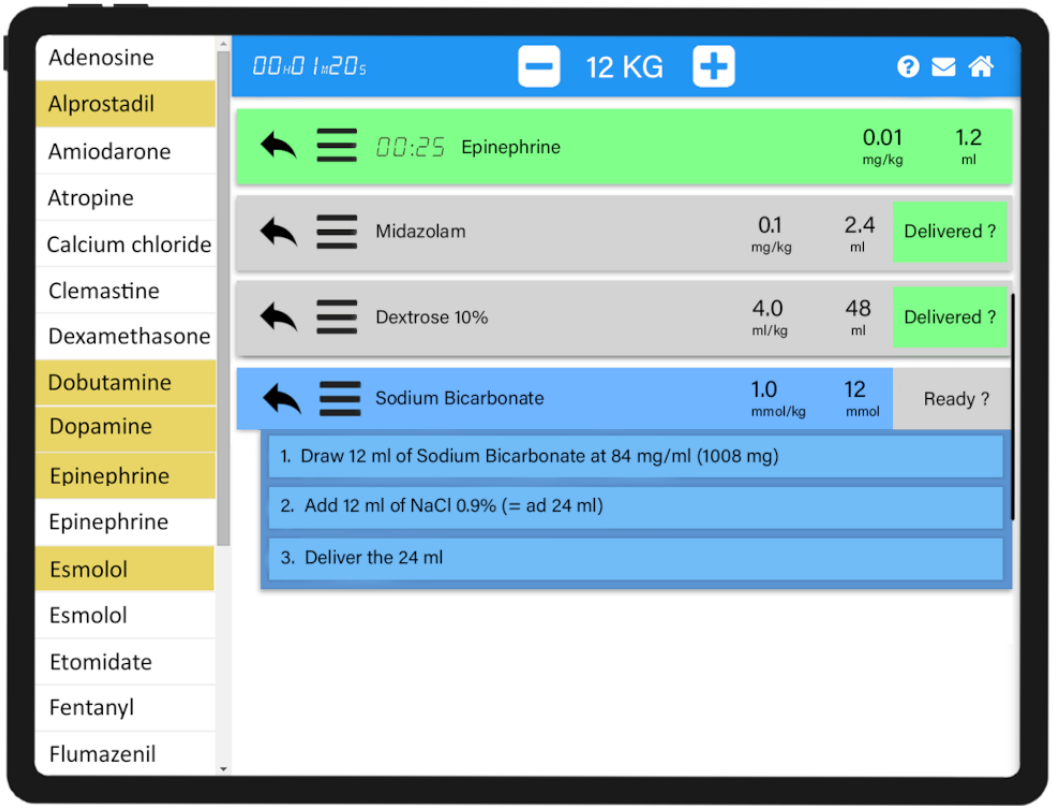
On the day of the participation after random allocation, each participating paramedic had to complete 2 self-administered questionnaires to measure their perceived stress at baseline (see below), and a heart rate (HR) monitor was placed on their wrist. They also had to complete a survey collecting data regarding their demographics, attend a standardized 5-minute training session on how to use the mobile app, and follow a presentation detailing the features of the simulation manikin. No conventional drug preparation training was provided in either group, as this was part of the paramedics' daily practice. Each participant was then exposed to a 20-minute, standardized, fully video-recorded, highly realistic pediatric OHCA CPR scenario of an 18-month-old child. The participants were asked to sequentially prepare and inject 4 different direct intravenous drugs of various degrees of preparation difficulties (epinephrine, midazolam, dextrose 10%, and sodium bicarbonate), either with the support of the app or following conventional methods (ie, without app support). Full details of the intervention and scenario have been previously published [38]. During the timed scenario, the resuscitation team maintained a stressful and realistic resuscitation atmosphere by frequently reporting vital signs aloud and asking the participant to promptly provide the drugs. The portable defibrillator, displaying real time vital signs, was placed in close proximity to the paramedics. The monitoring

alarms were turned on, and the investigator who played the father's role repeatedly verbalized his dismay.

The PedAMINES Mobile App

This app provided an exhaustive and editable list of intravenous drugs, either for direct injections or continuous infusions [37,38,42]. The drugs were displayed in alphabetical order with doses automatically adapted to the weight or age of the patient. When selected, a detailed drug preparation procedure based on a standardized and simplified pathway was provided to the user. The app was developed using a user-centered, evidence-based approach. Emergency department caregivers, ergonomists, psychologists, and computer scientists from research and development services conceived the app. Its interface was designed considering design principles aimed at minimizing cognitive load [43-45]. By always displaying the most important information in a larger font and sorting the most recent information at the top of the screen, hierarchization of information was taken into consideration. Interaction choices were limited to their strictest utility through the interface to facilitate decision making. Only a predefined sequence of actions was proposed; thus, users did not need to make complicated choices. Display conventions were defined to display drug dosages with conventional and consistent units to avoid confusion. Fits laws were complied with by placing interaction buttons at the edge of the screen [46]. This minimized the distance to reach them and focused on the interaction zone in a limited area. The app also complied with the progressive disclosure principle [47], where complex information is sequenced across several smaller chunks to reduce the feeling of being overwhelmed by the user. Therefore, complex instructions such as weight-based drug dose preparation was sequenced and ordered in several lines to streamline information and facilitate its comprehension (Figure 1). Feedback mechanisms have also been integrated for specific actions, such as canceling or modifying the patient's weight. In both cases, the system provides feedback to ensure that users are aware of their actions. Acceptance and use of the app were assessed through self-administrated electronic surveys using a modified version of the Unified Theory of Acceptance and Use of Technology (UTAUT) model [48] and the System Usability Scale [49,50]. They will be the subject of detailed analysis in another study nested within the parent trial.

Figure 1. PedAMINES (Pediatric Accurate Medication in Emergency Situations) app screenshot.



Outcomes and Measures

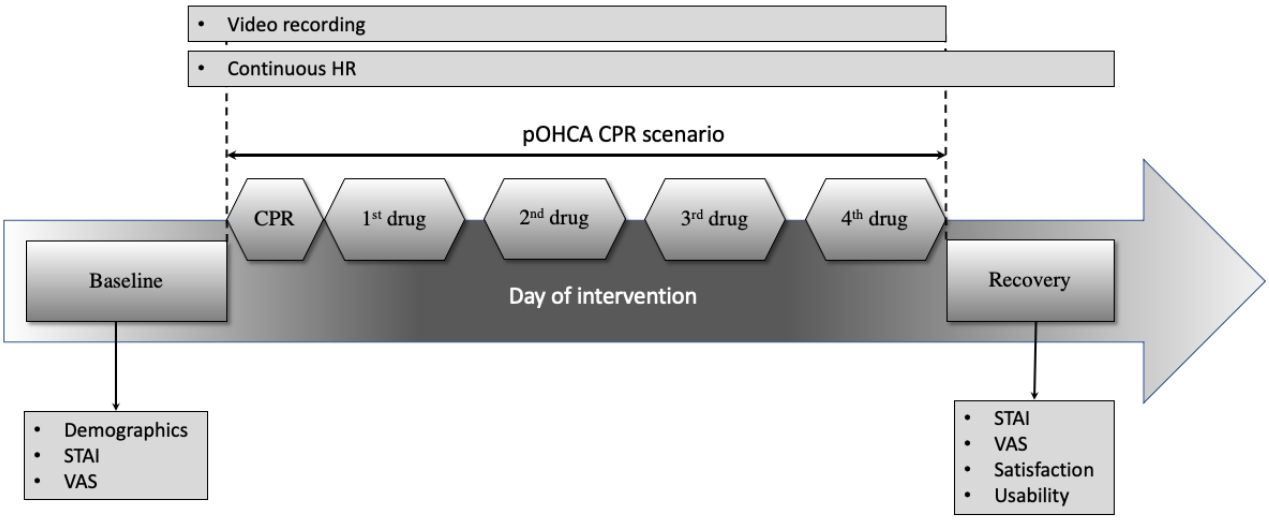
Primary Outcome: Perceived Stress

Participants' self-assessed psychological stress was measured using the Gauthier and Bouchard's French-Canadian adaptation [51] of Spielberger's psychometric State-Trait Anxiety Inventory (STAI) questionnaire [52,53]. STAI was provided with permission from the copyright owner (Mind Garden, Inc [54]. Copyright prevents reproduction of the full scale). It is one of the most commonly used subjective measures of stress in health research, including emergency care [13,14,55-58]. This questionnaire is composed of two 20-item self-report subscales to measure 2 distinct anxiety concepts: (1) the temporary *state of anxiety* at the time of reporting (STAI form Y-1), which can be affected by stressful situations and 2) the more stable and long-standing presence of *trait anxiety* (form Y-2) [59]. Both forms can be used alone or as a complement. Form Y-1 was used in this study. To avoid interrupting the scenario and influence its veracity and inherent stress, no STAI was

administered during the scenario. It was administered just before the scenario began to assess the stress at this moment, and again just after the end of the scenario by asking the paramedics to assess their maximum perceived stress during the scenario (Figure 2). Each item was mandatory to avoid missing values and answered on a 4-point Likert scale ranging from 1 (not at all) to 4 (very much). After reversing the scores for stress-absent items (ie, items 1, 2, 5, 8, 10, 11, 15, 16, 19, and 20) according to Spielberger's instructions [59], the total score was calculated by summing up the weighted scores for the 20 items. STAI ranges from 20 to 80, with higher scores being positively correlated with greater stress [59]. A score greater than 40 is commonly used to define a clinical state of stress.

Perceived stress was also assessed by self-assessment using a numerical 10-point Likert visual analogue scale (VAS) [60]. Values ranged from 1 (totally unstressed) to 10 (totally stressed) to avoid neutral answers. To prevent any anticipation bias, participants were not informed that they would have to complete the STAI and VAS questionnaires after scenario completion (Figure 2). The questionnaires were provided onsite with the necessary precautions so that participants could not communicate with each other. No interaction other than detailing an item upon request of a participant occurred between the participants and investigators during the completion of the questionnaires.

Figure 2. Course of the intervention. CPR: cardiopulmonary resuscitation; HR: heart rate; pOHCA: pediatric out-of-hospital cardiac arrest; STAI: State-Trait Anxiety Inventory; VAS: visual analogue scale.



Secondary Outcome: Physiologic Stress

HR was measured as a surrogate of the physiological sympathetic response to stress [61]. A single continuous measurement at 1-second interval was recorded during the scenario with optical photoplethysmography using a Polar A360 wrist-worn HR monitor (Polar Electro Oy). This wearable sensor has been previously validated for HR assessment [62,63], although it is not sensitive enough to track HR variability. The A360 was tightly attached to the participants' wrist in accordance with the manufacturer's specifications to avoid motion artifacts that could lead to inaccurate HR measurements. Data locally stored on the wristwatch itself during the scenarios was thereafter synchronized with the dedicated Polar FlowSync web service for later offline analysis. In line with previous research [64], several time-points of cardiovascular activity were measured: (1) the minimal HR measured within the 5 minutes before the scenario starts ($HR_{baseline}$) while participants were not performing mental or physical exercise; (2) peak HR (HR_{peak}) for each drug, defined as the maximal HR reached during the sequence from drug prescription by the physician to drug delivery; and (3) an additional $HR_{recovery}$ was also measured as the minimal HR measured during the 5 minutes immediately following scenario completion (ie, at the stop of the timed period of the scenario represented by the patient's arousal, but before debriefing).

Data Collection

All scenarios were video-recorded by 3 GoPro Hero 5 and 7 Black Edition (GoPro Inc) video cameras mounted on participants and dispatched around them for later analysis. The camera setup was standardized. The investigators double-checked whether the questionnaires were fully and accurately completed. Data was collected using a REDCap database web app (REDCap, Vanderbilt University) hosted at Geneva University Hospitals and interfaced on an iPad Pro iOS 12.4 (Apple Inc). Neither follow up nor retention plans were required.

Statistical Analysis

Assuming a two-sided α risk of .05, the number of participants enrolled in the parent study was sufficient to detect an effect size of 0.50, which corresponds to a medium effect of the app compared with the conventional methods on perceived stress (STAI-Y score), with a power of 80%. Outcomes with a single measurement per participant (including HR_{peak} per drug and

postintervention STAI-Y and VAS) were compared between trial arms by using linear regression models with adjustment for the preintervention or baseline value and on centers to account for the randomized stratification. To compare the overall HR_{peak} between the arms, a linear regression model with mixed effects and adjusted for centers and $HR_{baseline}$ was used to account for the multiple measures per participant (one per drug). With this model, the intercept was random with crossed random effects at the participant and drug levels. This model was adjusted for the centers and $HR_{baseline}$. The Spearman correlation coefficient was used to assess the correlation between outcomes. Analyses were carried out using R software version 4.0.2, and the R package lme4 [65]. All statistical tests were two-sided, with an α risk of .05.

Ethics Approval

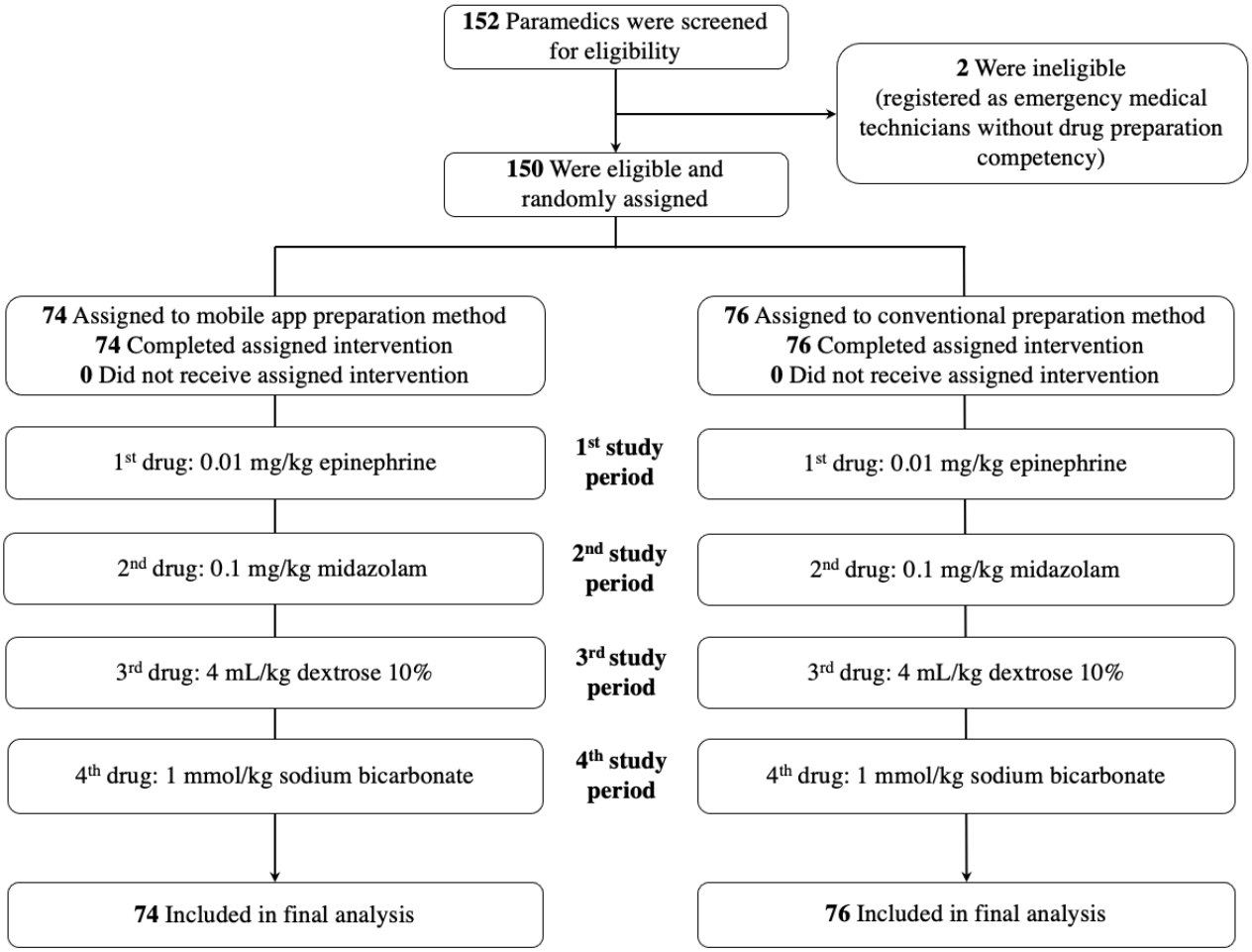
This trial received a declaration of no objection by the Geneva Cantonal Ethics Committee and Swiss Ethics on March 29, 2018, as its purpose was to examine the effect of the intervention on health care providers. The study was conducted in accordance with the principles of the Declaration of Helsinki [66], Good Clinical Practice guidelines [67], the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth (CONSORT-EHEALTH) [68] guidelines, and the Reporting Guidelines for Health Care Simulation Research [69].

Results

Overview

Of the 150 paramedics enrolled between September 3, 2019, and January 21, 2020, 74 were assigned to the intervention group and 76 to the control group. A total of 600 drug preparations were prepared. One participant's HR recording data were missing in the intervention group due to a technical problem with the watch. No dropouts occurred (Figure 3). The baseline characteristics are detailed in the published parent trial [38]. They were balanced in the 2 groups, and recruitment was balanced across the centers.

Figure 3. Screening, randomization, and analysis.



Stress Response: Perceived Stress

Both the STAI Form Y-1 and VAS questionnaires were completed by all 150 participants. Baseline STAI-perceived stress levels before the start of the scenario did not differ between the allocation groups (Table 1). The mean STAI-perceived stress scores of the participants supported by conventional methods then increased significantly at the time of drug preparations, whereas no significant increase was observed with the app support during the scenario. A higher stress increase was observed during the scenario using the

conventional methods than the app (P<.001; Table 1 and Figure 4).

Similarly, on the VAS questionnaire, participants rated the mean perceived stress before the scenario as not different with or without the app support (Table 1 and Figure 4). After scenario completion, they reported a higher increase in stress at the peak of the scenario using the conventional method than the app. The paramedics' gender, age, and years of experience did not modify the intervention effect, although for paramedics with more than 10 years of experience, the effect of the app seems to weaken (Multimedia Appendix 1).

Table 1. Comparison of STAI Form Y-1 and VAS scores before and after scenario completion (N=150).

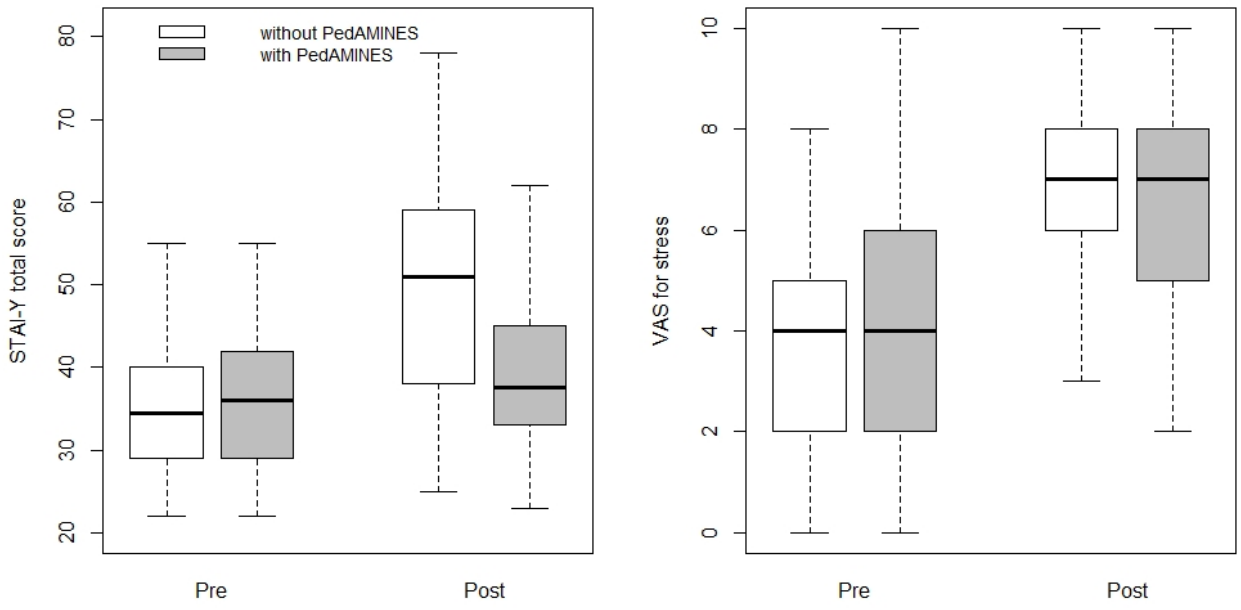
	Mobile app (n=73), mean (SD)	Conventional method (n=76), mean (SD)	Difference (95% CI)[a]	*P* value
STAI[b]				
Preintervention	36.1 (8.1)	35.4 (8.2)	0.8 (−1.7 to 3.3)	.55
Postintervention	39.0 (8.4)	49.8 (13.2)	−11.4 (−14.7 to −8.0)	<.001
Relative change (% of preintervention)	12.3 (29)	43.1 (35)	−30.1 (−39.8 to −20.5)	<.001
VAS[c]				
Preintervention	4.2 (2.5)	3.9 (2.2)	0.3 (−0.4 to 1.0)	.35
Postintervention	6.4 (1.9)	7.1 (1.8)	−0.8 (−1.3 to −0.2)	.006

[a]Linear regression models adjusted for center; in addition, differences in postintervention and relative change were adjusted for the preintervention values.

[b]STAI: State-Trait Anxiety Inventory.

[c]VAS: visual analogue scale.

Figure 4. State-Trait Anxiety Inventory Form Y-1 and Visual Analogic Score box plots per study arm. PedAMINES: Pediatric Accurate Medication in Emergency Situations; STAI: State Trait Anxiety Inventory; VAS: visual analogue scale.



Stress Response: Heart Rate

The mean $HR_{baseline}$ before the scenario started was approximately 80 beats per minute and similar between the allocation groups (Table 2). Maximal HR achieved during cardiac compressions was 122.5 (95% CI 118-127) in the app group and 123 (95% CI 117-128) in the control group and decreased in both groups before the drug preparation phase. During the scenario, the overall HR_{peaks} for the 4 drugs increased to 118 (95% CI 111 to 125) beats per minute with conventional methods and to 119.7 (95% CI 115 to 124) beats per minute with the app. This increase in HR from baseline represented approximately 50% of the mean $HR_{baseline}$ (Figure 5). The

difference between groups in HR_{peaks} during the scenario and over the 4 drugs was 1.4 (95% CI −1.8 to 4.6; *P*=.43) beats per minute, which was not statistically significant. The same was true for each drug (Table 2). After completion of the scenarios, HR declined to recovery values similar to $HR_{baseline}$ in both groups.

Table 2. Heart rates before, during, and after scenario completion, per study group (N=150).

	Heart rate (bpm[a])			
	Mobile app (n=73), mean (SD)	Conventional method (n=76), mean (SD)	Difference (95% CI)[b]	*P* value
Baseline	79.3 (14.4)	78.5 (12.7)	1.0 (−3.3 to 5.3)	.64
First drug, HR^{c}_{peak}	123.1 (9.2)	124.1 (12.2)	−1.1 (−4.7 to 2.4)	.53
Second drug, HR_{peak}	121.1 (10.9)	119.9 (13.3)	1.0 (−2.8 to 4.8)	.61
Third drug, HR_{peak}	120.4 (11.4)	117.9 (13.3)	2.4 (−1.4 to 6.2)	.22
Fourth drug, HR_{peak}	114.1 (13.5)	110.5 (13.7)	3.4 (−0.9 to 7.6)	.12
Recovery	79.3 (15.0)	76.8 (13.7)	1.9 (−2.3 to 6.1)	.37
Maximal HR_{peak}[d]	126.1 (10.3)	126.0 (12.1)	−0.1 (−3.8 to 3.5)	.95

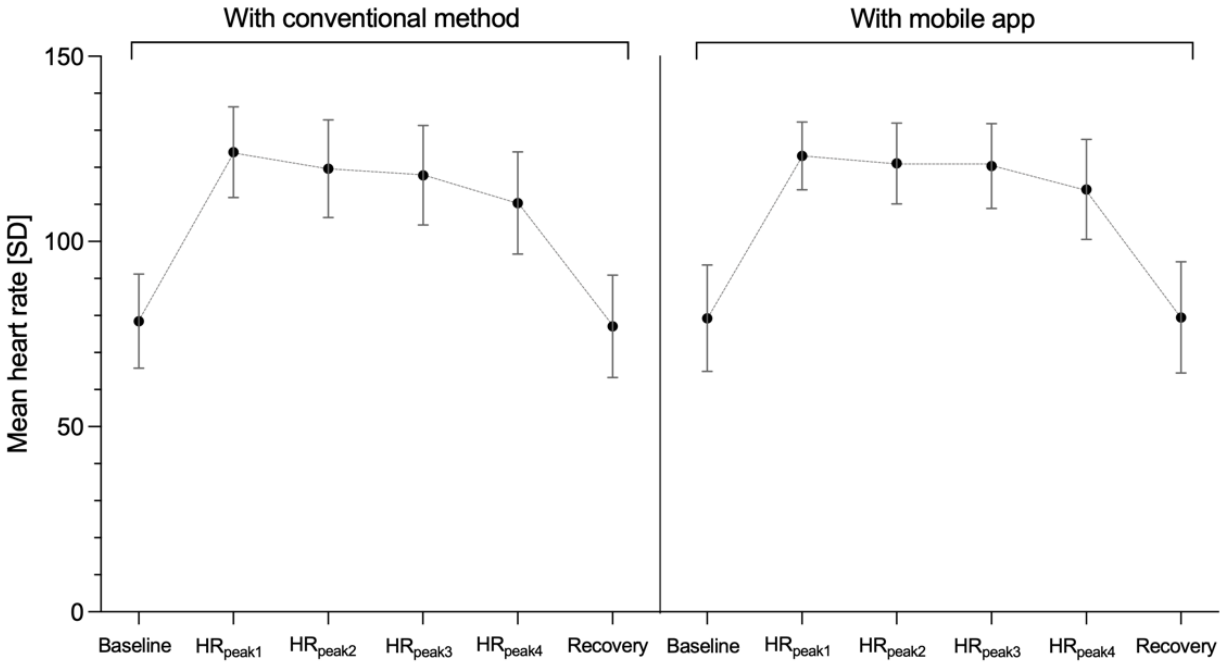
[a]bpm: beats per minute.

[b]Linear regression models adjusted for center and baseline heart rate using linear regression models (except for baseline heart rate adjusted for center only). A negative difference means that the average value is lower with than without PedAMINES.

[c]HR: heart rate.

[d]Highest value of HR_{peak} among the 4 drugs.

Figure 5. Mean heart rate (error bars=SD) in the baseline, the 4 consecutive heart rate peaks numbered according to the sequential prescription of each drug, and recovery time points over the course of the scenario. HR: heart rate.



Correlation Between Perceived Stress and Heart Rate

A strong correlation was found between STAI-Y and the VAS scores for the perceived stress by the participants before the intervention (Spearman *r*=0.72) and after the intervention (Spearman *r*=0.68). However, the STAI-Y and the VAS scores before the intervention were poorly correlated with HR at baseline (Spearman *r*=0.14 and *r*=0.22, respectively). The STAI-Y and VAS scores after the intervention were not correlated with the maximal HR_{peak} (Spearman *r*=0.02 and *r*=−0.04, respectively)

Discussion

Principal Findings

The unpredictable out-of-hospital environment and high-stake CPR situations lead to major stressful experiences for involved rescuers, which can adversely affect patient safety [5]. In this

randomized controlled trial, we report a significant increase in both perceived and physiological stress in paramedics during the preparation of emergency drugs for pediatric OHCA patients in a simulated model. However, perceived stress was 27% lower among those who used a specific mHealth app designed to facilitate drug preparation compared with those who only had access to conventional preparation methods. This result was observed irrespective of paramedics' years of experience, age, or gender, suggesting a worthwhile benefit of its use in the prehospital setting.

Quantification of the acute stress response of rescuers during CPR has been the subject of many studies, although no single stress-specific marker for its measurement has been definitively validated [70]. Thus, previous studies have used, alone or in conjunction, different surrogate stress markers (ie, biological, electrophysiological or psychological) to assess the relationship between stress and CPR performance [11]. Those considering subjective stress markers by means of self-reported questionnaires, such as the STAI, showed the strongest association with lower performance [6,11]. On the other hand, stress-coping strategies such as leadership and stress management training, cognitive aids (eg, checklists and algorithms), and mindfulness meditation have been shown to enhance CPR performance [5,71]. However, these studies mostly focused primarily on residents [11], a less experienced and likely more stress-prone population, thus providing little evidence regarding more experienced rescuers. Evidence regarding paramedics is scarce. Few studies in the field have yielded conflicting results. LeBlanc et al [56] compared paramedics' acute perceived and salivary cortisol stress responses and performance during simulated low- and high-stress scenarios. They observed impairments in some aspects of clinical performance in response to a high-stress scenario. Among these aspects, stress significantly impaired drug dosage calculation and was associated with a greater risk of error [14]. Interestingly, these authors reported a subjective mean STAI-rated overall stress of 46.1 for paramedics under high-stress conditions, corroborating our own findings. Conversely, Bjørshol et al [72] observed no performance impairment associated with higher self-reported stress in a randomized trial comparing paramedics under calm CPR conditions with those under stressful CPR conditions. Unfortunately, in these studies, no evaluation of interventions aimed at specifically reducing stress during CPR and studying their consecutive impact on performance was carried out. To date, the bulk of the research and development for interventions to reduce stress has mostly focused on long-lasting mental illnesses using mobile self-management apps in the field of health psychology [73]. To our knowledge, this trial is the first of its kind to report reduced perceived stress during multiple drug dosage calculations by advanced care paramedics supported by a mobile app in acute life-threatening situations. Prehospital dosing errors, although probably underreported due to failure to recognize them or reluctance to report them [70,74], affect approximately 56,000 children treated by EMS each year in the United States, with drugs administered outside of the proper dose range reported in up to 39.8% of more than 5500 children [31,75]. Facilitating emergency drug preparation at pediatric dosage while reducing stress might reduce these errors.

Acute stress situations elicit not only a psychological but also an adaptive, generally transient, physiological stress response carried out by regulatory pathways through the activation of the cardiovascular, endocrine, immune, and autonomic nervous systems [76,77]. The overall stress response is a combination of these complex and relatively independent pathways, without necessarily showing correlations with each other [57,58,78]. Additionally, there are interindividual differences in perceived and physiological stress responses in the same situation [76,78]. In this study, although perceived acute stress was reduced by the use of the app, such a reduction was not observed in HR that remained high in both groups during the scenario. This finding is consistent with that of the existing literature. At the individual level, the relationship between self-reported perceived stress and physiological stress measured by objective parameters such as HR has been shown to be somewhat inconsistent [57,79]. Clarke et al [80] examined the relationship between emergency medicine residents' self-reported stress before and after a simulation exercise and HR throughout the scenario. They observed that HR elevation alone correlated poorly with both perceived stress and clinical performance. Similar results were observed in another study where varying stress levels in simulated trauma scenarios elicited higher subjective stress and cortisol levels and poorer performance among residents exposed to high-stress conditions, whereas HR elevation was not significantly different between low- or high-stress conditions [81]. Among the reasons that may explain this phenomenon, it has been speculated that strenuous physical activity could be a confounding factor for HR, limiting its value as a marker of mental stress in acute situations [6,11]. In this study, physical activity was limited to the initial hands-on resuscitation period. Thereafter, physical activity was restricted to the preparation of drugs in close proximity to the patient, thus limiting exertion. Hence, although indicative of a state of mental stress in the absence of physical exertion, it appears from this study that sustained HR at high levels cannot be used alone as a physiological expression of perceived stress. An alternative explanation for the variation in HR observed over the course of the scenario may lie in the level of stress-induced physiological activation desirable for optimal task performance, with higher HR levels observed during drug preparations. This is consistent with the relationship between stress and performance that has been described theoretically by Yerkes-Dodson [82] and follows an inverted U-shaped curve. At low levels of stress, performance is poor but increases with increasing stress levels until a certain optimal point is reached, beyond which task performance and decision making abilities could become impaired, followed by a decline in performance quality [83]. Although we cannot comment on whether the observed HR levels represented the optimal HR zone for best performance, our results suggest that an adaptive physiological response to stress, at least in terms of HR levels, was similarly triggered in both study groups.

Limitations

This study had several limitations. First, as drug preparation times varied considerably from one participant to another and could have influenced any stress attenuation or exaggeration over time, we decided not to measure average HRs per drug. Second, the collection of additional physiological stress markers

was not carried out, which could also be seen as a limitation. However, measuring paramedics' cortisol levels would have required complex consideration of intra and interindividual circadian rhythms among day and night shift paramedics, as well as serial collection of blood, saliva, or sweat samples to capture baseline and in-scenario values, making the procedure impractical and potentially prone to preparation bias [84]. On the other hand, measuring paramedics' HR variability, which has been shown to be influenced by stress during emergency care and a reliable means of monitoring it [11,57,85], would require signal acquisition by an electrocardiogram. This could have potentially restrained paramedics' movements and involvement, thus hindering their ability to prepare the drugs according to usual practice. Considering the use of wearable photoplethysmography to measure pulse rate variability may be a promising alternative to overcome this limitation in future studies [86,87]. Another limitation of our trial includes its simulated setting, which raises the concern of its generalizability to real-life situations. High-fidelity simulation has been shown to be as realistic and stressful as real-life situations, both at the psychological and physiological levels [6,11,88]. Furthermore, simulation is an essential method for assessing research questions and technologies that cannot be answered during real CPR as the diversity among patients and their diseases makes such studies difficult to standardize in critical situations [89]. In addition, in the voluntary absence of correlational analysis in this study due to a nonexhaustive investigation of all stress response axes, our data does not allow causal inferences to be made about the relationship between psychological and physiological stress reactivity and emergency drugs preparation. They only provide a sketch in this direction, which requires further research. Finally, we acknowledge that this study did not distinguish whether and to what extent emergency drugs preparation had a greater influence on stress than the pediatric OHCA situation itself or vice versa.

Conclusions

In this nested randomized controlled trial, paramedics confronted with a high-stress pediatric OHCA scenario reported a lower level of perceived stress when using a mobile app designed to help pediatric drug preparation compared with conventional calculation methods. As acute perceived stress is associated with lower CPR performance in calculating drug dosages and as the children are particularly vulnerable to such errors, we suggest that dedicated medical mobile apps have the potential to reduce stress while improving medication safety. This could change the prehospital clinical practice in the area of pediatric emergency medicine. The next step would be to determine in real-life studies whether stress reduction through the app translates into improved clinical outcomes to contrast the results of this actual research.

Acknowledgments

We are grateful to Rosemary Sudan (Freelance Technical Editor) for providing editorial assistance. We are also grateful to Mr. Lorenz Vogt and Mr. Alexandre Glasner for their assistance in translation in some emergency medical services centers. These 3 people were compensated for their work.

This trial was funded by the Swiss National Science Foundation (grant SNSF_32003B_182374. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

ML was responsible for the literature search and reading the articles, performed data curing, interpreted the data, and drafted the manuscript together with JNS. LB, SM, and JNS were responsible for the concept and design of the study and critical review of the manuscript content, with the support of CC, LS, and FR. LB, SM, and JNS were responsible for the data acquisition. CC performed the data analyses. KH and FH provided technical and material support. FE was responsible for the development of the project software with the support of SM and JNS. CL and AG oversaw the development of the project software. SM was the trial coordinator and procured funding. JNS was responsible for the overall conduct of this study and oversaw the writing of this manuscript. All authors have contributed to, seen, and approved the final submitted version of the manuscript; had full access to all the data, including statistical reports and tables in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author (JNS) confirms that he had full access to the participants' data and endorsed the final responsibility for the submission. He further affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted, and that any deviations from the study plan have been explained. The collaborating members of the PedAMINES Prehospital Group are as follows (for the complete list of the group members, see Multimedia Appendix 2): Marec Saillant, EMS-p (Geneva Team Ambulances (GTA), Geneva, Switzerland), Renaud Grandjean EMS-p (SK Ambulances, Geneva, Switzerland), Annick Leuenberger EMS-p (Secours Ambulances Genève (SAG), Geneva, Switzerland), Pascal Donnet, EMS-p, and Philippe Hauck, EMS-p (Service de Sauvetage et de Lutte contre les Incendies Aéroportuaires (SSLIA), Geneva, Switzerland), Sébastien Pappalardo EMS-p (Service d'Incendie et de Secours (SIS), Geneva, Switzerland), Philippe Nidegger EMS-p (Ambulance Riviera, La Tour-de-Peilz, Switzerland), David Neel, Flight EMS-p (Air Zermatt SA, Zermatt, Switzerland), Stephan Steinhauser, MD (Höhere Fachschule für Rettungsberufe (HFRB), Zürich, Switzerland), Michel Ceschi, EMS-p, and Bruno Belli EMS-p (Servizio Ambulanza Locarnese e Valli (SALVA), Ticino, Switzerland), Sébastien Ottet EMS-p, and Wenceslao Garcia, MD (Ambulances du Sud Fribourgeois, Vaulruz, Switzerland),

Yoan Mollier EMS-p, Yves Vollenweider EMS-p, and Pierre Voumard, EMS-p (Service Communal de la Sécurité (SCS), Neuchâtel, Switzerland), Karine Corbat, EMS-p, and Philippe Robadey, EMS-p (Service de Protection et Sauvetage Lausanne (SPSL), Lausanne, Switzerland), Joël Bauer, EMS-p, and Cyril Berger, EMS-p (Centre de Secours et d'Urgences (CSU) Morges-Aubonne, Aubonne, Switzerland). These contributors coordinated the trial in their respective EMS centers. None of them received compensation for their role in the study.

Conflicts of Interest

Geneva University Hospitals are owners of the PedAMINES app. The app is currently commercially available on Google Play Store and App Store (Apple) for research and educational purposes. JNS, CL, AG, FE, and SM declare individual intellectual property rights on this app and, as employees of Geneva University Hospitals, indirect institutional rewarding through its commercialization (ie, without personal enrichment). The authors declare no other relationships or activities that could appear to have influenced the submitted work. All authors have completed the International Committee of Medical Journal Editors uniform disclosure form and declare no support from commercial entities for the submitted work, and no financial relationships with any commercial entities that might have an interest in the submitted work in the previous 3 years.

Multimedia Appendix 1
Subgroup analysis for the primary outcome by gender, age, and years since paramedic certification and State-Trait Anxiety Inventory perceived stress score.
[DOCX File , 17 KB - mhealth_v9i10e31748_app1.docx]

Multimedia Appendix 2
PedAMINES Prehospital Trial Group.
[DOCX File , 16 KB - mhealth_v9i10e31748_app2.docx]

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):139-596. [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
2. Yan S, Gan Y, Jiang N, Wang R, Chen Y, Luo Z, et al. The global survival rate among adult out-of-hospital cardiac arrest patients who received cardiopulmonary resuscitation: a systematic review and meta-analysis. Crit Care 2020 Feb 22;24(1):61 [FREE Full text] [doi: 10.1186/s13054-020-2773-2] [Medline: 32087741]
3. Genbrugge C, Eertmans W, Salcido DD. Monitor the quality of cardiopulmonary resuscitation in 2020. Curr Opin Crit Care 2020 Jun;26(3):219-227. [doi: 10.1097/MCC.0000000000000726] [Medline: 32332284]
4. Krage R, Zwaan L, Len LT, Kolenbrander MW, van Groeningen D, Loer SA, et al. Relationship between non-technical skills and technical performance during cardiopulmonary resuscitation: does stress have an influence? Emerg Med J 2017 Nov;34(11):728-733 [FREE Full text] [doi: 10.1136/emermed-2016-205754] [Medline: 28844039]
5. Groombridge CJ, Kim Y, Maini A, Smit DV, Fitzgerald MC. Stress and decision-making in resuscitation: A systematic review. Resuscitation 2019 Nov;144:115-122 [FREE Full text] [doi: 10.1016/j.resuscitation.2019.09.023] [Medline: 31562904]
6. Hunziker S, Semmer NK, Tschan F, Schuetz P, Mueller B, Marsch S. Dynamics and association of different acute stress markers with performance during a simulated resuscitation. Resuscitation 2012 May;83(5):572-578. [doi: 10.1016/j.resuscitation.2011.11.013] [Medline: 22115935]
7. Hunziker S, Laschinger L, Portmann-Schwarz S, Semmer NK, Tschan F, Marsch S. Perceived stress and team performance during a simulated resuscitation. Intensive Care Med 2011 Sep;37(9):1473-1479. [doi: 10.1007/s00134-011-2277-2] [Medline: 21695475]
8. Keitel A, Ringleb M, Schwartges I, Weik U, Picker O, Stockhorst U, et al. Endocrine and psychological stress responses in a simulated emergency situation. Psychoneuroendocrinology 2011 Jan;36(1):98-108. [doi: 10.1016/j.psyneuen.2010.06.011] [Medline: 20650570]
9. Cumming S, Harris L. The impact of anxiety on the accuracy of diagnostic decision-making. Stress Health 2001 Oct 27;17(5):281-286 [FREE Full text] [doi: 10.1002/smi.909]
10. Schull MJ, Ferris LE, Tu JV, Hux JE, Redelmeier DA. Problems for clinical judgement: 3. Thinking clearly in an emergency. Can Med Asso J 2001 Apr 17;164(8):1170-1175 [FREE Full text] [Medline: 11338805]
11. Vincent A, Semmer NK, Becker C, Beck K, Tschan F, Bobst C, et al. Does stress influence the performance of cardiopulmonary resuscitation? A narrative review of the literature. J Crit Care 2021 Jun;63:223-230. [doi: 10.1016/j.jcrc.2020.09.020] [Medline: 33046274]

12. Lauria MJ, Gallo IA, Rush S, Brooks J, Spiegel R, Weingart SD. Psychological skills to improve emergency care providers' performance under stress. Ann Emerg Med 2017 Dec;70(6):884-890. [doi: 10.1016/j.annemergmed.2017.03.018] [Medline: 28460863]
13. Dias R, Neto A. Stress levels during emergency care: A comparison between reality and simulated scenarios. J Crit Care 2016 Jun;33:8-13. [doi: 10.1016/j.jcrc.2016.02.010] [Medline: 26987261]
14. LeBlanc VR, MacDonald RD, McArthur B, King K, Lepine T. Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. Prehosp Emerg Care 2005;9(4):439-444. [doi: 10.1080/10903120500255255] [Medline: 16263679]
15. Brindley PG, O'Dochartaigh D, Volney C, Ryan S, Douma MJ. Time delays associated with vasoactive medication preparation and delivery in simulated patients at risk of cardiac arrest. J Crit Care 2017 Aug;40:149-153. [doi: 10.1016/j.jcrc.2017.04.003] [Medline: 28402925]
16. Flannery AH, Parli SE. Medication errors in cardiopulmonary arrest and code-related situations. Am J Crit Care 2016 Jan;25(1):12-20. [doi: 10.4037/ajcc2016190] [Medline: 26724288]
17. Kaufmann J, Laschat M, Wappler F. Medication errors in pediatric emergencies: a systematic analysis. Dtsch Arztebl Int 2012 Sep;109(38):609-616 [FREE Full text] [doi: 10.3238/arztebl.2012.0609] [Medline: 23093991]
18. Moreira ME, Hernandez C, Stevens AD, Jones S, Sande M, Blumen JR, et al. Color-coded prefilled medication syringes decrease time to delivery and dosing error in simulated emergency department pediatric resuscitations. Ann Emerg Med 2015 Aug;66(2):97-106 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.12.035] [Medline: 25701295]
19. McDowell SE, Ferner HS, Ferner RE. The pathophysiology of medication errors: how and where they arise. Br J Clin Pharmacol 2009 Jun;67(6):605-613 [FREE Full text] [doi: 10.1111/j.1365-2125.2009.03416.x] [Medline: 19594527]
20. Stucky ER, American Academy of Pediatrics Committee on Drugs, American Academy of Pediatrics Committee on Hospital Care. Prevention of medication errors in the pediatric inpatient setting. Pediatrics 2003 Aug;112(2):431-436. [doi: 10.1542/peds.112.2.431] [Medline: 12897304]
21. Cushman JT, Fairbanks RJ, O'Gara KG, Crittenden CN, Pennington EC, Wilson MA, et al. Ambulance personnel perceptions of near misses and adverse events in pediatric patients. Prehosp Emerg Care 2010;14(4):477-484 [FREE Full text] [doi: 10.3109/10903127.2010.497901] [Medline: 20662679]
22. Luten R, Wears RL, Broselow J, Croskerry P, Joseph MM, Frush K. Managing the unique size-related issues of pediatric resuscitation: reducing cognitive load with resuscitation aids. Acad Emerg Med 2002 Aug;9(8):840-847 [FREE Full text] [Medline: 12153892]
23. Seidel JS, Henderson DP, Ward P, Wayland BW, Ness B. Pediatric prehospital care in urban and rural areas. Pediatrics 1991 Oct;88(4):681-690. [Medline: 1896270]
24. Kaji AH, Gausche-Hill M, Conrad H, Young KD, Koenig WJ, Dorsey E, et al. Emergency medical services system changes reduce pediatric epinephrine dosing errors in the prehospital setting. Pediatrics 2006 Oct;118(4):1493-1500. [doi: 10.1542/peds.2006-0854] [Medline: 17015540]
25. Shah MN, Cushman JT, Davis CO, Bazarian JJ, Auinger P, Friedman B. The epidemiology of emergency medical services use by children: an analysis of the National Hospital Ambulatory Medical Care Survey. Prehosp Emerg Care 2008;12(3):269-276 [FREE Full text] [doi: 10.1080/10903120802100167] [Medline: 18584491]
26. Su E, Schmidt TA, Mann NC, Zechnich AD. A randomized controlled trial to assess decay in acquired knowledge among paramedics completing a pediatric resuscitation course. Acad Emerg Med 2000 Jul;7(7):779-786 [FREE Full text] [doi: 10.1111/j.1553-2712.2000.tb02270.x] [Medline: 10917328]
27. Benjamin L, Frush K, Shaw K, Shook JE, Snow SK. Pediatric medication safety in the emergency department. Pediatrics 2018 Mar 01;141(3):e20174066. [doi: 10.1542/peds.2017-4066]
28. Gonzales K. Medication administration errors and the pediatric population: a systematic search of the literature. J Pediatr Nurs 2010 Dec;25(6):555-565. [doi: 10.1016/j.pedn.2010.04.002] [Medline: 21035020]
29. Kaushal R, Bates DW, Landrigan C, McKenna KJ, Clapp MD, Federico F, et al. Medication errors and adverse drug events in pediatric inpatients. J Am Med Assoc 2001 Apr 25;285(16):2114-2120. [Medline: 11311101]
30. Marcin JP, Dharmar M, Cho M, Seifert LL, Cook JL, Cole SL, et al. Medication errors among acutely ill and injured children treated in rural emergency departments. Ann Emerg Med 2007 Oct;50(4):361-367. [doi: 10.1016/j.annemergmed.2007.01.020] [Medline: 17433496]
31. Hoyle JD, Crowe RP, Bentley MA, Beltran G, Fales W. Pediatric prehospital medication dosing errors: a national survey of paramedics. Prehosp Emerg Care 2017;21(2):185-191. [doi: 10.1080/10903127.2016.1227001] [Medline: 28257249]
32. Lammers R, Byrwa M, Fales W. Root causes of errors in a simulated prehospital pediatric emergency. Acad Emerg Med 2012 Jan;19(1):37-47 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01252.x] [Medline: 22251191]
33. Hall C, Robertson D, Rolfe M, Pascoe S, Passey ME, Pit SW. Do cognitive aids reduce error rates in resuscitation team performance? Trial of emergency medicine protocols in simulation training (TEMPIST) in Australia. Hum Resour Health 2020 Jan 08;18(1):1 [FREE Full text] [doi: 10.1186/s12960-019-0441-x] [Medline: 31915029]
34. Gausche-Hill M, Krug S, Wright J. Emergency medical services (EMS) 2050: a vision for the future of pediatric prehospital care. Prehosp Emerg Care 2021;25(1):91-94. [doi: 10.1080/10903127.2020.1734123] [Medline: 32091291]

35. Roncero A, Marques G, Sainz-De-Abajo B, Martín-Rodríguez F, Vegas C, Garcia-Zapirain B, et al. Mobile health apps for medical emergencies: systematic review. JMIR Mhealth Uhealth 2020 Dec 11;8(12):e18513 [FREE Full text] [doi: 10.2196/18513] [Medline: 33306037]
36. Siebert JN, Ehrler F, Combescure C, Lacroix L, Haddad K, Sanchez O, et al. A mobile device app to reduce time to drug delivery and medication errors during simulated pediatric cardiopulmonary resuscitation: a randomized controlled trial. J Med Internet Res 2017 Feb 01;19(2):e31 [FREE Full text] [doi: 10.2196/jmir.7005] [Medline: 28148473]
37. Siebert JN, Ehrler F, Combescure C, Lovis C, Haddad K, Hugon F, PedAMINES Trial Group. A mobile device application to reduce medication errors and time to drug delivery during simulated paediatric cardiopulmonary resuscitation: a multicentre, randomised, controlled, crossover trial. Lancet Child Adolesc Health 2019 May;3(5):303-311. [doi: 10.1016/S2352-4642(19)30003-3] [Medline: 30797722]
38. Siebert JN, Bloudeau L, Combescure CB, Haddad K, Hugon F, Suppan L, Pediatric Accurate Medication in Emergency Situations (PedAMINES) Prehospital Group. Effect of a mobile app on prehospital medication errors during simulated pediatric resuscitation: a randomized clinical trial. JAMA Netw Open 2021 Aug 02;4(8):e2123007 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.23007] [Medline: 34459905]
39. Siebert JN, Bloudeau L, Ehrler F, Combescure C, Haddad K, Hugon F, et al. A mobile device app to reduce prehospital medication errors and time to drug preparation and delivery by emergency medical services during simulated pediatric cardiopulmonary resuscitation: study protocol of a multicenter, prospective, randomized controlled trial. Trials 2019 Nov 20;20(1):634 [FREE Full text] [doi: 10.1186/s13063-019-3726-4] [Medline: 31747951]
40. Stenner K, van Even S, Collen A. Early adopters of paramedic prescribing: a qualitative study. Br Paramed J 2019 Dec 01;4(3):57 [FREE Full text] [doi: 10.29045/14784726.2019.12.4.3.57] [Medline: 33447155]
41. Kiguchi T, Okubo M, Nishiyama C, Maconochie I, Ong ME, Kern KB, et al. Out-of-hospital cardiac arrest across the World: First report from the International Liaison Committee on Resuscitation (ILCOR). Resuscitation 2020 Jul;152:39-49. [doi: 10.1016/j.resuscitation.2020.02.044] [Medline: 32272235]
42. Ehrler F, Siebert JN. PedAMINES: a disruptive mHealth app to tackle paediatric medication errors. Swiss Med Wkly 2020 Aug 24;150:w20335 [FREE Full text] [doi: 10.4414/smw.2020.20335] [Medline: 32920794]
43. Oviatt S. Human-centered design meets cognitive load theory: designing interfaces that help people think. In: Proceedings of the 14th ACM international conference on Multimedia. New York, NY, United States: Association for Computing Machinery; 2006 Oct 23 Presented at: MM '06: 14th ACM international conference on Multimedia; October 23 - 27, 2006; Santa Barbara, CA, USA p. 871-880. [doi: 10.1145/1180639.1180831]
44. Durrani S, Durrani Q. Applying cognitive psychology to user interfaces. In: Proceedings of the First International Conference on Intelligent Human Computer Interaction. New Delhi: Springer India; 2009 Presented at: First International Conference on Intelligent Human Computer Interaction; January 20–23, 2009; Allahabad, India p. 156-168.
45. Harper S, Michailidou E, Stevens R. Toward a definition of visual complexity as an implicit measure of cognitive load. ACM Trans Appl Percept 2009 Feb 15;6(2):1-18. [doi: 10.1145/1498700.1498704]
46. MacKenzie IS. Fitts' law as a research and design tool in human-computer interaction. Hum Comput Interact 2009 Nov 11;7(1):91-139. [doi: 10.1207/s15327051hci0701_3]
47. Spillers F. Progressive disclosure: the glossary of human computer interaction. Denmark: Interaction Design Foundation. 2016. URL: https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/progressive-disclosure [accessed 2021-09-19]
48. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
49. Brooke J. SUS- 'A quick and dirty' usability scale. In: Usability Evaluation in Industry. London, UK: Taylor and Francis; 1996:4-7.
50. Sousa VE, Lopez KD. Towards usable e-Health. A systematic review of usability questionnaires. Appl Clin Inform 2017 Dec 10;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]
51. Gauthier J, Bouchard S. Adaptation canadienne-française de la forme révisée du State–Trait Anxiety Inventory de Spielberger. Can J Behav Sci / Revue canadienne des sciences du comportement 1993 Oct;25(4):559-578. [doi: 10.1037/h0078881]
52. Spielberger CD. State-trait anxiety inventory for adults™. Mind Garden, Inc. URL: http://www.mindgarden.com/products/staisad.htm [accessed 2021-09-19]
53. Spielberger CD. Manual for the State-Trait Anxiety Inventory STAI (form Y) (“self-evaluation questionnaire”). Palo Alto, CA, USA: Consulting Psychologists Press; 1983.
54. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. State-trait anxiety inventory for adults: self-evaluation questionnaire (STAI Form Y-1 and Form Y-2). Mind Garden, Redwood City, CA. 1968. URL: https://oml.eular.org/sysModules/obxOml/docs/ID_150/State-Trait-Anxiety-Inventory.pdf [accessed 2021-09-22]
55. Harvey A, Nathens AB, Bandiera G, Leblanc VR. Threat and challenge: cognitive appraisal and stress responses in simulated trauma resuscitations. Med Educ 2010 Jun;44(6):587-594. [doi: 10.1111/j.1365-2923.2010.03634.x] [Medline: 20604855]
56. Leblanc VR, Regehr C, Tavares W, Scott AK, Macdonald R, King K. The impact of stress on paramedic performance during simulated critical events. Prehosp Disaster Med 2012 Aug;27(4):369-374. [doi: 10.1017/S1049023X12001021] [Medline: 22831965]

57. Ghazali DA, Darmian-Rafei I, Nadolny J, Sosner P, Ragot S, Oriot D. Evaluation of stress response using psychological, biological, and electrophysiological markers during immersive simulation of life threatening events in multidisciplinary teams. Aust Crit Care 2018 Jul;31(4):226-233. [doi: 10.1016/j.aucc.2017.07.001] [Medline: 28756943]
58. Bong C, Fraser K, Oriot D. Cognitive load and stress in simulation. In: Grant V, Cheng A, editors. Comprehensive Healthcare Simulation: Pediatrics. Cham: Springer International Publishing; Jun 16, 2016:3-17.
59. Spielberger C, Gorsuch R, Lushene R, Vagg R, Jacobs G. State-Trait Anxiety Inventory for Adults. Manual, Instrument, and Scoring Guide. Redwood City, CA, USA: Mind Garden, Inc; 1983.
60. Lesage F, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med (Lond) 2012 Dec;62(8):600-605 [FREE Full text] [doi: 10.1093/occmed/kqs140] [Medline: 22965867]
61. Furedy JJ, Szabo A, Péronnet F. Effects of psychological and physiological challenges on heart rate, T-wave amplitude, and pulse-transit time. Int J Psychophysiol 1996;22(3):173-183. [doi: 10.1016/0167-8760(96)00025-6] [Medline: 8835625]
62. Rider BC, Conger SA, Ditzenberger GL, Besteman SS, Bouret CM, Coughlin AM. Examining the accuracy of the Polar A360 Monitor. J Strength Cond Res 2021 Aug 01;35(8):2165-2169. [doi: 10.1519/JSC.0000000000003136] [Medline: 34398076]
63. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
64. Baldwin S, Bennell C, Andersen JP, Semple T, Jenkins B. Stress-activity mapping: physiological responses during general duty police encounters. Front Psychol 2019;10:2216 [FREE Full text] [doi: 10.3389/fpsyg.2019.02216] [Medline: 31636582]
65. 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]
66. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
67. International Conference on Harmonisation E9 Expert Working Group. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. Stat Med 1999 Aug 15;18(15):1905-1942. [Medline: 10532877]
68. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
69. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, International Network for Simulation-based Pediatric Innovation, Research,Education (INSPIRE) Reporting Guidelines Investigators. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. Simul Healthc 2016 Aug;11(4):238-248. [doi: 10.1097/SIH.0000000000000150] [Medline: 27465839]
70. Ramadanov N, Klein R, Schumann U, Aguilar AD, Behringer W. Factors, influencing medication errors in prehospital care: a retrospective observational study. Medicine (Baltimore) 2019 Dec;98(49):e18200 [FREE Full text] [doi: 10.1097/MD.0000000000018200] [Medline: 31804342]
71. Hunziker S, Pagani S, Fasler K, Tschan F, Semmer NK, Marsch S. Impact of a stress coping strategy on perceived stress levels and performance during a simulated cardiopulmonary resuscitation: a randomized controlled trial. BMC Emerg Med 2013 Apr 22;13(8):1-9 [FREE Full text] [doi: 10.1186/1471-227X-13-8] [Medline: 23607331]
72. Bjørshol CA, Myklebust H, Nilsen KL, Hoff T, Bjørkli C, Illguth E, et al. Effect of socioemotional stress on the quality of cardiopulmonary resuscitation during advanced life support in a randomized manikin study. Crit Care Med 2011 Feb;39(2):300-304. [doi: 10.1097/CCM.0b013e3181ffe100] [Medline: 21076285]
73. Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mHealth: systematic search in App stores. JMIR Mhealth Uhealth 2018 May 09;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
74. Walker D, Moloney C, SueSee B, Sharples R. Contributing factors that influence medication errors in the prehospital paramedic environment: a mixed-method systematic review protocol. BMJ Open 2019 Dec 23;9(12):e034094 [FREE Full text] [doi: 10.1136/bmjopen-2019-034094] [Medline: 31874897]
75. Hoyle JD, Davis AT, Putman KK, Trytko JA, Fales WD. Medication dosing errors in pediatric patients treated by emergency medical services. Prehosp Emerg Care 2012;16(1):59-66. [doi: 10.3109/10903127.2011.614043] [Medline: 21999707]
76. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
77. Epel E, Crosswell A, Mayer S, Prather A, Slavich G, Puterman E, et al. More than a feeling: a unified view of stress measurement for population science. Front Neuroendocrinol 2018 Apr;49:146-169 [FREE Full text] [doi: 10.1016/j.yfrne.2018.03.001] [Medline: 29551356]
78. Can YS, Chalabianloo N, Ekiz D, Ersoy C. Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. Sensors 2019 Apr 18;19(8):1849. [doi: 10.3390/s19081849]

79. Oldehinkel AJ, Ormel J, Bosch NM, Bouma EM, Van Roon AM, Rosmalen JG, et al. Stressed out? Associations between perceived and physiological stress responses in adolescents: the TRAILS study. Psychophysiology 2011 Apr;48(4):441-452. [doi: 10.1111/j.1469-8986.2010.01118.x] [Medline: 21361964]
80. Clarke S, Horeczko T, Cotton D, Bair A. Heart rate, anxiety and performance of residents during a simulated critical clinical encounter: a pilot study. BMC Med Educ 2014 Jul 27;14:153 [FREE Full text] [doi: 10.1186/1472-6920-14-153] [Medline: 25064689]
81. Harvey A, Bandiera G, Nathens AB, LeBlanc VR. Impact of stress on resident performance in simulated trauma scenarios. J Trauma Acute Care Surg 2012 Feb;72(2):497-503. [doi: 10.1097/ta.0b013e31821f84be] [Medline: 22439221]
82. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 1908 Nov;18(5):459-482. [doi: 10.1002/cne.920180503]
83. Frazier SE, Parker SH. Measurement of physiological responses to acute stress in multiple occupations: a systematic review and implications for front line healthcare providers. Transl Behav Med 2019 Jan 01;9(1):158-166. [doi: 10.1093/tbm/iby019] [Medline: 29522140]
84. Zamkah A, Hui T, Andrews S, Dey N, Shi F, Sherratt RS. Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors. Biosensors (Basel) 2020 Apr 16;10(4):40 [FREE Full text] [doi: 10.3390/bios10040040] [Medline: 32316280]
85. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
86. Mejía-Mejía E, Budidha K, Abay TY, May JM, Kyriacou PA. Heart rate variability (HRV) and pulse rate variability (PRV) for the assessment of autonomic responses. Front Physiol 2020;11:779 [FREE Full text] [doi: 10.3389/fphys.2020.00779] [Medline: 32792970]
87. Pinheiro N, Couceiro R, Henriques J, Muehlsteff J, Quintal I, Goncalves L, et al. Can PPG be used for HRV analysis? Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:2945-2949. [doi: 10.1109/EMBC.2016.7591347] [Medline: 28268930]
88. Ghazali DA, Breque C, Sosner P, Lesbordes M, Chavagnat J, Ragot S, et al. Stress response in the daily lives of simulation repeaters. A randomized controlled trial assessing stress evolution over one year of repetitive immersive simulations. PLoS One 2019;14(7):e0220111 [FREE Full text] [doi: 10.1371/journal.pone.0220111] [Medline: 31344077]
89. Cheng A, Auerbach M, Hunt EA, Chang TP, Pusic M, Nadkarni V, et al. Designing and conducting simulation-based research. Pediatrics 2014 Jun;133(6):1091-1101. [doi: 10.1542/peds.2013-3267] [Medline: 24819576]

Abbreviations

CPR: cardiopulmonary resuscitation
EMS: emergency medical services
mHealth: mobile health
OHCA: out-of-hospital cardiac arrest
PedAMINES: Pediatric Accurate Medication in Emergency Situations
STAI: State-Trait Anxiety Inventory
VAS: visual analogue scale

Edited by G Eysenbach; submitted 02.07.21; peer-reviewed by A Benis, Y Kim, B Nievas Soriano; comments to author 23.07.21; revised version received 05.08.21; accepted 08.08.21; published 07.10.21.

Please cite as:
Lacour M, Bloudeau L, Combescure C, Haddad K, Hugon F, Suppan L, Rodieux F, Lovis C, Gervaix A, Ehrler F, Manzano S, Siebert JN, PedAMINES Prehospital Group
Impact of a Mobile App on Paramedics' Perceived and Physiologic Stress Response During Simulated Prehospital Pediatric Cardiopulmonary Resuscitation: Study Nested Within a Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(10):e31748
URL: https://mhealth.jmir.org/2021/10/e31748
doi:10.2196/31748
PMID:34617916

Original Paper

Smartphone-Based Artificial Intelligence–Assisted Prediction for Eyelid Measurements: Algorithm Development and Observational Validation Study

Hung-Chang Chen[1,2*], MD; Shin-Shi Tzeng[1,2*], MD; Yen-Chang Hsiao[1,2*], MD; Ruei-Feng Chen[3*], MD; Erh-Chien Hung[1,2*], MD; Oscar K Lee[4,5*], MD, PhD

[1]Department of Plastic and Reconstructive Surgery, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[2]College of Medicine, Chang Gung University, Taoyuan, Taiwan

[3]Vendome Plastic Clinic, Taipei, Taiwan

[4]Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[5]Department of Orthopedics, China Medical University Hospital, Taichung, Taiwan

[*]all authors contributed equally

Corresponding Author:
Oscar K Lee, MD, PhD
Institute of Clinical Medicine
National Yang Ming Chiao Tung University
No 155, Section 2, Li-Nong Street, Beitou District
Taipei, 112
Taiwan
Phone: 886 2 28757391
Email: oscarlee9203@gmail.com

Abstract

Background: Margin reflex distance 1 (MRD1), margin reflex distance 2 (MRD2), and levator muscle function (LF) are crucial metrics for ptosis evaluation and management. However, manual measurements of MRD1, MRD2, and LF are time-consuming, subjective, and prone to human error. Smartphone-based artificial intelligence (AI) image processing is a potential solution to overcome these limitations.

Objective: We propose the first smartphone-based AI-assisted image processing algorithm for MRD1, MRD2, and LF measurements.

Methods: This observational study included 822 eyes of 411 volunteers aged over 18 years from August 1, 2020, to April 30, 2021. Six orbital photographs (bilateral primary gaze, up-gaze, and down-gaze) were taken using a smartphone (iPhone 11 Pro Max). The gold-standard measurements and normalized eye photographs were obtained from these orbital photographs and compiled using AI-assisted software to create MRD1, MRD2, and LF models.

Results: The Pearson correlation coefficients between the gold-standard measurements and the predicted values obtained with the MRD1 and MRD2 models were excellent (r=0.91 and 0.88, respectively) and that obtained with the LF model was good (r=0.73). The intraclass correlation coefficient demonstrated excellent agreement between the gold-standard measurements and the values predicted by the MRD1 and MRD2 models (0.90 and 0.84, respectively), and substantial agreement with the LF model (0.69). The mean absolute errors were 0.35 mm, 0.37 mm, and 1.06 mm for the MRD1, MRD2, and LF models, respectively. The 95% limits of agreement were –0.94 to 0.94 mm for the MRD1 model, –0.92 to 1.03 mm for the MRD2 model, and –0.63 to 2.53 mm for the LF model.

Conclusions: We developed the first smartphone-based AI-assisted image processing algorithm for eyelid measurements. MRD1, MRD2, and LF measures can be taken in a quick, objective, and convenient manner. Furthermore, by using a smartphone, the examiner can check these measurements anywhere and at any time, which facilitates data collection.

(JMIR Mhealth Uhealth 2021;9(10):e32444) doi:10.2196/32444

KEYWORDS

artificial intelligence; AI; deep learning; margin reflex distance 1; margin reflex distance 2; levator muscle function; smartphone; measurement; eye; prediction; processing; limit; image; algorithm; observational

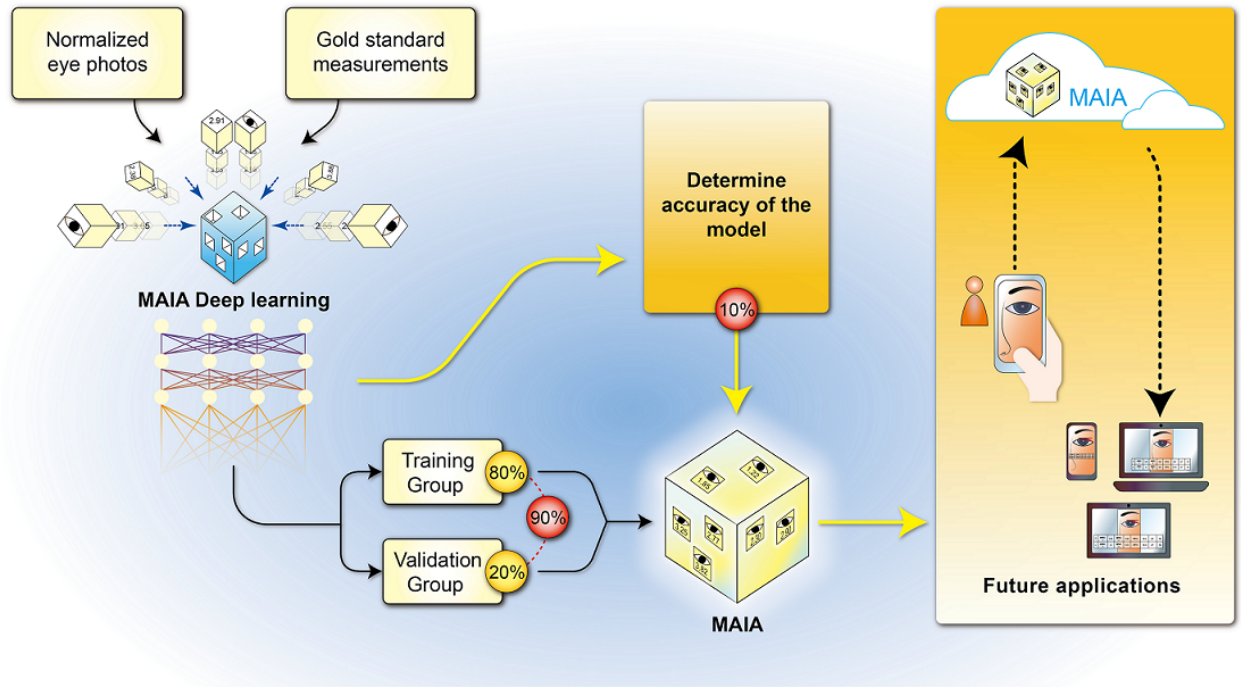
Introduction

Margin reflex distance 1 (MRD1), margin reflex distance 2 (MRD2), and levator muscle function (LF) are crucial for the evaluation and management of ptosis, a condition in which the upper eyelid droops over the eye [1]. MRD1 is defined as the distance between the upper eyelid margin and the center of the pupillary light reflex, whereas MRD2 is defined as the distance between the lower eyelid margin and the center of the pupillary light reflex. The LF is defined as the distance the upper eyelid margin moves from down-gaze to up-gaze without any eyebrow movement. According to a normal MRD1 of 4-5 mm, ptosis can be classified as mild (MRD1: 3-4 mm), moderate (MRD1: 2-3 mm), or severe (MRD1: 0-2 mm).

Manual measurements of MRD1, MRD2, and LF are time-consuming, subjective, and prone to human error [2]. More accurate measurements may be determined using a slit-lamp biomicroscope [3], and several automatic and semiautomatic photographic analysis techniques have been developed to obtain relatively objective measurements of MRD1 and MRD2 [4-6]. However, in these studies, a standardized environment is required for taking the photographs. The Volk Eye Check System measures MRD1 automatically using photographs taken by an integrated camera; however, this system tends to overestimate MRD1 in patients with ptosis [7]. To the best of our knowledge, there are no automatic photographic analysis techniques available for LF measurements.

A smartphone is more portable and convenient than a traditional photography room and slit-lamp biomicroscope. Artificial intelligence (AI), specifically deep learning (also known as deep neural network learning), is a new and popular area of research that yields impressive results and is growing rapidly. Smartphone-based deep learning image processing is a potential solution to overcome these limitations for measurements of MRD1, MRD2, and LF (Figure 1). We developed the first smartphone-based AI-assisted image processing algorithm for MRD1, MRD2, and LF measurements, which was validated in comparison with gold-standard measurements in an observational study.

Figure 1. Smartphone-based artificial intelligence–assisted prediction of eyelid measurements. MAIA: medical artificial intelligence assistant (Muen Biomedical and Optoelectronic Technologist, Inc; Version 1.2.0).



Methods

Study Design

This observational study included 822 eyes of 411 volunteers aged over 18 years who were referred to a plastic surgery clinic for blepharoplasty between August 1, 2020, and April 30, 2021. The study was approved by the institutional review board of Chang Gung Memorial Hospital. Volunteers with eyelid defects or deformities, history of corneal injury, enophthalmos, and anophthalmia were excluded.

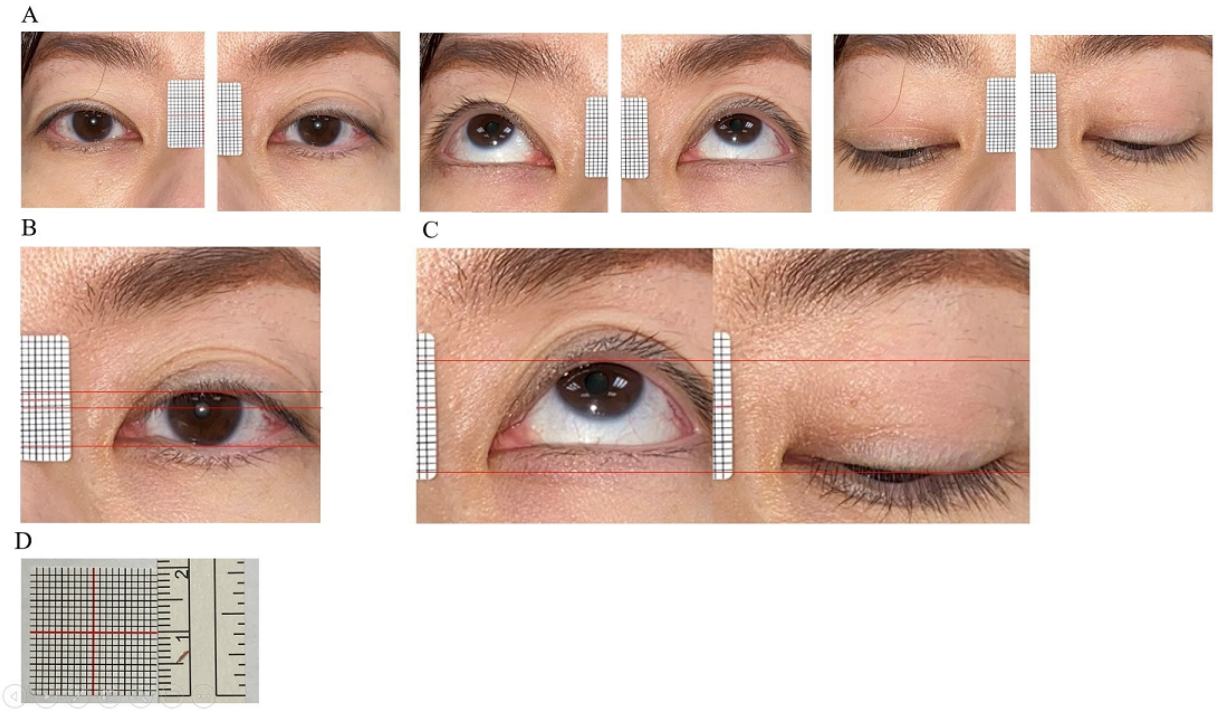
Photographs and Gold-Standard Measurements (Actual Values)

A 20×20-mm scale was placed on the nasal dorsum as a reference. The scale was only necessary for gold-standard measurements and was not required for deep learning model training or for determining the accuracy of the model.

Bilateral orbital photographs of each patient (standing or sitting; total 6 photographs including bilateral primary gaze, up-gaze, and down-gaze) were taken using a smartphone (iPhone 11 Pro Max, with flash and a 1:1 ratio) held at the same level between the patient's eyes at a distance of approximately 20-30 cm, which simulated the distance between the patient and doctor when the doctor uses a handheld ruler to measure MRD1, MRD2, and LF in the clinic.

The photographs were magnified on the computer, and MRD1, MRD2, and LF measurements were taken by two doctors independently (measured in increments of 0.25 mm). The doctors drew a horizontal line across the upper eyelid margin, light reflex, and lower eyelid margin to the 20×20-mm scale to obtain the MRD1, MRD2, and LF measurements. The mean value of measurements obtained by the two doctors was taken as the gold-standard measurement (actual value), which served as the input data for deep learning model training (Figure 2).

Figure 2. Photographs and gold-standard measurements (real values) (A) Six orbital photographs, including bilateral primary gaze, up-gaze, and down-gaze, were taken by a smartphone. (B) The primary gaze photograph was then magnified for margin reflex distance 1 (MRD1) and MRD2 measurements. (C) The up-gaze and down-gaze photographs were then magnified for levator muscle function (LF) measurements. (D) A 20×20-mm scale.

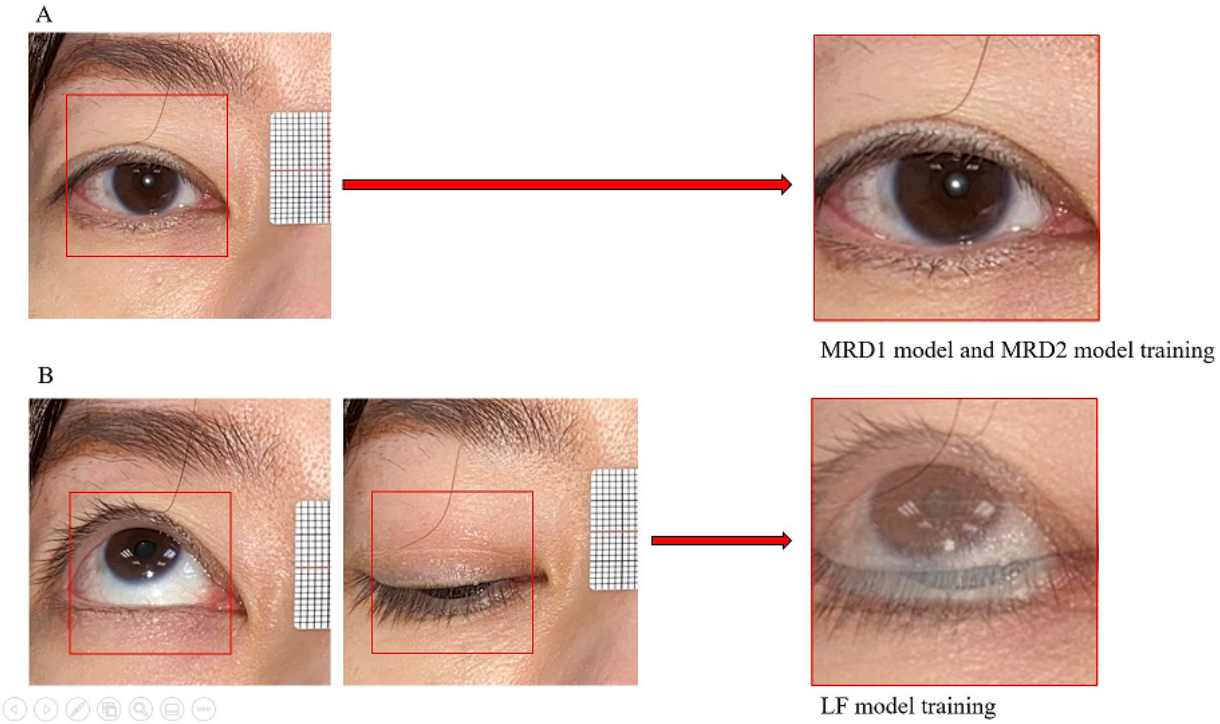


Usually, in ptotic eyelids without a corneal light reflex, the distance (in millimeters) that the eyelid must be lifted is recorded as a negative value, which is the MRD1. However, the distance the examiner lifts the eyelid is very subjective and therefore cannot be used as a gold-standard measurement. Accordingly, in this study, all MRD1 measurements in ptotic eyelids without a corneal light reflex were set to 0.

Photograph Normalization

Segmentation of primary-gaze orbital photographs (a square region around the light reflex as a center) was automatically performed by our software algorithm. We used LabelImage [8] to label the pupil light reflex location (X, Y), and then built a MobilenetV2 [9] model to train a regression model that can find a pupil light reflex coordinate. Square orbital pictures were automatically cropped (image size/4) using the light reflex coordinate extension after determining the light reflex coordinate. These segmented square photographs were considered the "normalized eye photographs," which were used as input data for MRD1 and MRD2 deep learning model training. Segmentations of up- and down-gaze orbital photographs were automatically merged into one photograph by our software algorithm. These segmented and merged photographs were considered the "normalized eye photographs" for LF deep learning model training (Figure 3).

Figure 3. Photograph normalization. (A) Autosegmentation of primary-gaze orbital photographs. These photographs are considered the "normalized eye photographs" for margin reflex distance 1 (MRD1) and MRD2 model training. (B) Autosegmentation of up- and down-gaze orbital photographs, which were then merged into one photograph. These photographs are considered the "normalized eye photographs" for levator muscle function (LF) model training.



Model Training: Image Analysis by Automatic Deep Learning Software

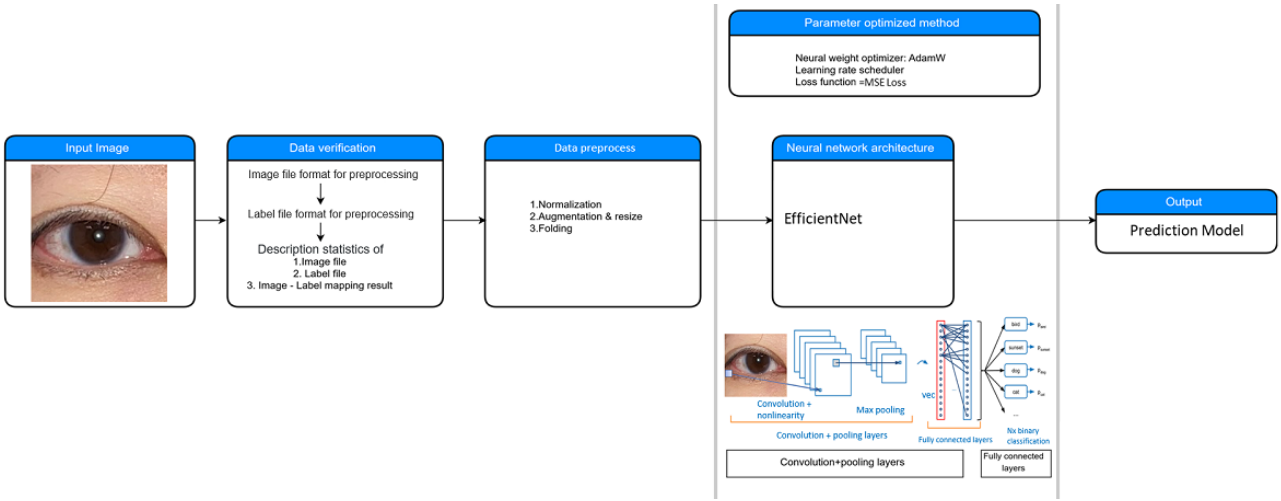
The normalized eye photographs and gold-standard measurements of MRD1, MRD2, and LF were compiled using medical artificial intelligence assistant (MAIA) software (Muen Biomedical and Optoelectronic Technologist Inc; Version 1.2.0) to analyze the image features and classify different situations. MAIA software automatically optimizes parameters for training models, including multiple convolutional neural network (CNN) models such as SE ResNet and EfficientNet [10,11].

The input data were processed with the following steps: (1) images were resized into 256×256 using a bilinear interpolation method, (2) images were augmented using horizontal flip and randomly rotated using the albumentations method [12], and (3) five-fold cross-validation was used to estimate the performance of the models.

The neural network architecture chooses an optimal network for memory consumption. We added the dropout function and applied different data augmentation methods to prevent the model from overfitting to our dataset [13,14]. The dropout rate was set from 0.25 to 0.5 for regularization. We then trained the model using minibatches of size 32, which were selected based on memory consumption [15]. The learning rate was tuned based on cosine annealing and a one-cycle policy strategy [16,17]. Using the cosine annealing schedule, the model repeatedly fits the gradient to the local minimum. The network was trained end-to-end using the Adam optimization algorithm, which optimized the mean square error as a loss function [18]. Lastly, we ensembled all of the models using the average output of the model to obtain a more robust result, minimize the bias of prediction error, and improve the prediction accuracy of the CNN models (Figure 4).

MAIA software was used with Python 3.x and PyTorch 1.1.x for Windows 10.

Figure 4. The convolutional neural network architecture of medical artificial intelligence assistant (MAIA) software.



Model Performance Evaluation

In total, three AI models, the MRD1, MRD2, and LF models, were trained. The photograph processing time for each model was recorded. The mean absolute error (MAE) and mean square error (MSE) were selected to evaluate the performance of model prediction. The Pearson correlation coefficient was used to assess the correlation between the deep learning model prediction and gold-standard measurements. The intraclass correlation coefficient (ICC) was used to compare the agreement between the deep learning model prediction and the gold-standard measurements. Statistical analyses were performed using R software (version 4.1.0; R Foundation). Bland-Altman analysis was used to compare the agreement between the deep learning model prediction and the gold-standard measurements. Statistical significance was set at $P<.05$.

Results

Data Characteristics

We collected 822 eye photographs from 411 volunteers, including 344 (83.7%) women and 67 (16.3%) men. The photographs were subsequently randomly divided into two groups: 90% as the training/validation group and 10% as the test group. Within the training/validation group, 80% of photographs were used as the training group and 20% were used as the validation group (Figure 5). The case numbers and sex ratios in the MRD1, MRD2, and LF models are shown in Table 1. In the LF model, 137 normalized eye photographs were excluded because the up- and down-gaze orbital photographs were not well merged.

Figure 5. Data organization for model evaluation. Ninety percent of the data were used as the training/validation group and 10% were used as the test group; 80% of the data from the training/validation group were used as the training group and 20% were used as the validation group.

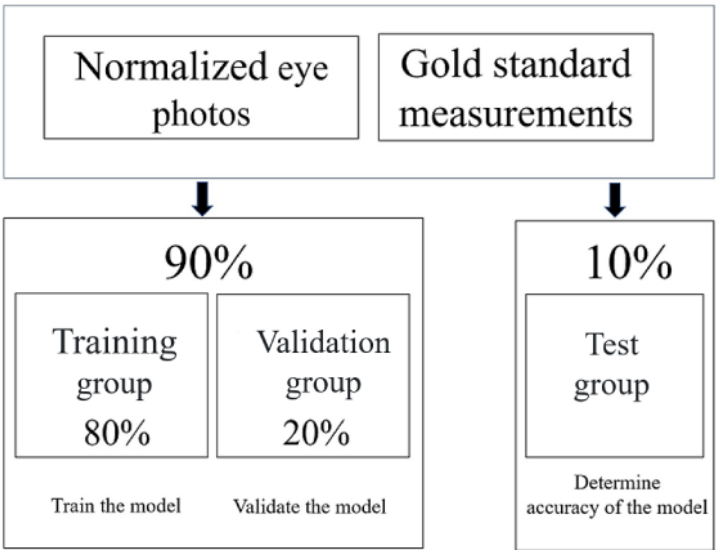


Table 1. Case numbers and sex ratios in each model.

Model	Cases, n (%)	Males, n (%)
MRD1[a]		
Total	822 (100.0)	154 (18.7)
Training group	740 (90.0)	142 (19.2)
Test group	82 (10.0)	12 (14.6)
MRD2[b]		
Total	822 (100.0)	154 (18.7%)
Training group	740 (90.0)	142 (9.2%)
Test group	82 (10.0)	12 (14.6%)
LF[c]		
Total	685 (100.0)[d]	122 (17.8)
Training group	617 (90.0)	113 (8.3)
Test group	68 (10.0)	9 (13.2)

[a]MRD1: marginal reflex distance 1.

[b]MRD2: marginal reflex distance 2.

[c]LF: levator muscle function.

[d]In the LF model, 137 normalized eye photographs were excluded because the up- and down-gaze orbital photographs were not well merged.

Reliability of Gold-Standard Measurements

The gold-standard measurements of MRD1, MRD2, and LF are summarized in Table 2. To determine the reliability, the measurements performed by the two doctors were evaluated using MAE, MSE, Pearson correlation coefficient, ICC, and Bland-Altman analysis. The reliability of the two doctors was excellent (Table 3, Figure 6).

Table 2. Summary of gold-standard measurements.

Measurements	N	Mean (SD)	Range
MRD1[a] (mm)	822	2.59 (1.21)	0.00-6.00
MRD2[b] (mm)	822	5.51 (0.83)	1.50-10.00
LF[c] selected (mm)	685[d]	12.1 (2.12)	3.50-18.00

[a]MRD1: marginal reflex distance 1.

[b]MRD2: marginal reflex distance 2.

[c]LF: levator muscle function.

[d]In the LF model, 137 normalized eye photographs were excluded because the up- and down-gaze orbital photographs were not well merged.

Table 3. Reliability of gold-standard measurements (actual values) manually performed by the two doctors.

Metric	MRD1[a]	MRD2[b]	LF[c]
MAE[d]	0.007	0.008	0.018
MSE[e]	0.005	0.001	0.002
Pearson correlation coefficient	0.999	0.998	0.999
ICC[f] (agreement)	0.999	0.998	0.999
ICC (consistency)	0.999	0.998	0.999

[a]MRD1: marginal reflex distance 1.

[b]MRD2: marginal reflex distance 2.

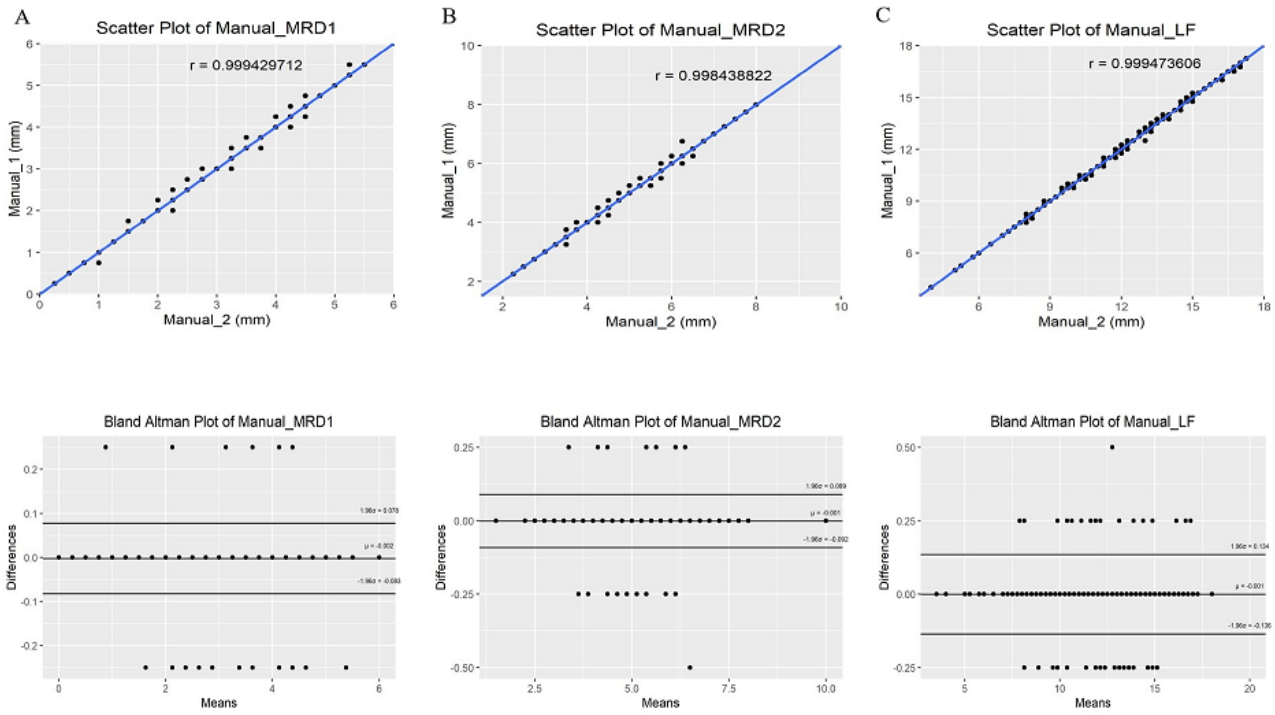
[c]LF: levator muscle function.

[d]MAE: mean absolute error.

[e]MSE: mean square error.

[f]ICC: intraclass correlation coefficient.

Figure 6. Scatter plots and Bland-Altman plots of gold-standard measurements (real values) for marginal reflex distance (MRD)1 (A), MRD2 (B), and levator muscle function (LF) (C) performed by two doctors.



Validation of the Training Model

There were 740 patients in the training/validation group included in the MRD1 and MRD2 models, and 617 patients included in the training/validation group in the LF model. The validation results based on MAE, MSE, Pearson correlation coefficient, ICC, and Bland-Altman analysis were good overall (Table 4, Figure 7).

Table 4. Validation and test results of the training model.

Metric	Validation			Test		
	MRD1[a]	MRD2[b]	LF[c]	MRD1	MRD2	LF
MAE[d] (mm)	0.087	0.158	0.290	0.349	0.375	1.059
MSE[e]	0.023	0.050	0.303	0.227	0.246	1.709
Pearson correlation coefficient	0.992	0.963	0.967	0.908	0.875	0.728
ICC[f] (Agreement)	0.992	0.962	0.966	0.903	0.837	0.692
ICC (Consistency)	0.992	0.963	0.966	0.902	0.837	0.689

[a]MRD1: marginal reflex distance 1.

[b]MRD2: marginal reflex distance 2.

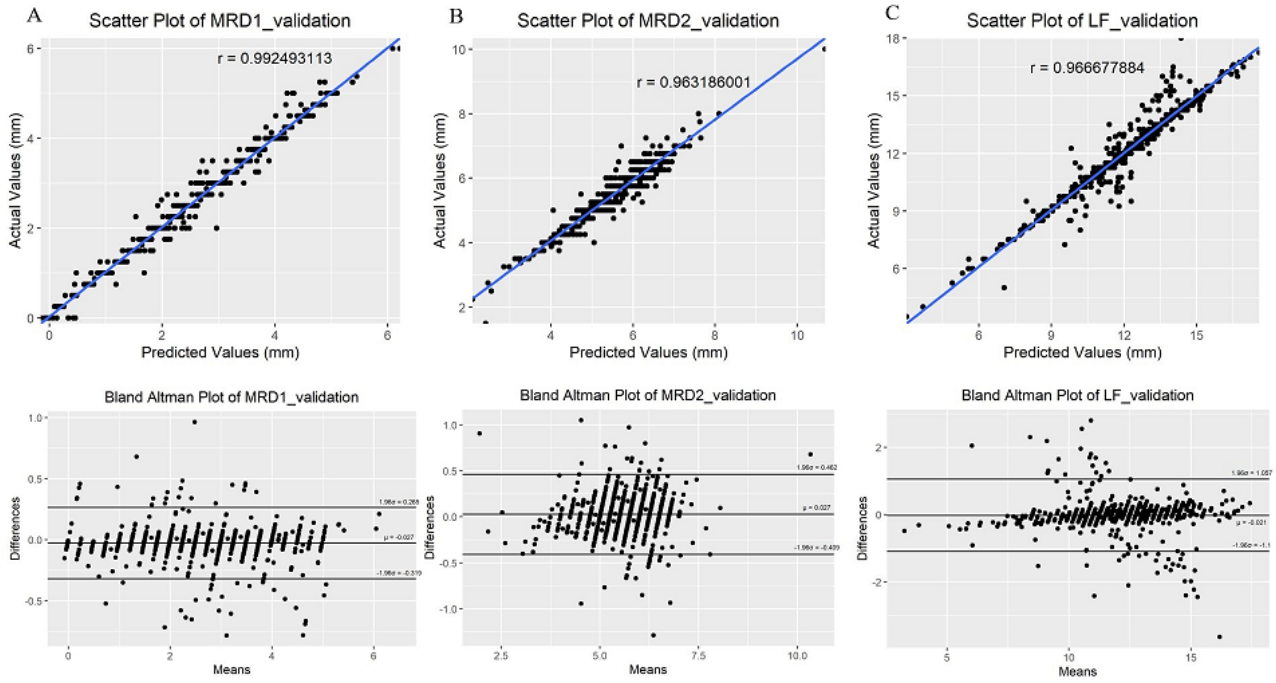
[c]LF: levator muscle function.

[d]MAE: mean absolute error.

[e]MSE: mean square error.

[f]ICC: intraclass correlation coefficient.

Figure 7. Scatter plots and Bland-Altman plots of validation results of the marginal reflex distance (MRD)1(A), MRD2 (B), and levator muscle function (LF) (C) training models.



Test Results of the MRD1, MRD2, and LF models

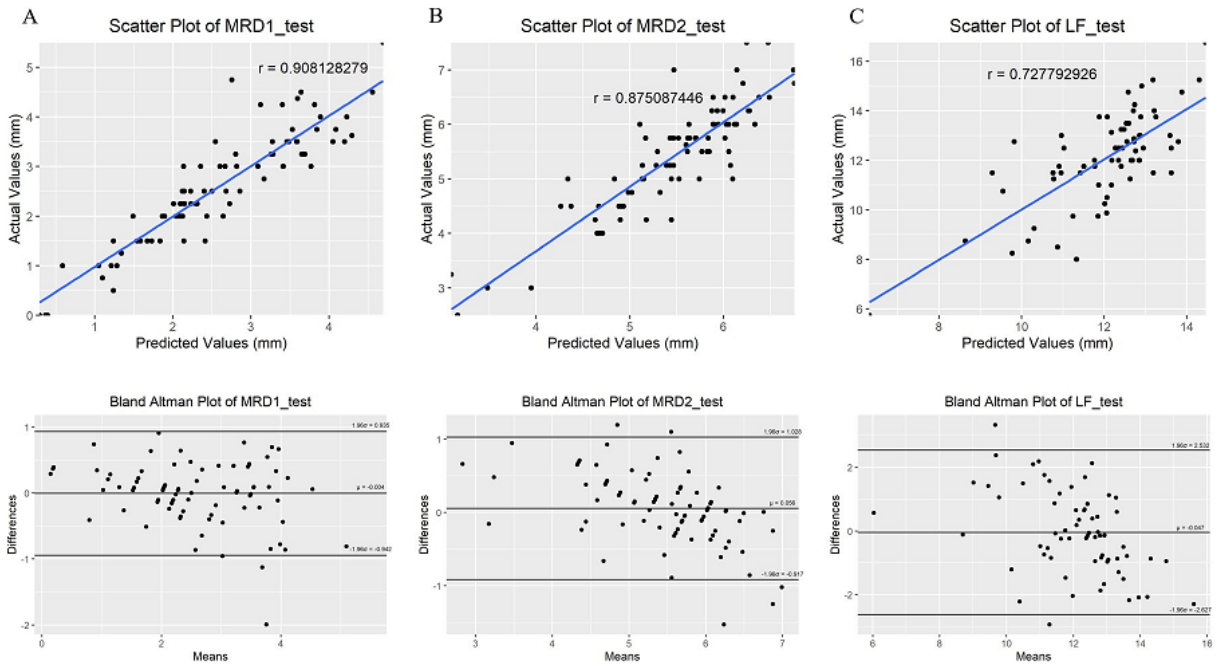
A total of 82 patients were used as the test group in the MRD1 and MRD2 models, and 68 patients were used as the test group in the LF model. The test results determine the accuracy of the model. It took 2.09 seconds and 2.15 seconds for the MRD1 and MRD2 models to respectively process 82 photos, and it took 1.97 seconds for the LF model to process 68 photos. The MAE of the predicted values to the gold-standard measurements of MRD1, MRD2, and LF were 0.35 mm, 0.37 mm, and 1.06 mm, respectively, and the MSE of the predicted values to the gold-standard measurements of MRD1, MRD2, and LF were 0.23 mm, 0.25 mm, and 1.71 mm, respectively.

The correlations between the gold-standard measurements and the values predicted by the MRD1 and MRD2 models were excellent (r=0.91 and 0.88, respectively). The correlation between the test results obtained with the LF model and gold-standard measurements was good (r=0.73).

The ICCs (agreement) between the gold-standard measurements and the values predicted with the MRD1, MRD2, and LF models were 0.90, 0.84, and 0.69, respectively. The ICCs (consistency) between the gold-standard measurements and the values predicted with the MRD1, MRD2, and LF models were 0.90, 0.84, and 0.69, respectively. These results indicate excellent agreement between the gold-standard measurements and the values predicted with the MRD1and MRD2 models, and substantial agreement with the LF model [19].

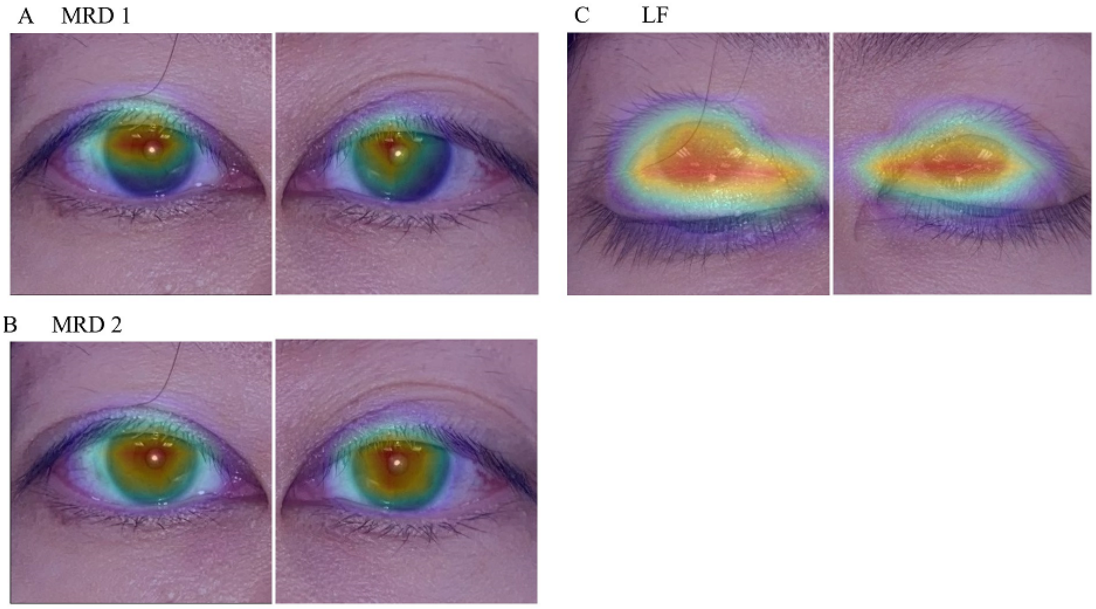
Bland-Altman analyses showed that the bias between the gold-standard measurements and the values predicted by the MRD1, MRD2, and LF models was –0.004 mm (95% CI –0.1090 to 0.1015 mm), 0.056 mm (95% CI –0.05347 to 0.1646 mm), and –0.047 mm (95% CI –0.3658 to 0.2713 mm), respectively. The 95% limits of agreement were −0.94 to 0.94 mm for the MRD1 model, –0.92 to 1.03 mm for the MRD2 model, and –2.63 to 2.53 mm for the LF model (Table 4, Figure 8).

Figure 8. Scatter plots and Bland-Altman plots of the test results of the marginal reflex distance (MRD)1(A), MRD2 (B), and levator muscle function (LF) (C) training models.



Representative heat maps in Figure 9 demonstrate the image region with the highest feature density and the most discriminative value (red), which was the region between the upper eyelid margin and light reflex in the MRD1 model, the region between the lower eyelid margin and light reflex in the MRD2 model, and the region between the upper eyelid margin in a merged up- and down-gaze in the LF model.

Figure 9. Representative heat maps of marginal reflex distance (MRD)1(A), MRD2 (B), and levator muscle function (LF) (C). The red color indicates regions with the highest discriminative value.



Discussion

Principal Findings

Since Putterman and Urist introduced the MRD, MRD1 has become an important tool for pre and postoperative ptosis evaluation [20,21]. From Putterman's description, the MRD is measured in millimeters, and to determine MRD1, the examiner uses one hand to hold a muscle light and the other hand to hold a ruler to measure the distance from the light reflex on the cornea to the upper eyelid margin. The examiner also needs another hand (a third hand) to hold the patient's eyebrow to prevent eyebrow elevation [22]. As a result, an examiner is less likely to perform the measurement on their own. Smartphones combine deep learning image processing as a solution to overcome this limitation.

Several automatic and semiautomatic photographic analysis approaches have been developed to provide a relatively objective assessment of MRD1 and MRD2 [4-6]. However, these studies compared their automatic and semiautomatic MRD1 and MRD2 assessments to manual measurements, not to gold-standard measurements, and the former are subjective and associated with a risk of human error. There are no automatic photographic analysis approaches for measuring the LF. To the best of our knowledge, ours is the first AI software algorithm capable of predicting MRD1, MRD2, and LF measurements with completely automated image processing and comparison of the prediction results with gold-standard measurements.

Manual MRD1, MRD2, and LF measurements are time-consuming, subjective, and have a limited precision of approximately 0.5 mm. According to Boboridis et al [2], the mean difference in measured MRD between doctors with varying degrees of experience was up to 0.5 mm, indicating poor repeatability. In this study, the correlations between the gold-standard measurements and the values predicted by the MRD1 and MRD2 models were excellent and the correlation for the values predicted by the LF model was good. The ICC results showed excellent agreement between the gold-standard measurements and the predicted values by the MRD1and MRD2 models, and substantial agreement with the values predicted by the LF models. The MAE values were 0.35 mm, 0.37 mm, and 1.06 mm for the MRD1, MRD2, and LF models, respectively, and the variance increased with length. The 95% limits of agreement were −0.94 to 0.94 mm for the MRD1 model, –0.92 to 1.03 mm for the MRD2 model, and –2.63 to 2.53 mm for the LF model. These results showed that the MRD1 and MRD2 models were equivalent and might even be better than manual measurements.

The performance of the LF model was not as excellent as that of the MRD1 and MRD2 models. One reason is that the longer the measurement, the greater the variance in the measurements. The second reason is the error during photograph normalization in the LF model. In some cases, the software algorithm could not merge the up- and down-gaze orbital photographs perfectly. The third reason is overfitting, which occurs when a model does not generalize adequately from observed data to unknown data [23]. The LF model in our study had good validation results in the training set but had limited success on the test set. An extended dataset might enhance the prediction accuracy, especially in a complex model such as the LF model used in this study [24].

Some AI models face a conundrum: their performance on the test set is good, but it is significantly lower when used in a clinical scenario. One issue is that the training data are collected under stringent conditions (such as a strictly controlled photography environment), which makes it difficult for the trained model to adjust to clinical situations (such as at the clinic). In this study, the models were created to simulate a clinical scenario. The ocular photos for model training were obtained by a smartphone to simulate the doctor checking patients' eyelid measurements in the clinic using a handheld ruler. Therefore, we believe that our model can adapt well to clinical use.

We used a deep learning algorithm to establish three models: the MRD1, MRD2, and LF models. We intend to integrate these models into a cloud-based service available on the internet. Based on these three models, we will also develop an app software contained within a smartphone, which can work offline. In the future, the examiner can use one hand to hold a smartphone and snap six images, including bilateral primary gaze, up-gaze, and down-gaze, while holding the patient's brow with the other. The MRD1, MRD2, and LF measurements can then be predicted by the deep learning app (Figure 1). This is a quick, objective, and convenient method for obtaining MRD1, MRD2, and LF measurements. Furthermore, the examiner can check these measurements anywhere and at any time using a smartphone, which also facilitates data collection.

Limitations

This study had some limitations. Mascara, false lashes, obvious eyelid creases, and the lack of well-merged orbital photographs interfered with the model prediction. Negative MRD1 and MRD2 levels could not be predicted, which is another limitation. In the training/validation group, the MRD1 measurements of 25 ptotic eyelids (25/740, 3.4%) without corneal light reflex were recorded as 0 mm in this study. Surprisingly, the MRD1 model predicted negative values in 18 eyelids (18/25, 72%) of these cases, implying that the algorithm may eventually learn to predict negative values on its own. When taking orbital photographs, fine movement of patients or the smartphone cannot be completely avoided, resulting in imperfectly merged images, which is a defect of our current algorithm. The merged photos will be displayed by the cloud-based service and app software in the future, so that examiners can discard the images that are not perfectly merged and retake orbital photographs to obtain better-merged images.

Conclusion

In this study, we developed the first smartphone-based AI-assisted image processing algorithm for eyelid measurements. MRD1, MRD2, and LF measurements can be taken in a quick, objective, and convenient manner. Furthermore, by using a smartphone, the examiner can check these measurements anywhere and at any time, which also makes data collection easier.

Conflicts of Interest

None declared.

References

1. Beard C. A new classification of blepharoptosis. Int Ophthalmol Clin 1989;29(4):214-246. [doi: 10.1097/00004397-198902940-00002] [Medline: 2592154]
2. Boboridis K, Assi A, Indar A, Bunce C, Tyers A. Repeatability and reproducibility of upper eyelid measurements. Br J Ophthalmol 2001 Jan;85(1):99-101 [FREE Full text] [doi: 10.1136/bjo.85.1.99] [Medline: 11133723]
3. Nemet A. Accuracy of marginal reflex distance measurements in eyelid surgery. J Craniofac Surg 2015 Oct;26(7):e569-e571. [doi: 10.1097/SCS.0000000000001304] [Medline: 26468822]
4. Bodnar ZM, Neimkin M, Holds JB. Automated ptosis measurements from facial photographs. JAMA Ophthalmol 2016 Feb 01;134(2):146-150. [doi: 10.1001/jamaophthalmol.2015.4614] [Medline: 26605967]
5. Chun YS, Park HH, Park IK, Moon NJ, Park SJ, Lee JK. Topographic analysis of eyelid position using digital image processing software. Acta Ophthalmol 2017 Nov;95(7):e625-e632. [doi: 10.1111/aos.13437] [Medline: 28391655]
6. Lou L, Yang L, Ye X, Zhu Y, Wang S, Sun L, et al. A novel approach for automated eyelid measurements in blepharoptosis using digital image analysis. Curr Eye Res 2019 Oct;44(10):1075-1079. [doi: 10.1080/02713683.2019.1619779] [Medline: 31148484]
7. Sinha K, Yeganeh A, Goldberg R, Rootman D. Assessing the accuracy of eyelid measurements utilizing the Volk Eye Check System and clinical measurements. Ophthalmic Plast Reconstr Surg 2018;34(4):346-350. [doi: 10.1097/IOP.0000000000000991] [Medline: 28863120]
8. LabelImage. GitHub. URL: https://github.com/tzutalin/labelImg [accessed 2020-09-02]
9. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. arXiv. 2019 Mar 21. URL: https://arxiv.org/abs/1801.04381 [accessed 2021-10-01]
10. Tan M, Le QV. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv. 2020 Sep 11. URL: https://arxiv.org/abs/1905.11946 [accessed 2021-10-01]
11. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. arXiv. 2019 May 16. URL: https://arxiv.org/abs/1709.01507 [accessed 2021-10-01]
12. Buslaev A, Iglovikov V, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: fast and flexible image augmentations. Information 2020 Feb 24;11(2):125 [FREE Full text] [doi: 10.3390/info11020125]
13. Wu H, Gu X. Towards dropout training for convolutional neural networks. Neural Netw 2015 Nov;71:1-10. [doi: 10.1016/j.neunet.2015.07.007] [Medline: 26277608]
14. Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv. 2017 Dec 13. URL: https://arxiv.org/abs/1712.04621 [accessed 2021-10-01]
15. Lym S, Behroozi A, Wen W, Li G, Kwon Y, Erez M. Mini-batch serialization: CNN training with inter-layer data reuse. arXiv. 2019 May 04. URL: https://arxiv.org/abs/1810.00307 [accessed 2021-10-01]
16. Smith LN. A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. arXiv. 2018 Apr 24. URL: https://arxiv.org/abs/1803.09820 [accessed 2021-10-01]
17. Huang G, Li Y, Pleiss G, Liu Z, Hopcroft JE, Weinberger KQ. Snapshot ensembles: Train 1, get M for free. arXiv. 2017 Apr 01. URL: https://arxiv.org/abs/1704.00109 [accessed 2021-10-01]
18. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv. 2017 Jan 30. URL: https://arxiv.org/abs/1412.6980 [accessed 2021-10-01]
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
20. Putterman AM, Urist MJ. Müller muscle-conjunctiva resection. Technique for treatment of blepharoptosis. Arch Ophthalmol 1975 Aug;93(8):619-623. [doi: 10.1001/archopht.1975.01010020595007] [Medline: 1156223]
21. Putterman AM. Basic oculoplastic surgery. In: Peyman D, Sanders D, Goldberg MF, editors. Principals and practice of ophthalmology. Philadelphia, PA: Saunders Company; 1980:2247-2250.
22. Putterman AM. Margin reflex distance (MRD) 1, 2, and 3. Ophthalmic Plast Reconstr Surg 2012;28(4):308-311. [doi: 10.1097/IOP.0b013e3182523b7f] [Medline: 22785597]
23. Overfitting in machine learning: what it is and how to prevent it. Elite Data Science. URL: https://elitedatascience.com/overfitting-in-machine-learning [accessed 2021-05-01]
24. Ying X. An overview of overfitting and its solutions. J Phys Conf Ser 2019 Mar 12;1168:022022. [doi: 10.1088/1742-6596/1168/2/022022]

Abbreviations

AI: artificial intelligence

CNN: convolutional neural network
ICC: intraclass correlation coefficient
LF: levator muscle function
MAE: mean absolute error
MAIA: medical artificial intelligence assistant
MRD: marginal reflex distance
MSE: mean square error

Edited by G Eysenbach; submitted 28.07.21; peer-reviewed by X Dong; comments to author 26.08.21; revised version received 02.09.21; accepted 19.09.21; published 08.10.21.

Please cite as:
Chen HC, Tzeng SS, Hsiao YC, Chen RF, Hung EC, Lee OK
Smartphone-Based Artificial Intelligence–Assisted Prediction for Eyelid Measurements: Algorithm Development and Observational Validation Study
JMIR Mhealth Uhealth 2021;9(10):e32444
URL: https://mhealth.jmir.org/2021/10/e32444
doi:10.2196/32444
PMID:34538776

Original Paper

Smartphone-Assisted High-Intensity Interval Training in Inflammatory Rheumatic Disease Patients: Randomized Controlled Trial

Håvard Haglo[1,2], MSc; Eivind Wang[1,3], PhD; Ole Kristian Berg[1], PhD; Jan Hoff[2,4], PhD; Jan Helgerud[2,5], PhD

[1]Faculty of Health and Social Sciences, Molde University College, Molde, Norway

[2]Myworkout, Medical Rehabilitation Clinic, Trondheim, Norway

[3]Department of Medicine, University of Utah, Salt Lake City, UT, United States

[4]Department of Physical Medicine and Rehabilitation, St. Olav's University Hospital, Trondheim, Norway

[5]Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding Author:
Håvard Haglo, MSc
Myworkout
Medical Rehabilitation Clinic
Ingvald Ystgaards veg 23
Trondheim
Norway
Phone: 47 92621619
Email: havard@treningsklinikken.no

Abstract

Background: Patients with inflammatory rheumatic diseases (IRDs) experience disease-related barriers to physical training. Compared with the general population, IRD patients are reported to have reduced maximal oxygen uptake (VO_{2max}) and physical activity levels. Supervised high-intensity interval training (HIIT) is documented to counteract the reduced VO_{2max} and poor cardiovascular health associated with IRDs. However, supervised HIIT is resource demanding.

Objective: This study sought to investigate if self-administered 4×4-min HIIT guided by a smartphone app (Myworkout GO) could yield similar HIIT-induced effects as standard 4×4-min HIIT performed under the guidance and supervision of health care professionals. The effects studied were on VO_{2max} and health-related quality of life (HRQoL).

Methods: Forty patients (33 female patients, mean age 48 years, SD 12 years; 7 male patients, mean age 52 years, SD 11 years) diagnosed with rheumatoid arthritis, spondyloarthritis, or systemic lupus erythematosus were randomized to a supervised group (SG) or an app group (AG). Both groups were instructed to perform 4×4-min intervals with a rate of perceived exertion of 16 to 17, corresponding to 85% to 95% of the maximal heart rate, twice a week for 10 weeks. Treadmill VO_{2max} and HRQoL measured using RAND-36 were assessed before and after the exercise period.

Results: VO_{2max} increased (P<.001) in both groups after 10 weeks of HIIT, with improvements of 3.6 (SD 1.3) mL/kg/min in the SG and 3.7 (SD 1.5) mL/kg/min in the AG. This was accompanied by increases in oxygen pulse in both groups (P<.001), with no between-group differences apparent for either measure. Improvements in the HRQoL dimensions of bodily pain, vitality, and social functioning were observed for both groups (P<.001 to P=.04). Again, no between-group differences were detected.

Conclusions: High-intensity 4×4-min interval training increased VO_{2max} and HRQoL, contributing to patients' reduced cardiovascular disease risk, improved health and performance, and enhanced quality of life. Similar improvements were observed following HIIT when IRD patients were guided using perceived exertion by health care professionals or the training was self-administered and guided by the app Myworkout GO. Utilization of the app may help reduce the cost of HIIT as a treatment strategy in this patient population.

Trial Registration: ClinicalTrials.gov NCT04649528; https://clinicaltrials.gov/ct2/show/NCT04649528

(JMIR Mhealth Uhealth 2021;9(10):e28124) doi:10.2196/28124

KEYWORDS

VO2max; maximal oxygen uptake; mobile app; cardiovascular health; quality of life; endurance training

Introduction

Patients with inflammatory rheumatic diseases (IRDs), such as rheumatoid arthritis (RA), spondyloarthritis (SpA), and systemic lupus erythematosus (SLE), are reported to have low cardiorespiratory fitness, commonly assessed as maximal oxygen uptake (VO_{2max}) [1-3].

Considering the characteristic symptoms of pain, joint swelling, fatigue, and stiffness [4-6], it is unsurprising that patients with IRDs face disease-related barriers for performing health-enhancing physical training [7-9]. Hence, patients with IRDs are observed to not only have decreased VO_{2max} compared to the general population, but also be less physically active [10], or perform physical activities with lower intensity [1]. Patients with RA, SpA, and SLE may all have these consequences [1-3]; however, SLE patients are reported to have a particularly great VO_{2max} reduction [3]. The latter IRD subpopulation is also, in general, predominantly incapacitated by physical fatigue [4], whereas for patients with SpA [6] and RA [5], pain and joint stiffness may more commonly be the leading causes of disability.

VO_{2max} is acknowledged as a strong predictor of cardiovascular disease (CVD)-related mortality and all-cause mortality [11], and it is thus unsurprising that the low VO_{2max} observed in patients with IRDs (eg, RA) is associated with an unfavorable cardiovascular profile and an increased risk of CVD [12]. IRD patients also have a higher mortality than the general population, with more than 50% of premature deaths being attributed to CVD [13].

In parallel with the relatively larger impairment of VO_{2max}, patients with SLE are at a higher risk of CVD-related events and increased mortality [13,14] than RA patients [10,13] and SpA patients, where the latter subpopulation has the relatively lowest risk [10,15]. Moreover, IRD patients are at greater risk of nonfatal ischemic heart disease, which more often goes unrecognized in this patient group compared to age- and sex-matched controls [13]. Another common finding is the development of accelerated atherosclerosis [16]. Ultimately, the poor physical health status observed in the IRD population is reflected in their self-perceived health-related quality of life (HRQoL) and is particularly manifested in their experience of physical function and performance [17]. Recognizing the poor physical health status of IRD patients, training interventions aiming to effectively improve VO_{2max} are sought after.

Aerobic training, tailored to improve VO_{2max}, may not only enhance physical function and performance, but also reduce the risk of CVD and all-cause mortality [11]. Furthermore, it has the potential to reduce symptom burden and inflammation in IRD patients [18]. In turn, a relief in symptoms could have a synergistic effect as disease-related factors (pain, stiffness, fatigue, disability, and quality of sleep) improve and may lower the barrier to engaging in general physical activity [8]. Aerobic training can be organized in terms of volume, frequency, and intensity, and in particular, intensity has been shown to be of critical importance for VO_{2max} improvements, with high intensity being superior to moderate or low training intensity [19]. A model for organizing aerobic high-intensity interval training (HIIT) can be 4 times 4-min work bouts carried out at 85%-95% of the maximal heart rate (HR_{max}), interspaced by active recovery phases of 3 min at approximately 70% HR_{max}. This model has been documented to yield effective increases in VO_{2max} in young [19], old [20], and different groups of patients [21-24]. In fact, for individuals with an aerobic capacity typical of what is observed in the general population, similar improvements in the range of approximately 0.3 to 0.4 L/min have been shown for various age groups following an 8-week HIIT intervention [25]. It has also been demonstrated to be effective and a well-tolerated mode of exercise in the IRD patient population, with similar VO_{2max} improvements as observed in healthy individuals [18,26,27]. Of importance, HIIT interventions have typically been supervised and carried out in a laboratory setting.

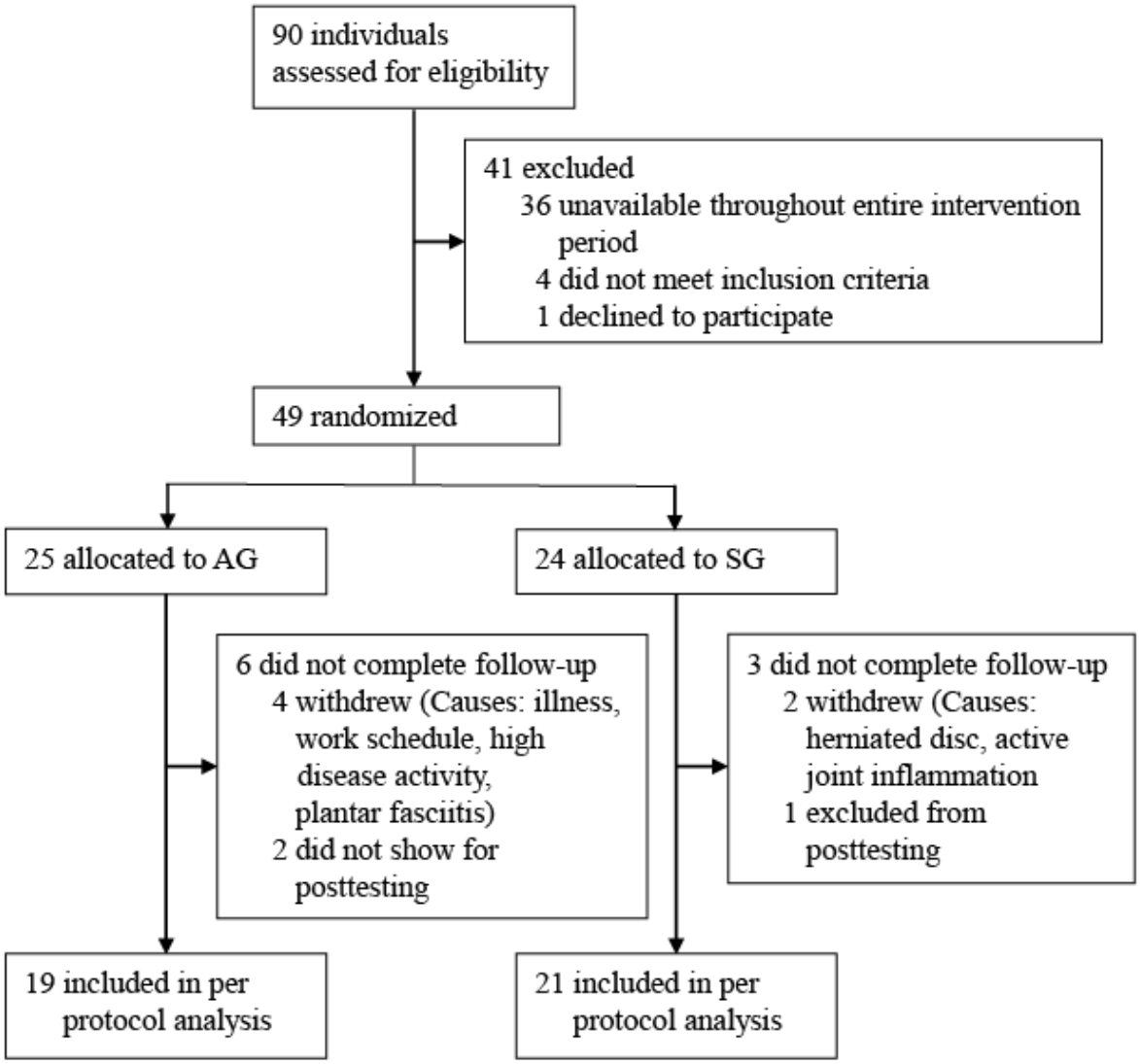
Unsurprisingly, supervised exercise demonstrates superior improvements in VO_{2max} compared to self-administered exercise after receiving exercise recommendations [28]. Even following initial personal instructions and free access to heart rate monitors, supervised HIIT increases VO_{2max} more than self-administered HIIT [29]. However, with recent advances in mobile technology and its availability, closer follow-up and instructions for exercising individuals may be possible. This could offer cost-effective modes of enticing patient self-management in chronic conditions [30]. Indeed, smartphone apps and associated notifications increase adherence to self-administered exercise [31]. Accordingly, apps could enhance the effectiveness of self-administered home-based exercise, and may offer a viable alternative to time- and resource-demanding supervised exercise. However, to date, results after app-guided exercise rehabilitation are equivocal. While some studies have documented improved cardiorespiratory fitness [32,33], others have not [34]. The unclear effect may, at least in part, be due to the various app designs and training interventions. Nevertheless, some studies are certainly promising with regard to the fact that app-guided training may yield some of the expected increase in VO_{2max}, which is typically observed following supervised training. Thus, the aim of this study was to investigate the effect of 10 weeks of self-administered HIIT guided by an app on VO_{2max} in IRD patients, and to compare this to standard supervised HIIT. Moreover, the aim was to explore if effects on VO_{2max} were reflected in the patients' HRQoL. Specifically, we hypothesized that both modes of exercise would increase directly assessed VO_{2max} and improve HRQoL, but that supervised HIIT would improve both outcomes more than self-administered app-guided HIIT.

Methods

Subjects

This 2-group randomized trial included 49 male and female volunteers (aged ≥18 years) diagnosed with at least one of the following IRDs: RA, SpA, and SLE. The subjects were not familiar with performing HIIT prior to inclusion in the study. They were recruited through the Norwegian Rheumatic Association and a rehabilitation clinic in Central Norway, and randomized into the following 2 groups (Figure 1): self-administered HIIT guided by an app group (AG) and standard supervised HIIT group (SG). All participants were encouraged to keep regular treatments given by the health care system, nutrition, and other physical activity habits constant throughout the study period. Sixteen of the subjects reported not engaging in any regular physical activity (AG: 7; SG: 9), while 24 of the subjects (AG: 12; SG: 12) reported being physically active 1 to 3 times per week. Prior to enrollment into the study, all participants were screened by a medical doctor for eligibility, and asked if they were able to get access to a training facility. Patients with various disease activities were included. The exclusion criteria were unstable ischemic heart disease, pregnancy, not owning a smartphone, planned surgeries influencing the training or testing, and inability to complete the testing and exercise protocol. Additionally, patients were excluded if they had other comorbid diseases, such as cerebrovascular disease, pulmonary disease, angina, diabetes type I, and hypertension, which were considered severe and/or the main limiting factor for training and testing. The participants reviewed and signed informed consent forms before participating in the study. The study was approved by the Regional Committee for Medical and Health Research Ethics in Norway and was performed in accordance with the Declaration of Helsinki.

Figure 1. Trial flow diagram. AG: app group; SG: supervised group.



Study Timeline

Prior to commencement of the study, sealed opaque allocation envelopes were prepared by a third party in a 1:1 ratio. After enrollment and pretesting, participants picked and opened allocation envelopes assigning them to the AG or SG. Participants performed pretesting 1 to 3 days before the 10-week HIIT period, and posttesting was completed 2 to 5 days after the last HIIT session. Participants were instructed to not perform intensive activity 48 hours before the test days. All testing was performed by the same personnel and using the same equipment and standardized protocol before and after the training period. Testing personnel were blinded for which of the groups the subjects had been randomized to, and participants were instructed to not provide any information of their allocation. Supervising health care professionals and subjects were aware of group assignment.

Testing of VO_{2max}

Pulmonary VO_{2max} was measured using a Metamax II portable gas analyzer (Cortex Biophysik) on a treadmill calibrated for speed and inclination (Gymsport TX200). Simultaneously, heart rate was continually registered during the test using Polar RS100 (Polar Electro). In addition to determining VO_{2max} and HR_{max}, performing an incremental cardiorespiratory exercise test is recommended to maximize patient safety and ensure exercise tolerance [35] before commencing a training program. Following a 6-min warm-up period at 4.0 km/h at 5% treadmill inclination, the workload was increased in increments of 1.0 km/h or 1% every minute until exhaustion. This implies that subjects with a low VO_{2max} typically performed the test walking, while subjects with a high VO_{2max} were running during the final minutes of the test. Participants received verbal encouragement and feedback from the tester throughout the test. A respiratory exchange ratio ≥1.05, in combination with a plateau in VO_2 despite increased work rate, was used as the criterion for reaching VO_{2max} [36]. If the criterion for achieving VO_{2max} was not reached, a retest was scheduled 3 days later. If the criterion was still not met, a VO_{2peak} was reported. VO_{2max} was calculated as the mean of the 3 highest consecutive 10-second measurements. The highest heart rate measured during the last minute of the test was used as $HR_{max.}$

HRQoL

All participants were given a generic HRQoL questionnaire (the self-administered Norwegian version of the RAND 36-Item Short-Form Health Survey [RAND-36]) at pretest and posttest. The validated Norwegian version [37] was translated from the Medical Outcome Study 36-Item Short-Form Health Survey (SF-36) [38]. RAND-36 has the same items as SF-36 with a slightly different scoring in the dimensions *bodily pain* and *general health*, and a correlation of 0.99 between the 2 scoring algorithms for these 2 items has been demonstrated [39]. The questionnaire is comprised of 8 dimensions, and scores are converted to a range from 0 to 100, with higher scores representing better health outcomes.

HIIT

The SG and AG were instructed to undergo 2 training sessions per week on nonconsecutive days for 10 weeks. The following instructions on how to conduct HIIT sessions were given by a health care professional (SG) or a smartphone app (AG):

1. Start each session with a 6-min warm-up period at ≥5% inclination, using moderate intensity. Talking in complete sentences should still be possible (talking speed) aiming to target a rate of perceived exertion (RPE) of 13 (approximately 70% HR_{max}).

2. Following warm-up, conduct the 4×4-min intervals at an intensity (speed and/or incline) that elicits heavy breathing within 2 min of each interval. It should be difficult to speak more than 2 to 3 words in a row, corresponding to an RPE of 16 to 17 (approximately 85%-95% of HR_{max}).

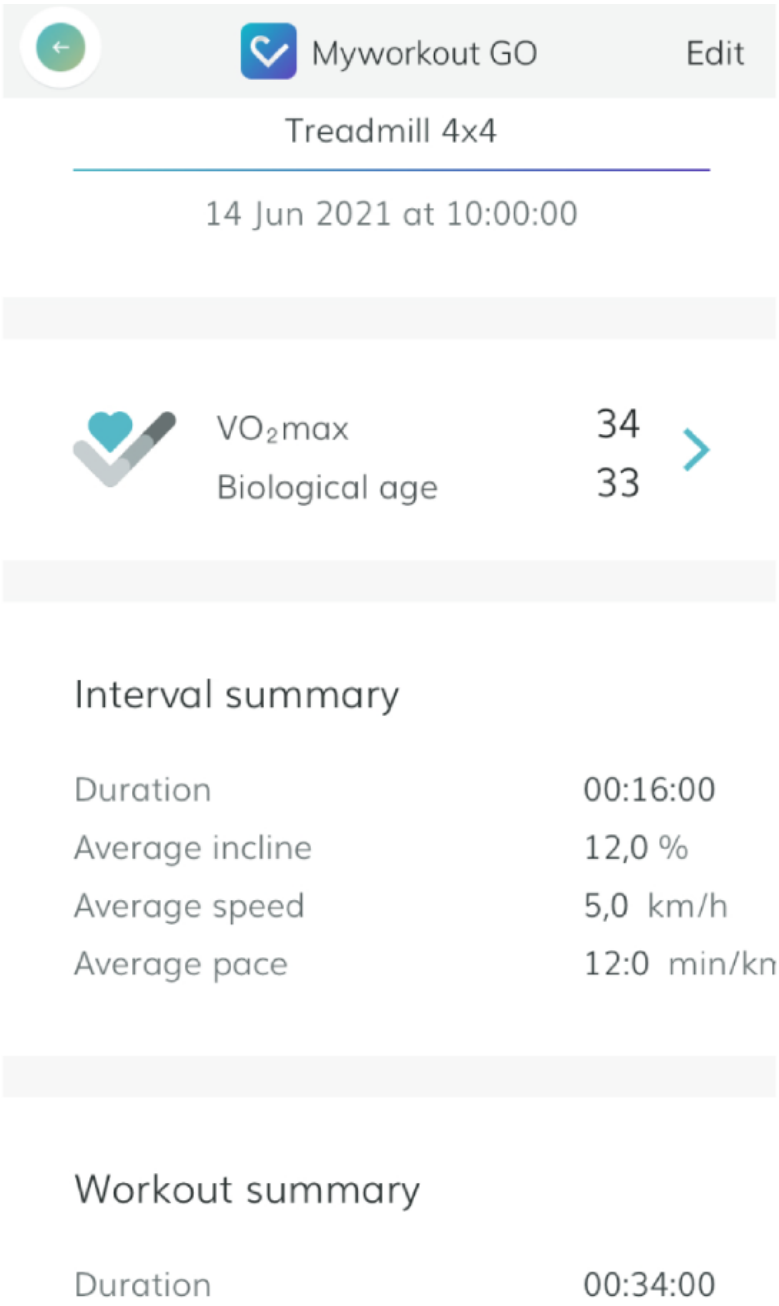
3. The high intensity intervals should be interspaced by 3 min of active recovery with a low to moderate intensity, similar to that applied during the warm-up.

In total, each session lasted 34 min. Additionally, the following 2 rules of thumb were given: (1) at the end of each 4-min interval, being able to continue 1 more interval minute should feel possible and (2) after the fourth 4-min interval, given an active recovery period, it should feel possible to complete a fifth interval. Depending on the patients' VO_{2max} levels, HIIT was carried out walking or running.

In order to maintain the same relative RPE intensity throughout the intervention period, absolute intensity was continuously adjusted to comply with the intensity guiding. These intensity instructions were given orally to individual subjects by a health care professional, a physiotherapist with specialization in exercise physiology, in the SG. The SG mainly trained on a treadmill in the rehabilitation clinic, which was integrated in a training facility open for the public. Some sessions (approximately one time every 2 weeks) were conducted outdoors. All intervals were carried out individually (uphill at ≥5% inclination and intensity adjusted) both indoors and outdoors to meet the targeted relative RPE. Similarly, with the intention to make the guidance type the only difference between the 2 groups, the AG conducted the training either on a treadmill indoors at a local training facility or uphill outdoors, at their own discretion, using an app (Myworkout GO, version 2.8) with written and standardized preprogrammed audio instructions, similar to the guidance given to the SG, during the HIIT. Moreover, visual display of individualized speed and incline that adapted to progression was provided by the app. After each session, the app presented performance feedback as interval work output (speed and incline), estimated $VO_{2max,}$ and estimated biological age (Figure 2). The AG was given a scheduled posttest date at baseline, and individuals in the group were contacted once by an automated email encouraging them to comply with the planned sessions halfway in the training period, if they had logged less than 70% of the scheduled sessions. This was the only time any type of monitoring of the AG took place during the 10-week training period. With regard to safety precautions during testing and training, health care personnel had cardiopulmonary resuscitation training, and a

defibrillator was available in the training facility. Although this was not available for the AG, the patients were instructed to carry out the training at daytime and contact available health care professionals and/or the medical doctor at the rehabilitation clinic by telephone if they had safety concerns. Importantly, they were instructed to contact emergency medical services if they experienced or suspected any HIIT-related adverse events. Notably, the HIIT intervention used in this study has previously been considered safe and recommendable in stable CVD patients, even in an unsupervised setting [29,40].

Figure 2. Screenshot of app (Myworkout GO) feedback following a treadmill 4×4 high-intensity interval training session for a female subject.



Considering that heart rate was not monitored during exercise sessions, intensity for both the AG and SG was estimated from work output (speed and incline) in 6 HIIT sessions and presented as a percentage of VO_{2max}. Oxygen uptake (VO_2) from the first 3 and last 3 HIIT sessions was calculated using the American College of Sports Medicine (ACSM) metabolic equations [41] and presented as a percentage of VO_{2max} from pretest and posttest. The ACSM has proposed that 60%, 80%, and 85% of VO_{2max} corresponds to 70%, 85%, and 90% of HR_{max}, respectively [41].

Statistics

The sample size calculation in this study was estimated based on the expected between-group difference in VO_{2max} at posttest. Assuming a SD of 0.2 L/kg/min with an expected mean difference of 0.2 L/kg/min between the groups, a sample of 16 subjects in each group would be required to maintain a statistical power of 0.80, with a 2-sided α of .05. However, as higher drop-out rates could be expected from patient populations, we planned to enroll 50 participants (25 in each group). Evaluation of normal distribution was performed using the Shapiro-Wilk

test and Q-Q plots. All variables except HRQoL exhibited normal distribution. Per protocol analyses were performed for all the outcome measures. For most variables, paired sample *t* tests were used to detect within-group differences, and 2×2 repeated measures analysis of variance (ANOVA) with time (pre and post) and group (AG and SG) as factors was used to identify differences between groups following the training period. HRQoL was analyzed using the Wilcoxon signed-rank test to detect within-group differences, and the Mann-Whitney *U* test was used to identify between-group differences. IBM SPSS statistics software (version 26; IBM Corp) was used for statistical analyses, and GraphPad Prism software (version 8; GraphPad Software, Inc) was used to create figures. At least 70% of the planned sessions had to be completed for inclusion in the analyses. Data in the tables and text are presented as mean (SD), and data in the figures are presented as mean (SEM).

Results

Adherence and Characteristics

No adverse CVD events were registered during or after the VO_{2max} testing or the HIIT intervention. A total of 4 subjects withdrew from the AG; 1 due to the illness, 1 for not finding enough time to train (busy work period) in the second half of the intervention period, 1 due to a period of high disease activity, and 1 due to plantar fasciitis (week 3) possibly associated with the training. Additionally, 2 dropouts occurred from the AG as the participants forgot to appear at the posttest. Both had logged at least 16 out of the 20 planned 4×4 sessions in the app, but were unable to reschedule the posttest. Two subjects from the SG withdrew; 1 due to a low back disk herniation not related to the study and 1 due to inflammation of the knee and ankle joint possibly related to the training, as well as disease activity. Furthermore, 1 subject from the SG was excluded from the data analyses as a period of illness resulted in not complying with the planned sessions. See Figure 1 for the trial flow diagram.

Descriptive characteristics of the included participants are presented in Table 1, with no significant difference between groups observed at baseline. However, VO_{2max} (P=.006) and oxygen pulse (P=.01) were both lower in the SG than in the AG at baseline (Table 2). Both groups complied well with the planned training sessions. Among the 20 sessions, the AG performed 18.4 sessions (92% [SD 13%]) and the SG performed 19.3 sessions (97% [SD 4%]), with no apparent between-group difference (Table 1). The calculated VO_2 from 6 HIIT sessions revealed that both the AG and SG performed these sessions with an intensity of 85%-90% of VO_{2max}, with no difference between the groups. No within- or between-group difference in body mass was observed at baseline or from pretraining to posttraining. Moreover, 5 subjects from the AG did not have complete HRQoL data sets, leaving 14 completed data sets for the analyses. Halfway through the intervention period, a random sample of 10 participants in the SG performed 1 training session with heart rate monitoring (same equipment as during testing), which confirmed that the target intensities (85%-95% HR_{max}) were met (Figure 3). Both the supervising physiologist and subject were blinded to heart rate data during the sampling session.

Table 1. Descriptive characteristics.

Characteristic	App group (n=19)	Supervised group (n=21)
Female, n (%)	14 (74)	19 (91)
Age (years), mean (SD)	48 (12)	50 (11)
Height (cm), mean (SD)	172 (9)	169 (6)
BMI (kg/m^2), mean (SD)	26.8 (4.3)	28.3 (6.1)
Diagnosis, n (%)		
Rheumatoid arthritis	4 (21)	8 (38)
Spondyloarthritis	11 (58)	10 (48)
Systemic lupus erythematosus	4 (21)	3 (14)
Disease duration (years), mean (SD)	13 (9)	10 (9)
Medication, n (%)		
Disease-modifying antirheumatic drugs	15 (79)	18 (86)
Nonsteroidal anti-inflammatory drugs	13 (68)	11 (52)
Completed sessions, mean (SD)	18 (3)	19 (1)

Table 2. Changes in physiological parameters from pretraining to posttraining.

Parameter	App group (n=19)			Supervised group (n=21)			Between-group comparison	
	Pretraining, mean (SD)	Posttraining, mean (SD)	*P* value[a]	Pretraining, mean (SD)	Posttraining, mean (SD)	*P* value[a]	*P* value[b]	*P* value[c]
VO_{2max}[d]								
Value (L/min)	2.90 (0.53)	3.15 (0.54)	<.001	2.46 (0.54)	2.74 (0.63)	<.001	.01	.46
Value (mL/kg/min)	36.8 (5.3)	40.5 (5.6)	<.001	31.1 (7.0)	34.7 (7.6)	<.001	.006	.97
HR_{max}[e] (bpm)	176 (10)	177 (11)	.67	171 (16)	172 (14)	.63	.30	.88
Oxygen pulse								
Value (mL/beat)	16.6 (3.3)	17.9 (3.3)	<.001	14.3 (2.7)	15.9 (3.2)	<.001	.02	.36
Value (mL/kg/beat)	0.21 (0.03)	0.23 (0.03)	<.001	0.18 (0.04)	0.20 (0.04)	<.001	.01	.65
V_E[f] (L/min)	81.8 (11.5)	93.6 (13.6)	<.001	73.7 (15.1)	87.8 (17.2)	<.001	.07	.29
R[g]	1.13 (0.08)	1.14 (0.07)	.47	1.15 (0.09)	1.17 (0.06)	.70	.47	.90
Body weight (kg)	79.1 (11.3)	78.2 (10.9)	.12	81.4 (19.6)	80.8 (19.2)	.10	.66	.65

[a]Within group difference from pretraining.

[b]Difference between groups at baseline.

[c]Difference between groups at posttest.

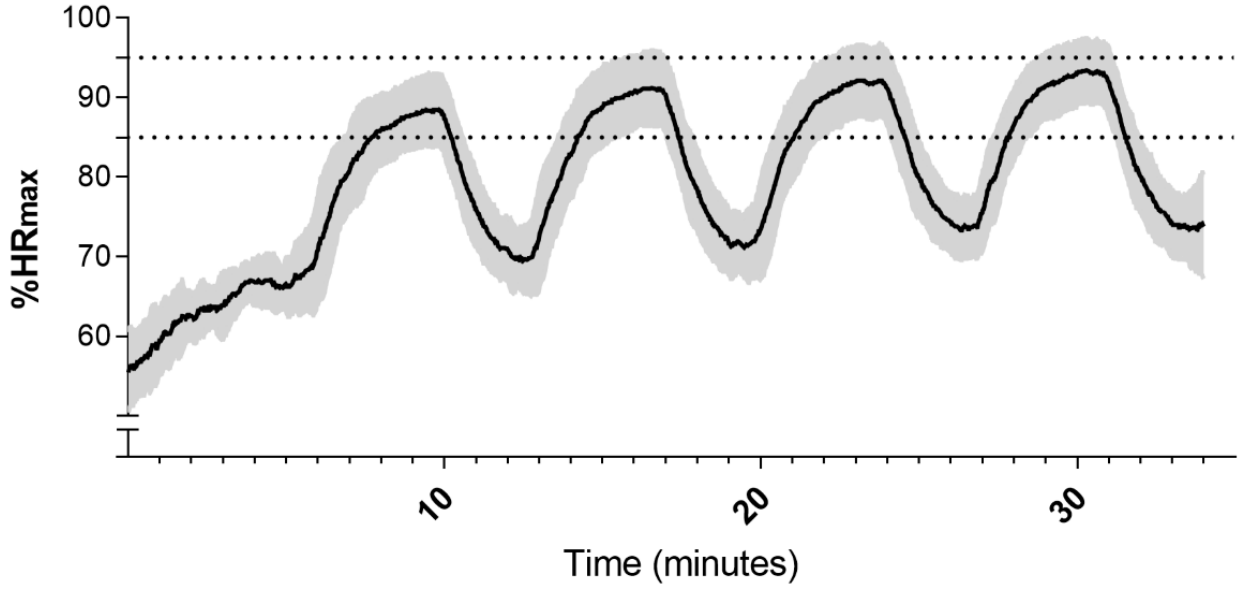
[d]VO_{2max}: maximal oxygen uptake.

[e]HR_{max}: maximal heart rate.

[f]V_E: pulmonary ventilation.

[g]R: respiratory exchange ratio.

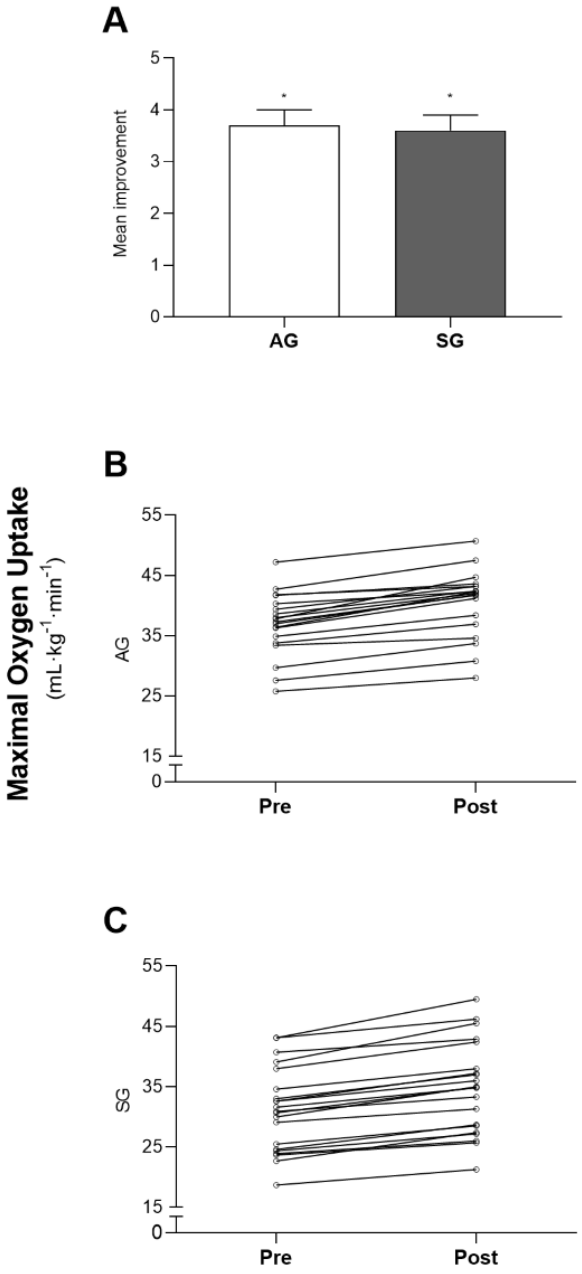
Figure 3. Time course of heart rate (HR) response during a 4×4 high-intensity interval training session halfway through the intervention. Subjects and the supervising physiologist were blinded to HR during the session. Intensity was guided by rate of perceived exertion. The n value is 10. The black line represents the mean. The gray error band represents SD. The area between dotted lines represents intended intensity during intervals.



VO_{2max} and Oxygen Pulse

In accordance with the protocol, all patients reached the VO_{2max} criterion. VO_{2max} increased in both groups from pretraining to posttraining (Table 2). The AG exhibited a 10% (SD 4%) increase (*P*<.001), while the SG exhibited a 12% (SD 4%) increase (*P*<.001). No difference in improvement was apparent between the groups (Figure 4 and Table 2). After 10 weeks of HIIT, the AG and SG showed increased oxygen pulse (both *P*<.001; Table 2), and this was accompanied by increases in ventilation (both *P*<.001; Table 2).

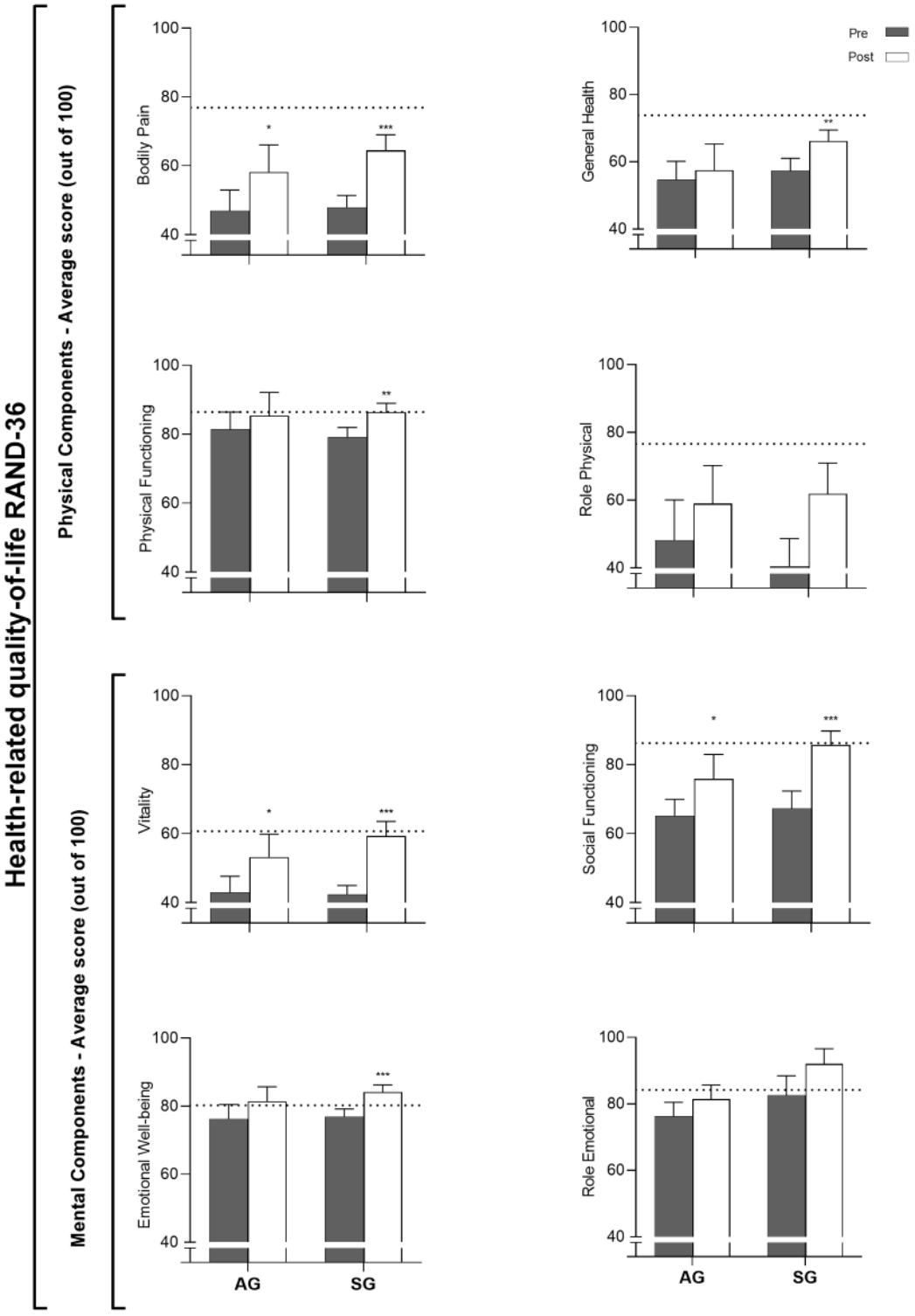
Figure 4. Changes in maximal oxygen uptake (VO_{2max}) in mL per kg of body weight per minute after 10 weeks of high-intensity interval training. (A) Mean (SEM) change from pretraining to posttraining. (B) and (C) Individual values. AG: app group; SG: supervised group. *P<.001, within group difference from pretraining to postraining.



HRQoL

Three HRQoL dimensions improved in both groups following training. *Bodily pain* improved by 11.3 (SD 17.4; *P*=.04) in the AG and 16.7 (SD 12.6; *P*<.001) in the SG, *vitality* improved by 10.4 (SD 13.1; *P*=.01) in the AG and 16.9 (SD 17.8; *P*=.001) in the SG, and *social functioning* improved by 10.7 (SD 18.3; *P*=.04) in the AG and 18.5 (SD 15.1; *P*<.001) in the SG (Figure 5) [42]. No differences were observed between the groups following training. Additionally, the dimensions *general health*, *physical functioning*, and *emotional well-being* increased by 8.8 (SD 10.5; *P*=.003), 7.4 (SD 9.7; *P*=.004), and 7.2 (SD 6.9; *P*=.001), respectively, in the SG (Figure 5). Again, no differences were observed in these dimensions between the groups following training.

Figure 5. Health-related quality of life before and after high-intensity interval training. Values are presented as mean (SEM). Horizontal dotted lines represent normative data. AG: app group (n=14); SG: supervised group (n=21). *P<.05, **P<.01, ***P≤.001; significant within group difference from pretraining.



Discussion

Principal Findings

HIIT has been documented to effectively increase VO_{2max} and reduce the risk of CVD. However, supervised exercise treatment is time and resource demanding. With recent advances in easily available mobile technology, this may provide an opportunity to design effective, low-cost, self-management training programs for patients with chronic conditions like IRD, aiming to improve their VO_{2max}. Thus, in this study, we sought to investigate if an app-guided HIIT intervention could yield some of the increase in VO_{2max} that is typically expected following supervised training. The main findings were as follows: (1) 10 weeks of HIIT (4 times with 4-min intervals) improved VO_{2max} in IRD

patients; (2) the VO_{2max} increase was similar after guidance by an app and individual supervision in the rehabilitation clinic; (3) the VO_{2max} increase was accompanied by an improvement in HRQoL, and the improvement was, again, similar in the 2 groups. In contrast to our initial hypothesis, HIIT guided by an app was as effective in improving VO_{2max} and HRQoL as supervised HIIT. Short-term digital rehabilitation appears to be an excellent cost-effective alternative to supervised clinic rehabilitation of IRD patients, implying reduced risk of CVD, improved performance, and enhanced quality of life.

VO_{2max} and the Magnitude of Improvement

Both the AG and SG showed increased VO_{2max} following 10 weeks of HIIT. However, somewhat surprisingly, the improvement in the AG (3.7 mL/kg/min) was of a similar magnitude as that observed in the SG (3.6 mL/kg/min). Importantly, both groups in this study exhibited a VO_{2max} increase comparable to what has been previously reported following supervised HIIT interventions with similar training intensity and volume in patients with active axial SpA (3.4 mL/kg/min) [43], psoriatic arthritis (3.7 mL/kg/min) [27], and RA and juvenile idiopathic arthritis (4.4 mL/kg/min) [26]. Similar HIIT-induced improvements (mean 3.8, SD 1.1 mL/kg/min) in VO_{2max} have also been reported in healthy men and women with an age comparable to that of our subjects [25].

The similar VO_{2max} improvements across the current app-based investigation and previous conventional, supervised, HIIT studies [25-27,43] strengthens the assumption that app-guided HIIT may be capable of producing an equally potent VO_{2max} increase as that of supervised HIIT. Of importance, the same instructions were given by the app as were orally provided by the instructor in the supervised training sessions. This likely contributed to the same execution of the intervals, as observed in the work output from 6 training sessions. These sessions revealed a similar HIIT intensity of 85%-90% of VO_{2max} in the AG and SG, and consequently, a similar increase in VO_{2max}. Although intensity control was carried out only for the first 3 and last 3 training sessions, it gave a good indication that the app instructions or supervised guiding were perceived well when matched with the controlled intensity at the pretest and posttest. The rationale for not obtaining heart rate data from the AG during the study period was that the intention was to keep contact with the group to a minimum. The similarity in VO_{2max} improvements between previous studies and this study also indicates that both the SG and AG reached the targeted intensity by instructions and the use of RPE. Intensity is crucial, as high intensity has been previously demonstrated to be superior compared to moderate intensity to elicit improvements in VO_{2max} in both healthy individuals [19] and patients [44]. Indeed, the blinded verification tests performed halfway through the intervention in the SG (Figure 3) confirmed that the patients were trained with the intended intensity of 85%-95% HR_{max}. This suggests that a heart rate monitor may not be necessary for training intensity adjustments. Accordingly, a previous study showed that the RPE during supervised heart rate monitor–guided HIIT was 16 (SD 3) on the Borg scale [27], corresponding well with our instructions targeting an RPE of 16 to 17.

VO_{2max}, Cardiac Function, CVD, and Mortality

The improvements in VO_{2max} were accompanied by increases in oxygen pulse in this study. Although an indirect measure, it may indicate a greater stroke volume of the heart [45], as previous observations have revealed that the stroke volume plays a key role when VO_{2max} is reduced or increased [46,47]. Moreover, VO_{2max} improvements following HIIT have previously been documented to be accompanied by improvements in cardiac output and stroke volume, with the arteriovenous oxygen difference remaining unchanged and no increase in HR_{max} [20]. Consequently, it is plausible that the improved oxygen pulse in IRD patients following HIIT in this study indicates improved cardiac function, which would suggest a lower risk of CVD and reduced cardiovascular and all-cause mortality [48]. Comparable to the VO_{2max} improvement in the AG and SG in this study, an increase in the VO_{2max} of 3.5 mL/kg/min has previously been documented to represent a 13% and 15% risk reduction for all-cause and CVD mortality, respectively [49].

VO_{2max} and HRQoL

In this study, IRD patients' VO_{2max} improvements were accompanied by greater quality of life. The HRQoL dimensions *bodily pain*, *vitality*, and *social functioning* were all enhanced in both the AG and SG following HIIT. In particular, *bodily pain* and *vitality* appeared to be severely reduced at baseline. However, following HIIT both dimensions increased closer to normative values (Figure 5), and the observed improvements indicate a reduction in symptom burden in IRD patients [50]. Interestingly, despite the SG involving social interaction between the patient and therapist, the AG exhibited a similar enhancement of these dimensions. Previous observations have shown a positive correlation between VO_{2max} and several aspects of HRQoL [51]. Thus, the positive effects of self-administered app-guided exercise on HRQoL may be explained by the increase in VO_{2max}. Conversely, a VO_{2max} decrease in IRD patients has previously been associated with an impaired quality of life [52]. Moreover, worse self-reported physical functioning has also been associated with poorer HRQoL in patients with SpA [50]. Although both groups similarly showed improved VO_{2max} and most aspects of HRQoL, only the SG showed improvement in the dimension *physical functioning*. In RA patients, a minimally clinically important difference of 7.7 points in *physical functioning* has previously been described [53]. However, others have described minimally clinically important differences of 3 to 5 points [54]. The mean improvement in *physical functioning* was 3.9 (SD 10.4) and 7.4 (SD 9.7) in the AG and SG, respectively. Thus, the improvements may be below clinical importance or of marginal clinical importance in both the AG and SG. Notably, both groups were close to the normative score for the Norwegian population at baseline. Therefore, the potential for improvement was likely limited.

HIIT: Man or Machine

Previous studies have typically reported self-administered exercise programs to yield smaller VO_{2max} improvements compared to supervised programs [29,55]. It is thus surprising that the AG and SG exhibited a similar VO_{2max} improvement in this study. The results indicate that the HIIT intervention was likely carried out similarly in the 2 training groups. Indeed, the compliance rate was not different between the 2 groups and revealed that both groups completed more than 90% of the planned sessions. This is in contrast to most previous studies documenting compliance rates to be lower after self-administered programs. For example, in a study by Cox et al [56], the compliance rate after a home-based exercise program was approximately 63% compared with approximately 84% after supervised center-based training. Interestingly, in accordance with the notion that compliance rates may explain the lower VO_{2max} response, the attenuated VO_{2max} improvement after a self-administered cardiac rehabilitation HIIT intervention in the study by Aamot et al [29] dissipated when subjects who completed less than 70% of scheduled sessions were excluded from the analysis.

The high compliance rate for HIIT in this study may be explained by the format of self-administered training. Although the AG was not supervised, automated audio instructions were given during the exercise along with written instructions within the app for guiding the intensity. Additionally, subjects consented to give researchers access to the logged sessions from the app's server, and the awareness that they could be monitored might have made the patients commit more to the training. Halfway through the intervention, an email was sent to the subjects who had logged less than 70% of the scheduled sessions, reminding them to keep up the exercise. Thus, the AG should not be considered nonsupervised; subjects in this group were guided by the instructions in the app and were given a push notification by email. Such remote support and virtual guidance through apps have improved compliance to exercise programs in other patient groups [31,57]. For example, patients with diabetes completed as much as 95% of the planned self-administered HIIT sessions over 6 weeks when heart rate monitoring was combined with an app and an email push notification [58]. Considering compliance rates of approximately 80% [43] and 78% [27] in previous studies with IRD patients, the app certainly appears to be a viable alternative to supervised sessions, indicating that both training program compliance and execution of each training session were good.

Safety

Previous studies in patients with IRD have reported no severe adverse cardiovascular events following incremental exercise testing or HIIT performed at an intensity of 85% to 95% HR_{max} [18,26,27,43]. In patients with coronary artery disease (CAD), the risk of cardiac incidents has been reported to be low when performing HIIT [40]. Similarly, when CAD patients followed unsupervised HIIT, no adverse cardiovascular events were observed [29]. The authors in the latter study stated that HIIT appears to be safe in CAD patients given that an incremental cardiorespiratory exercise test establishing exercise tolerance is performed before the HIIT starts. Considering previous literature and that there were no adverse events following the supervised VO_{2max} testing in either group, the initiation of both HIIT interventions was considered relatively safe in this study.

Clinical Implications

The increasing availability of smartphone technology has introduced digital possibilities for delivering physical rehabilitation interventions. As demonstrated in this study, it appears to yield similar effects on VO_{2max} for IRD patients as effective, conventional, supervised training carried out in a rehabilitation clinic. However, it may be necessary to combine it with interactive feedback to provide successful self-administered exercise treatment. Although carried out outside the rehabilitation clinic, the utilization of apps can offer therapists detailed information on how the training is performed and the opportunity to provide detailed adjustments; however, the latter was not done in this study. Moreover, this study demonstrates that instructions and in-app information on RPE-guided exercise intensity may be an excellent alternative to heart rate monitoring of training sessions, making the administration even simpler. Rehabilitation clinics may also reach more patients through digital rehabilitation, as travel time and physical presence may limit some patients from attending treatment. Digital rehabilitation also offers a more cost-effective approach to exercise rehabilitation and may even result in enhanced patient satisfaction [59].

Interestingly, all patients in this study were able to reach VO_{2max} at pretest, implying a good tolerance for intensive training before the initiation of HIIT. For patients with confirmed or high risk of CVD or other severe concomitant diseases, unsupervised HIIT may be used in collaboration with health care professionals. Incorporating self-administered HIIT in such a way may increase patient self-efficacy and, at the same time, free up time and resources for both the patient and the treating health care professional.

Strengths and Limitations

This study had both strengths and limitations. Allocation of patients to the AG or SG was randomized, and testers were blinded to subject allocation, which helped prevent possible selection bias. On the other hand, motivation to volunteer for participation in a research study might result in inclusion of subjects with higher internal motivation to adhere to the treatment. Notably, the group characteristics (Table 1) and VO_{2max} (Table 2) appear to be similar to data for RA and SpA populations [1,2,18,26,27], indicating that the findings may be representative of these IRD populations. Although SLE patients had typical patient characteristics in this study [3,4,60], their VO_{2max} was relatively high. Importantly, our data revealed that all 3 IRD subpopulations exhibited similar responses to HIIT and no adverse events were documented. Furthermore, this study was designed with the scope of a per protocol analysis. Hence, from an ethical perspective, posttests were only conducted with patients who were to be included in the analysis. Though this type of analysis gives great insights into the effects when subjects adhere to treatment, information that more closely reflects the clinical setting might have been obtained with an intention-to-treat analysis. This represents a limitation in this

study and needs to be investigated in future research. Another limitation in this study is the lack of direct monitoring of intensity in the AG. Postexercise RPE reporting in a diary could have been done. However, such self-recorded variables often result in missing data owing to low compliance. Heart rate monitoring during home exercise would have been a viable option. Such information could be collected directly through the app, but it requires the subject to learn about wearing a monitor and to wear the monitor for each session. This may pose an added barrier to complying with the intervention. Importantly, the findings of this study indicate that such equipment or monitoring is not necessary to induce similar effects as that of supervised HIIT. Another possible limitation in this investigation is the missing data of 5 subjects in the AG for the secondary outcome (HRQoL). The reasons for not responding are unknown. Supervising self-reported questionnaire administration and requiring subjects to answer questions they do not want to or to give reasoning for not answering some questions may raise ethical concerns. Thus, controlling for such missing data is a challenge. Providing sufficient information and instructions prior to handing out questionnaires was emphasized, and further emphasis on the importance of submitting complete forms and ensuring that subjects are satisfied with their anonymity should be considered.

Conclusion

Similar increases in VO_{2max} were observed after HIIT in IRD patients who were individually supervised or guided by oral and written instructions in the app in combination with an email reminder. The VO_{2max} improvements likely contributed to reduced risk of CVD and were accompanied by improvements in HRQoL, again with similar results between the AG and SG. Digital rehabilitation, at least in a short-term perspective, appears to be an excellent cost-effective strategy to improve the health and performance of IRD patients, and should be considered in clinical practice in the future.

Acknowledgments

The study was supported by the Norwegian Research Council (grant number 305506). The funding organization had no role in the design and execution of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

HH, J Helgerud, and J Hoff conceived and designed the experiment. HH and J Helgerud conducted the experiment. HH, J Helgerud, OKB, and EW analyzed the data, interpreted the results, and wrote the manuscript. J Hoff provided critical inputs and contributed in writing the manuscript.

Conflicts of Interest

HH is employed by Myworkout, Medical Rehabilitation Clinic owned by Myworkout AS, the developer of the smartphone app Myworkout GO. J Helgerud and J Hoff are board members and shareholders of Myworkout AS. Due to the potential conflict of interest, EW and OKB monitored adherence to the design and statistical analysis in the study. There are no further disclosures or potential conflicts of interest to report.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 605 KB - mhealth_v9i10e28124_app1.pdf]

References

1. Munsterman T, Takken T, Wittink H. Are persons with rheumatoid arthritis deconditioned? A review of physical activity and aerobic capacity. BMC Musculoskelet Disord 2012 Oct 18;13:202 [FREE Full text] [doi: 10.1186/1471-2474-13-202] [Medline: 23078261]
2. Halvorsen S, Vøllestad NK, Fongen C, Provan SA, Semb AG, Hagen KB, et al. Physical fitness in patients with ankylosing spondylitis: comparison with population controls. Phys Ther 2012 Feb;92(2):298-309. [doi: 10.2522/ptj.20110137] [Medline: 22095208]
3. Keyser RE, Rus V, Cade WT, Kalappa N, Flores RH, Handwerger BS. Evidence for aerobic insufficiency in women with systemic Lupus erythematosus. Arthritis Rheum 2003 Feb 15;49(1):16-22 [FREE Full text] [doi: 10.1002/art.10926] [Medline: 12579589]
4. Ayán C, Martín V. Systemic lupus erythematosus and exercise. Lupus 2007;16(1):5-9. [doi: 10.1177/0961203306074795] [Medline: 17283578]
5. Sparks JA. Rheumatoid Arthritis. Ann Intern Med 2019 Jan 01;170(1):ITC1-ITC16. [doi: 10.7326/AITC201901010] [Medline: 30596879]
6. Perrotta FM, Lories R, Lubrano E. To move or not to move: the paradoxical effect of physical exercise in axial spondyloarthritis. RMD Open 2021 Feb;7(1) [FREE Full text] [doi: 10.1136/rmdopen-2020-001480] [Medline: 33547227]

7. Veldhuijzen van Zanten JJCS, Rouse PC, Hale ED, Ntoumanis N, Metsios GS, Duda JL, et al. Perceived Barriers, Facilitators and Benefits for Regular Physical Activity and Exercise in Patients with Rheumatoid Arthritis: A Review of the Literature. Sports Med 2015 Oct;45(10):1401-1412 [FREE Full text] [doi: 10.1007/s40279-015-0363-2] [Medline: 26219268]
8. Fongen C, Sveaas SH, Dagfinrud H. Barriers and Facilitators for Being Physically Active in Patients with Ankylosing Spondylitis: A Cross-sectional Comparative Study. Musculoskeletal Care 2015 Jun;13(2):76-83. [doi: 10.1002/msc.1088] [Medline: 25370164]
9. Mancuso CA, Perna M, Sargent AB, Salmon JE. Perceptions and measurements of physical activity in patients with systemic lupus erythematosus. Lupus 2011 Mar;20(3):231-242. [doi: 10.1177/0961203310383737] [Medline: 21183562]
10. Agca R, Heslinga SC, van Halm VP, Nurmohamed MT. Atherosclerotic cardiovascular disease in patients with chronic inflammatory joint disorders. Heart 2016 May 15;102(10):790-795. [doi: 10.1136/heartjnl-2015-307838] [Medline: 26888573]
11. Harber MP, Kaminsky LA, Arena R, Blair SN, Franklin BA, Myers J, et al. Impact of cardiorespiratory fitness on all-cause and disease-specific mortality: advances since 2009. Prog Cardiovasc Dis 2017 Jul;60(1):11-20. [doi: 10.1016/j.pcad.2017.03.001] [Medline: 28286137]
12. Metsios GS, Koutedakis Y, Veldhuijzen van Zanten JJCS, Stavropoulos-Kalinoglou A, Vitalis P, Duda JL, et al. Cardiorespiratory fitness levels and their association with cardiovascular profile in patients with rheumatoid arthritis: a cross-sectional study. Rheumatology (Oxford) 2015 Dec;54(12):2215-2220. [doi: 10.1093/rheumatology/kev035] [Medline: 26209790]
13. Symmons DPM, Gabriel SE. Epidemiology of CVD in rheumatic disease, with a focus on RA and SLE. Nat Rev Rheumatol 2011 May 31;7(7):399-408. [doi: 10.1038/nrrheum.2011.75] [Medline: 21629241]
14. Mason JC, Libby P. Cardiovascular disease in patients with chronic inflammation: mechanisms underlying premature cardiovascular events in rheumatologic conditions. Eur Heart J 2015 Feb 21;36(8):482-49c [FREE Full text] [doi: 10.1093/eurheartj/ehu403] [Medline: 25433021]
15. Liew JW, Ramiro S, Gensler LS. Cardiovascular morbidity and mortality in ankylosing spondylitis and psoriatic arthritis. Best Pract Res Clin Rheumatol 2018 Jun;32(3):369-389. [doi: 10.1016/j.berh.2019.01.002] [Medline: 31171309]
16. Prasad M, Hermann J, Gabriel SE, Weyand CM, Mulvagh S, Mankad R, et al. Cardiorheumatology: cardiac involvement in systemic rheumatic disease. Nat Rev Cardiol 2015 Mar;12(3):168-176 [FREE Full text] [doi: 10.1038/nrcardio.2014.206] [Medline: 25533796]
17. Salaffi F, Di Carlo M, Carotti M, Farah S. The Patient-Reported Outcomes Thermometer-5-Item Scale (5T-PROs): Validation of a New Tool for the Quick Assessment of Overall Health Status in Painful Rheumatic Diseases. Pain Res Manag 2018;2018:3496846 [FREE Full text] [doi: 10.1155/2018/3496846] [Medline: 30425755]
18. Sveaas SH, Bilberg A, Berg IJ, Provan SA, Rollefstad S, Semb AG, et al. High intensity exercise for 3 months reduces disease activity in axial spondyloarthritis (axSpA): a multicentre randomised trial of 100 patients. Br J Sports Med 2020 Mar;54(5):292-297. [doi: 10.1136/bjsports-2018-099943] [Medline: 30745314]
19. Helgerud J, Høydal K, Wang E, Karlsen T, Berg P, Bjerkaas M, et al. Aerobic high-intensity intervals improve VO2max more than moderate training. Med Sci Sports Exerc 2007 Apr;39(4):665-671. [doi: 10.1249/mss.0b013e3180304570] [Medline: 17414804]
20. Wang E, Næss MS, Hoff J, Albert TL, Pham Q, Richardson RS, et al. Exercise-training-induced changes in metabolic capacity with age: the role of central cardiovascular plasticity. Age (Dordr) 2014 Apr;36(2):665-676 [FREE Full text] [doi: 10.1007/s11357-013-9596-x] [Medline: 24243396]
21. Helgerud J, Karlsen T, Kim WY, Høydal KL, Støylen A, Pedersen H, et al. Interval and strength training in CAD patients. Int J Sports Med 2011 Jan;32(1):54-59. [doi: 10.1055/s-0030-1267180] [Medline: 21072747]
22. Gjellesvik TI, Brurok B, Hoff J, Tørhaug T, Helgerud J. Effect of high aerobic intensity interval treadmill walking in people with chronic stroke: a pilot study with one year follow-up. Top Stroke Rehabil 2012;19(4):353-360. [doi: 10.1310/tsr1904-353] [Medline: 22750965]
23. Støa EM, Meling S, Nyhus L, Glenn S, Mangerud KM, Helgerud J, et al. High-intensity aerobic interval training improves aerobic fitness and HbA1c among persons diagnosed with type 2 diabetes. Eur J Appl Physiol 2017 Mar;117(3):455-467. [doi: 10.1007/s00421-017-3540-1] [Medline: 28160083]
24. Bjørgen S, Hoff J, Husby VS, Høydal MA, Tjønna AE, Steinshamn S, et al. Aerobic high intensity one and two legs interval cycling in chronic obstructive pulmonary disease: the sum of the parts is greater than the whole. Eur J Appl Physiol 2009 Jul;106(4):501-507. [doi: 10.1007/s00421-009-1038-1] [Medline: 19337746]
25. Støren Ø, Helgerud J, Sæbø M, Støa EM, Bratland-Sanda S, Unhjem RJ, et al. The Effect of Age on the V˙O2max Response to High-Intensity Interval Training. Med Sci Sports Exerc 2017 Jan;49(1):78-85. [doi: 10.1249/MSS.0000000000001070] [Medline: 27501361]
26. Sandstad J, Stensvold D, Hoff M, Nes BM, Arbo I, Bye A. The effects of high intensity interval training in women with rheumatic disease: a pilot study. Eur J Appl Physiol 2015 Oct;115(10):2081-2089. [doi: 10.1007/s00421-015-3186-9] [Medline: 26013051]

27. Thomsen RS, Nilsen TIL, Haugeberg G, Bye A, Kavanaugh A, Hoff M. Effect of high-intensity interval training on cardiovascular disease risk factors and body composition in psoriatic arthritis: a randomised controlled trial. RMD Open 2018;4(2):e000729 [FREE Full text] [doi: 10.1136/rmdopen-2018-000729] [Medline: 30402265]
28. Bichay AAF, Ramírez JM, Núñez VM, Lancho C, Poblador MS, Lancho JL. Efficacy of treadmill exercises on arterial blood oxygenation, oxygen consumption and walking distance in healthy elderly people: a controlled trial. BMC Geriatr 2016 May 25;16:110 [FREE Full text] [doi: 10.1186/s12877-016-0283-5] [Medline: 27225602]
29. Aamot I, Forbord SH, Gustad K, Løckra V, Stensen A, Berg AT, et al. Home-based versus hospital-based high-intensity interval training in cardiac rehabilitation: a randomized study. Eur J Prev Cardiol 2014 Sep;21(9):1070-1078. [doi: 10.1177/2047487313488299] [Medline: 23613224]
30. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
31. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
32. Bonato M, Turrini F, DE Zan V, Meloni A, Plebani M, Brambilla E, et al. A mobile application for exercise intervention in people living with HIV. Med Sci Sports Exerc 2020 Feb;52(2):425-433. [doi: 10.1249/MSS.0000000000002125] [Medline: 31415448]
33. Nabutovsky I, Ashri S, Nachshon A, Tesler R, Shapiro Y, Wright E, et al. Feasibility, Safety, and Effectiveness of a Mobile Application in Cardiac Rehabilitation. Isr Med Assoc J 2020 Jun;22(6):357-363 [FREE Full text] [Medline: 32558441]
34. Berglind D, Yacaman-Mendez D, Lavebratt C, Forsell Y. The Effect of Smartphone Apps Versus Supervised Exercise on Physical Activity, Cardiorespiratory Fitness, and Body Composition Among Individuals With Mild-to-Moderate Mobility Disability: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Feb 04;8(2):e14615 [FREE Full text] [doi: 10.2196/14615] [Medline: 32014846]
35. Mezzani A, Hamm LF, Jones AM, McBride PE, Moholdt T, Stone JA, European Association for Cardiovascular PreventionRehabilitation, American Association of CardiovascularPulmonary Rehabilitation, Canadian Association of Cardiac Rehabilitation. Aerobic exercise intensity assessment and prescription in cardiac rehabilitation: a joint position statement of the European Association for Cardiovascular Prevention and Rehabilitation, the American Association of Cardiovascular and Pulmonary Rehabilitation and the Canadian Association of Cardiac Rehabilitation. Eur J Prev Cardiol 2013 Jun;20(3):442-467. [doi: 10.1177/2047487312460484] [Medline: 23104970]
36. Wang E, Solli GS, Nyberg SK, Hoff J, Helgerud J. Stroke volume does not plateau in female endurance athletes. Int J Sports Med 2012 Sep;33(9):734-739. [doi: 10.1055/s-0031-1301315] [Medline: 22581683]
37. Loge JH, Kaasa S, Hjermstad MJ, Kvien TK. Translation and Performance of the Norwegian SF-36 Health Survey in Patients with Rheumatoid Arthritis. I. Data Quality, Scaling Assumptions, Reliability, and Construct Validity. Journal of Clinical Epidemiology 1998 Nov;51(11):1069-1076. [doi: 10.1016/s0895-4356(98)00098-5] [Medline: 9817124]
38. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
39. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ 1993 Oct;2(3):217-227. [doi: 10.1002/hec.4730020305] [Medline: 8275167]
40. Rognmo Ø, Moholdt T, Bakken H, Hole T, Mølstad P, Myhr NE, et al. Cardiovascular risk of high- versus moderate-intensity aerobic exercise in coronary heart disease patients. Circulation 2012 Sep 18;126(12):1436-1440. [doi: 10.1161/CIRCULATIONAHA.112.123117] [Medline: 22879367]
41. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, PA: Wolters Kluwer; 2018.
42. Garratt AM, Stavem K. Measurement properties and normative data for the Norwegian SF-36: results from a general population survey. Health Qual Life Outcomes 2017 Mar 14;15(1):51 [FREE Full text] [doi: 10.1186/s12955-017-0625-9] [Medline: 28292292]
43. Sveaas SH, Berg IJ, Provan SA, Semb AG, Hagen KB, Vøllestad N, et al. Efficacy of high intensity exercise on disease activity and cardiovascular risk in active axial spondyloarthritis: a randomized controlled pilot study. PLoS One 2014;9(9):e108688 [FREE Full text] [doi: 10.1371/journal.pone.0108688] [Medline: 25268365]
44. Rognmo Ø, Hetland E, Helgerud J, Hoff J, Slørdahl SA. High intensity aerobic interval exercise is superior to moderate intensity exercise for increasing aerobic capacity in patients with coronary artery disease. Eur J Cardiovasc Prev Rehabil 2004 Jun;11(3):216-222. [doi: 10.1097/01.hjr.0000131677.96762.0c] [Medline: 15179103]
45. Van Iterson EH, Baker SE, Wheatley CM, Morgan WJ, Olson TP, Snyder EM. Exercise Stroke Volume in Adult Cystic Fibrosis: A Comparison of Acetylene Pulmonary Uptake and Oxygen Pulse. Clin Med Insights Circ Respir Pulm Med 2018;12:1179548418790564 [FREE Full text] [doi: 10.1177/1179548418790564] [Medline: 30083061]
46. Saltin B, Blomqvist G, Mitchell JH, Johnson RL, Wildenthal K, Chapman CB. Response to exercise after bed rest and after training. Circulation 1968 Nov;38(5 Suppl):VII1-VI78. [Medline: 5696236]

47. Saltin B, Calbet JAL. Point: in health and in a normoxic environment, VO2max is limited primarily by cardiac output and locomotor muscle blood flow. J Appl Physiol (1985) 2006 Feb;100(2):744-745 [FREE Full text] [doi: 10.1152/japplphysiol.01395.2005] [Medline: 16421282]
48. Laukkanen JA, Araújo CGS, Kurl S, Khan H, Jae SY, Guazzi M, et al. Relative peak exercise oxygen pulse is related to sudden cardiac death, cardiovascular and all-cause mortality in middle-aged men. Eur J Prev Cardiol 2018 May;25(7):772-782. [doi: 10.1177/2047487318761679] [Medline: 29488810]
49. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA 2009 May 20;301(19):2024-2035. [doi: 10.1001/jama.2009.681] [Medline: 19454641]
50. Macfarlane GJ, Rotariu O, Jones GT, Pathan E, Dean LE. Determining factors related to poor quality of life in patients with axial spondyloarthritis: results from the British Society for Rheumatology Biologics Register (BSRBR-AS). Ann Rheum Dis 2020 Feb 29;79(2):202-208. [doi: 10.1136/annrheumdis-2019-216143] [Medline: 31662321]
51. Oscar García López O, Duarte Bedoya Á, Jiménez Gutiérrez A, Burgos Postigo S. Relationship between Short-Form Health SF36 Questionnaire and oxygen uptake in healthy workers. J Sports Med Phys Fitness 2016 Mar;56(3):280-286. [Medline: 26609963]
52. Çağliyan Türk A, Arslan S, Karavelioğlu Y, Kalcik M, Özel S, Musmul A, et al. Pulmonary Function, Aerobic Capacity and Related Variables in Patients With Ankylosing Spondylitis. Arch Rheumatol 2019 Sep;34(3):317-325 [FREE Full text] [doi: 10.5606/ArchRheumatol.2019.7145] [Medline: 31598598]
53. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis & Rheumatism 2001 Jul;43(7):1478-1487 [FREE Full text] [doi: 10.1002/1529-0131(200007)43:7<1478::AID-ANR10>3.0.CO;2-M] [Medline: 10902749]
54. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. Pharmacoeconomics 1999 Feb;15(2):141-155. [doi: 10.2165/00019053-199915020-00003] [Medline: 10351188]
55. Regensteiner JG, Meyer TJ, Krupski WC, Cranford LS, Hiatt WR. Hospital vs home-based exercise rehabilitation for patients with peripheral arterial occlusive disease. Angiology 1997 Apr;48(4):291-300. [doi: 10.1177/000331979704800402] [Medline: 9112877]
56. Cox KL, Burke V, Gorely TJ, Beilin LJ, Puddey IB. Controlled comparison of retention and adherence in home- vs center-initiated exercise interventions in women ages 40-65 years: The S.W.E.A.T. Study (Sedentary Women Exercise Adherence Trial). Prev Med 2003 Jan;36(1):17-29. [doi: 10.1006/pmed.2002.1134] [Medline: 12473421]
57. Xu L, Li F, Zhou C, Li J, Hong C, Tong Q. The effect of mobile applications for improving adherence in cardiac rehabilitation: a systematic review and meta-analysis. BMC Cardiovasc Disord 2019 Jul 12;19(1):166 [FREE Full text] [doi: 10.1186/s12872-019-1149-5] [Medline: 31299903]
58. Scott SN, Shepherd SO, Andrews RC, Narendran P, Purewal TS, Kinnafick F, et al. A Multidisciplinary Evaluation of a Virtually Supervised Home-Based High-Intensity Interval Training Intervention in People With Type 1 Diabetes. Diabetes Care 2019 Dec;42(12):2330-2333. [doi: 10.2337/dc19-0871] [Medline: 31530660]
59. Kraal JJ, Van den Akker-Van Marle ME, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: Results of the FIT@Home study. Eur J Prev Cardiol 2017 Aug;24(12):1260-1273 [FREE Full text] [doi: 10.1177/2047487317710803] [Medline: 28534417]
60. Boström C, Dupré B, Tengvar P, Jansson E, Opava CH, Lundberg IE. Aerobic capacity correlates to self-assessed physical function but not to overall disease activity or organ damage in women with systemic lupus erythematosus with low-to-moderate disease activity and organ damage. Lupus 2008 Feb;17(2):100-104. [doi: 10.1177/0961203307085670] [Medline: 18250132]

Abbreviations

ACSM: American College of Sports Medicine
AG: app group
CAD: coronary artery disease
CVD: cardiovascular disease
HIIT: high-intensity interval training
HRmax: maximal heart rate
HRQoL: health-related quality of life
IRD: inflammatory rheumatic disease
RA: rheumatoid arthritis
RPE: rate of perceived exertion
SG: supervised group

SLE: systemic lupus erythematosus
SpA: spondyloarthritis
VO_{2max}: maximal oxygen uptake

Edited by L Buis; submitted 22.02.21; peer-reviewed by A Bilberg, W Stut, A Williams; comments to author 01.04.21; revised version received 27.05.21; accepted 25.06.21; published 21.10.21.

Please cite as:
Haglo H, Wang E, Berg OK, Hoff J, Helgerud J
Smartphone-Assisted High-Intensity Interval Training in Inflammatory Rheumatic Disease Patients: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(10):e28124
URL: https://mhealth.jmir.org/2021/10/e28124
doi:10.2196/28124
PMID:34673536

Original Paper

A Mobile Health Salt Reduction Intervention for People With Hypertension: Results of a Feasibility Randomized Controlled Trial

Sarah Payne Riches[1], MPH, DPhil; Carmen Piernas[1], MSc, PhD; Paul Aveyard[1,2], PhD; James P Sheppard[1], BSc, PhD; Mike Rayner[3], BA, DPhil; Charlotte Albury[1], BA, MSc, DPhil; Susan A Jebb[1,2], DPhil

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]National Institute for Health Research, Oxford Biomedical Research Centre, Oxford University Hospitals, Oxford, United Kingdom

[3]Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Susan A Jebb, DPhil
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865 289300
Email: susan.jebb@phc.ox.ac.uk

Abstract

Background: A high-salt diet is a risk factor for hypertension and cardiovascular disease; therefore, reducing dietary salt intake is a key part of prevention strategies. There are few effective salt reduction interventions suitable for delivery in the primary care setting, where the majority of the management and diagnosis of hypertension occurs.

Objective: The aim of this study is to assess the feasibility of a complex behavioral intervention to lower salt intake in people with elevated blood pressure and test the trial procedures for a randomized controlled trial to investigate the intervention's effectiveness.

Methods: This feasibility study was an unblinded, randomized controlled trial of a mobile health intervention for salt reduction versus an advice leaflet (control). The intervention was developed using the Behavior Change Wheel and comprised individualized, brief advice from a health care professional with the use of the SaltSwap app. Participants with an elevated blood pressure recorded in the clinic were recruited through primary care practices in the United Kingdom. Primary outcomes assessed the feasibility of progression to a larger trial, including follow-up attendance, fidelity of intervention delivery, and app use. Secondary outcomes were objectively assessed using changes in salt intake (measured via 24-hour urine collection), salt content of purchased foods, and blood pressure. Qualitative outcomes were assessed using the think-aloud method, and the process outcomes were evaluated.

Results: A total of 47 participants were randomized. All progression criteria were met: follow-up attendance (45/47, 96%), intervention fidelity (25/31, 81%), and app use (27/31, 87%). There was no evidence that the intervention significantly reduced the salt content of purchased foods, salt intake, or blood pressure; however, this feasibility study was not powered to detect changes in secondary outcomes. Process and qualitative outcomes demonstrated that the trial design was feasible and the intervention was acceptable to both individuals and practitioners and positively influenced salt intake behaviors.

Conclusions: The intervention was acceptable and feasible to deliver within primary care; the trial procedures were practicable, and there was sufficient signal of potential efficacy to change salt intake. With some improvements to the intervention app, a larger trial to assess intervention effectiveness for reducing salt intake and blood pressure is warranted.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 20910962; https://www.isrctn.com/ISRCTN20910962

(JMIR Mhealth Uhealth 2021;9(10):e26233) doi:10.2196/26233

KEYWORDS

salt reduction; behavior change; mobile health; mHealth; smartphone app; mobile phone

Introduction

Background

Cardiovascular disease (CVD) is the leading cause of premature mortality worldwide [1], and one of the leading risk factors for CVD is hypertension [2-4]. Hypertension affects a quarter of the adult population in the United Kingdom [5]. A high-salt diet increases blood pressure [6-14], and reducing salt intake has consistently been shown to reduce blood pressure in those with hypertension [15]. Recent estimates have identified a high-sodium diet as the largest *dietary* risk factor for global mortality [16]. Therefore, reducing dietary salt intake is a key factor in CVD prevention. In the United Kingdom, for example, adults consume, on average, 40% more salt than that recommended by the national and international dietary guidelines [17,18], with the majority of salt derived from processed foods rather than salt added during cooking or at the table [19].

A population-level salt reduction strategy was implemented in the United Kingdom in 2003, which achieved some success in reducing the population-level salt intake [20]. However, there has been no significant reduction in average population salt intake since 2009 [17]. Despite good evidence for the effectiveness of individual-level interventions centered on dietary advice and acquisition of skills to achieve and maintain a low-salt diet [7,8,21], effective interventions tend to be intensive or require specialist staff. There are few interventions suitable for delivery in the primary care setting [22], where the majority of the management and diagnosis of hypertension occurs [23]. In the United Kingdom, clinicians are encouraged, through national guidelines, to advise reduced salt intake when managing hypertension, but they have no tools, structured advice, or guidance for doing so effectively [23,24]. *Brief interventions* are those which can be delivered by clinical staff without specialist nutrition or behavior change knowledge and require minimal health professional contact [25]. Such interventions could offer significant and sustainable impacts with less participant burden and greater reach than more intensive approaches. Behaviorally informed, brief interventions for dietary behavior change have been used successfully for weight loss or maintenance and cholesterol reduction, using techniques such as self-regulation or habit control to drive behavior change [26-30] but not, to our knowledge, for salt reduction in a clinically at-risk population.

Mobile apps have been shown to support dietary changes for weight loss or increased fruit and vegetable consumption [31,32]. Although high-quality evidence for the use of mobile apps in salt reduction is limited, recent studies have shown promising effects. A study investigating a daily diet monitoring app in a university population in the United States showed a significant reduction in salt intake over 4 weeks [33]; a study using an app to encourage the purchase of lower-salt food in New Zealand found a significant reduction in recorded salt purchases but no evidence of reductions in the overall salt intake [34]. A further study in the United States, using geofencing to send targeted messages on salt reduction based on location, demonstrated a reduction in the estimated salt intake from spot urine samples but no significant change in daily urinary sodium excretion [35]. The potential benefits of using apps to deliver health interventions include the low relative cost of delivery, the opportunity to deliver at scale, the low cost compared with face-to-face interventions, the flexibility to allow tailoring to individuals, and the ability to deliver an intervention with consistency. The fact that individuals often have their devices with them when they are undertaking the target behavior could make the intervention more salient. We propose that mobile apps could complement and augment brief interventions for salt reduction delivered in primary care—the principal setting for the diagnosis and management of hypertension in the United Kingdom.

Objectives

We developed a behaviorally informed brief intervention, incorporating a mobile app, to motivate and support individuals to reduce their salt intake by choosing lower-salt foods when grocery shopping. This paper describes the design of the SaltSwap intervention and reports the results of a randomized controlled trial (RCT) that incorporates qualitative and process outcomes. The trial examined whether the intervention is feasible to deliver in primary care and if it has the potential to reduce salt intake and blood pressure and tested the procedures for a future effectiveness trial.

Methods

Study Design and Participants

This feasibility study was an individually randomized, parallel, 2-arm, controlled trial. After providing consent and completing the baseline assessment, participants entered a 2-week run-in period, and those who provided a baseline urine sample and grocery shopping data during this period were randomized. After the 6-week intervention period, the participants attended a single follow-up visit. The study was reviewed and approved on March 17, 2017, by the National Health Service Research Ethics Committee and the Health Research Authority (reference 17/SC/0098). The trial protocol was registered on April 5, 2017. The study methods are published elsewhere [36] and are summarized below.

Recruitment

Participants were recruited through 5 general practitioner (GP) surgeries in Oxfordshire, United Kingdom. GP surgery electronic health records were searched for adults with a recent blood pressure reading (systolic blood pressure in the past 2 years above 130 mm Hg if currently prescribed a stable-dose, antihypertensive medication or above 140 mm Hg if not prescribed medication), and these patients were invited by letter to participate. Letters included a web link to web-based study information. Major exclusion criteria included the following: secondary, previous accelerated or malignant hypertension,

currently being assessed for diagnosis of hypertension; existing or recent cardiovascular conditions; or following a clinician-supervised diet or a restricted diet. The principal investigator screened interested participants by phone according to the full inclusion and exclusion [36] criteria and provided study information. Eligible participants were booked into a baseline study visit, where a study researcher provided written study information and confirmed consent in person.

Randomization

Participants were randomly allocated in a 2:1 ratio (intervention:control) using a computer-generated allocation sequence stratified by GP surgery. An independent researcher generated a random number sequence using a web-based random sequence generator and informed the principal investigator of the intervention allocation for each participant. Participants and investigators were unaware of the treatment allocation unaware of the treatment allocation prior to participant consent. Owing to the nature of the intervention and trial procedures, it was not possible to blind participants, clinicians delivering the intervention, or the outcome assessor to the treatment group.

Intervention

The SaltSwap intervention aimed to reduce dietary salt intake by encouraging individuals to swap to lower-salt alternatives when grocery shopping, buy fewer high-salt foods, and use less salt when cooking or at the table. We designed the intervention using the Behavior Change Wheel, a theoretical framework widely used to develop health behavior change interventions [37] (Multimedia Appendix 1 [38]). We used a co-design approach, recruiting a panel of individuals diagnosed with hypertension via a preexisting database of volunteers and advertisements in GP surgeries in Oxford. This panel informed the behavioral analysis, intervention content, mode of delivery, and key questions for intervention evaluation. Panel members also reviewed the final intervention design and conducted the initial pilot testing. The intervention comprised behavioral advice and support, provided by a health care professional (typically a nurse or health care assistant), in a 20-minute face-to-face intervention visit. This visit aimed to educate and persuade individuals about the importance of salt reduction and how they can reduce their salt intake and provide the motivation to do this using established behavior change techniques (BCTs; Figure 1) [38]. These BCTs were largely drawn from the BCT taxonomy groups *goals and planning* and *feedback and monitoring*, which have significant evidence to support their effectiveness in dietary behavior change [38-40]. The intervention visit also introduced the SaltSwap app to help individuals identify lower-salt options when grocery shopping, provide feedback on swaps made, and enable users to share successful swaps with their social network (Figure 2; Multimedia Appendix 2 [41,42]). The app enabled users to record their purchased products and any swaps made. Health care professionals were trained in a 1-hour face-to-face session covering the intervention content, theoretical background, and familiarity with the app and were provided training videos for future reference. All other clinical care was continued as usual.

Figure 1. The SaltSwap intervention and included behavior change techniques. BCT: behavior change technique.

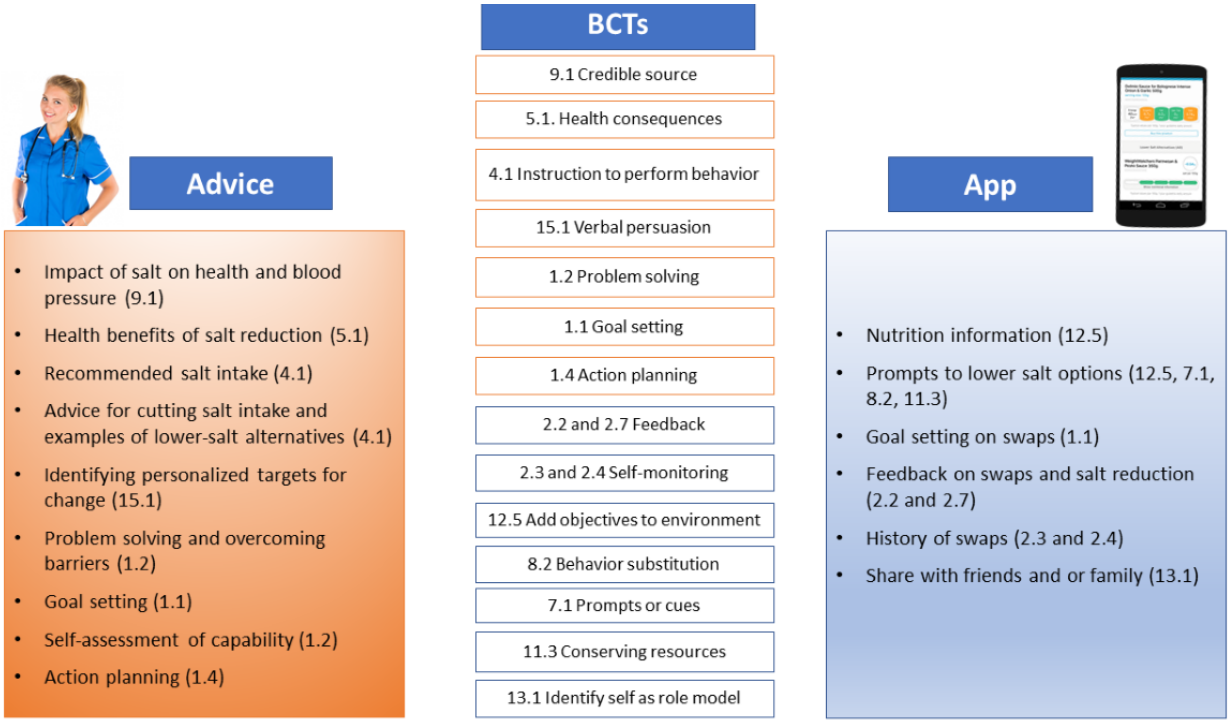
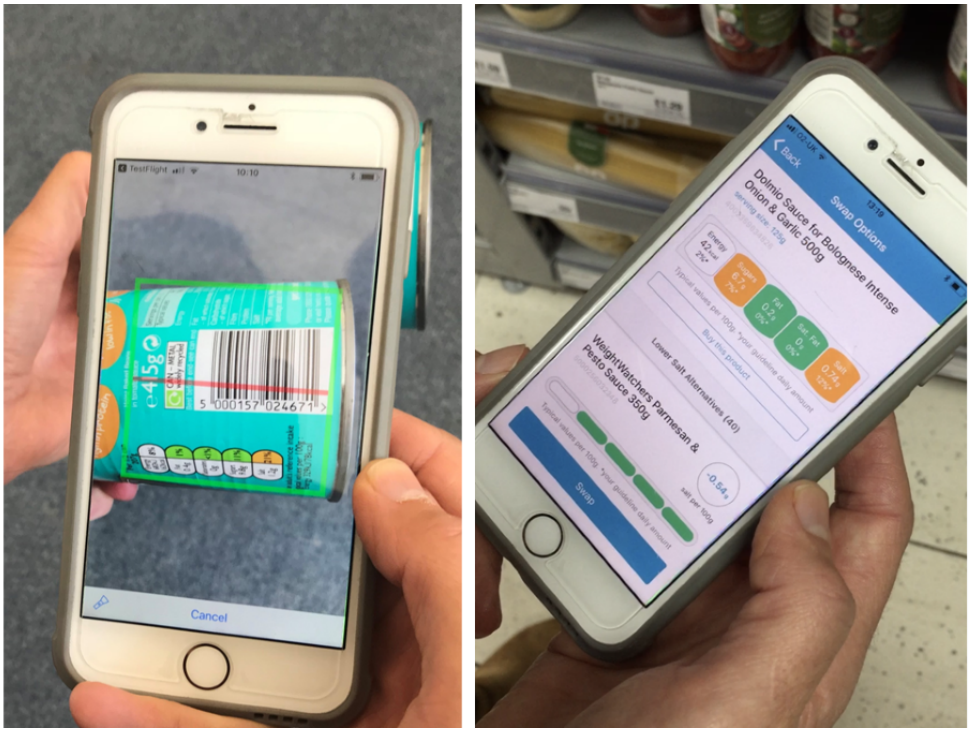


Figure 2. The SaltSwap app screenshots.



Comparator

Participants randomized to the control arm received *usual care* in the form of a postal copy of the publicly available British Heart Foundation Cut Down on Salt booklet [43] or its successor Taking Control of Salt [44]. On completion of their follow-up visit, control participants were given the SaltSwap booklet and were shown how to download the SaltSwap app. Control participants were not aware that they would receive the intervention materials until the end of the follow-up visit. All other clinical care was continued as usual in both arms.

Procedures

Participants attended a baseline study visit where demographic information, medical history (including medications), and clinical measures (height, weight, and blood pressure) were collected. Participants completed a questionnaire on their shopping behaviors and were asked to collect a single 24-hour urine sample for urinary sodium assessment (written instructions provided). They were also asked to collect household grocery shopping receipts and record all grocery purchases using a purchase scanning app for a 2-week baseline run-in period (Multimedia Appendix 3). Only participants who returned baseline data were randomized. Participants were informed of their baseline blood pressure readings but no other baseline measures.

Participants randomized to the intervention group were booked an intervention appointment with a health care professional; control participants were sent a copy of the control booklet by post. The intervention appointment was audio recorded to assess the intervention fidelity. All participants were asked to record their grocery shopping purchases and collect grocery shopping receipts over a 6-week intervention period. SMS text messaging prompts were sent to remind participants to collect their shopping data if the app was not accessed within 7 days and when the app was not used for 10 days after the first use. Participants were scheduled for a follow-up visit with a researcher 6 weeks post randomization, where blood pressure was measured. Participants were asked to collect a second 24-hour urine sample, on any week or weekend day, within 7 days of this appointment. At this follow-up visit, participants completed a questionnaire about changes in their dietary salt knowledge and behaviors.

A convenience sample of intervention participants was invited to participate in an accompanied shopping trip whereby the lead author observed them doing their usual supermarket shopping, using the think-aloud method that attempts to access and record participants' inner speech, the *transformation of thought processes into words* [45]. The think-aloud method has been widely used in the design and evaluation of digital interventions [46,47] and to explore consumers' food choice behaviors [48-51]. The accompanied shopping trip was followed by a one-on-one interview about the participants' experiences of the intervention. Data were captured through audio recordings, which were transcribed by an external transcriber and supplemented by field notes.

Outcome Measures

Primary Outcomes

The primary outcomes were prespecified progression criteria:

1. Follow-up rate: At least 65% of the randomized participants attend the single follow-up visit.
2. Fidelity of intervention delivery: Health care professionals deliver at least 4 of the 6 prespecified essential intervention elements during the advice session.
3. Use of the SaltSwap app: At least 50% of the randomized participants use the SaltSwap app to scan products on at least 1 occasion in the first month.

The principal investigator and a second researcher coassessed fidelity of delivery of the first 5 intervention sessions using audio recordings of the intervention delivery. The assessment involved determining whether 6 prespecified essential elements were delivered: (1) provided advice on the health consequences of a high-salt diet, including recommended daily intake; (2) discussed the main sources of salt; (3) discussed and set goals to reduce salt intake; (4) discussed and set an action plan to achieve these goals; (5) delivered the SaltSwap leaflet; and (6) introduced and helped download the SaltSwap app. Following a discussion of discrepancies in the interpretation of the assessment framework, the remaining audio recordings were assessed by the principal investigator using the refined framework.

Secondary Outcomes

Secondary outcomes were the salt content of household food purchases recorded in the app (g/100 g), salt intake (g/day) estimated from 24-hour urinary sodium excretion, and blood pressure at 6-week follow-up. Although the nutrient of interest is sodium, nutrient labeling in the United Kingdom and public health messaging use the term *salt*. Therefore, the aim of this intervention was to encourage participants to reduce their *salt* intake, and the intervention highlighted the differences in the *salt* content of products not *sodium* content. Therefore, we measured the salt content, which was converted from urinary sodium excretion. For consistency, we referred to salt instead of sodium in this paper, unless referring to a study that has explicitly investigated sodium or when referring to urinary sodium excretion.

Process Outcomes

Process outcomes included recruitment rates; use and acceptability of the SaltSwap app; acceptability of the intervention by health care professionals, assessed through semistructured interviews; feasibility of outcome data collection; changes in participants' knowledge about the effects of salt on health, use of nutrition labels for salt, and dietary salt behaviors, including the use of salt in cooking or at the table as well as consumption of high-salt foods; and contamination.

Qualitative Outcomes

Accompanied shopping and interview data were used in a qualitative analysis to explore the impact of the SaltSwap intervention on grocery shopping behavior, the use and acceptability of the SaltSwap app, and experiences of attempting to reduce salt intake.

Statistical Analysis and Sample Size

Baseline characteristics were summarized using descriptive statistics by the trial arm. Statistical analyses were conducted using Stata IC 16.0 (StataCorp) [52]. Descriptive statistics were reported with 95% CIs. Progression criteria analysis used data from all available randomized participants. Secondary outcomes were analyzed on an intention-to-treat basis using a complete case analysis. Before calculating the salt content of purchased foods, missing data on product weight or volume or salt content were estimated where possible by cross-referencing web-based retailer websites; otherwise, products were excluded (considered *missing data*). Items categorized as fresh fruit and vegetables, alcohol, or household nonfood items were excluded from the analysis, as they included no or minimal salt.

We used linear regression models to calculate differences in means at follow-up between the intervention and control groups, with 95% CIs, adjusted for baseline values of the dependent variable and recruitment site. We examined the sensitivity of the results to confounding owing to differences in those baseline characteristics which were plausible confounders and imbalanced between groups. We also examined sensitivity of the results due to outliers. We tested the model fit using likelihood ratio tests. Where the assumptions of linear regression were not met, we applied the Wilcoxon rank-sum test to assess the outcomes. We also described the proportion of participants who showed a reduction in salt intake and purchase and assessed any differences in this proportion by trial group, using 2-tailed *t* tests for proportions.

Process measures used all data available, regardless of whether participants completed the trial, and were reported using descriptive statistics with 95% CIs. Qualitative data were analyzed using the Braun and Clarke [53] approach to thematic analysis. The lead author, an applied health services researcher specializing in public health interventions, analyzed the data with input from CA, a qualitative researcher with experience in thematic analysis. The coding framework was based on topics purposefully explored in interviews and those that developed from the data. Transcripts were coded line by line and then grouped into broad topics across cases. Interpretation of the data was informed by the COM-B model, the theoretical model of behavior change underpinning the intervention design [37]. Data quality and validity were enhanced by applying the techniques of Lincoln and Guba [54] to establish trustworthiness (Multimedia Appendix 4). Data were stored, coded, and managed using NVivo 12 (QSR International) for PC [55].

This study was a feasibility study and was not powered to detect any significant intervention effects. We calculated that a sample size of 40 would be sufficient to estimate progression outcomes within acceptably narrow CIs to enable robust testing of the trial methods. For the trial to be considered feasible, 80% follow-up attendance and fidelity of intervention delivery would need to be achieved. Therefore, allowing for a 95% CI around this point estimate, we set the minimum criteria for progression at 65%. For app use, a minimum of 70% was required; therefore, we set the minimum progression criteria at 50% based on the CI around this estimate. The think-aloud sample size of 12 was guided by the model of information power by Malterud et al [56], and sample selection was informed by recruitment site, participant gender, timing within the intervention period, and participant availability.

Patient and Public Involvement

Before submission for ethical approval, we involved several members of an existing Patient and Public Involvement panel interested in weight loss and diet to inform and advise on the study objectives, procedures, and patient materials. We also convened a panel of members of the public with diagnosed hypertension to co-design the intervention.

Adverse Events

All serious adverse events were reported during the trial. Participants were asked about adverse events at follow-up or at the point of withdrawal from the trial.

Results

Participants

Participants were recruited between October 2018 and April 2019. A total of 2028 patients were deemed eligible based on blood pressure criteria and major exclusion criteria. Following a letter from the GP inviting participation in the trial, 107 patients registered their interest in taking part and completed phone-based screening with a researcher. Of the 107 patients, 51 (47.7%) did not meet the inclusion criteria. The main reasons for that were the following: not owning a smartphone (29/107, 27.1%), not being responsible for household grocery shopping (7/107, 6.6%), or being away during the study period (6/107, 5.7%). Out of the 56 patients who met the inclusion criteria, 6 (11%) did not attend the baseline assessment visit and 50 (89%) were enrolled (Figure 3). Participants had a mean age of 65 (SD 11) years, and the majority of the participants (35/47, 74%) reported a diagnosis of hypertension (Table 1). Only 13% (6/47) of all the participants had previously been advised by a health care professional to reduce their salt intake, although 48% (15/31) of the intervention group and 69% (11/16) of the controls reported that they had previously attempted to reduce their salt intake.

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) diagram. GP: general practitioner; HCP: health care professional.

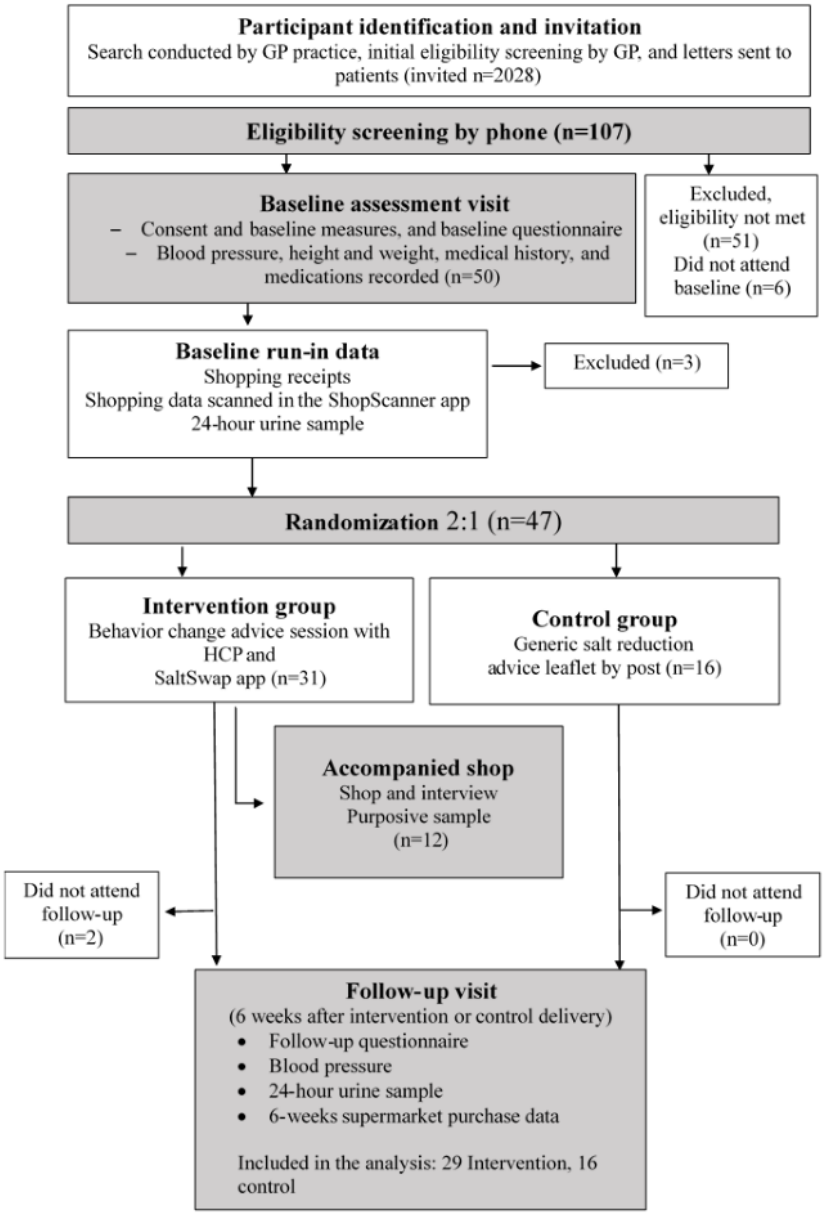


Table 1. Baseline characteristics of randomized study participants (N=47).

Characteristics	Control (n=16)	Intervention (n=31)	Total
Age (years), mean (SD)	67 (7)	64 (12)	65 (11)
Sex (female), n (%)	10 (63)	20 (65)	30 (64)
BMI (kg/m^2), mean (SD)	29 (5)	29 (6)	29 (6)
Blood pressure (mm Hg), mean (SD)			
Systolic	137 (15)	134 (16)	135 (15)
Diastolic	80 (8)	81 (10)	81 (9)
Estimated daily salt intake[a] (grams), mean (SD)	6.8 (2.7)	6.5 (3.9)	6.6 (3.5)
Smoking, n (%)			
Current	0 (0)	1 (3)	1 (2)
Ex-smoker	7 (44)	11 (35)	18 (38)
Never	9 (56)	19 (61)	28 (60)
Ethnic group, n (%)			
White	15 (94)	29 (94)	44 (94)
Asian or Asian British	0 (0)	1 (3)	1 (2)
Black or Black British	0 (0)	0 (0)	0 (0)
Mixed or other or Chinese	1 (6)	1 (3)	2 (4)
Education, n (%)			
No formal qualifications	4 (25)	0 (0)	4 (9)
Secondary education	6 (38)	10 (32)	16 (34)
Higher education	6 (38)	21 (68)	27 (57)
Household size, median (IQR)	2 (2-2)	2 (1-2)	2 (2-2)
Weekly grocery shopping > £25 (US $34)/trip, n (%)			
More than once a week	5 (31)	12 (39)	17 (36)
Once a week	9 (56)	15 (48)	24 (51)
Once a fortnight	2 (13)	3 (10)	5 (11)
Once a month	0 (0)	0 (0)	0 (0)
Less than once a month	0 (0)	1 (3)	1 (2)
Factors influencing grocery shopping decisions, n (%)[b]			
Price	6 (38)	20 (65)	26 (55)
Appearance	0 (0)	5 (16)	5 (11)
Taste	13 (81)	20 (65)	33 (70)
Habits	5 (31)	7 (23)	12 (26)
Health	9 (56)	17 (55)	26 (55)
Convenience	1 (6)	5 (16)	6 (13)
Special offers	7 (44)	8 (26)	15 (32)
Organic	2 (13)	3 (10)	5 (11)
Special diet	2 (13)	2 (6)	4 (9)
Other	4 (25)	6 (19)	10 (21)
Frequency of using nutrition labels, n (%)			
Salt	10 (63)	14 (45)	24 (51)
Sugar	14 (88)	20 (65)	34 (72)

Characteristics	Control (n=16)	Intervention (n=31)	Total
Fat (total or saturated)	13 (81)	21 (68)	34 (72)
Energy (calories)	13 (81)	21 (68)	34 (72)
Been advised by a health care professional to reduce their salt intake, n (%)	2 (13)	4 (13)	6 (13)
Previously tried to reduce salt intake, n (%)	11 (69)	15 (48)	26 (55)
Knowledge of the effect of salt on blood pressure?, n (%)[c]			
Yes	8 (50)	11 (35)	19 (40)
No	1 (6)	7 (23)	8 (17)
Do not know	7 (44)	13 (42)	20 (43)
Eating breakfast out, median (IQR)	0 (0-0.5)	0 (0-0)	0 (0-0)
Eating lunch out, median (IQR)	0 (0-1)	1 (0-2)	1 (0-1)
Eating dinner out, median (IQR)	1 (0-1)	0.5 (0-1)	1 (0-1)
Relevant health history, n (%)			
CVD[d]	0 (0)	4 (13)	4 (9)
Diagnosed hypertension	14 (88)	21 (68)	35 (74)
Diabetes	4 (16)	3 (10)	7 (15)
Atrial fibrillation	0 (0)	0 (0)	0 (0)
Chronic kidney disease	0 (0)	1 (3)	1 (2)
Peripheral vascular disease	1 (7)	1 (3)	2 (4)
Other (related to CVD)	5 (31)	5 (16)	10 (21)
Antihypertensive medication	12 (75)	19 (61)	30 (64)

[a]Converted from 24-hour urinary sodium excretion.

[b]Participants were asked to choose the top three factors.

[c]Participants were asked, “Do you think the amount of salt you eat affects your blood pressure?”

[d]CVD: cardiovascular disease.

Primary Outcomes

All 3 progression criteria were met well above the set thresholds (Table 2). The total attendance at follow-up was 96% (45/47; 29/31, 94% in the intervention group and 16/16, 100% in the control group). There were 2 participants who did not attend follow-up; 1 withdrew because of difficulty using the study app and 1 because of an unrelated health problem. Of the intervention sessions that were adequately recorded (25/31, 81%), all included at least 4 of the 6 essential elements. All 6 elements were delivered in 21 (84%) recorded sessions; the element most commonly not delivered was action planning (3/25). A total of 27 out of 31 (87%) intervention participants used the SaltSwap app to scan products on at least 1 occasion by the end of month 1, above the progression threshold of 50%.

Table 2. Participant follow-up rate, fidelity of intervention delivery, and use of the SaltSwap app (n=47).

Progression criteria	Total		Control (n=16)		SaltSwap (n=31)	
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
Follow-up rate	45 (96)	85-99	16 (100)	79-100	29 (94)	79-99
Fidelity of the intervention (advice) session[a]	N/A[b]	N/A	N/A	N/A	25 (81)	63-93
Use of the SaltSwap app	N/A	N/A	N/A	N/A	27 (87)	70-96

[a]Audio recordings for the assessment of intervention fidelity were available for 25 of the 31 advice sessions delivered.

[b]N/A: not applicable.

Secondary Outcomes

Baseline and follow-up data on the salt content of the purchased products were available for 78% (37/47) of participants. The analysis included all the purchased items recorded in the app. The number of products purchased was slightly lower at follow-up in both groups, and missing data were comparable between groups (Multimedia Appendix 5).

There was no evidence that the intervention significantly reduced the salt content of purchased foods, salt intake, or blood pressure, and there was no significant between-group difference (Table 3; Figure 4). In the intervention group, 72% (18/25) of the participants reduced the salt content of purchased foods (mean −0.15 g/100 g, SD 0.13 g/100 g) compared with 66% (8/12) of the control group (mean −0.4 g/100 g, SD 0.27 g/100 g). Similarly, 50% (14/28) of intervention participants reduced their salt intake from baseline to follow-up (mean reduction −2.44 g/day; 95% CI −8.71 to −0.20 g/day) compared with 66% (10/15) of the control group (mean reduction −2.42 g/day; 95% CI −6.76 to −0.08 g/day). The differences in these percentages were not statistically significant. There were nonsignificant reductions in expenditure on food at follow-up for both groups, with no significant between-group differences.

Table 3. Changes in the mean salt intake, blood pressure, and purchased salt from baseline and estimates of differences between the intervention and control groups.

	Baseline values, mean (SD)		Follow-up values, mean (SD)		Change, mean (95% CI)		Group difference[a], mean (95% CI)	*P* value
	Control	SaltSwap	Control	SaltSwap	Control	SaltSwap	SaltSwap vs control	
Salt intake (g/day)	6.8 (2.7)	6.5 (3.9)	6.0 (2.1)	6.2 (3.7)	−1.0 (−2.4 to 0.4)	−0.2 (−1.4 to 0.9)	−0.4 (−2.3 to 1.5)	.68
Number of participants, n	16	31	15	28	15	28	43	N/A[c]
BP[b] (mm Hg)								
Systolic BP	137.3 (14.8)	134.2 (15.8)	136.2 (15.9)	133.2 (16.9)	−1.1 (−6.7 to 4.4)	−1.0 (−5.5 to 3.6)	N/A	.82[d]
Diastolic BP	79.7 (8.0)	81.2 (10.2)	82.0 (10.5)	80.2 (10.7)	2.3 (−2.7 to 7.2)	−1.0 (−4.3 to 2.2)	−3.0 (−8.0 to 2.0)	.23
Number of participants, n	16	31	16	29	16	29	45	N/A
Purchased salt (g/100g)	0.8 (0.5)	0.5 (0.2)	0.6 (0.4)	0.5 (0.4)	−0.1 (−0.5 to 0.2)	0.0 (−0.1 to 0.2)	N/A	.16[d]
Number of participants, n	12	25	12	25	12	25	N/A	N/A

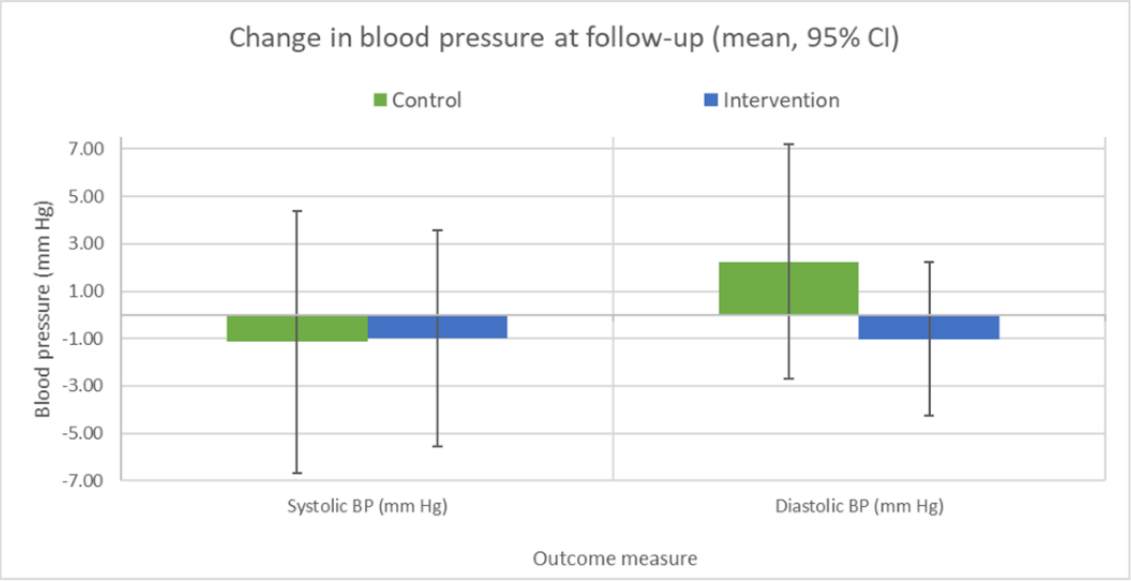
[a]Model 2. Regression analysis adjusted for baseline values, general practitioner surgery, and level of education.

[b]BP: blood pressure.

[c]N/A: not applicable.

[d]*P* value for the Wilcoxon rank-sum test for between-group differences. Assumptions of linear regression were not met for systolic blood pressure and salt content of purchased foods; therefore, we analyzed these outcomes using the Wilcoxon rank-sum test.

Figure 4. Changes in blood pressure from baseline to follow-up. BP: blood pressure.



Process Outcomes

The intervention was acceptable to both individuals and practitioners, and outcome data were successfully collected via trial methods. The SaltSwap app was used (ie, at least one product scanned) in 75% of all shopping trips by intervention participants, and 79% (23/29) of intervention participants who used the app recorded at least one swap to a lower-salt product in the app. The average salt reduction per swap recorded in the app was −0.6 g/100 g. Participants in both groups reported eating high-salt foods less frequently and swapping to lower-salt foods (Multimedia Appendix 5).

Qualitative Outcomes

A total of 11 intervention participants attended an accompanied shopping trip and postshopping interview, and 1 participant attended only a postshopping interview (Multimedia Appendix 4). Overall, the qualitative study sample reflected the demographic profile of the total sample. The average age of the participants in this substudy was 67 (SD 7) years compared with 65 (SD 11) years in the total sample, representative of the age of the population with hypertension. The qualitative study comprised 58% (7/12) female participants compared with 64% (30/47) in the total sample, and all but 2 of the qualitative participants were of White British ethnicity, reflecting the lack of ethnic diversity in the overall study sample. Shopping trips varied in duration from 10 to 44 minutes and interviews from 17 to 40 minutes, with an average combined duration of 48 minutes per person. Key themes were developed around 3 main areas: SaltSwap intervention, salt consumption and purchase behaviors, and the wider environment. Detailed theme descriptions and example quotes are presented in the Multimedia Appendix 4.

Participants reported improved knowledge about the major sources of salt and heightened awareness of the importance of salt intake in relation to blood pressure. They described greater attention to the salt content of foods than before the study, showing surprise at the high salt content of some foods they routinely buy and at the availability of lower-salt options. All these aspects were explicitly targeted by the intervention:

> *...doing this it's heightened my awareness of salt in the diet...I was horrified at the levels of salt in crisps and things, so have gone for the no adding salt to that. Sauces, I've been surprised how much salt is in those. [ID0205; female, 68 years]*

The advice session helped participants identify specific high-salt foods to swap and successfully encouraged them to set goals to make changes to their purchases. There was little evidence of action planning during the intervention session or mention of doing this afterward. Interview data suggested that the intervention did not sufficiently generate motivation to reduce salt intake. Some participants indicated that they would be more motivated if they received feedback on their salt intake from baseline 24-hour urine measurement:

> *I guess I'd like to know where I am on the spectrum. So, as a result of the urine tests. Where do I sit...in terms of a male of my age? Because that may provide something...motivation. [ID0309; male, 71 years]*

All intervention participants were observed or described using the SaltSwap app to identify lower-salt alternatives with the app. Most participants found the app usable and acceptable and noted that it helped focus their attention on the salt content of foods. Few participants reported using the feedback function in the app, showing a reduction in the salt content of selected items through making swaps.

Barriers to successfully identifying lower-salt products and continued use of the app were the lack of comprehensive coverage of products within the app database, unsuitable suggestions for lower-salt alternatives, and lack of availability of suitable lower-salt options in some stores or for specific food groups:

> *This is...tomato paste. See, I think being [store] they only do one. [ID0206; male, 64 years]*

Participants reported changing their shopping behaviors, for example, swapping to lower-salt foods, avoiding buying high-salt foods entirely, and changing their cooking practices to reduce their salt intake (Table 4). This sometimes resulted in participants avoiding or reducing the frequency of purchasing foods, which were also higher in other nutrients associated with poor diet quality, such as saturated fat. Participants reported increasing their use of nutrition labels for salt following the intervention, often in conjunction with the SaltSwap app, although some participants still reported challenges interpreting nutrition labels:

> *I would say since I've started the salt study, I've looked at those labels more than in the past. In the past I used to only look at the sugar content. [ID0403; male, 70 years]*

Table 4. Examples of participants-reported behavior change and the related COM-B components (intervention participants).

Behaviors	Example quotes	COM-B components
Swapping to a lower-salt product	• "Now we used to buy the deals here of different sorts of ham – but when I tried to do a salt swap on them it was quite difficult to find an alternative. So, what we do now is we buy sliced turkey" [ID0403; male, 70 years] • "I used to buy crisps without thinking about salt. And then I started the salt file, I just switched to these." [ID0313; female, 57 years] • "I used to buy a tomato sort of sauce. And in fact, because that was quite high in salt, I started using Passata little boxes. Because you get the sort of intense tomato flavour, but you don't get the salt." [ID0410; female, 74 years] • "Like I haven't been buying so much ham. I have bought chicken instead." [ID0415; female, 71 years]	• Psychological capability—knowledge • Reflective motivation • Automatic motivation—disrupting habit • Environmental opportunity—making it easy to identify lower-salt options
Avoiding high-salt foods	• "They were like a ripple type crisp, and I was surprised how much salt was in that. And in fact, I don't think I bought them in the end. Because I thought, 'Wow, I thought these were supposed to be, you know, healthy." [ID0206; male, 64 years] • "Anyway, I used to buy those particularly when it was the two for one type offers. Just stopped buying those completely, yeh" [ID0408; male, 65 years] • "I do like salted peanuts and I haven't been able to find anything that's low in salt...I probably would have bought them most weeks. Whereas now I think, 'No, it is a treat.'... I won't have those." [ID0205; female, 68 years]	• Psychological capability—knowledge • Reflective motivation—intentions or persuasion • Automatic motivation—disrupting habit
Changing cooking practices to reduce salt intake	• "And I've also...we've also this last week haven't cooked with salt either. Yeh, previously if we used in cooking potatoes; boiling potatoes up for mashed potatoes; always put a bit of salt in there. Even with your, you know, vegetables, put a bit of salt in there, but we haven't this time." [ID0207; male, 58 years] • "So, instead of using stock cubes. Which we know are high in salt, I tend to use Bouillon because I can use just a tiny bit of that. I will sometimes use that. Or what I do is I add that after I've taken my portion. And that, because it's not a stock cube has to be whatever, and what have you. With that I can stir that in...So, that works quite well." [ID0319; female, 69 years]	• Psychological capability—knowledge and skills • Automatic motivation—disrupting habit
Swapping from a store-bought product to home-made, with no added salt	• "After I'd seen those things and how high they were in salt, when I went and did my own, I didn't put any salt whatsoever in it at all, and I just used the Indian spices" [ID0218; female, 76 years]	• Psychological capability—knowledge and skills • Reflective motivation—persuasion
Reducing frequency of consumption or portion size	• "I probably would have bought them most weeks. Whereas now I think, 'No, it is a treat.' And I would also not sit and eat the whole packet." [ID0205; female, 68 years]	• Psychological capability—knowledge • Reflective motivation—intentions

The time and effort required to find lower-salt alternatives was commonly mentioned as a barrier to changing their purchasing behavior, but others described making new habits and planning their shopping as ways to help make and stick to lower-salt choices.

Some participants found it challenging to balance salt reduction with other nutritional concerns, such as saturated fat or sugar. They also expressed a desire for feedback on the outcome of their behavioral change, that is, explicit measurement of their salt intake, highlighting the potentially important role of feedback in increasing motivation.

Participants commented on the choice of products available, both that there was too much choice to make it easy to choose low-salt options and that some stores do not offer lower-salt options. They generally felt that there was a lack of support in making dietary changes to reduce their salt intake but that their GP surgery would be a good place to provide support of this nature.

Adverse Events

One serious adverse event was reported during the study. The lead GP for the surgery concluded that this adverse event was unrelated to the study. The participant did not continue in the study or attend the follow-up, but no further action was taken.

Discussion

Principal Findings

This trial intended to assess the feasibility of the SaltSwap intervention and trial procedures to determine if a larger trial to assess the effectiveness of the intervention for reducing blood pressure is warranted. The results showed high levels of follow-up, fidelity of intervention delivery, and use of the

intervention app, demonstrating that progression criteria were met. Health care professionals were positive about the intervention and confident in delivering it, indicating that the SaltSwap intervention can be delivered by nonspecialist staff in primary care without intensive training. Furthermore, intervention participants used the app in store in most shopping trips, and the majority used the app to identify and select lower-salt alternatives, demonstrating high acceptability among individuals with high blood pressure. The app product database provided insufficient coverage of in-store products, and product categorization resulted in some unsuitable lower-salt suggestions, hindering effective use.

Although there were no significant effects on purchased salt, salt intake, or blood pressure, CIs around each point estimate included clinically meaningful reductions, and the intervention was successful in changing key salt reduction behaviors, suggesting that the intervention could be potentially successful if tested in a larger trial. The trial procedures to evaluate the effect of the intervention on salt intake and blood pressure were shown to be robust.

Strengths and Limitations

A key strength of this study is the in-depth qualitative analysis and process evaluation to assess the proposed *mechanisms of impact* of this novel intervention, providing insight into context-specific use of the app and barriers to the intervention's potential effectiveness. Participants described increased knowledge, intention to reduce salt intake, and ability to identify lower-salt options—key targets of the intervention based on our COM-B behavioral analysis. Process outcomes demonstrated that participants swapped to lower-salt items predominantly in the food categories of bread, cheese, and processed meats, foods that contribute most to total daily salt intake (accounting for 34%) among those aged 65-74 years [57], suggesting that the intervention successfully targeted the main sources of dietary salt.

Another strength is the development and proven feasibility of a behaviorally informed smartphone app tailored to the UK grocery market, which can support individuals to make lower-salt choices when grocery shopping and provide feedback on salt reduction. Smartphone use in the United Kingdom is ubiquitous and has grown dramatically; a 2019 report found that most adults in the United Kingdom use a smartphone phone (84%), and this is true for all age groups, including 93% of 35- to 54-year-olds and 64% of those aged ≥55 years [58]. However, the use of mobile health apps is not as widespread among older adults [59], and app-based interventions may still present specific age-related barriers to the effective use of smartphone apps [60]. This is an important concern when addressing a population with hypertension, including many older adults [61]. This trial successfully engaged a population of adults with an average age of 64 years with a mobile app–based intervention with few usability issues, demonstrating the feasibility of delivering app-based health behavior change interventions to older adults. A previous study investigating barriers to app use in older adults (aged >60 years) found the main barriers to be concerns about data privacy, lack of trust, and fear of misdiagnosis [59]. The support of a trusted health care professional in the use of the app could potentially overcome barriers of this nature. Moreover, as the app does not require any personal data or any health care data and does not communicate information about any individual to any third party, these common barriers are unlikely to be significant concerns for SaltSwap. Health care professionals successfully delivered the intervention advice session. Although audio recording of the intervention may have acted as a prompt to deliver the advice elements, introducing a Hawthorne effect [62], the audio recordings provided evidence that the intervention content was within their expertise and skill set to deliver competently.

The trial response rate of only 5% was a limitation. Although comparable with other trials of a similar nature [35,63], it is lower than that for other dietary intervention trials, such as those targeting weight loss [64]. The low response rate raises two issues. First, it may indicate less interest in reducing salt intake than other dietary goals, presenting a challenge for the implementation of the intervention. In clinical practice, brief interventions are likely to be given in consultation, regardless of patients' inclination to reduce salt intake or interest in using an app. Such respondents may differ from those we enrolled, meaning that the results here, indicating the acceptability of the intervention and potential for changing salt intake behaviors, may not generalize to the wider population of people with hypertension.

Second, the low response rate may indicate that this trial engaged a group of individuals with high agency and motivation to address their diet. This intrinsic motivation could lead to participants actively engaging with the control intervention (self-help booklet) and may explain why salt intake reduced in both arms, although not significantly, reducing the apparent impact of the SaltSwap intervention. Indeed, salt intake at baseline was lower than the age-adjusted national average, and many participants reported previous attempts to eat less salt. In addition, measurement reactivity may have resulted in participants being particularly careful to reduce salt intake on measurement days, biasing results toward the null. A larger effectiveness trial would do well to exclude people with low baseline salt intake and to consider multiple measurements at both baseline and follow-up, including both weekdays and weekends, to ensure a more representative salt intake assessment. Furthermore, urine samples could be used to assess potassium levels in a future trial to explore any salt substitution effects.

The study population was self-selected—for reasons of convenience, drawn solely from Oxfordshire, an area with lower deprivation than the national average [65]. Socioeconomic status is known to influence salt intake and adherence to wider dietary guidelines [66-68], and the impact in this group may not be representative of the wider population. Moreover, the study participants were predominantly of White ethnicity. Salt intake varies across ethnic groups [69,70], and culture is a major determinant of food choice and consumption [70,71]. However, the database includes a wide range of foods available in supermarkets and should cover most dietary preferences, making the app universally applicable. The literature on the differential impacts of salt reduction interventions across ethnic groups is

sparse, and there is a lack of evidence on specific barriers and facilitators for food purchasing and dietary behavior change among minority ethnic groups to adequately inform intervention adaption [72]. An important consideration for any future trial of the SaltSwap intervention is the recruitment of a study sample with greater ethnic and socioeconomic diversity.

A further limitation was the lack of comprehensive product coverage within the SaltSwap app, despite it containing more than 95,000 unique products. Participants noted this as a key barrier to successfully identifying lower-salt options in stores, which may have affected the resulting salt reduction. This also resulted in a high level of missing data on the salt content of purchased foods, thus limiting the interpretation of this secondary outcome; however, this is unlikely to have biased the outcome differently by arm [73]. Participants used the app as asked; therefore, an improved database would reduce missing purchase data in future trials. Data on other important dietary components that may be affected by this intervention were also not captured through the SaltSwap app; however, with improvements to the app database, these could be included as secondary outcomes in a larger trial.

Comparison With Other Literature

There are few RCTs of the impact of apps alone or combined with other interventions to reduce salt intake [74], but the available evidence suggests that this is a promising strategy. One trial showed that a standalone app to encourage the purchase of lower-salt foods reduced salt purchasing but not salt intake or blood pressure [34]. Another RCT in a small sample of healthy adults showed short-term reductions in sodium consumption in participants using an app to record and monitor their daily sodium intake [33]. A recent pilot study of an app using geolocation to prompt users to choose lower-sodium options when shopping or eating out or at home showed no significant reduction in 24-hour sodium compared with usual care and no change in confidence adhering to a low-sodium diet (including reading food labels, shopping at grocery stores, or choosing low-sodium options at restaurants) but did show a significant reduction in spot urine as an estimate of sodium intake and sodium intake measured by a food frequency questionnaire [35]. These trials all investigated standalone app interventions as compared with the multicomponent nature of SaltSwap. A review of behavior change apps for dietary change highlighted that multicomponent interventions had, on average, larger effect sizes than single-component interventions [32]. Furthermore, existing evidence shows that brief behavioral support from a health care professional can generate the necessary capability and motivation to make dietary changes [28-30], which could be supported by an app at the point of choice. The theoretical approach to intervention development highlighted the need for a salient and persuasive message to motivate individuals to change their diet, and SaltSwap aimed to deliver this via in-person advice from a credible source. We found no other studies in the literature of a multicomponent intervention using an app to enhance a brief intervention for salt reduction; therefore, this feasibility study provides valuable insight into the feasibility and acceptability of such an approach.

For the SaltSwap intervention to have a meaningful effect on blood pressure, it would need to generate a reduction in salt intake of at least 1 to 2 g/day. A meta-analysis of RCTs investigating the effect of salt reduction on blood pressure reported an average reduction in salt intake of −4.6 g/day, which led to reductions in systolic (5.0 mm Hg) and diastolic blood pressure (2.7 mm Hg) [75]. However, a later RCT showed that a low-salt diet resulted in a modest salt reduction of 1.7 g/day, with a corresponding reduction in systolic blood pressure of 3.7 mm Hg, among participants with high blood pressure [76], indicating the potential for clinically meaningful benefit from a more modest salt reduction. In fact, individual counseling interventions for salt reduction, delivered by specialist dieticians, have been shown to be cost-effective based on expected reductions in salt intake of 0.5 g/day [77]. A reduction of 1 to 2 g/day falls within the CI for the salt reduction achieved here, in a sample of individuals with comparatively low baseline salt intake, who may have a limited capacity for further reduction. The magnitude of change in objectively measured salt intake and blood pressure, along with the behavioral changes identified in our qualitative evaluation, provides a sufficient signal of efficacy to warrant progression to a full trial.

In the United Kingdom, as with many other high-income countries, where most of the salt consumed is already in the food chain and the food environment promotes high-salt, highly processed foods, providing advice, knowledge, and motivation is unlikely to be enough on its own to provide sufficient reduction in salt intake to achieve national recommendations. Concurrent structural interventions, such as product reformulation and nutrition labeling, are important salt reduction strategies, creating an environment conducive to individual-level behavior change [78]. For people to habitually choose lower-salt options, these need to be readily available, acceptable, and identifiable without undue effort. There is a synergistic relationship between individual-level and structural, population-level interventions. In the United Kingdom, a voluntary reformulation strategy has resulted in significant reductions in the mean salt content of many food categories that contribute substantially to salt intake (eg, cereals, breads, and snack foods) [79]. However, one result of this approach is that some manufacturers have made ambitious progress, whereas others have lagged behind, leading to a wide variation in salt content across some food categories [80,81]. With such a variation in salt content across foods, the full benefit of the reformulation strategy can only be achieved when individuals are motivated and enabled to choose available lower-salt options, reinforcing the need for individual-level interventions that can support people to do this. This, in turn, could give companies that reduce salt in their products a competitive edge, stimulating a virtuous cycle of salt reduction.

Conclusions

A behaviorally informed, brief intervention to reduce salt intake in people with hypertension, delivered in primary care and supported by an app to inform purchasing decisions, was acceptable to those delivering and those receiving it, and the trial approaches are considered feasible. A full trial to assess the effectiveness of the SaltSwap intervention for reducing salt intake and blood pressure is warranted following improvements

to the SaltSwap app highlighted as important during this feasibility study.

Acknowledgments

The authors would like to acknowledge the Thames Valley Clinical Research Network for supporting the delivery of the trial, the Oxfordshire GP surgeries involved in the study and their wonderfully accommodating staff, the study participants, the Patient and Public Involvement panel members who contributed so valuably to the intervention design and study materials, and The George Institute for Global Health who allowed the authors to use their product database as part of the SaltSwap app. JPS was supported by the Wellcome Trust and Royal Society via a Sir Henry Dale Fellowship (ref. 211182/Z/18/Z) and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. PA and SAJ are NIHR senior investigators and funded by the NIHR Oxford Biomedical Research Centre and the NIHR Oxford Applied Research Collaboration. CP was funded by the NIHR Oxford Applied Research Collaboration. SPR was supported by a British Heart Foundation Clinical Research Training Fellowship (FS/16/34/32211), the Medical Research Council. The NIHR Oxford Applied Research Collaboration supported the app development. No funders had a role in the review and approval of the manuscript for publication.

Authors' Contributions

All authors contributed to the study design and critically revised and approved the final manuscript.

Conflicts of Interest

The University of Oxford is the owner of the SaltSwap app.

Multimedia Appendix 1
Intervention development.
[DOCX File , 31 KB - mhealth_v9i10e26233_app1.docx]

Multimedia Appendix 2
The SaltSwap app description and technical information.
[DOCX File , 693 KB - mhealth_v9i10e26233_app2.docx]

Multimedia Appendix 3
The ShopScanner app.
[DOCX File , 301 KB - mhealth_v9i10e26233_app3.docx]

Multimedia Appendix 4
Qualitative methods and outcomes.
[DOCX File , 38 KB - mhealth_v9i10e26233_app4.docx]

Multimedia Appendix 5
Process outcomes.
[DOCX File , 30 KB - mhealth_v9i10e26233_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 93 KB - mhealth_v9i10e26233_app6.pdf]

References

1. The top 10 causes of death. World Health Organization. 2020. URL: http://www.who.int/mediacentre/factsheets/fs310/en/ [accessed 2020-05-09]
2. Mozaffarian D, Fahimi S, Singh GM, Micha R, Khatibzadeh S, Engell RE, et al. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014 Aug 14;371(7):624-634 [FREE Full text] [doi: 10.1056/nejmoa1304127]
3. Ettehad D, Emdin CA, Kiran A, Anderson SG, Callender T, Emberson J, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet 2016 Mar;387(10022):957-967 [FREE Full text] [doi: 10.1016/s0140-6736(15)01225-8]
4. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. Br Med J 2009 May 19;338:b1665 [FREE Full text] [doi: 10.1136/bmj.b1665] [Medline: 19454737]

5. Tackling high blood pressure. From evidence into action. Public Health England. 2014 Nov. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375985/20141018_Tackling_high_blood_pressure_-_FINAL_INCL_LINK_CHANGES.pdf [accessed 2021-09-02]
6. Matyas E, Jeitler K, Horvath K, Semlitsch T, Hemkens LG, Pignitter N, et al. Benefit assessment of salt reduction in patients with hypertension: systematic overview. J Hypertens 2011 May;29(5):821-828. [doi: 10.1097/HJH.0b013e3283442840] [Medline: 21475042]
7. Dickinson HO, Mason JM, Nicolson DJ, Campbell F, Beyer FR, Cook JV, et al. Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials. J Hypertens 2006 Mar;24(2):215-233 [FREE Full text] [doi: 10.1097/01.hjh.0000199800.72563.26] [Medline: 16508562]
8. Gay H, Rao S, Vaccarino V, Ali MA. Effects of different dietary interventions on blood pressure. Hypertension 2016 Apr;67(4):733-739 [FREE Full text] [doi: 10.1161/hypertensionaha.115.06853]
9. Graudal NA, Hubeck-Graudal T, Jurgens G. Effects of low sodium diet versus high sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterol, and triglyceride. Cochrane Database Syst Rev 2011 Nov 09(11):CD004022. [doi: 10.1002/14651858.CD004022.pub3] [Medline: 22071811]
10. Adler AJ, Taylor F, Martin N, Gottlieb S, Taylor RS, Ebrahim S. Reduced dietary salt for the prevention of cardiovascular disease. Cochrane Database Syst Rev 2014 Dec 18(12):CD009217 [FREE Full text] [doi: 10.1002/14651858.CD009217.pub3] [Medline: 25519688]
11. He FJ, Li J, Macgregor GA. Effect of longer term modest salt reduction on blood pressure: Cochrane systematic review and meta-analysis of randomised trials. Br Med J 2013 Apr 03;346:f1325 [FREE Full text] [doi: 10.1136/bmj.f1325] [Medline: 23558162]
12. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: systematic review and meta-analyses. Br Med J 2013 Apr 03;346:f1326 [FREE Full text] [doi: 10.1136/bmj.f1326] [Medline: 23558163]
13. Arcand J, Wong MM, Santos JA, Leung AA, Trieu K, Thout SR, et al. More evidence that salt increases blood pressure and risk of kidney disease from the Science of Salt: A regularly updated systematic review of salt and health outcomes (April-July 2016). J Clin Hypertens (Greenwich) 2017 Aug;19(8):813-823 [FREE Full text] [doi: 10.1111/jch.13049] [Medline: 28834222]
14. Mente A, O'Donnell MJ, Rangarajan S, McQueen MJ, Poirier P, Wielgosz A, et al. Association of urinary sodium and potassium excretion with blood pressure. N Engl J Med 2014 Aug 14;371(7):601-611 [FREE Full text] [doi: 10.1056/nejmoa1311989]
15. Graudal NA, Hubeck-Graudal T, Jurgens G. Effects of low sodium diet versus high sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterol, and triglyceride. Cochrane Database Syst Rev 2020 Dec 12;12(12):CD004022. [doi: 10.1002/14651858.CD004022.pub5] [Medline: 33314019]
16. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/s0140-6736(19)30041-8]
17. National Diet and Nutrition Survey: Assessment of salt intake from urinary sodium in adults (aged 19 to 64 years) in England, 2018 to 2019. Public Health England. 2020 Mar 25. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/876252/Report_England_Sodium_Survey_2018-to-2019__3_.pdf [accessed 2021-09-02]
18. Global Action Plan for the Prevention and Control of NCDs 2013-2020. In: World Health Organization. Geneva: World Health Organization; May 31, 2013:1-103.
19. Anderson CA, Appel LJ, Okuda N, Brown IJ, Chan Q, Zhao L, et al. Dietary sources of sodium in China, Japan, the United Kingdom, and the United States, women and men aged 40 to 59 years: the INTERMAP study. J Am Diet Assoc 2010 May;110(5):736-745 [FREE Full text] [doi: 10.1016/j.jada.2010.02.007] [Medline: 20430135]
20. He FJ, Pombo-Rodrigues S, Macgregor GA. Salt reduction in England from 2003 to 2011: its relationship to blood pressure, stroke and ischaemic heart disease mortality. BMJ Open 2014 Apr 14;4(4):e004549 [FREE Full text] [doi: 10.1136/bmjopen-2013-004549] [Medline: 24732242]
21. Hooper L, Bartlett C, Smith G, Ebrahim S. Advice to reduce dietary salt for prevention of cardiovascular disease. Cochrane Database Syst Rev 2004 Jan 26(1):1-54 [FREE Full text] [doi: 10.1002/14651858.cd003656.pub2]
22. Ruzicka M, Hiremath S, Steiner S, Helis E, Szczotka A, Baker P, et al. What is the feasibility of implementing effective sodium reduction strategies to treat hypertension in primary care settings? A systematic review. J Hypertens 2014 Jul;32(7):1388-1394. [doi: 10.1097/HJH.0000000000000182] [Medline: 24694380]
23. Hypertension in adults: diagnosis and management. NICE guideline [NG136]. National Institute for Health and Care Excellence. 2019 Aug 28. URL: https://www.nice.org.uk/guidance/ng136 [accessed 2020-03-23]
24. Scottish Intercollegiate Guidelines Network (SIGN). Risk Estimation and the Prevention of Cardiovascular Disease. Edinburgh: Healthcare Improvement Scotland; 2017:1-118 URL: https://www.sign.ac.uk/media/1085/sign149.pdf

25. Werch CC, Grenard JL, Burnett J, Watkins JA, Ames S, Jobli E. Translation as a function of modality: the potential of brief interventions. Eval Health Prof 2006 Mar;29(1):89-125 [FREE Full text] [doi: 10.1177/0163278705284444] [Medline: 16510881]
26. Frie K, Hartmann-Boyce J, Jebb SA, Aveyard P. Effectiveness of a self-regulation intervention for weight loss: A randomized controlled trial. Br J Health Psychol 2020 Sep;25(3):652-676 [FREE Full text] [doi: 10.1111/bjhp.12436] [Medline: 32489005]
27. Aveyard P, Lewis A, Tearne S, Hood K, Christian-Brown A, Adab P, et al. Screening and brief intervention for obesity in primary care: a parallel, two-arm, randomised trial. Lancet 2016 Nov;388(10059):2492-2500 [FREE Full text] [doi: 10.1016/s0140-6736(16)31893-1]
28. Kulick D, Langer RD, Ashley JM, Gans KM, Schlauch K, Feller C. Live well: a practical and effective low-intensity dietary counseling intervention for use in primary care patients with dyslipidemia--a randomized controlled pilot trial. BMC Fam Pract 2013 May 12;14:59 [FREE Full text] [doi: 10.1186/1471-2296-14-59] [Medline: 23663789]
29. Beeken RJ, Leurent B, Vickerstaff V, Wilson R, Croker H, Morris S, et al. A brief intervention for weight control based on habit-formation theory delivered through primary care: results from a randomised controlled trial. Int J Obes (Lond) 2017 Feb;41(2):246-254 [FREE Full text] [doi: 10.1038/ijo.2016.206] [Medline: 27867204]
30. Mason F, Farley A, Pallan M, Sitch A, Easter C, Daley AJ. Effectiveness of a brief behavioural intervention to prevent weight gain over the Christmas holiday period: randomised controlled trial. Br Med J 2018 Dec 10;363:k4867 [FREE Full text] [doi: 10.1136/bmj.k4867] [Medline: 30530821]
31. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
32. Schoeppe S, Alley S, Van Lippevelde W, 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]
33. Ipjian ML, Johnston CS. Smartphone technology facilitates dietary change in healthy adults. Nutrition 2017 Jan;33:343-347 [FREE Full text] [doi: 10.1016/j.nut.2016.08.003] [Medline: 27742102]
34. Eyles H, McLean R, Neal B, Jiang Y, Doughty RN, McLean R, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: Findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Sep;24(13):1435-1444 [FREE Full text] [doi: 10.1177/2047487317715713] [Medline: 28631933]
35. 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]
36. Payne Riches S, Piernas C, Aveyard P, Sheppard JP, Rayner M, Jebb SA. The Salt Swap intervention to reduce salt intake in people with high blood pressure: protocol for a feasibility randomised controlled trial. Trials 2019 Oct 11;20(1):584 [FREE Full text] [doi: 10.1186/s13063-019-3691-y] [Medline: 31604477]
37. Michie S, West R, Atkins L. The Behaviour Change Wheel: A Guide To Designing Interventions. London: Silverback Publishing; 2014:1-329.
38. 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 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
39. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701 [FREE Full text] [doi: 10.1037/a0016136] [Medline: 19916637]
40. Behaviour change: individual approaches. National Institute for Health and Care Excellence. 2014 Jan 02. URL: https://www.nice.org.uk/guidance/ph49 [accessed 2020-03-23]
41. Brandview. 2020. URL: https://www.brandview.com/ [accessed 2020-05-08]
42. The George Institute for Global Health. URL: https://www.georgeinstitute.org.au/ [accessed 2021-02-08]
43. Cut down on salt. British Heart Foundation, London. 2013. URL: https://www.bhf.org.uk/informationsupport/publications/healthy-eating-and-drinking/cut-down-on-salt-pdf [accessed 2018-10-23]
44. Taking control of salt. British Heart Foundation. 2018 Dec 19. URL: https://www.bhf.org.uk/informationsupport/publications/healthy-eating-and-drinking/taking-control-of-salt [accessed 2020-03-11]
45. Charters E. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. Brock Edu J 2003 Jul 01;12(2):68-82. [doi: 10.26522/brocked.v12i2.38]
46. Sin J, Woodham LA, Henderson C, Williams E, Hernández AS, Gillard S. Usability evaluation of an eHealth intervention for family carers of individuals affected by psychosis: A mixed-method study. Digit Health 2019;5:2055207619871148 [FREE Full text] [doi: 10.1177/2055207619871148] [Medline: 31489205]

47. Akhloufi H, Verhaegh SJ, Jaspers MW, Melles DC, van der Sijs H, Verbon A. A usability study to improve a clinical decision support system for the prescription of antibiotic drugs. PLoS One 2019;14(9):e0223073 [FREE Full text] [doi: 10.1371/journal.pone.0223073] [Medline: 31553785]
48. Henry H, Reicks M, Smith C, Reimer K, Atwell J, Thomas R. Identification of factors affecting purchasing and preparation of fruit and vegetables by stage of change for low-income African American mothers using the think-aloud method. J Am Diet Assoc 2003 Dec;103(12):1643-1646 [FREE Full text] [doi: 10.1016/j.jada.2003.09.039] [Medline: 14647093]
49. Robson SM, DeLuccia R, Baker S, Bodt B, Trabulsi J. Qualitative research on the real-time decision making of wic participants while food shopping: use of think-aloud methodology. J Acad Nutr Diet 2020 Jan;120(1):111-119 [FREE Full text] [doi: 10.1016/j.jand.2019.05.009] [Medline: 31307943]
50. Rayner M, Boaz A, Higginson C. Consumer use of health-related endorsements on food labels in the United Kingdom and Australia. J Nutrit Edu 2001 Jan;33(1):24-30 [FREE Full text] [doi: 10.1016/s1499-4046(06)60006-7]
51. Tanner SA, McCarthy MB, O'Reilly SJ. Exploring the roles of motivation and cognition in label-usage using a combined eye-tracking and retrospective think aloud approach. Appetite 2019 Apr 01;135:146-158 [FREE Full text] [doi: 10.1016/j.appet.2018.11.015] [Medline: 30482412]
52. Stata Statistical Software: Release 16. StataCorp LLC, College Station, TX. 2019. URL: https://www.stata.com/products/ [accessed 2020-03-23]
53. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
54. Cohen D, Crabtree B. Lincoln and Guba's Evaluative Criteria. Qualitative Research Guidelines Project. 2006 Jul. URL: http://www.qualres.org/HomeLinc-3684.html [accessed 2020-11-11]
55. NVivo qualitative data analysis software. QSR International Pty Ltd. 2018. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-09-03]
56. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760 [FREE Full text] [doi: 10.1177/1049732315617444] [Medline: 26613970]
57. Results of the National Diet and Nutrition Survey (NDNS) rolling programme for 2014 to 2015 and 2015 to 2016 (combined data tables). Public Health England. 2018. URL: https://www.gov.uk/government/statistics/ndns-results-from-years-7-and-8-combined [accessed 2019-08-25]
58. Adults: Media use and attitudes report 2019. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf [accessed 2020-03-23]
59. Rasche P, Wille M, Bröhl CL, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: National survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26 [FREE Full text] [doi: 10.2196/mhealth.8619] [Medline: 29362211]
60. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
61. Joffres M, Falaschetti E, Gillespie C, Robitaille C, Loustalot F, Poulter N, et al. Hypertension prevalence, awareness, treatment and control in national surveys from England, the USA and Canada, and correlation with stroke and ischaemic heart disease mortality: a cross-sectional study. BMJ Open 2013 Aug 30;3(8):e003423 [FREE Full text] [doi: 10.1136/bmjopen-2013-003423] [Medline: 23996822]
62. Spencer E, Mahtani K. Hawthorne bias. Catalogue Of Bias. 2017. URL: https://catalogofbias.org/biases/hawthorne-effect/ [accessed 2020-03-12]
63. Piernas C, Aveyard P, Lee C, Tsiountsioura M, Noreik M, Astbury NM, et al. Evaluation of an intervention to provide brief support and personalized feedback on food shopping to reduce saturated fat intake (PC-SHOP): A randomized controlled trial. PLoS Med 2020 Nov;17(11):e1003385 [FREE Full text] [doi: 10.1371/journal.pmed.1003385] [Medline: 33151934]
64. Morris E, Aveyard P, Dyson P, Noreik M, Bailey C, Fox R, et al. A food-based, low-energy, low-carbohydrate diet for people with type 2 diabetes in primary care: A randomized controlled feasibility trial. Diabetes Obes Metab 2020 Apr;22(4):512-520 [FREE Full text] [doi: 10.1111/dom.13915] [Medline: 31709697]
65. Indices of Multiple Deprivation 2019. Oxfordshire County Council, JSNA Bitesize. 2019. URL: https://insight.oxfordshire.gov.uk/cms/system/files/documents/20191004_bitesize_IMD2019overview.pdf [accessed 2021-09-09]
66. Inglis V, Ball K, Crawford D. Why do women of low socioeconomic status have poorer dietary behaviours than women of higher socioeconomic status? A qualitative exploration. Appetite 2005 Dec;45(3):334-343 [FREE Full text] [doi: 10.1016/j.appet.2005.05.003] [Medline: 16171900]
67. Ji C, Cappuccio FP. Socioeconomic inequality in salt intake in Britain 10 years after a national salt reduction programme. BMJ Open 2014 Aug 14;4(8):e005683 [FREE Full text] [doi: 10.1136/bmjopen-2014-005683] [Medline: 25161292]
68. de Mestral C, Mayén A, Petrovic D, Marques-Vidal P, Bochud M, Stringhini S. Socioeconomic determinants of sodium intake in adult populations of high-income countries: a systematic review and meta-analysis. Am J Public Health 2017 Apr;107(4):1-12 [FREE Full text] [doi: 10.2105/ajph.2016.303629]

69. Fulgoni VL, Agarwal S, Spence L, Samuel P. Sodium intake in US ethnic subgroups and potential impact of a new sodium reduction technology: NHANES Dietary Modeling. Nutr J 2014 Dec 18;13(1):120 [FREE Full text] [doi: 10.1186/1475-2891-13-120] [Medline: 25522786]
70. Leung G, Stanner S. Diets of minority ethnic groups in the UK: influence on chronic disease risk and implications for prevention. Nutr Bull 2011;36(2):161-198 [FREE Full text] [doi: 10.1111/j.1467-3010.2011.01889.x]
71. Smeaton D, Drape A, Vowden K, Durante L. Development work for wave 2 of Food and You Survey: Executive summary. Policy Studies Institute. 2011 Mar. URL: https://www.food.gov.uk/sites/default/files/media/document/food-and-you-2012-development-report-summary_0.pdf [accessed 2020-03-23]
72. Patel N, Ferrer HB, Tyrer F, Wray P, Farooqi A, Davies MJ, et al. Barriers and facilitators to healthy lifestyle changes in minority ethnic populations in the UK: a narrative review. J Racial Ethn Health Disparities 2017 Dec;4(6):1107-1119 [FREE Full text] [doi: 10.1007/s40615-016-0316-y] [Medline: 27928772]
73. Dong Y, Peng C. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
74. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of mobile health technologies aimed at salt reduction: systematic review. JMIR Mhealth Uhealth 2019 Apr 17;7(4):e13250 [FREE Full text] [doi: 10.2196/13250] [Medline: 30994467]
75. He FJ, MacGregor GA. Effect of modest salt reduction on blood pressure: a meta-analysis of randomized trials. Implications for public health. J Hum Hypertens 2002 Nov;16(11):761-770 [FREE Full text] [doi: 10.1038/sj.jhh.1001459] [Medline: 12444537]
76. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Writing Group of the PREMIER Collaborative Research Group. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. J Am Med Assoc 2003;289(16):2083-2093 [FREE Full text] [doi: 10.1001/jama.289.16.2083] [Medline: 12709466]
77. Nghiem N, Blakely T, Cobiac LJ, Pearson AL, Wilson N. Health and economic impacts of eight different dietary salt reduction interventions. PLoS One 2015;10(4):e0123915 [FREE Full text] [doi: 10.1371/journal.pone.0123915] [Medline: 25910259]
78. World Health Organization. The SHAKE Technical Package for Salt Reduction. Geneva: World Health Organization; 2016.
79. Eyles H, Webster J, Jebb S, Capelin C, Neal B, Ni Mhurchu C. Impact of the UK voluntary sodium reduction targets on the sodium content of processed foods from 2006 to 2011: analysis of household consumer panel data. Prev Med 2013 Nov;57(5):555-560 [FREE Full text] [doi: 10.1016/j.ypmed.2013.07.024] [Medline: 23954183]
80. Salt targets 2017: Progress report. Public Health England. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/765571/Salt_targets_2017_progress_report.pdf [accessed 2020-03-23]
81. Brinsden HC, He FJ, Jenner KH, Macgregor GA. Surveys of the salt content in UK bread: progress made and further reductions possible. BMJ Open 2013 Jun 20;3(6):1-7 [FREE Full text] [doi: 10.1136/bmjopen-2013-002936] [Medline: 23794567]

Abbreviations

BCT: behavior change technique
CVD: cardiovascular disease
GP: general practitioner
NIHR: National Institute for Health Research
RCT: randomized controlled trial

Edited by L Buis; submitted 03.12.20; peer-reviewed by P Zhang, S Ali, M Dorsch; comments to author 25.12.20; revised version received 11.03.21; accepted 07.05.21; published 21.10.21.

Please cite as:
Payne Riches S, Piernas C, Aveyard P, Sheppard JP, Rayner M, Albury C, Jebb SA
A Mobile Health Salt Reduction Intervention for People With Hypertension: Results of a Feasibility Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(10):e26233
URL: https://mhealth.jmir.org/2021/10/e26233
doi:10.2196/26233
PMID:34673535

Original Paper

Impact of an eHealth Smartphone App on the Mental Health of Patients With Psoriasis: Prospective Randomized Controlled Intervention Study

Lena Domogalla[1*], MD; Alena Beck[1*], Cand Med; Theresa Schulze-Hagen[1], MD; Raphael Herr[2], PhD; Johannes Benecke[1], MD; Astrid Schmieder[1,3], MD

[1]Department of Dermatology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany

[2]Mannheim Institute of Public Health, Social and Preventive Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[3]Department of Dermatology, University Hospital Würzburg, Würzburg, Germany

[*]these authors contributed equally

Corresponding Author:
Astrid Schmieder, MD
Department of Dermatology
University Hospital Würzburg
Josef-Schneider-Straße 2, Haus D8
Würzburg
Germany
Phone: 49 931 201 26235
Fax: 49 931 201 2670
Email: astrid.schmieder@umm.de

Abstract

Background: Psoriasis has a negative impact on patients' physical and mental health and can lead to anxiety and depression. Disease management strategies, including educational programs and eHealth devices, have been shown to improve health care for several chronic diseases. However, such disease management strategies are lacking in the routine care of patients with psoriasis.

Objective: This study aims to study the impact of a novel intervention that combines an educational program with a disease management smartphone app on the mental health of patients with psoriasis.

Methods: Patients with psoriasis in the intervention group received an educational program; attended visits on weeks 0, 12, 24, 36, and 60; and had access to the study app. Patients in the control group only attended the visits. The primary endpoint was a significant reduction of scores on the Hospital Anxiety and Depression Scale (HADS). Secondary end points were reductions in Dermatology Life Quality Index score, Psoriasis Area and Severity Index score, pruritus, and pain, as well as improvements in mood and daily activities. In addition, modulating effects of sex, age, disease duration, and app use frequency were evaluated.

Results: A total of 107 patients were included in the study and randomized into the control group (53/107, 49.5%) or intervention group (54/107, 50.5%). Approximately 71.9% (77/107) of the patients completed the study. A significant reduction in HADS-Depression (HADS-D) in the intervention group was found at weeks 12 (P=.04) and 24 (P=.005) but not at weeks 36 (P=.12) and 60 (P=.32). Patient stratification according to app use frequency showed a significant improvement in HADS-D score at weeks 36 (P=.004) and 60 (P=.04) and in HADS-Anxiety (HADS-A) score at weeks 36 (P=.04) and 60 (P=.05) in the group using the app less than once every 5 weeks. However, in patients using the app more than once every 5 weeks, no significant reduction in HADS-D (P=.84) or HADS-A (P=.20) score was observed over the 60-week study period compared with that observed in patients in the control group. All findings were independent of sex, age, and disease duration.

Conclusions: These findings support the use of a disease management smartphone app as a valid tool to achieve long-term improvement in the mental health of patients with psoriasis if it is not used too frequently. Further studies are needed to analyze the newly observed influence of app use frequency.

Trial Registration: Deutsches Register Klinischer Studien DRKS00020755; https://tinyurl.com/nyzjyvvk

(JMIR Mhealth Uhealth 2021;9(10):e28149) doi:10.2196/28149

KEYWORDS

psoriasis; eHealth; mHealth; telemedicine; teledermatology; patient educational program; disease management; smartphone app; mental health; mobile phone

Introduction

Background

Psoriasis is a common, chronic inflammatory skin disease that affects approximately 2.53% of the German population and has a significant impact on physical and mental health [1-3]. In addition to physical symptoms such as pain and itching, patients with psoriasis report stigmatization, shame, lack of self-confidence, depression, and anxiety as major impairments [4-6]. Moreover, low therapy adherence in patients and consequent poor disease control results from a lack of knowledge regarding the disease [7].

For several different chronic diseases, educational programs are part of routine care because they improve patients' self-management skills and reduce disease activity [8]. A routine educational program is not in place to cater to patients who wish to be better informed about their illness and more involved in therapy decisions [7-9]. The pilot study conducted by Bubak et al [10] showed that an educational program for patients with psoriasis alone leads to a significant improvement in the knowledge of their disease and to an amelioration of general health, but not to an improvement in mental health. Several other studies have reported educational programs as valid tools to improve self-expertise, therapy adherence, and quality of life in patients with psoriasis; encourage lifestyle changes; and reduce disease severity [11-14].

Another possibility to improve patient care is using eHealth devices. Since 2019, Germany has allowed the prescription of scientifically validated digital health apps. In several studies, eHealth devices have shown positive effects on common chronic diseases such as diabetes, hypertension, chronic heart failure, and asthma [15-18]. For psoriasis, the data are still limited; however, digital care seems to be as safe and effective as regular in-person care [19-21]. So far, no studies that evaluate smartphone apps for long-term disease management in psoriasis exist.

Objectives

In this study, we combined the educational program already presented by Bubak et al [10] with a disease management smartphone app, generating a new possibility to strengthen patients' self-management, reduce disease burden, and build a more trusting doctor-patient relationship. In a first interim analysis of this study, after 6 months, a significant improvement in depressive symptoms was found in the intervention group receiving the educational program and the study app. In addition, a clear impact of the app use frequency was assessed. Patients using the app <20% (ie, less than once every 5 weeks) had a significantly stronger reduction in depressive and anxiety symptoms [22]. This study investigates the long-term effects over 60 weeks of a patient educational program combined with a psoriasis app on the mental health of patients with psoriasis and examines the mediating effect of app use frequency more closely.

Methods

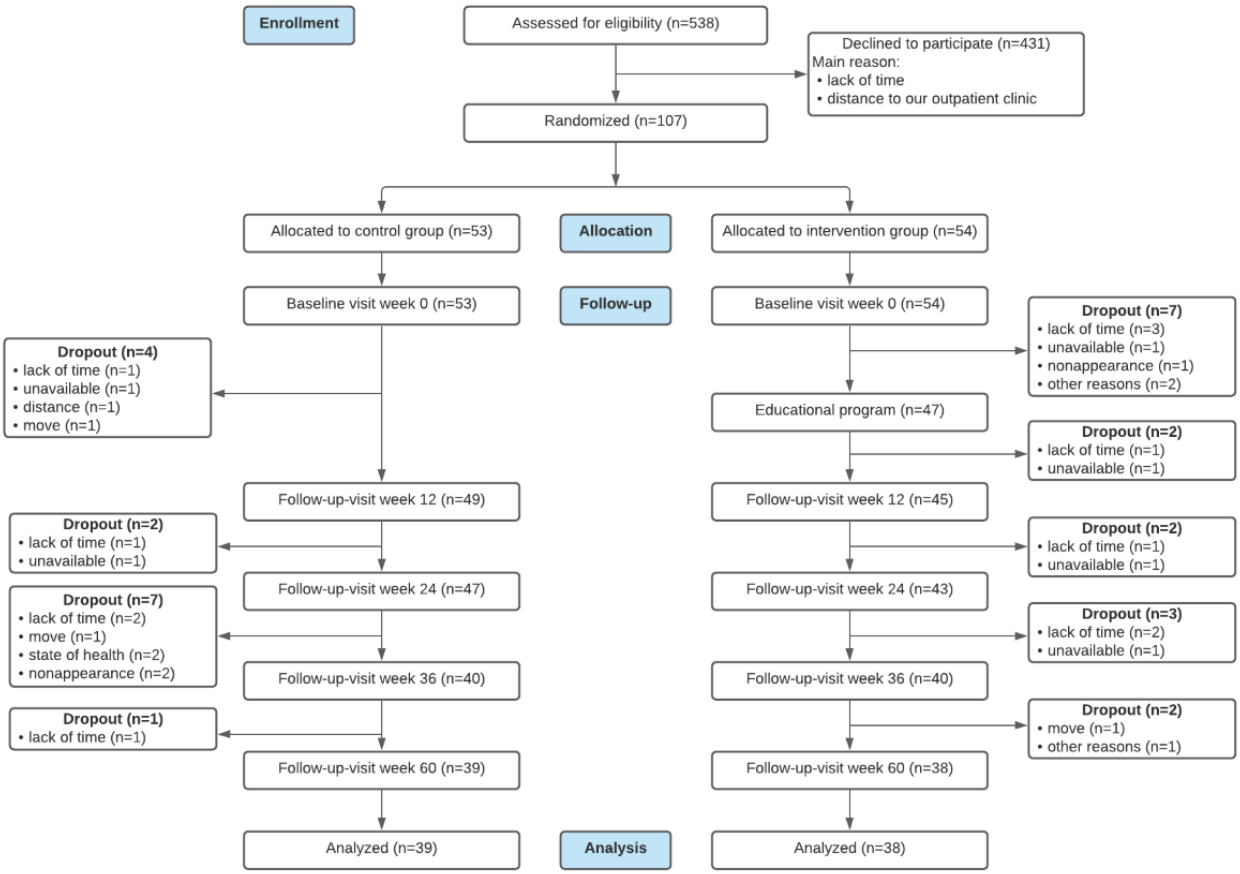
Study Design and Patients

This intervention study was undertaken at the Department of Dermatology, Venereology, and Allergology at the University Medical Center Mannheim, Germany, between January 2018 and June 2020. Patients were recruited at the outpatient clinic, where they were asked to participate in the study while attending their regular doctors' appointments. Eligible patients were aged 18 to 75 years, able to provide written informed consent, had a physician-confirmed diagnosis of moderate-to-severe psoriasis (defined as Psoriasis Area and Severity Index [PASI] score >10; body surface area >10; Dermatology Life Quality Index [DLQI] score >10; psoriatic involvement of the scalp, palms, soles, and genital area; or psoriatic nail involvement of at least 2 nails). The exclusion criteria were the inability to provide written informed consent and no access to a smartphone.

The study was conducted in accordance with the Declaration of Helsinki and approved by the Medical Ethics Committee of the Medical Faculty Mannheim, Heidelberg University (ethics approval 2017-655N-MA). The trial is registered at Deutsches Register Klinische Studien (registration number: DRKS00020755). Written informed consent was provided by each patient before their participation in the study. Eligible participants were randomized into an intervention or control group in a 1:1 ratio. For this purpose, sealed envelopes labeled *intervention group* or *control group* were randomly assigned to each patient during the baseline visit.

The control group (n=53) started the 60-week study period with their baseline visit in week 0, when they indicated sociodemographic and psoriasis-related data. In addition, the PASI (range 0-72), DLQI (range 0-30), Hospital Anxiety and Depression Scale–Anxiety/Depression (HADS-A/D; range 0-21 for anxiety and depression each), and numeric rating scales for skin pain and pruritus (range 0-10 each) were assessed. Furthermore, the patients indicated how much psoriasis negatively affected their mood and daily activity levels during the past weeks on a Likert scale (range 0-3 each; 0=not at all, 1=a little, 2=fairly, and 3=very). In-person follow-up visits with the same assessments took place at weeks 12, 24, and 36 and the final visit at week 60 (Figure 1).

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of the study cohort.



The intervention group (n=54) participated in the same baseline and follow-up visits as the control group. In addition, they attended a 2-hour-long educational program on the topic of psoriasis, which was held by specialists in dermatology (AS and JB) after the baseline visit (Figure 1). The educational program was held in a group setting and included information on psoriasis pathogenesis, therapy options, and comorbidities (Multimedia Appendix 1) [10]. Each participant received a personal anonymized access code and an introduction to our psoriasis monitoring app DermaScope Mobile (Multimedia Appendices 2 and 3). For 60 weeks, the patients were able to document their psoriasis regularly by photodocumentation of their skin and answering health questionnaires regarding their quality of life, mood, activity, pain, and pruritus via the app. App use frequency for adding new data was limited to once a week. Furthermore, the patients could contact specialized dermatologists unrestrictedly (AS and JB) via a chat feature within the app. Consort eHealth checklist can be found in the Multimedia Appendix 4.

Outcomes

The primary end point was a significant reduction in the HADS-D/-A in the intervention group at the follow-up visits compared with the control group. Secondary endpoints were the reduction in the PASI and numeric rating scale for pruritus and pain and an improvement in the DLQI and mood and daily activity Likert Scales at the follow-up. Modulating effects of sex, age, disease duration, and app use frequency were evaluated for each time point.

Statistical Analysis

Linear panel data regression analyses estimated the trajectories of the outcomes. Random-effect regressions determined the main and interaction effects of group membership (intervention vs control group; interaction *week x group* in Multimedia Appendix 5) and visit time point (interaction *week* in Multimedia Appendix 5) on HADS-D, HADS-A, DLQI, mood, daily activity, PASI, pruritus, and pain (Multimedia Appendix 5). Two models of adjustment were calculated. One was unadjusted, whereas the other was adjusted for sex, age, and disease duration. In additional analyses, the effects of the app use frequency over 60 weeks were included (group membership: control vs <20% vs ≥20% use; interaction *week x group* in Multimedia Appendix 6). The chosen cutoff of 20% equals one-time use every 5 weeks. Variables were transformed to approach normal distribution (power transform square root of HADS-D and HADS-A, log10 of DLQI, and square of mood; no transformation of daily activity, pruritus, and pain). All statistical analyses were performed using STATA SE 14 (StataCorp LLC). A statistically significant difference was assumed at $P \leq .05$.

Results

Study Cohort

This study was conducted between January 2018 and June 2020. Of the 538 patients assessed, 431 (80.1%) were found to be ineligible because of lack of time, distance to our outpatient

clinic, and no access to a smartphone. Of the 538 patients, 107 (19.9%) patients joined the study. After randomization to the control (53/107, 49.5%) or intervention (54/107, 50.5) groups, all patients participated in the baseline visit. After the baseline visit, 10.3% (11/107) of patients discontinued the study (control group: 4/11, 36%; intervention group: 7/11, 64%). The remaining 87% (47/54) of patients in the intervention group attended the educational program. Until week 60, 20% (10/49) of patients from the control group and 19% (9/47) of patients from the intervention group discontinued the study (Figure 1). The most common reasons for withdrawal were lack of time (12/107, 11.2%) and unavailability or nonappearance (10/107, 9.3%). Other reasons included distance to our outpatient clinic or relocating (4/107, 3.7%), poor health status (2/107, 1.9%), and other reasons (3/107, 2.8%). Of the 107 patients, 77 (72.0%) completed the study; 74% (39/53) from the control group and 70.4% (38/54) from the intervention group completed the study. Over the entire study period, the demographic, socioeconomic, and psoriasis-related characteristics were well balanced between the control and intervention groups (Tables 1 and 2), except for a slightly older control group (P=.04). At week 60, 9% (7/77) of patients were treated with disease-modifying antirheumatic drugs, and 64% (49/77) were treated with biologics. Of the 77 patients, 16 (21%) received topical or phototherapy, 2 (3%) received other systemic drugs, and 3 (4%) were not treated with any antipsoriatic therapy.

Table 1. Characteristics of the study cohort at weeks 0 (N=107).

Characteristic at week 0[a]	Overall	Control group (n=53)	Intervention group (n=54)	*P* value
Sex, n (%)				
Female	42 (39.3)	23 (43)	19 (35)	.38[b]
Male	65 (60.7)	30 (57)	35 (65)	.38[b]
Age (years)				
Mean (SD)	49.1 (12.1)	51.7 (11.8)	46.5 (11.9)	.03[c]
Median (IQR)	51.0 (42-57)	53.0 (49-59)	48.0 (36-54)	.03[c]
BMI (kg/m^2)				
Mean (SD)	29.1 (5.7)	29.3 (6.1)	28.9 (5.3)	.72[b]
Median (IQR)	28.3 (24.8-32.5)	28.3 (24.7-32.2)	28.2 (24.9-32.5)	.72[b]
Alcohol (days/week)				
Mean (SD)	1.1 (1.5)	1.1 (1.5)	1.2 (1.5)	.68[c]
Median (IQR)	1.0 (0-2)	0.0 (0-2)	1.0 (0-1.5)	.68[c]
Smoker, n (%)	35 (32.7)	20 (38)	15 (28)	.27[b]
Duration of psoriasis (years)				
Mean (SD)	18.9 (14.6)	20.4 (15.5)	17.5 (13.7)	.31[c]
Median (IQR)	15.0 (6-28)	18.0 (7-34)	13.0 (6-26)	.31[c]
Psoriatic arthritis, n (%)	46 (43.0)	24 (45)	22 (42)	.76[b]
Antipsoriatic therapy, n (%)				
Topical or UV therapy	31 (29.0)	14 (26)	17 (31)	.43[b]
DMARDS[d]	17 (15.9)	6 (11)	11 (20)	.43[b]
Others	1 (0.9)	1 (2)	0 (0)	.43[b]
Biologicals	51 (47.7)	29 (55)	22 (41)	.43[b]
No therapy	7 (6.5)	3 (6)	4 (7)	.43[b]
HADS-D[e] (range 0-21)				
Mean (SD)	5.3 (4.7)	5.2 (5.0)	5.3 (4.3)	.94[c]
Median (IQR)	4.0 (1-8)	3.0 (1-8)	4.5 (2-8)	.94[c]
HADS-A[f] (range 0-21)				
Mean (SD)	6.9 (4.4)	7.0 (4.9)	6.7 (3.8)	.66[c]
Median (IQR)	6.0 (3-10)	6.0 (3-10)	6.0 (4-9)	.66[c]
DLQI[g] (range 0-30)				
Mean (SD)	8.2 (8.0)	8.5 (8.5)	7.9 (7.6)	.69[c]
Median (IQR)	5.0 (2-13)	5.0 (3-12)	5.0 (2-13)	.69[c]
Mood (range 0-3)				
Mean (SD)	1.2 (1.2)	1.2 (1.2)	1.2 (1.2)	.92[c]
Median (IQR)	1.0 (0-2)	1.0 (0-2)	1.0 (0-2)	.92[c]
Daily activity (range 0-3)				

Characteristic at week 0[a]	Overall	Control group (n=53)	Intervention group (n=54)	*P* value
Mean (SD)	1.3 (1.1)	1.4 (1.1)	1.1 (1.1)	.30[c]
Median (IQR)	1.0 (0-2)	1.0 (0-2)	1.0 (0-2)	.30[c]
PASI[h] (range 0-72)				
Mean (SD)	5.1 (5.2)	5.1 (5.1)	5.0 (5.4)	.98[c]
Median (IQR)	3.0 (1.4-7.6)	3.6 (1.4-7.2)	3.0 (1.5-7.6)	.98[c]
Pain (range 0-10)[i]				
Mean (SD)	2.0 (2.4)	2.2 (2.7)	1.9 (2.2)	.53[c]
Median (IQR)	1.0 (0-4)	1.0 (0-3)	1.0 (0-4)	.53[c]
Pruritus (range 0-10)				
Mean (SD)	3.0 (2.7)	3.2 (3.1)	2.8 (2.2)	.35[c]
Median (IQR)	2.0 (1-5)	2.0 (1-5)	2.0 (1-4)	.35[c]
App use frequency, n (%)				
<20%	20 (18.7)	0 (0)	20 (37)[j]	<.001[b]
≥20%	20 (18.7)	0 (0)	20 (37)[j]	<.001[b]
No app use	67 (62.6)	53 (100)	14 (26)[j]	<.001[b]

[a]Data for sex, age, BMI, smoking, alcohol consumption, psoriasis duration, psoriasis arthritis, and therapy were collected at week 0 only.

[b]Categorical variables were analyzed using the chi-square test.

[c]Continuous variables were analyzed using the *t* test.

[d]DMARDs: disease-modifying antirheumatic drugs.

[e]HADS-D: Hospital Anxiety and Depression Scale–Depression.

[f]HADS-A: Hospital Anxiety and Depression Scale–Anxiety.

[g]DLQI: Dermatology Life Quality Index.

[h]PASI: Psoriasis Area and Severity Index.

[i]For 1 patient, data for weight and height were missing.

[j]App use frequency for week 0 was identified retrospectively in weeks 0 to 12. From the intervention group, 14 patients did not download the app at all or dropped out of the study. Therefore, they could not be divided into app use frequency subgroups.

Table 2. Characteristics of the study cohort at weeks 60 (N=77).

Characteristics week 60	Overall (n=77)	Control group (n=39)	Intervention group (n=38)	*P* value
Sex, n (%)				
Female	28 (36)	13 (33)	15 (39)	.58[a]
Male	49 (64)	26 (67)	23 (61)	.58[a]
Age (years)				
Mean (SD)	49.6 (11.7)	52.3 (10.5)	46.9 (12.3)	.04[b]
Median (IQR)	52.0 (45-57.0)	54.0 (50-59)	48.0 (39-54)	.04[b]
BMI (kg/m^2)				
Mean (SD)	28.3 (5.6)	28.9 (6.3)	27.6 (4.7)	.33[b]
Median (IQR)	26.6 (24.5-31.7)	26.6 (24.5-31.2)	26.0 (24.2-32.5)	.33[b]
Alcohol (days/week)				
Mean (SD)	1.3 (1.6)	1.3 (1.6)	1.3 (1.7)	.99[b]
Median (IQR)	1.0 (0-2.5)	1.0 (0-3)	1.0 (0-2)	.99[b]
Smoker n (%)	26 (34)	16 (41)	10 (27)	.17[a]
Duration of psoriasis (years)				
Mean (SD)	19.2 (14.8)	21.2 (16.1)	17.1 (13.1)	.22[b]
Median (IQR)	15.0 (6-28)	18.0 (7-37)	13.5 (6-24)	.22[b]
Psoriatic arthritis, n (%)	35 (45.5)	20 (51.3)	15 (39.5)	.35[a]
Antipsoriatic therapy, n (%)				
Topical or UV therapy	16 (21)	5 (13)	11 (29)	.07[a]
DMARDS[c]	7 (9)	3 (8)	4 (11)	.07[a]
Others	2 (3)	0 (0)	2 (5)	.07[a]
Biologicals	49 (64)	28 (72)	21 (55)	.07[a]
No therapy	3 (4)	3 (8)	0 (0)	.07[a]
HADS-D (range 0-21)[d]				
Mean (SD)	4.0 (4.4)	4.1 (4.2)	4.0 (4.5)	.94[b]
Median (IQR)	2.0 (0-7)	3.0 (0-7)	2.0 (0-7)	.94[b]
HADS-A (range 0-21)[e]				
Mean (SD)	5.5 (3.8)	6.0 (4.0)	4.9 (3.5)	.22[b]
Median (IQR)	5.0 (2-8)	6.0 (2-8)	4.0 (2-7)	.22[b]
DLQI (range 0-30)[f]				
Mean (SD)	4.1 (5.5)	3.7 (4.1)	4.4 (5.5)	.56[b]
Median (IQR)	1.0 (1-7)	1.0 (1-7)	1.0 (1-8)	.56[b]
Mood (range 0-3)				
Mean (SD)	0.7 (0.7)	0.7 (0.7)	0.6 (0.8)	.71[b]
Median (IQR)	1.0 (0-1)	1.0 (0-1)	0.5 (0-1)	.71[b]
Daily activity (range 0-3)				
Mean (SD)	0.6 (0.7)	0.6 (0.7)	0.5 (0.7)	.59[b]

Characteristics week 60	Overall (n=77)	Control group (n=39)	Intervention group (n=38)	*P* value
Median (IQR)	0.0 (0-1)	0.0 (0-1)	0.0 (0-1)	.59[b]
PASI (range 0-72)[g]				
Mean (SD)	2.6 (3.0)	2.6 (2.9)	2.7 (3.3)	.90[b]
Median (IQR)	1.5 (0.4-3.4)	1.7 (0-3.4)	1.4 (0.6-3.4)	.90[b]
Pain (range 0-10)[h]				
Mean (SD)	1.5 (2.5)	2.0 (2.8)	1.0 (2.0)	.10[b]
Median (IQR)	0.0 (0-2.5)	0.0 (0-3)	0.0 (0-1)	.10[b]
Pruritus (range 0-10)				
Mean (SD)	1.8 (2.1)	1.8 (2.1)	1.8 (2.2)	.95[b]
Median (IQR)	1.0 (0-3)	1.0 (0-3)	1.0 (0-3)	.95[b]
App use frequency, n (%)				
<20%	15 (20)	0 (0)	15 (39)	<.001[a]
≥20%	18 (23)	0 (0)	18 (47)	<.001[a]
No app use	44 (57)	39 (100)	5 (13)	<.001[a]

[a]Categorical variables were analyzed using the chi-square test.

[b]Continuous variables were analyzed using the *t* test.

[c]DMARDs: disease-modifying antirheumatic drugs.

[d]HADS-D: Hospital Anxiety and Depression Scale–Depression.

[e]HADS-A: Hospital Anxiety and Depression Scale–Anxiety.

[f]DLQI: Dermatology Life Quality Index.

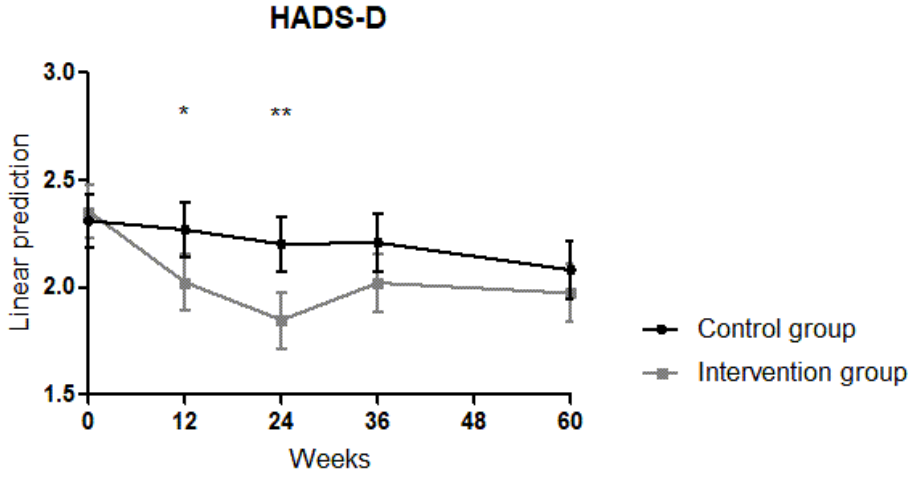
[g]PASI: Psoriasis Area and Severity Index.

[h]For 1 patient, data for pain were missing at week 60.

Outcomes

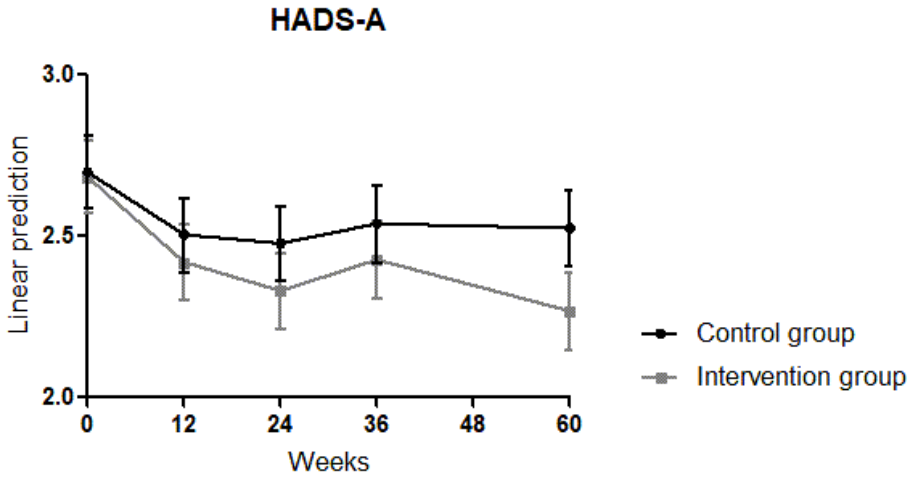
The significant reduction in the HADS-D in the intervention group compared with that in the control group found at weeks 12 and 24 did not persist in weeks 36 and 60 (interaction week x intervention: week 12: coefficient=–0.289, *P*=.04; week 24: coefficient=–0.397, *P*=.005; week 36: coefficient=0.231, *P*=.12; week 60: coefficient=0.15, *P*=.32; Model 0 in Multimedia Appendix 5 and Figure 2). The HADS-D score was significantly lower in both groups at week 60 compared with the score at the baseline visit (interaction week 60: coefficient=0.23, *P*=.03; Model 0 in Multimedia Appendix 5).

Figure 2. Significant improvement of the HADS-D score in patients with psoriasis by combining an educational program with a psoriasis app over 24 weeks. HADS-D: Hospital Anxiety and Depression Scale–Depression. *$P \leq .05$, **$P \leq .01$.



The improvement found in the HADS-A in all patients until week 24 also did not persist in weeks 36 and 60 (interaction week: week 12: coefficient=–0.194, *P*=.02; week 24: coefficient=–0.221, *P*=.01; week 36: coefficient=–0.16, *P*=.08; week 60: coefficient=–0.17, *P*=.06; Model 0 in Multimedia Appendix 5). A tendency for a stronger reduction of the HADS-A in the intervention group compared with the control group was assessed at week 60 (interaction week x intervention: week 60: coefficient=–0.24, *P*=.06; Model 0 in Multimedia Appendix 5; Figure 3).

Figure 3. Effects of an educational program combined with a disease management psoriasis app on the HADS-D score over 60 weeks. HADS-D: Hospital Anxiety and Depression Scale–Depression.



A significant reduction in the DLQI was observed at weeks 24 and 60 in all patients but not in weeks 12 and 36 (interaction week: week 12: coefficient=–0.24, *P*=.08; week 24: coefficient=–0.37, *P*=.007; week 36: coefficient=–0.26, *P*=.07; week 60: coefficient=–0.56, *P*<.001; Model 0 in Multimedia Appendix 5). There were no significant differences assessed between the study groups (Model 0 in Multimedia Appendix 5).

In both study groups, the mood of the patients with psoriasis was ameliorated. Compared with scores in the baseline visit, significantly lower scores on the Likert scale were found in all the following visits in all patients (interaction week: week 12: coefficient=–1.80, *P*=.006; week 24: coefficient=–2.12, *P*=.001; week 36: coefficient=–1.83, *P*=.009; week 60: coefficient=–2.64, *P*<.001; Model 0 in Table 3). However, group membership had no significant effect (Model 0 in Table 3).

Table 3. Characteristics of the study cohort divided by app use frequency of more or less than 20% at week 60 (N=38).[a]

Characteristics week 60	Overall	App use frequency <20% (n=15)	App use frequency ≥20% (n=18)	*P* value
Sex, n (%)				
Female	15 (39)	5 (33)	9 (50)	.50[b]
Male	23 (61)	10 (67)	9 (50)	.50[b]
Age (years)				
Mean (SD)	46.9 (12.3)	45.0 (14.1)	47.7 (10.8)	.99[c]
Median (IQR)	48.0 (39-54)	47.0 (28-53)	49.0 (40-54)	.99[c]
BMI[d] (kg/m^2)				
Mean (SD)	27.6 (4.7)	27.5 (5.1)	27.7 (4.4)	.99[c]
Median (IQR)	26.0 (24.2-32.5)	27.5 (22.5-32.5)	26.0 (24.8-32.7)	.99[c]
Alcohol (days/week)				
Mean (SD)	1.3 (1.7)	1.5 (1.3)	1.2 (1.8)	.99[c]
Median (IQR)	1.0 (0-2.0)	1.0 (1-2.0)	1.0 (0-1)	.99[c]
Smoker, n (%)	10 (26.3)	2 (13.3)	5 (27.8)	.13[b]
Employed, n (%)				
Yes	34 (89)	13 (87)	17 (94)	.15[b]
No	3 (8)	1 (7)	1 (6)	.15[b]
Retired	1 (3)	1 (7)	0 (0)	.15[b]
Graduation level, n (%)				
No final school examination	1 (3)	0 (0)	1 (6)	.23[b]
Final school examination after 9 years[e]	5 (13)	1 (7)	3 (17)	.23[b]
Final school examination after 10 years[f]	9 (24)	4 (27)	4 (22)	.23[b]
Vocational diploma[g]	4 (11)	1 (7)	2 (11)	.23[b]
General qualification for university entrance[h]	5 (13)	2 (13)	2 (11)	.23[b]
University degree	14 (37)	7 (47)	6 (33)	.23[b]
Duration of psoriasis (years)				
Mean (SD)	17.1 (13.1)	19.7 (14.3)	14.3 (12.1)	.91[c]
Median (IQR)	13.5 (6-24)	23.0 (6-28)	9.0 (6-20)	.91[c]
Psoriatic arthritis, n (%)	15 (39.5)	8 (53.3)	6 (35.3)	.49[b]
Antipsoriatic therapy, n (%)				
Topical or UV therapy	11 (29)	2 (13)	7 (39)	.12[b]
DMARDS[i]	4 (11)	2 (13)	0 (0)	.12[b]
Others	2 (5)	1 (7)	1 (6)	.12[b]
Biologicals	21 (55)	10 (67)	10 (56)	.12[b]
No therapy	0 (0)	0 (0)	0 (0)	.12[b]
HADS-D[j] (range 0-21)				
Mean (SD)	4.0 (4.5)	2.3 (3.4)	5.4 (5.2)	.15[c]

Characteristics week 60	Overall	App use frequency <20% (n=15)	App use frequency ≥20% (n=18)	*P* value
Median (IQR)	2.0 (0-7)	1.0 (0-3)	4.0 (1-9)	.15[c]
HADS-A[k] (range 0-21)				
Mean (SD)	4.9 (3.5)	4.2 (3.1)	5.6 (3.5)	.83[c]
Median (IQR)	4.0 (2-7)	4.0 (1-6)	5.0 (3-7)	.83[c]
DLQI[l] (range 0-30)				
Mean (SD)	4.4 (5.5)	3.4 (4.9)	5.4 (6.2)	.70[c]
Median (IQR)	1.0 (1-8)	1.0 (1-6)	3.0 (1-9)	.70[c]
Mood (range 0-3)				
Mean (SD)	0.6 (0.8)	0.5 (0.6)	0.8 (0.8)	.61[c]
Median (IQR)	0.5 (0-1)	0.0 (0-1)	1.0 (0-1)	.61[c]
Daily activity (range 0-30)				
Mean (SD)	0.5 (0.7)	0.4 (0.6)	0.7 (0.8)	.63[c]
Median (IQR)	0.0 (0-1)	0.0 (0-1)	1.0 (0-1)	.63[c]
PASI[m] (range 0-72)				
Mean (SD)	2.7 (3.3)	2.3 (3.7)	2.5 (2.6)	.99[c]
Median (IQR)	1.4 (0.6-3.4)	1.2 (0.2-1.5)	1.85 (0.6-3.4)	.99[c]
Pain (range 0-10)[d]				
Mean (SD)	1.0 (2.0)	0.8 (1.2)	1.6 (2.6)	.99[c]
Median (IQR)	0.0 (0-1)	0.0 (0-2)	0.5 (0-2)	.99[c]
Pruritus (range 0-10)				
Mean (SD)	1.8 (2.2)	2.1 (2.7)	1.9 (2.0)	.99[c]
Median (IQR)	1.0 (0-3)	1.0 (0-3)	1.5 (0-3)	.99[c]

[a]Data for sex, age, BMI, smoking, alcohol consumption, employment status, graduation level, psoriasis duration, psoriasis arthritis, and therapy were collected at week 0 only.

[b]Categorical variables were analyzed using the chi-square test.

[c]Continuous variables were analyzed using the Bonferroni-test.

[d]For 1 patient, data for weight and height were missing, and for 1 patient, data for pain was missing at week 60. In the intervention group, 5 patients did not download the app at all. Therefore, they could not be divided into the app use frequency subgroups.

[e]Lowest Certificate of Secondary Education in Germany.

[f]Roughly equal to the General Certificate of Secondary Education (no university entrance qualification).

[g]Final school examination after 12 years in combination with vocational education (university entrance qualification for specific degrees only).

[h]Final school examination after 13 years.

[i]DMARD: disease-modifying antirheumatic drug.

[j]HADS-D: Hospital Anxiety and Depression Scale–Depression.

[k]HADS-A: Hospital Anxiety and Depression Scale–Anxiety.

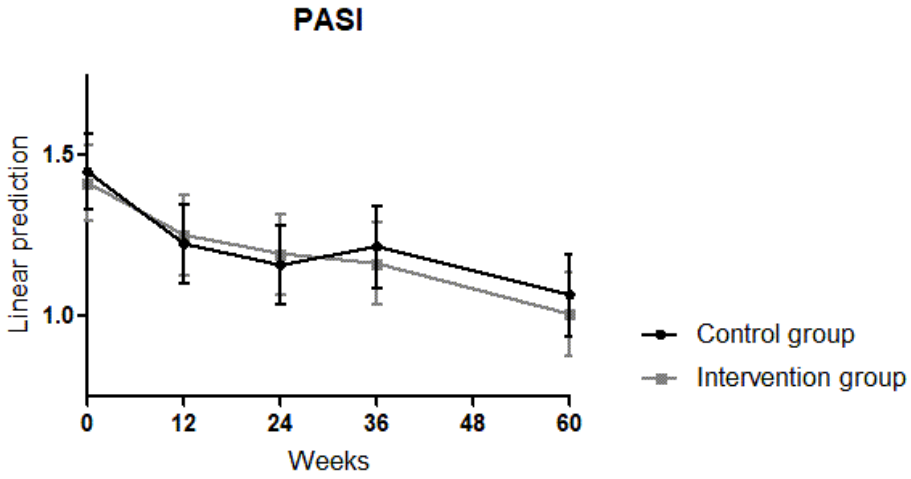
[l]DLQI: Dermatology Life Quality Index.

[m]PASI: Psoriasis Area and Severity Index.

Likewise, a significant reduction in the impairment of daily activities on the Likert scale was found in all follow-up visits compared with baseline visits in both groups, with no significant difference between the study groups (interaction week: week 12: coefficient=–0.51, *P*<.001; week 24: coefficient=–0.53, *P*<.001; week 36: coefficient=–0.55, *P*<.001; week 60: coefficient=–0.68, *P*<.001; Model 0 in Table 3).

There was also a significant reduction in the PASI scores observed over the 60 weeks compared with those at week 0 in all patients (interaction week: week 12: coefficient=–0.24, *P*=.04; week 24: coefficient=–0.29, *P*=.008; week 26: coefficient=–0.23, *P*=.04; week 60: coefficient=–0.38, *P*=.001; Model 0 in Table 3). Group membership had no significant effect (Model 0 in Table 3; Figure 4).

Figure 4. Effects of an educational program combined with a disease management psoriasis app on the PASI score over 60 weeks. PASI: Psoriasis Area and Severity Index.



A significant reduction in pruritus was assessed in the control and intervention groups at week 60 compared with that at baseline, with no significant difference between the 2 groups (interaction week: week 60: coefficient=–0.97, *P*=.02; Model 0 in Table 3).

No effects were found on the numerical rating scale for skin pain (Model 0 in Table 3).

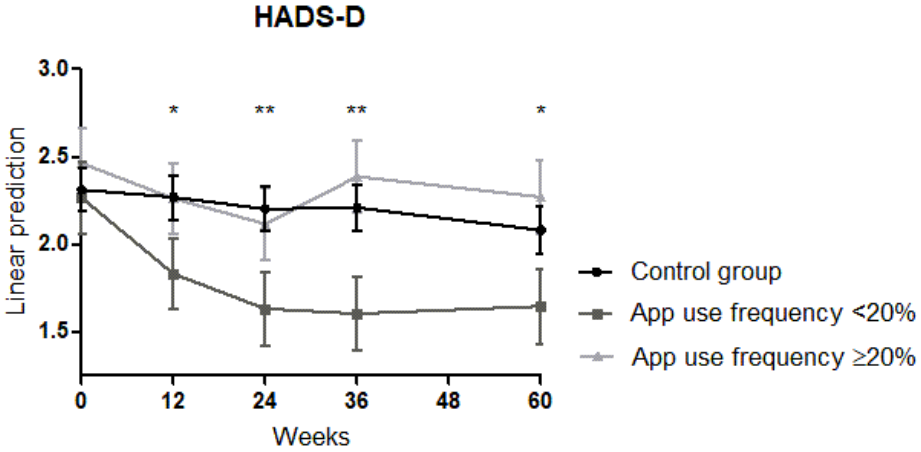
These findings did not differ if adjusted for sex, age, and disease duration (matching coefficients and *P* values under Model 1 in Table 3).

App Use Frequency Subgroup Analysis

The intervention group was divided into 2 groups, with patients using the app more or less frequently than 20% (equals one-time use every 5 weeks). Both groups showed similar characteristics (Table 3). Neither demographic nor socioeconomic or disease-related characteristics had any influence on app use frequency.

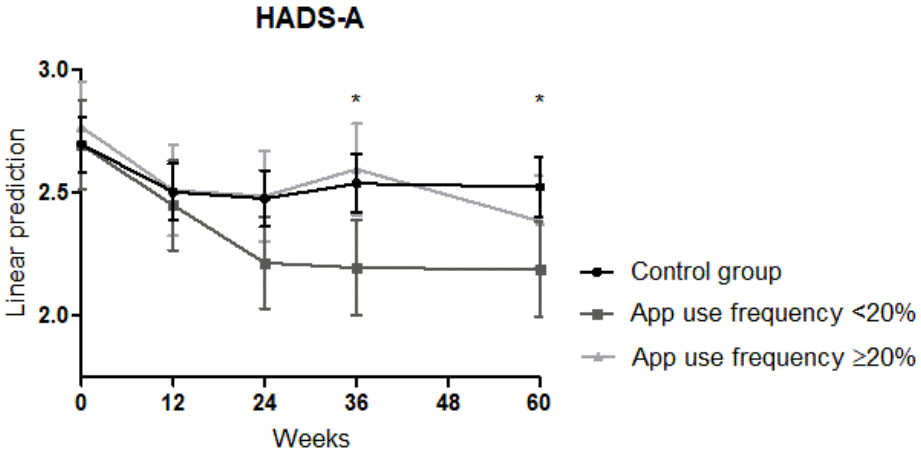
The significant reduction in the HADS-D at weeks 12 and 24 persisted until week 60 in patients, with an app use frequency <20% compared with that of the control group (interaction week x <20%: week 12: coefficient=–0.39, *P*=.03; week 24: coefficient=–0.53, *P*=.004; week 36: coefficient=–0.56, *P*=.004; week 60: coefficient=–0.39, *P*=.04; Model 0 in Multimedia Appendix 6; Figure 5).

Figure 5. Significant improvement of the HADS-D score in patients with psoriasis using the psoriasis app less than once every 5 weeks (<20%) compared with the control group over 60 weeks. HADS-D: Hospital Anxiety and Depression Scale–Depression. **P*≤.05, ***P*≤.01.



Furthermore, patients with an app use frequency of <20% showed a significant reduction in the HADS-A at weeks 36 and 60 (interaction week x <20%: week 36: coefficient=–0.34; *P*=.04; week 60: coefficient=–0.33; *P*=.05; Model 0 in Multimedia Appendix 6; Figure 6).

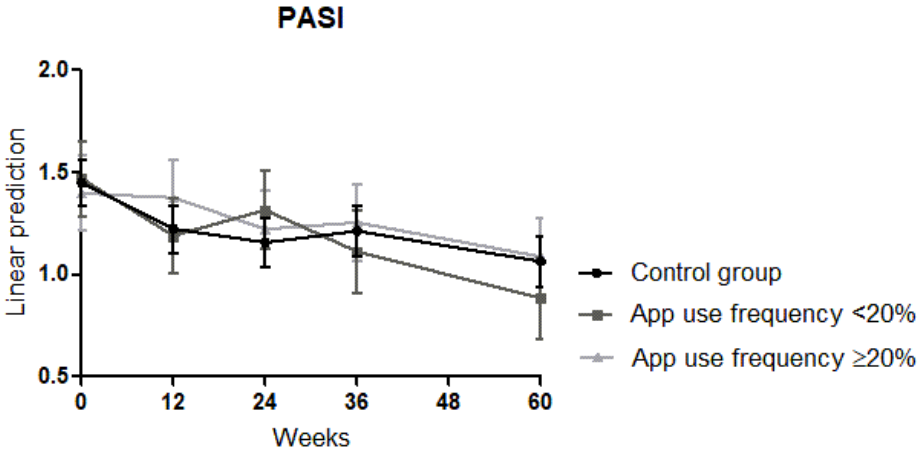
Figure 6. Significant improvement of the HADS-D score in patients with psoriasis using the psoriasis app less than once every 5 weeks (<20%) compared with the control group over 60 weeks. HADS-D: Hospital Anxiety and Depression Scale–Depression. *$P \leq .05$.



In patients using the app >20%, no significant reduction in the HADS-D/-A was observed over the 60-week study period (Model 0 in Multimedia Appendix 6; Figures 5 and 6).

For the DLQI, mood, daily activity, PASI, and skin pruritus and pain, no differences in the subgroup analysis were found compared with the analysis of the whole group (Model 0 in Multimedia Appendix 6; PASI in Figure 7). These results were independent of sex, age, and disease duration (matching coefficients and *P* values in Model in Multimedia Appendix 6).

Figure 7. Effects of an educational program combined with a disease management psoriasis app on the PASI score of the subgroups using the psoriasis app more (≥20%) or less (<20%) than once every 5 weeks. PASI: Psoriasis Area and Severity Index.



Discussion

Principal Findings

In this study, we analyzed the effects of an educational program combined with an eHealth smartphone app on the clinical outcomes of patients with psoriasis. We were able to show that this intervention can significantly assist in improving depression and anxiety in patients with psoriasis. Educational programs have proven to be effective in increasing knowledge, quality of life, and self-efficacy in patients with psoriasis in the past [11-14]. To analyze the additional benefits of a disease management app, we tested the educational program applied in this study in a pilot trial first. The trial showed that the educational program per se could lead to a significant improvement in knowledge and self-expertise about the disease and to an amelioration of general health but not to an improvement in patients' mental health [10]. Therefore, we are convinced that the smartphone app used in this study provides an additional psychological benefit. Improvements in quality of life were also reported by Armstrong et al [23] in an American randomized controlled equivalency trial, in which 296 patients were randomly assigned to either a web-based or in-person care group. eHealth devices such as the one used in this study can probably increase the sense of security in patients by giving them the opportunity to always contact the treating physician using the chat function. In addition, the app might have assisted in building a long-term trusting relationship with the doctor, allowing patients to feel more comfortable with the questions, concerns, and fears about their disease. Furthermore, the combined intervention might have led to a better self-perception of patients' mental health and disease status by the regular self-report of life quality and mood as well as photodocumentation of the skin. This could have given patients

a more structured overview of their disease and thus a greater sense of control. A study by Blome et al [24] showed that greater control is stated as a relevant treatment goal by 92.3% of patients with psoriasis. The fulfillment of this goal can lead to a higher level of satisfaction, which has the potential to reduce depression.

Others have noted that some effects achieved by educational programs vanish after about 6 months [11,13]. In accordance with this, we also observed a significant reduction in the HADS-D score in the intervention group only in the first 6 months. Further studies are needed to determine whether it is possible to lengthen this effect with another educational program after 6 months or by providing regular educational information via the app. Interestingly, a stronger reduction in depression and anxiety levels persisting over the entire 60-week study period was seen in patients using the app less than once every 5 weeks. More frequent app use did not worsen mental health, and there was no further benefit compared with the control group. In accordance with the findings of Ancker et al [25] and Seppen et al [26], we assume that chronically ill patients such as patients with psoriasis do not want to be reminded about their disease too often. New and highly efficient psoriasis treatments such as biologics often lead to lesion-free skin, helping patients feel that they are not sick anymore. Too frequent app use could counteract this effect. We are convinced that an optimal app use frequency needs to be identified for patients. Our data indicate that an app use frequency of approximately once every 5 weeks could be the most beneficial; however, more scientific data are needed.

Web-based care models reduce health care system–related costs and can improve patients' lives by decreasing the number of clinic visits [23,27,28]. On the patients' side, this means less absence at work and saved time. Therefore, the implementation of scientifically validated eHealth devices for chronic diseases such as psoriasis seems to be favorable. It will be important to determine how to encourage patients to engage in continuous app use over a long period and whether telemedicine can replace in-person visits or simply support patients' care [19-21,24].

Limitations

The major limitations of our study are the monocentric design, small study cohort, and limited generalizability of the results. In particular, the number of patients in the subgroups divided by app use frequency was quite low, which could have led to missed or overinterpreted differences between the groups. In addition, the age of our patients was slightly higher than that in other psoriasis studies and differed between the control and intervention groups. The initial PASI score was lower than in other studies, and the percentage of patients treated with systemic therapy was quite high, which could have led to undetected effects. Further studies are necessary to verify our findings on a broader scale and in a multicenter setting.

Conclusions

In conclusion, the educational program combined with the psoriasis app had a positive impact on the mental health of patients with psoriasis if not used too frequently. The provision of valid and comprehensible knowledge about their own disease and web-based support has the potential to improve health care for patients with psoriasis. Therefore, educating and supporting patients with psoriasis using digital health devices seems to be a promising additional component in the disease management of psoriasis in the long run.

Acknowledgments

The authors are thankful for the financial support of Novartis GmbH for implementing this study and the support of the physicians and nursing staff of the Department of Dermatology of the University Medical Center Mannheim. All named authors and Novartis GmbH had full access to the study data and approved the manuscript. The patients in this study have provided written informed consent for the publication of their case details.

Conflicts of Interest

LD, AB, and TSH obtained financial compensation for poster presentations at congresses from Novartis GmbH. RH declares no conflicts of interest. AS conducted clinical trials for AbbVie, Boehringer-Ingelheim, Celgene, Eli Lilly, Janssen-Cilag, LEO Pharma, Merck, Novartis GmbH, and Pfizer; is a consultant for LEO Pharma; and received financial support from Janssen-Cilag and Novartis GmbH and support for conferences from AbbVie, Janssen-Cilag, Novartis GmbH, and Pfizer. AS and JB are the CEOs and owners of Derma Intelligence GmbH, which programmed DermaScope Mobile.

Multimedia Appendix 1
Content overview of the psoriasis educational program.
[PNG File , 336 KB - mhealth_v9i10e28149_app1.png]

Multimedia Appendix 2
App design of the study app DermaScope Mobile.
[PNG File , 755 KB - mhealth_v9i10e28149_app2.png]

Multimedia Appendix 3
Clinic dashboard of the study app DermaScope Mobile.

[PNG File , 1325 KB - mhealth_v9i10e28149_app3.png]

Multimedia Appendix 4
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials eHealth) checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1201 KB - mhealth_v9i10e28149_app4.pdf]

Multimedia Appendix 5
Random effect regression models over 60 weeks; model 0, unadjusted, and model 1, adjusted for age, sex, and disease duration.
[DOCX File , 33 KB - mhealth_v9i10e28149_app5.docx]

Multimedia Appendix 6
Random effect regression models of the app use frequency subgroups <20% and ≥20% over 60 weeks (N=93; observed=411).
[DOCX File , 20 KB - mhealth_v9i10e28149_app6.docx]

References

1. Schäfer I, Rustenbach SJ, Radtke M, Augustin J, Glaeske G, Augustin M. [Epidemiology of psoriasis in Germany--analysis of secondary health insurance data]. Gesundheitswesen 2011 May;73(5):308-313. [doi: 10.1055/s-0030-1252022] [Medline: 20544588]
2. Augustin M, Krüger K, Radtke MA, Schwippl I, Reich K. Disease severity, quality of life and health care in plaque-type psoriasis: a multicenter cross-sectional study in germany. Dermatol 2008;216(4):366-372. [doi: 10.1159/000119415] [Medline: 18319601]
3. Rapp SR, Feldman SR, Exum ML, Fleischer AB, Reboussin DM. Psoriasis causes as much disability as other major medical diseases. J Am Acad Dermatol 1999 Sep;41(3):401-407. [doi: 10.1016/s0190-9622(99)70112-x] [Medline: 10459113]
4. Kurd SK, Troxel AB, Crits-Christoph P, Gelfand JM. The risk of depression, anxiety, and suicidality in patients with psoriasis: a population-based cohort study. Arch Dermatol 2010 Aug;146(8):891-895 [FREE Full text] [doi: 10.1001/archdermatol.2010.186] [Medline: 20713823]
5. Ljosaa TM, Mork C, Stubhaug A, Moum T, Wahl AK. Skin pain and skin discomfort is associated with quality of life in patients with psoriasis. J Eur Acad Dermatol Venereol 2012 Jan;26(1):29-35. [doi: 10.1111/j.1468-3083.2011.04000.x] [Medline: 21385220]
6. Schmid-Ott G, Schallmayer S, Calliess IT. Quality of life in patients with psoriasis and psoriasis arthritis with a special focus on stigmatization experience. Clin Dermatol 2007;25(6):547-554. [doi: 10.1016/j.clindermatol.2007.08.008] [Medline: 18021891]
7. Renzi C, Di Pietro C, Tabolli S. Participation, satisfaction and knowledge level of patients with cutaneous psoriasis or psoriatic arthritis. Clin Exp Dermatol 2011 Dec;36(8):885-888. [doi: 10.1111/j.1365-2230.2011.04126.x] [Medline: 21790723]
8. Larsen MH, Hagen KB, Krogstad A, Aas E, Wahl AK. Limited evidence of the effects of patient education and self-management interventions in psoriasis patients: a systematic review. Patient Educ Couns 2014 Feb;94(2):158-169. [doi: 10.1016/j.pec.2013.10.005] [Medline: 24184041]
9. Dressler C, Lambert J, Grine L, Galdas P, Paul C, Zidane M, et al. Therapeutic patient education and self-management support for patients with psoriasis - a systematic review. J Dtsch Dermatol Ges 2019 Jul;17(7):685-695. [doi: 10.1111/ddg.13840] [Medline: 31012994]
10. Bubak C, Schaarschmidt M, Schöben L, Peitsch WK, Schmieder A. Analyzing the value of an educational program for psoriasis patients: a prospective controlled pilot study. BMC Public Health 2019 Nov 19;19(1):1535 [FREE Full text] [doi: 10.1186/s12889-019-7778-x] [Medline: 31744488]
11. Bostoen J, Bracke S, De Keyser S, Lambert J. An educational programme for patients with psoriasis and atopic dermatitis: a prospective randomized controlled trial. Br J Dermatol 2012 Nov;167(5):1025-1031. [doi: 10.1111/j.1365-2133.2012.11113.x] [Medline: 22709422]
12. Balato N, Megna M, Di Costanzo L, Balato A, Ayala F. Educational and motivational support service: a pilot study for mobile-phone-based interventions in patients with psoriasis. Br J Dermatol 2013 Jan;168(1):201-205. [doi: 10.1111/j.1365-2133.2012.11205.x] [Medline: 23240729]
13. Lora V, Gisondi P, Calza A, Zanoni M, Girolomoni G. Efficacy of a single educative intervention in patients with chronic plaque psoriasis. Dermatol 2009;219(4):316-321. [doi: 10.1159/000250826] [Medline: 19844077]
14. Larsen MH, Krogstad AL, Aas E, Moum T, Wahl AK. A telephone-based motivational interviewing intervention has positive effects on psoriasis severity and self-management: a randomized controlled trial. Br J Dermatol 2014 Dec;171(6):1458-1469. [doi: 10.1111/bjd.13363] [Medline: 25143061]
15. Bonini M. Electronic health (e-health): emerging role in asthma. Curr Opin Pulm Med 2017 Jan;23(1):21-26. [doi: 10.1097/MCP.0000000000000336] [Medline: 27763999]

16. Giordano A, Scalvini S, Zanelli E, Corrà U, Longobardi GL, Ricci VA, et al. Multicenter randomised trial on home-based telemanagement to prevent hospital readmission of patients with chronic heart failure. Int J Cardiol 2009 Jan 09;131(2):192-199. [doi: 10.1016/j.ijcard.2007.10.027] [Medline: 18222552]
17. Harno K, Kauppinen-Mäkelin R, Syrjäläinen J. Managing diabetes care using an integrated regional e-health approach. J Telemed Telecare 2006;12 Suppl 1:13-15. [doi: 10.1258/135763306777978380] [Medline: 16884565]
18. McManus RJ, Mant J, Bray EP, Holder R, Jones MI, Greenfield S, et al. Telemonitoring and self-management in the control of hypertension (tasminh2): a randomised controlled trial. Lancet 2010 Jul 17;376(9736):163-172. [doi: 10.1016/S0140-6736(10)60964-6] [Medline: 20619448]
19. Chambers CJ, Parsi KK, Schupp C, Armstrong AW. Patient-centered online management of psoriasis: a randomized controlled equivalency trial. J Am Acad Dermatol 2012 Jun;66(6):948-953 [FREE Full text] [doi: 10.1016/j.jaad.2011.05.047] [Medline: 21890236]
20. Armstrong AW, Chambers CJ, Maverakis E, Cheng MY, Dunnick CA, Chren M, et al. Effectiveness of online vs in-person care for adults with psoriasis: a randomized clinical trial. JAMA Netw Open 2018 Oct 05;1(6):e183062 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3062] [Medline: 30646223]
21. Armstrong AW, Ford AR, Chambers CJ, Maverakis E, Dunnick CA, Chren M, et al. Online care versus in-person care for improving quality of life in psoriasis: a randomized controlled equivalency trial. J Invest Dermatol 2019 May;139(5):1037-1044 [FREE Full text] [doi: 10.1016/j.jid.2018.09.039] [Medline: 30481495]
22. Beck A, Schulze-Hagen T, Domogalla L, Herr R, Benecke J, Schmieder A. Effect of a disease-monitoring smartphone application in combination with a patient educational program on mental health of patients with psoriasis: a randomized intervention study. J Am Acad Dermatol 2020 Dec 25;S0190-9622(20):33246. [doi: 10.1016/j.jaad.2020.12.048] [Medline: 33359785]
23. Parsi K, Chambers CJ, Armstrong AW. Cost-effectiveness analysis of a patient-centered care model for management of psoriasis. J Am Acad Dermatol 2012 Apr;66(4):563-570. [doi: 10.1016/j.jaad.2011.02.022] [Medline: 21835497]
24. Young PM, Chen AY, Ford AR, Cheng MY, Lane CJ, Armstrong AW. Effects of online care on functional and psychological outcomes in patients with psoriasis: a randomized controlled trial. J Am Acad Dermatol 2019 Jun 05;S0190-9622(19):30893. [doi: 10.1016/j.jaad.2019.05.089] [Medline: 31175908]
25. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug 19;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
26. Seppen BF, Wiegel J, L'ami MJ, Rico SD, Catarinella FS, Turkstra F, et al. Feasibility of self-monitoring rheumatoid arthritis with a smartphone app: results of two mixed-methods pilot studies. JMIR Form Res 2020 Sep 21;4(9):e20165 [FREE Full text] [doi: 10.2196/20165] [Medline: 32955447]
27. Ford AR, Gibbons CM, Torres J, Kornmehl HA, Singh S, Young PM, et al. Access to dermatological care with an innovative online model for psoriasis management: results from a randomized controlled trial. Telemed J E Health 2019 Jul;25(7):619-627 [FREE Full text] [doi: 10.1089/tmj.2018.0160] [Medline: 30222518]
28. Oostveen AM, Beulens CA, van de Kerkhof PC, de Jong EM, Seyger MM. The effectiveness and safety of short-contact dithranol therapy in paediatric psoriasis: a prospective comparison of regular day care and day care with telemedicine. Br J Dermatol 2014 Feb;170(2):454-457. [doi: 10.1111/bjd.12621] [Medline: 24024678]

Abbreviations

DLQI: Dermatology Life Quality Index
HADS-A: Hospital Anxiety and Depression Scale–Anxiety
HADS-D: Hospital Anxiety and Depression Scale–Depression
PASI: Psoriasis Area and Severity Index

Edited by L Buis; submitted 23.02.21; peer-reviewed by C Liebram, HL Tam; comments to author 13.05.21; revised version received 27.06.21; accepted 22.07.21; published 25.10.21.

Please cite as:
Domogalla L, Beck A, Schulze-Hagen T, Herr R, Benecke J, Schmieder A
Impact of an eHealth Smartphone App on the Mental Health of Patients With Psoriasis: Prospective Randomized Controlled Intervention Study
JMIR Mhealth Uhealth 2021;9(10):e28149
URL: https://mhealth.jmir.org/2021/10/e28149
doi:10.2196/28149
PMID:34431478

Original Paper

A Mobile App (FallSA) to Identify Fall Risk Among Malaysian Community-Dwelling Older Persons: Development and Validation Study

Devinder Kaur Ajit Singh[1*], PhD; Jing Wen Goh[1*], BSc; Muhammad Iqbal Shaharudin[1,2*], MSc; Suzana Shahar[1*], PhD

[1]Center for Healthy Ageing and Wellness, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[2]Faculty of Health Sciences, Cawangan Pulau Pinang, Kampus Bertam, Universiti Teknologi Majlis Amanah Rakyat, Penang, Malaysia

[*]all authors contributed equally

Corresponding Author:
Devinder Kaur Ajit Singh, PhD
Center for Healthy Ageing and Wellness
Faculty of Health Sciences
Universiti Kebangsaan Malaysia
Jalan Raja Muda Abdul Aziz
Kuala Lumpur, 50300
Malaysia
Phone: 60 392897000
Email: devinder@ukm.edu.my

Abstract

Background: Recent falls prevention guidelines recommend early routine fall risk assessment among older persons.

Objective: The purpose of this study was to develop a Falls Screening Mobile App (FallSA), determine its acceptance, concurrent validity, test-retest reliability, discriminative ability, and predictive validity as a self-screening tool to identify fall risk among Malaysian older persons.

Methods: FallSA acceptance was tested among 15 participants (mean age 65.93 [SD 7.42] years); its validity and reliability among 91 participants (mean age 67.34 [SD 5.97] years); discriminative ability and predictive validity among 610 participants (mean age 71.78 [SD 4.70] years). Acceptance of FallSA was assessed using a questionnaire, and it was validated against a comprehensive fall risk assessment tool, the Physiological Profile Assessment (PPA). Participants used FallSA to test their fall risk repeatedly twice within an hour. Its discriminative ability and predictive validity were determined by comparing participant fall risk scores between fallers and nonfallers and prospectively through a 6-month follow-up, respectively.

Results: The findings of our study showed that FallSA had a high acceptance level with 80% (12/15) of older persons agreeing on its suitability as a falls self-screening tool. Concurrent validity test demonstrated a significant moderate correlation (r=.518, P<.001) and agreement (k=.516, P<.001) with acceptable sensitivity (80.4%) and specificity (71.1%). FallSA also had good reliability (intraclass correlation .948; 95% CI .921-.966) and an internal consistency (α=.948, P<.001). FallSA score demonstrated a moderate to strong discriminative ability in classifying fallers and nonfallers. FallSA had a predictive validity of falls with positive likelihood ratio of 2.27, pooled sensitivity of 82% and specificity of 64%, and area under the curve of 0.802.

Conclusions: These results suggest that FallSA is a valid and reliable fall risk self-screening tool. Further studies are required to empower and engage older persons or care givers in the use of FallSA to self-screen for falls and thereafter to seek early prevention intervention.

(JMIR Mhealth Uhealth 2021;9(10):e23663) doi:10.2196/23663

KEYWORDS

fall risk; self-screening; mobile app; older person

Introduction

Falls among older persons are a major health and socioeconomic concern globally [1]. Fall prevalence ranged from 4.2% to 61% in Malaysian older persons in 2018 [2]. This range is similar to other older Asians in Japan and United Arab Emirates (18% to 57%) [3,4]. This could possibly stem from the similar research methodology in which history of falls in the past 12 months is commonly used in studies. In our earlier large-scale population-based longitudinal study Long-Term Research Grant Scheme—Toward Useful Ageing (LRGS TUA), we reported a retrospective and prospective fall prevalence of 15% to 18% and 27%, respectively, among Malaysian community-dwelling older persons [5-7]. Single and repeated fall incidence rates were 8.47 and 3.21 per 100 person-years, respectively [8]. In our local context, community-dwelling older persons are defined as older adults aged 60 years and above who are living independently in the community [9]. Identified risk factors associated with falls were arthritis, diabetes, urinary incontinence, decreased handgrip strength, higher BMI, and poor self-rated health [7], while the predictors consisted of a history of falls and decreased muscle strength for both occasional and recurrent falls [8].

Falls comprise multifactorial etiology, and they occur as a result of interactions between multiple intrinsic and extrinsic factors. Current evidence-based practice guidelines recommend early falls screening using a multifactorial fall risk assessment tool [1]. The combination of sociodemographic factors (gender, joint pain and cataract/glaucoma), self-rated multifactorial questionnaire (previous fall history and worrying about falls), and a physical performance test (Timed Up and Go [TUG]) in a fall risk model was proposed to identify fall risk among Malaysian community-dwelling older persons [6]. The combination of multifactorial evidence-based assessments is expected to be more robust compared to a single test [10]. There may be a potential to use an app-based fall risk multifactorial assessment.

Currently, there are several app-based fall risk assessment tools, including Lindera, Steady, and Aachen fall prevention apps. Lindera is a smartphone-based app designed to facilitate the health care professional (nursing staff) to perform a structured multifactorial fall risk assessment among older adults [11]. Lindera consists of a mobility test (TUG test) and fall risk–related questionnaire. It was found to have a pooled sensitivity of 93% and specificity of 58%, with an overall accuracy of 73% [11]. In Steady, 5 progressively more challenging mobility tests (30s balance and sit to stand tests) and a medical history questionnaire were used to assess individual fall risk [12]. Steady was found to be valid and reliable in facilitating fall risk self-screening among older persons in home settings [13]. The Aachen fall prevention app has a combination of a balance test with simple questionnaire to improve fall risk awareness among older persons [14]. It has a sensitivity and specificity of 57% and 76.7%, respectively [14]. While Aachen and Steady were developed as fall self-screening tools, Lindera is used to assist health care professionals (nursing staff) in conducting multifactorial fall risk assessment among older persons.

Although multifactorial app-based fall risk assessment tools are currently available, they lack comprehensive information regarding their reliability and validity as self-assessment tools. Prior to the use of fall risk apps, their diagnostic accuracy in discriminating between fallers and nonfallers and predicting actual falls among at-risk older persons must be demonstrated. Existing fall risk assessment tools in clinical settings were found to have only moderate diagnostic accuracy [15]. Moreover, there is a need to develop a culturally specific fall risk assessment tool for the multiethnic Malaysian older population. Therefore, in this study, we aimed to develop the Fall Risk Screening App (FallSA) as a self-screening tool for assessing fall risk among Malaysian older persons using the combination of questionnaires and physical tests. Thereafter, we determined its acceptance, validity, reliability, discriminative ability, and predictive validity. We hope that with the early self-screening tool FallSA we are able to empower and engage older persons or their caregivers to be aware of falls and adopt fall prevention behavior.

Methods

Study Design

This study regarding FallSA was divided into 4 phases comprising (1) development, (2) acceptance among older persons, (3) concurrent validity and test-retest reliability, and (4) discriminative ability and predictive validity.

Ethical approval was obtained from the secretariat for research of ethics of Universiti Kebangsaan Malaysia (UKM 1.5.3.5/244/NN-060-2013 and UKM PPI/111/8/JEP-2018-559). Prior to all studies, participants were given information about the study and were required to provide informed verbal consent.

Phase 1: Development

Prior to the development of FallSA, a literature review was conducted to identify current gaps in the literature regarding fall risk mobile screening apps. Next, several group discussions and meetings were conducted to identify the intended features, interface, and design to meet the functional needs of older persons. After which, preparation of the proposed product features and design using international guidelines for ease of use and usability of graphic user interface for older persons (ISO/IEC 2001) was done.

FallSA was developed based on a model established from our team's large-scale study report (Neuroprotective Model for Healthy Longevity), which was designed to evaluate the magnitude of cognitive decline and its risk factors through comprehensive multifactorial assessment [16]. The significant predictive fall risk factors were used for the fall risk calculations in FallSA [5]. This fall risk model included the combination of sociodemographic information (gender, joint pain, and cataract/glaucoma), self-rated multifactorial questionnaire (previous fall history and worrying about falls), and physical performance test (TUG test) [5]. TUG normative values from the study by Ibrahim et al [5] were used to generate graphs in comparing user TUG test with 50th population norms-based age groups and gender. An instructional video for TUG test performance was provided within the FallSA app. The

step-by-step procedure to perform the TUG self-test was based on the original version by Podsiadlo and Richardson [17]. Several modifications were done later to fulfill the self-screening feature of a mobile app. The designs of the icons were obtained from the readily available Google Advanced Image Search. While the icons were derived from an online source of free icons [18] and made readily available under the terms of the end user license agreement, their selection and modification were based on international guidelines (ISO/IEC Guide 71:2001 [E]). The icons were modified using Microsoft Paint, Microsoft Photos, and the Microsoft Snipping Tool. With respect to the guidelines, all text was written in black on a white-colored background to achieve maximum contrast. An adequate icon size was ensured for easy system navigation and indication of the current interface.

The input button, navigation, and arrangements were developed following the same guidelines. The navigation buttons of FallSA were structured in a rectangle with some simple instruction text at the bottom of the interface; it is displayed as an icon shape on the current user section. In addition, colored navigation buttons were employed in questions sections with a selection of yes in green and no in red. In order to increase ease of use among older persons, a simple navigation system flow was used by having only 2 main selections: new user and current user. The app features for wording, questions, and instructions were bilingual in English and Malay languages, which are commonly used among Malaysian population.

Prior to the use of FallSA, participants were shown an informed consent on data protection policy at the beginning of the app interface. Personal information or data were only saved in the user's device and could only be assessed by the user and researcher. It is also noteworthy that the information stored in FallSA will be handled in a similar way to hard copies of medical records, as declared in the Data Protection Disclaimer and Laws of Malaysia Act 709 and Personal Data Protection Act 2010. Since FallSA is a self-screening mobile app for identifying fall risk and it does not provide any diagnosis or suggestion to change current medical dosages, it is not categorized as a medical device, which would require an approval from health authorities.

FallSA was developed by a freelance mobile software developer with more than 3 years' experience in a related field, based on the proposed product features and design of the researcher team. The development of FallSA was conducted according to the waterfall software development process using agile principles for better quality, time, and cost-effective mobile app development.

Phase 2: User Acceptance Testing

FallSA user acceptance testing was tested in the real world by 15 older persons.

Participants

A convenience sampling method was used to recruit the participants from 2 senior citizen clubs via letters of invitation. Participants were community-dwelling older persons aged 60 years and over and were from the main ethnicities in Malaysia (Malay, Chinese, and Indian). Inclusion criteria included being 60 years and above, able to comprehend the Malay or English language, able to use a smartphone, and able to ambulate with or without assistive devices with minimal supervision. Older persons with acute illnesses (unstable heart diseases and vestibular disorders) were excluded.

Procedure and Instrumentation

Participants were given information regarding the procedures and consent to participate in this study. Participant level of smartphone proficiency was determined verbally before conducting the study. Participants were presented with FallSA on a researcher's mobile, and they were asked to open the app and follow its instructions with minimal guidance from the researcher. They then needed to complete both levels of navigation (new user and return user) in order to finish a real-life environment response of the software system. The FallSA test took approximately 15 minutes to complete. Upon completion, participants were required to provide feedback using a series of closed- and open-ended questions regarding their understanding of FallSA in relation to its features and design: (1) color contrast, (2) graphics or illustration, (3) font size, (4) presentation of instructional video, and (5) overall FallSA suitability. A modified version of a technology acceptance model survey was also used. The technology acceptance model is rated on 7-point Likert scale and has a high internal reliability (Cronbach =.96) and positive correlation between each determinant (perceived usefulness, perceived ease of use, intention and attitude toward use of mobile technology) [19].

Phase 3: Concurrent Validity And Test-Retest Reliability

Phase 3 was a cross-sectional study to determine the concurrent validity and test-retest reliability of FallSA among community-dwelling older persons.

Participants

Participants were recruited among community-dwelling older persons at two other senior citizen clubs. A total of 91 community-dwelling older persons participated in this study. The sampling method and its inclusion and exclusion criteria were similar to the phase 2 study.

Procedure and Instrumentation

In order to identify the concurrent validity of FallSA, the researcher validated FallSA with an existing fall risk assessment tool. To date, there is no gold standard assessment for fall risk [17]. Thus, the Physiological Profile Assessment (PPA) was selected because of its robustness in identifying fall risk among older persons by assessing their impairments in main physiological measurements irrespective of health conditions. Despite this, studies found that the PPA had 75% accuracy and moderate reliability of ≥0.50 in identifying risk of falls among older persons [20]. A short-form PPA comprising 5 questions was used to assess fall risk and was based on the composite (z) score calculated using online software. Fall risk in the PPA is categorized as follows: (1) very low, (2) low, (3) mild, (4) moderate, (5) marked, and (6) very marked [20].

Upon screening based on inclusion and exclusion criteria, participants were briefed on the assessments and informed

consent was given. Next, participants proceeded with anthropometry measurements (height and weight), followed by collecting their sociodemographic data, fall history, and associated characteristics. Participants were then asked to use the latest version of the FallSA with minimal guidance from the trained enumerator. Participants selected their language preference (Malay/English) and then could proceed to the first section, registration of sociodemographic information. Upon completion of the registration section, participants were directed to the page on physical performance test instruction. Participants were asked to use all methods of instruction: wording, audio, and video. Following this, the interface changed to the physical performance test, with the start and stop timer being a big red button for ease of use among older persons. Two trials were attempted, and the mean was calculated. Next, participants were directed to 4 closed-ended questions regarding fall risk. Report of the overall result was presented upon completion with the whole assessment taking approximately 15 minutes to accomplish.

Participants were given a break of at least 15 minutes before proceeding to the short-form PPA. Trained physiotherapists with more than 1 year of experience were employed to administer the PPA. Participants were asked to perform the 5 tests included in the PPA: (1) edge contrast sensitivity (vision), (2) peripheral sensation (proprioception), (3) finger press (reaction time), (4) standing on the medium-density foam rubber mat (body sway), and (5) knee extension (lower limb strength). On average, the PPA was completed within 30 minutes.

After a 15-minute break upon completion of the PPA, participants repeated the FallSA test to determine its reliability. After verifying their identity card number and date of birth, participants directly proceeded to the physical performance test. Each participant took 10 minutes to complete this FallSA test.

Phase 4: Discriminative Ability and Predictive Validity

Phase 4 was a cross-sectional study to identify the discriminative ability of FallSA followed by a 6-month prospective follow-up to examine its predictive validity in identifying risk of falls among the older persons in Peninsular Malaysia.

Participants

A total of 610 community-dwelling older persons from Peninsular Malaysia (Johor, Selangor, Perak, and Kelantan) participated in this cross-sectional validation study. Participants were selected using a multistage random sampling method. Participants aged 60 years and over able to ambulate independently with or without assistive devices were included in this study, whereas those unable to comprehend and follow instructions, having severe medical conditions, having severe vision or hearing impairments, or having cognitive impairments (dementia or depression) were excluded from this study.

Procedure and Instrumentation

Participant sociodemographic data, fall history, and medical condition were obtained. FallSA assessment was performed after a demo and trial session to assess their risk of fall. In the discriminative study, FallSA scores were compared among fallers and nonfallers according to their past 1-year fall history data. A faller was defined as someone who had one or more falls in the past year and nonfaller as someone without any falls [21].

Next, the validation study was conducted whereby participants were provided with a fall diary to document fall incidence monthly for a period of 6 months. In addition, participants were contacted via phone monthly to obtain their fall incidence and any other feedback and to remind participants about documenting their falls if any occur. The diaries were collected after 6 months to determine the predictive validity of FallSA. At the end of the 6-month follow-up, participants who had a risk of falls were advised verbally to seek further fall risk assessment and management at their primary health care settings.

Statistical Analysis

The data were analyzed using SPSS (version 22.0, IBM Corp). Researchers performed the normality test for continuous variables in advance by using a Shapiro-Wilk test, Kolmogorov-Smirnov test, kurtosis, skewness ratio, histogram, stem and leaf, or box plot.

Descriptive analysis was performed on the sociodemographic profile. In phase 2, a Pearson correlation analysis was conducted to determine the acceptance level of FallSA among Malaysian community-dwelling older persons. In phase 3, the concurrent validity of FallSA was analyzed using Spearman correlation, kappa for agreement, sensitivity, and specificity. The test-retest reliability was conducted to identify internal consistency (Cronbach alpha), and the intraclass correlation (ICC) and Bland-Altman agreement were determined between 2 trials of the physical performance test. In the final phase, the discriminative ability of FallSA was analyzed using an independent *t* test, and the receiver operating characteristic (ROC) curve was used to identify the predictive validity of FallSA, with its sensitivity, specificity value, and cutoff point.

Results

Phase 1: FallSA Final Version

The figures below depict the screen shots of FallSA (final version): selection of the Malay or English language, which are commonly used in the Malaysian population (Figure 1), TUG self-conduct test results (Figure 2), items from the sociodemographic and multifactorial questionnaire (Figure 3), and fall risk report (Figure 4).

Figure 1. User initial navigation and language selection.

Figure 2. Physical Performance Test (TUG) results.

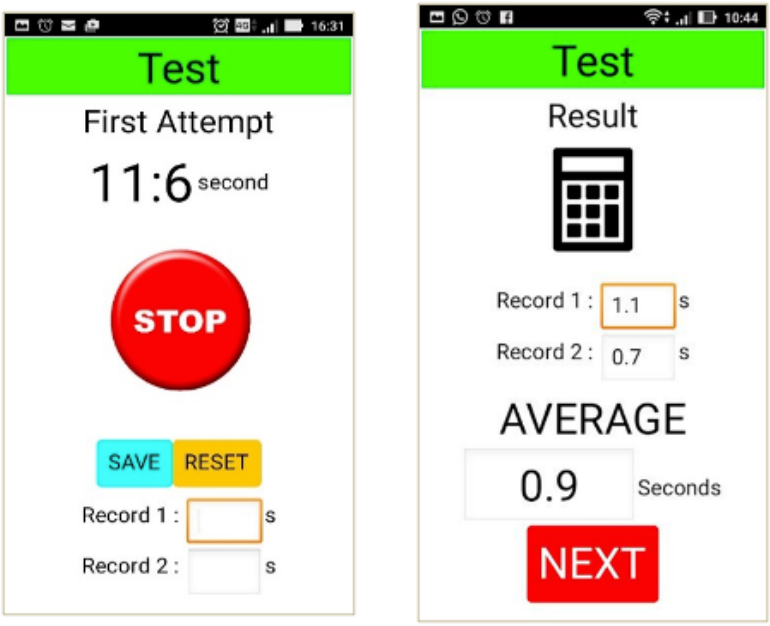
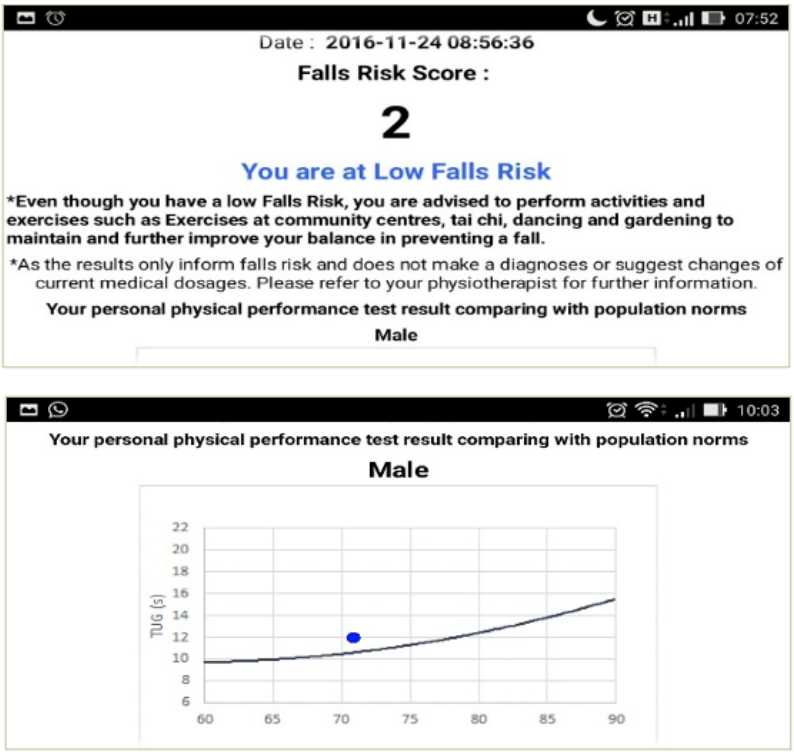


Figure 3. Sociodemographic and self-rated multifactorial questionnaire.

Figure 4. Fall risk reports.



Phase 2: User Acceptance Testing Results

Table 1 depicts the sociodemographic data and user acceptance survey of FallSA based on gender. The results of our study found that more than 90% (14/15) of participants were able to comprehend the contents of FallSA and the instructional video provided in the app and agreed with its suitability of graphics and color combinations. However, approximately 30% (3/15) of participants reported that the font size used in FallSA was not suitable for older persons. Overall, 80% (12/15) of participants found that FallSA is suitable as a self-screening mobile app to identify fall risk among Malaysian community-dwelling older persons.

Participants listed 4 aspects in the FallSA app that were found to be not suitable. First was the physical performance test as the users were required to learn and understand how to perform the TUG test independently. However, most of the participants agreed that the TUG test was easy to use after minimal guidance. Second, participants found the font, especially in the disclaimer interface section, to be small for them. Participants stated that they had difficulty reading the instructions and information when using the app. Third, the system hanged occasionally, probably due to an unstable server and nonsynchronization between the apps and online database. Last, 15% (3/15) of participants requested more interesting graphics or illustrations. Technology Acceptance Model (TAM) survey results showed

the presence of a high correlation (0.70 and above) between all determinants (perceived usefulness, perceived ease of use, intention and attitude towards usage of mobile technology).

Table 1. Sociodemographic information and user evaluation of different aspects of FallSA based on gender.

Characteristic	Males (n=6), n (%)	Females (n=9), n (%)	Total (n=15), n (%)
Sociodemographic			
Age (years)			
60-69	4 (36)	7 (64)	11 (73)
≥70	2 (50)	2 (50)	4 (27)
Race			
Malay	2 (29)	5 (71)	7 (47)
Chinese	3 (60)	2 (40)	5 (33)
Indian	1 (33)	2 (67)	3 (20)
Education level			
None	3 (75)	1 (25)	4 (27)
Primary	2 (33)	4 (67)	6 (40)
Secondary	—[a]	3 (100)	3 (20)
Tertiary	1 (50)	1 (50)	2 (13)
Marital status			
Married	2 (25)	6 (75)	8 (53)
Widowed	2 (50)	2 (50)	4 (27)
Divorced	1 (50)	1 (50)	2 (13)
Single	1 (100)	—	1 (7)
User evaluation of the different aspects of FallSA			
Contents			
Understand	5 (36)	9 (64)	14 (93)
Did not understand	1 (100)	—	1 (7)
Suitability of graphics			
Suitable	5 (36)	9 (64)	14 (93)
Not suitable	1 (100)	—	1 (7)
Color combination			
Suitable	6 (43)	8 (57)	14 (93)
Not suitable	—	1 (100)	1 (7)
Font size			
Easy to read	4 (40)	6 (60)	10 (67)
Hard to read	2 (40)	3 (60)	5 (33)
Instructional video			
Easy to understand	6 (40)	9 (60)	15 (100)
Hard to understand	—	—	—
Overall suitability of FallSA			
Suitable	5 (42)	7 (58)	12 (80)
Not suitable	1 (33)	2 (67)	3 (20)

[a]Not applicable.

Phase 3: Concurrent Validity (Against PPA) and Test-Retest Reliability Results

Participant characteristics are shown in Table 2.

The concurrent validity results between FallSA and PPA are shown in Table 3. PPA results were categorized in dichotomous data using cutoff points (high and low risk of fall). There was a significant moderate correlation (r=.518, P<.001) found between FallSA and PPA. All test parameters area under the ROC curve and Cohen kappa were statistically acceptable with sensitivity and specificity at 80.4% and 71.1%, respectively. There was a stronger correlation between FallSA and PPA in males (r=.538, P<.001) compared to females (r=.502, P<.001), with a higher sensitivity value of 88.9%. In addition, older persons with higher (secondary and tertiary) education (r=.427, P<.001) had a lower correlation between FallSA and PPA compared to those with lower (none and primary) education level (r=.511, P<.001).

As for test-retest reliability, there was a significant high reliability between repeated FallSA tests (P<.001; ICC .948, 95% CI .921-.966) as shown in Table 4. There was high agreement with small mean differences and narrow limits of agreement between repeated FallSA scores (Figure 5).

Table 2. Sociodemographic data of the participants based on gender.

Characteristics	Males (n=40), n (%)	Females (n=51), n (%)	Total (n=91), n (%)
Age (years)			
60-69	19 (32)	41 (68)	60 (66)
≥70	21 (68)	10 (32)	31 (34)
Race			
Malay	29 (47)	33 (53)	62 (68)
Chinese	9 (36)	16 (64)	25 (28)
Indian	2 (50)	2 (50)	4 (4)
Education level			
None	—[a]	2 (100)	2 (2)
Primary	8 (53)	7 (47)	15 (17)
Secondary	19 (41)	27 (59)	46 (51)
Tertiary	13 (46)	15 (54)	28 (31)
Marital status			
Married	37 (56)	29 (44)	66 (73)
Widowed	1 (6)	17 (94)	18 (20)
Divorced	—	2 (100)	2 (2)
Single	2 (40)	3 (60)	5 (6)

[a]Not applicable.

Table 3. Concurrent validity of FallSA against Physiological Profile Assessment (PPA).

Variables	*r* value	Kappa	Sensitivity (%)	Specificity (%)	ROC[a]
Gender					
Males	.538	.494	88.9	74.2	0.749
Females	.502	.429	70.7	90.0	0.790
Education level					
Low	.511	.628	81.8	83.3	0.902
High	.427	.489	80.0	69.2	0.776
Total score	.518	.516	80.4	71.1	0.794

[a]ROC: receiver operating characteristic.

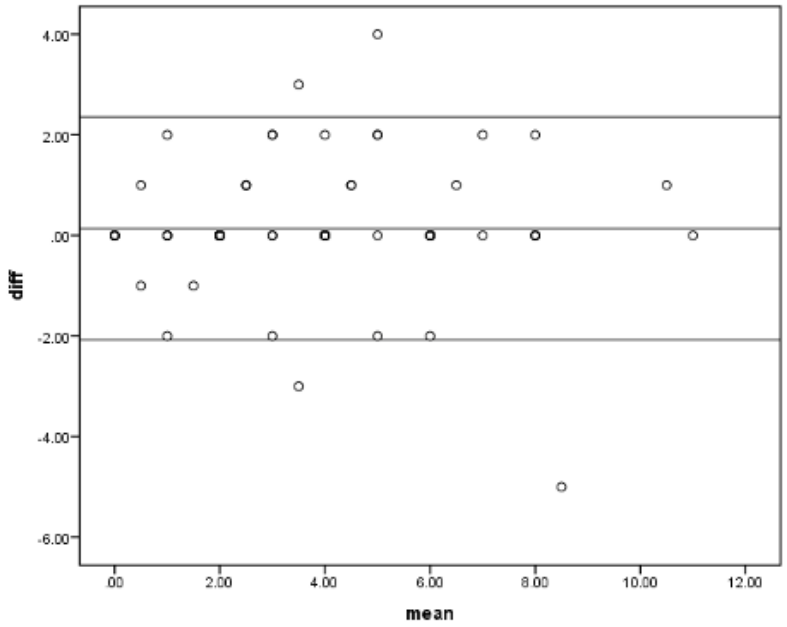
Table 4. Reliability of the FallSA total score.

Characteristic	Cronbach alpha	ICC[a]	95% CI	SEM[b]
Total score FallSA	.948	.948	.921-.966	1.11
Gender				
Male	.920	.920	.849-.958	0.98
Female	.929	.929	.875-.959	1.21
Education				
Low	.974	.974	.929-.991	0.88
High	.942	.942	.908-.963	1.15

[a]ICC: intraclass correlation.

[b]SEM: standard error of the mean.

Figure 5. Limit of agreement.



Phase 4: Discriminative Ability and Predictive Validity Results

A total of 1005 community-dwelling older persons aged 60 years and above within Peninsular Malaysia (Selangor, Perak, Johor, and Kelantan states) were screened for the inclusion criteria; 395 were excluded due to having a score ≥5 on the Geriatric Depression Scale, a score ≤21 on the Mini Mental State Examination, or failing to complete the FallSA test. The sociodemographic data of the participants is shown in Table 5. Participants with a past history of falls had significantly higher FallSA scores (7.33 [SD 1.77]) at baseline as compared to those without any falls (4.55 [SD 1.86]; *P*<.001). This indicates the discriminative ability of FallSA.

After 6 months, 74.4% (454/610) of participants were successfully followed up via fall diary and monthly phone calls. The 26.6% (156/610) who were dropouts were excluded from the follow-up analysis. About 14.5% (66/454) of the community-dwelling older persons had a fall after the 6-month follow-up. FallSA scores were compared with actual falls reported after the 6-month follow-up; 3% (3/66) of participants categorized as low risk and 18% (63/66) of participants categorized as at-risk experienced a fall.

The cutoff, sensitivity and specificity values, and positive and negative likelihood ratios of FallSA are presented in Table 6. The fall risk score of FallSA ranges from 0 to 11. The results of this study suggest a FallSA cutoff score of >5 is the best predicted cutoff point of fall among older persons. With this cutoff, the sensitivity and specificity values of the FallSA score were 81.82% (95% CI 70.4-90.2) and 63.92% (95% CI 58.9-68.7), respectively, with a positive likelihood ratio of 2.27, meaning those community-dwelling older persons were 2.27 times more likely to fall compared to those who scored ≤5.0. The Youden index shown in analysis was 0.47.

The ROC of FallSA is shown in Figure 6. With an average area under the curve (AUC) of 0.802, FallSA is demonstrated to have a good discriminative ability. An excellent result for AUC is indicated if close to 1 (0.8-0.9) [22].

Table 5. Baseline sociodemographic data classified based on fallers and nonfallers.

Variables	Total (n=610)	Fallers (n=111; 18.2%)	Nonfallers (n=499; 81.8%)	*P* value
Age, mean (SD)	71.78 (4.7)	72.01 (4.6)	71.73 (4.8)	.58
MMSE[a], mean (SD)	26.48 (2.4)	26.43 (2.4)	26.49 (2.4	.82
GDS[b], mean (SD)	1.85 (1.3)	1.96 (1.3)	1.83 (1.3)	.32
PASE[c], mean (SD)	125.76 (54.4)	125.49 (50.8)	125.82 (55.3)	.95
Gender, n (%)	__[d]	—	—	.001
Male	332 (54.4)	45 (13.6)	287 (86.4)	—
Female	278 (45.6)	66 (23.7)	212 (76.3)	—
Race, n (%)	—	—	—	.85
Malay	361 (59.2)	65 (18.0)	296 (82.0)	—
Chinese	218 (35.7)	40 (18.3)	178 (81.7)	—
Indian	31 (5.1)	6 (19.4)	25 (80.6)	—
Education level, n (%)	—	—	—	.79
None	63 (10.3)	10 (15.9)	53 (84.1)	—
Primary	306 (50.2)	60 (19.6)	246 (80.4)	—
Secondary	199 (32.6)	34 (17.1)	165 (82.9)	—
Tertiary	31 (5.1)	7 (22.6)	24 (77.4)	—
Other	11 (1.8)	0	11 (100)	—
Chronic illness, n (%)	—	—	—	—
Hypertension	83 (13.6)	18 (21.7)	65 (78.3)	.83
Diabetes	163 (26.7)	34 (20.9)	129 (79.1)	.84
Heart disease	62 (10.2)	15 (24.2)	47 (75.8)	.55
Arthritis	209 (34.3)	44 (21.1)	165 (78.9)	.15
Falls history in past 12 months, n (%)	—	—	—	.001
No falls	499 (81.8)	0	499 (100)	—
1 fall	68 (11.1)	68 (100)	0	—
≥2 falls	43 (7.1)	43 (100)	0	—
Medication, n (%)	—	—	—	.47
< meds	440 (72.1)	75 (17.0)	365 (83.0)	—
≥ meds	170 (27.9)	36 (21.2)	134 (78.8)	—
Eye problems, n (%)	—	—	—	.56
Yes	105 (17.2)	17 (16.2)	88 (83.8)	—
No	505 (82.8)	94 (18.6)	411 (81.4)	—
Use of assistive devices, n (%)	—	—	—	.05
Yes	31 (5.1)	11 (35.5)	20 (64.5)	—
No	579 (94.9)	100 (17.3)	479 (82.7)	—
FallSA score, mean (SD)	5.05 (2.15)	7.33 (1.77)	4.55 (1.86)	.001
FallSA Fall Risk, n (%)	—	—	—	.001
Low risk	132 (21.6)	1 (0.8)	131 (99.2)	—
At risk	478 (78.4)	110 (23.0)	368 (77.0)	—

[a]MMSA: Mini Mental State Examination.

[b]GDS: Geriatric Depression Scale.

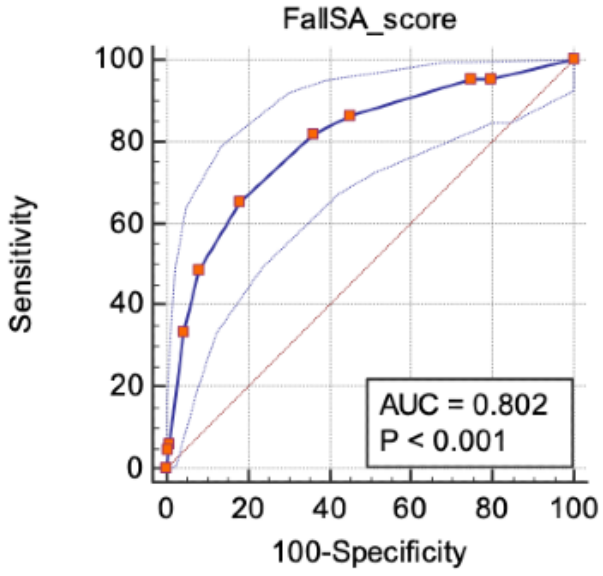
[c]PASE: Physical Activity Scale for Elderly.
[d]Not applicable.

Table 6. Criterion values and coordinates of the receiver operating characteristic curve in the FallSA fall risk score.

Criterion	Sn[a]	Sp[b]	+LR[c]	–LR[d]	PPV[e]	NPV[f]
≥2	100	0	1	—[g]	14.5	—
>2	95.45	20.36	1.20	0.22	16.9	96.3
>3	95.45	25.26	1.28	0.18	17.8	97.0
>4	82.35	54.90	1.91	0.25	24.6	95.9
>5[h]	81.82	63.92	2.27	0.28	27.8	95.4
>6	65.15	81.70	3.56	0.43	37.7	93.2
>7	48.48	91.75	5.88	0.56	50.0	91.3
>8	33.33	95.62	7.61	0.70	56.4	89.4
>9	6.06	99.23	7.84	0.95	57.1	86.1
>10	4.55	99.48	8.82	0.96	60.0	86.0
>11	0	100	—	1	—	85.5

[a]Sn: sensitivity.
[b]Sp: specificity.
[c]LR+: positive likelihood ratio.
[d]LR–: negative likelihood ratio.
[e]PPV: positive predictive value.
[f]NPV: negative predictive value.
[g]Not applicable.
[h]Cutoff score.

Figure 6. Receiver operating characteristic of FallSA based on the 6-month follow-up.



Discussion

Principal Findings

In our study, we successfully developed an accepted and validated mobile app, FallSA, that has the potential to identify a risk of falls among a multiethnic Malaysian older population. The FallSA score manifested a moderate to high discriminative ability and predictive validity in classifying fallers and nonfallers and predicting falls among older persons. A cutoff score of >5.0 is recommended in our FallSA predictive results. With this cutoff, FallSA had a positive likelihood ratio of 2.27, pooled sensitivity of 82%, and specificity of 64% with an average AUC of 0.802. However, only 4.5% of older persons screened to have a low fall risk using FallSA had experienced falls in the previous 6 months in our study. This suggests that the present sensitivity

and specificity of the FallSA test would be more useful for nonfallers.

Our findings showed that 80% of the participants agreed that FallSA is suitable for use as a self-screening tool among Malaysian community-dwelling older persons. Similarly, the Aachen fall prevention app [14] was found to be well accepted and useful when offered in the mobile stores. These results suggest it is possible to facilitate the use of a fall risk mobile app among older persons and their caregivers to self-screen for fall risk. FallSA as a mobile health technology with automated reports would be useful as a self-administered fall risk assessment tool. Such tools can be administered at home or in community settings and save time [23].

There were several rounds of adjustments in the development stages (requirement design, implementation, and verification) of the FallSA app that included changes in concept, design, graphics, contents, navigation flow, and technical and language corrections. In the user acceptance stage, the design of FallSA was well accepted and older persons had a positive attitude toward FallSA's adoption with some minor use issues. We addressed all the highlighted issues in the acceptance test before proceeding to conduct reliability and validity tests. For example, font size in FallSA was increased within the limitations of smartphone display size. In addition, we provided bilingual and simplified video instructions for the TUG test. Increasing the number of icons used in the graphic user interface and rectifying problems in the coding system that caused the system to hang during use were also addressed.

Upon addressing participant feedback regarding FallSA, we examined the concurrent validity of FallSA against PPA among 91 community-dwelling older persons. To date, there is no gold standard assessment tool to measure fall risk. However, PPA was identified as the most comprehensive evidence-based practice fall risk assessment tool available. Our study results showed that there is a significant moderate correlation between FallSA and PPA fall risk measurements (P<.001, k=0.875, r=.518), with good sensitivity (80.4%) and specificity (71.1%) and good ROC (0.794) association. The plausible explanation of this moderate correlation could be because PPA assessed different aspects of fall risk factors (proprioception, contrast sensitivity, postural sway, reaction time, and lower limb strength) objectively. Besides the TUG test in FallSA, the rest are based on questionnaires and are subjective. However, these findings support the concurrent validity of FallSA against PPA, suggesting that FallSA is able to identify fall risk in Malaysian community-dwelling older persons. Although only a moderate correlation was found between FallSA and PPA, it is sufficient and in line with the use of FallSA as an early fall risk self-screening tool.

Compared to the concurrent validity of other fall screening tools such as the Fall Risk Questionnaire [24] and Austin Health Falls Risk Screening Tool [25], FallSA's concurrent validity is slightly lower. This can be speculated as we only included 4 significant predictor variables to keep FallSA simple for self-use in older persons. It is also noteworthy that we tested its concurrent validity using comprehensive fall risk assessment tool rather than against another questionnaire. Furthermore, FallSA was designed to support an early screening that can be preceded upon inquiring for further comprehensive clinical assessments by trained health care professionals.

Test-retest analysis of first and second FallSA score results showed that all parameters had an excellent test-retest reliability with Cronbach =.948 (ICC .948, 95% CI .921-.966; SEM 1.11), suggesting FallSA is consistent in assessing fall risk among older persons. In addition, a good agreement was demonstrated with Bland-Altman analysis, having a small mean difference and narrow limits of agreement between the first and second FallSA assessments.

The fall prevalence demonstrated in our phase 4 study was 18.2%, which corresponded to earlier local studies conducted among community-dwelling older persons in Malaysia (15% to 18%) [5,6]. However, this value is much lower compared to prevalence rates on a global scale, which are found to be 12% to 63% [26]. This variation could be possibly due to having active and younger older persons in the local studies. Further, this study was conducted among older persons in the community, of which the prevalence is expected to be lower than clinical settings or in residential homes [27].

There was a moderate to high discriminative ability and predictive validity in discriminating fallers and nonfallers and predicting falls among older persons using FallSA. The pooled predictive sensitivity (82%) and specificity (64%) of FallSA are comparable to the Aachen fall prevention app (sensitivity 57.0%; specificity 76.7%) in discriminating between fallers and nonfallers. It is noteworthy that the FallSA predictive validity was conducted prospectively based on actual falls, while the Aachen fall prevention app used its primary outcome in a cross-sectional manner. Using a FallSA cutoff score of >5.0, 95.5% of falls after the 6-month follow-up among community-dwelling older persons were predicted. The number of false negative results can be reduced along with a lower negative likelihood ratio and further reinforced with a higher sensitivity value [28]. Only 4.5% of older persons screened to have low fall risk using FallSA had experienced falls in the previous 6 months in our study.

Comparably, the AUC (0.84) and sensitivity value (93%) of Lindera were much higher than FallSA. This discrepancy is possibly due to the variation in study methodology whereby the discriminative ability of Lindera was based on retrospective fall information, whereas for FallSA it was conducted prospectively. Moreover, FallSA is developed as a self-screening fall risk assessment tool to support the older persons as an early fall screening tool, while Lindera is specifically designed to assist health care professionals in clinical settings to identify fall risk among the older people. FallSA having a higher average sensitivity compared to specificity value is more suitable for early fall screening and prevention. This will allow a greater proportion of older persons to be screened and participate in fall prevention programs. This is in line to support the call for routine fall risk assessment and early management in the updated National Institute for Health and Care Excellence (NICE) fall prevention statements [1]. In terms of advanced diagnostic testing in clinical settings, a high specificity value is needed to

avoid the unnecessary, costly, and tiring management among the nonfallers.

Limitations and Strengths

FallSA has a higher sensitivity value when compared to traditional fall risk assessment using the TUG test. TUG test sensitivity as a fall risk assessment was reported to range from 30.5% to 67.5% [29,30]. As a stand-alone fall risk assessment tool, the sensitivity of TUG is lower as fall risk is multifactorial in nature. Moreover, although mobility and balance status could be assessed using TUG [31,32], it may not be comprehensive enough to determine the multiple interacting fall risk factors. This deficiency has been addressed in FallSA as the TUG test is combined with fall-related multifactorial questions in the fall risk calculation model.

The model used in FallSA was derived from both urban and rural community-dwelling older populations. Hence FallSA's use among older persons living in institutions and with frailty will require adaptations to the present model used. Another limitation of FallSA is that it does not provide information for specific impairment areas for tailored interventions as PPA does. However, FallSA is meant for preliminary fall risk self-screening. Last, the duration of follow-up for FallSA may be relatively short in this study, and the changes of health status among older persons may not have been accounted for.

The main strength of FallSA is that it was developed systematically and followed by testing its acceptance concurrent validity, reliability, and discriminative and predictive validity. In addition, we used prospective fall monitoring to identify the predictive validity of FallSA. Note that our study comprised older persons from Peninsular Malaysia and all 3 main ethnicities (Malays, Chinese and Indians) and the results can be generalized to the entire community-dwelling older population.

Clinically, FallSA has the potential to be used as a self-screening, caregiver administered, or at primary health care settings as an early fall risk screening tool. This will assist in the annual fall risk screening among older adults as outlined in the NICE updated fall prevention guidelines [1]. Packaged as a mobile app, FallSA is accessible anywhere anytime and is simple, fast, and easy to administer. Hence, it is also suitable to be used for large-scale community-based fall screenings. Early fall detection can assist in targeting for early fall prevention interventions in older persons at risk of falls. However, FallSA could be improved by enhancing the information in the instructional video for the TUG test, upgrading it with an educational video of fall risk, and adding a sit to stand test with its normative values.

Conclusions

In this study, we successfully developed a mobile app (FallSA) to identify fall risk among community-dwelling older persons that was accepted, valid, and reliable. FallSA is a short multifactorial assessment tool as it integrates sociodemographic, clinical, and physical fall risk factors. Although NICE updated fall prevention guidelines have recommended an annual fall risk screening, community-dwelling older persons in our local setting tend to visit primary health care settings more often. Since this test is self-administered, we suggest at least a biannual fall risk screening among older persons. To the best of our knowledge, FallSA is one of the most comprehensively tested fall risk self-assessment tools. FallSA has the potential to be used as one of the fall prevention strategies with the ultimate aim of maintaining independence and improving quality of life as long as possible among older persons. Future studies are required to empower and engage older persons or care givers in the use of FallSA to self-screen for falls and thereafter seek early prevention intervention.

Acknowledgments

The authors would like to thank the participants for their support. The work presented in this paper was supported by grant LRGS/BU/2012UKM/L/01 (LRGS TUA) awarded by the Ministry of Education and grant Inovasi-2019-001 from Universiti Kebangsaan Malaysia.

Conflicts of Interest

None declared.

References

1. The assessment and prevention of falls in older people. Nationel Institute for Health and Care Excellence. 2013. URL: https://www.nice.org.uk/guidance/cg161/resources/falls-in-older-people-assessing-risk-and-prevention-pdf-35109686728645 [accessed 2021-07-20]
2. Shaharudin MI, Ajit Singh DK, Shahar S. Falls prevalence and its risk assessment tools among Malaysian community-dwelling older adults: a review. Malay J Public Heal Med 2018 Jan 15;16(si):223. [doi: 10.17576/jskm-2018-34]
3. Sharif SI, Al-Harbi AB, Al-Shihabi AM, Al-Daour DS, Sharif RS. Falls in the elderly: assessment of prevalence and risk factors. Pharm Pract (Granada) 2018;16(3):1206. [doi: 10.18549/PharmPract.2018.03.1206] [Medline: 30416623]
4. Tomita Y, Arima K, Tsujimoto R, Kawashiri S, Nishimura T, Mizukami S, et al. Prevalence of fear of falling and associated factors among Japanese community-dwelling older adults. Medicine 2018;97(4):e9721. [doi: 10.1097/md.0000000000009721]
5. Ibrahim A, Singh DKA, Shahar S, Omar MA. Timed up and go test combined with self-rated multifactorial questionnaire on fall risk and sociodemographic factors predicts falls among community-dwelling older adults better than the timed up and go test on its own. J Multidiscip Healthc 2017 Oct;10:409-416. [doi: 10.2147/jmdh.s142520]

6. Ibrahim A, Singh D, Shahar S. "Timed Up and Go" test: age, gender and cognitive impairment stratified normative values of older adults. PLoS 2017;12(10):e0185641. [doi: 10.1371/journal.pone.0185641]
7. Singh D, Shahar S, Vanoh D, Kamaruzzaman S, Tan M. Diabetes, arthritis, urinary incontinence, poor self-rated health, higher body mass index and lower handgrip strength are associated with falls among community-dwelling middle-aged and older adults: pooled analyses from two cross-sectional Malaysian datasets. Geriatr Gerontol Int 2019 Aug;19(8):798-803. [doi: 10.1111/ggi.13717] [Medline: 31237103]
8. Ooi TC, Singh DKA, Shahar S, Rajab NF, Vanoh D, Sharif R, et al. Incidence and multidimensional predictors of occasional and recurrent falls among Malaysian community-dwelling older persons. BMC Geriatr 2021 Mar 02;21(1):154 [FREE Full text] [doi: 10.1186/s12877-021-02103-2] [Medline: 33653312]
9. Steultjens E, Dekker J, Bouter L, Jellema S, Bakker E, van den Ende CHM. Occupational therapy for community dwelling elderly people: a systematic review. Age Ageing 2004 Sep;33(5):453-460. [doi: 10.1093/ageing/afh174] [Medline: 15315918]
10. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining Risk of falls in community dwelling older adults: a systematic review and meta-analysis using posttest probability. J Geriatr Phys Ther 2017;40(1):1-36. [doi: 10.1519/jpt.0000000000000099]
11. Rabe S, Azhand A, Pommer W, Müller S, Steinert A. Descriptive evaluation and accuracy of a mobile app to assess the fall risk in seniors: retrospective case control study. JMIR Aging 2020 Feb 14;3(1):e16131 [FREE Full text] [doi: 10.2196/16131] [Medline: 32130111]
12. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A fall risk mHealth app for older adults: development and usability study. JMIR Aging 2018 Nov 20;1(2):e11569 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]
13. Hsieh K, Fanning J, Sosnoff J. A smartphone fall risk application is valid and reliable in older adults during real-world testing. Gerontechnology 2019 Mar 31;18(1):29-35. [doi: 10.4017/gt.2019.18.1.003.0]
14. Rasche P, Mertens A, Bröhl C, Theis S, Seinsch T, Wille M, et al. The "Aachen fall prevention App"—a smartphone application app for the self-assessment of elderly patients at risk for ground level falls. Patient Saf Surg 2017;11:14 [FREE Full text] [doi: 10.1186/s13037-017-0130-4] [Medline: 28503199]
15. Park S. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. Aging Clin Exp Res 2018 Jan;30(1):1-16. [doi: 10.1007/s40520-017-0749-0] [Medline: 28374345]
16. Shahar S, Omar A, Vanoh D, Hamid TA, Mukari SZM, Din NC, et al. Approaches in methodology for population-based longitudinal study on neuroprotective model for healthy longevity (TUA) among Malaysian older adults. Aging Clin Exp Res 2016 Dec;28(6):1089-1104. [doi: 10.1007/s40520-015-0511-4] [Medline: 26670602]
17. Podsiadlo D, Richardson S. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
18. flaticon. URL: http://www.flaticon.com/ [accessed 2021-07-20]
19. Alharbi S, Drew S. Using the technology acceptance model in understanding academics' behavioural intention to use learning management systems. Int J Adv Comput Sci Appl 2014;5(1):143-155. [doi: 10.14569/ijacsa.2014.050120]
20. Lord SR, Menz HB, Tiedemann A. A physiological profile approach to falls risk assessment and prevention. Phys Ther 2003 Mar;83(3):237-252. [Medline: 12620088]
21. Dai W, Tham Y, Chee M, Tan NYQ, Wong K, Majithia S, et al. Falls and recurrent falls among adults in a multi-ethnic Asian population: the Singapore Epidemiology of Eye Diseases Study. Sci Rep 2018 May 15;8(1):7575 [FREE Full text] [doi: 10.1038/s41598-018-25894-8] [Medline: 29765104]
22. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013;4(2):627-635 [FREE Full text] [Medline: 24009950]
23. Sprint G, Cook DJ, Weeks DL. Towards automating clinical assessments: a survey of the timed up and go (TUG). IEEE Rev Biomed Eng 2015;8:64-77 [FREE Full text] [doi: 10.1109/RBME.2015.2390646] [Medline: 25594979]
24. Rubenstein LZ, Vivrette R, Harker JO, Stevens JA, Kramer BJ. Validating an evidence-based, self-rated fall risk questionnaire (FRQ) for older adults. J Safety Res 2011 Dec;42(6):493-499. [doi: 10.1016/j.jsr.2011.08.006] [Medline: 22152267]
25. Said CM, Churilov L, Shaw K. Validation and inter-rater reliability of a three item falls risk screening tool. BMC Geriatr 2017 Nov 23;17(1):273 [FREE Full text] [doi: 10.1186/s12877-017-0669-z] [Medline: 29169328]
26. Schoene D, Heller C, Aung YN, Sieber CC, Kemmler W, Freiberger E. A systematic review on the influence of fear of falling on quality of life in older people: is there a role for falls? Clin Interv Aging 2019;14:701-719 [FREE Full text] [doi: 10.2147/CIA.S197857] [Medline: 31190764]
27. Maneeprom N, Taneepanichskul S, Panza A. Fall among physically active elderly in senior housings, Bangkok, Thailand: situations and perceptions. Clin Interv Aging 2018 Oct;13:2149-2159. [doi: 10.2147/cia.s175896]
28. Wrisley DM, Kumar NA. Functional gait assessment: concurrent, discriminative, and predictive validity in community-dwelling older adults. Phys Ther 2010 May;90(5):761-773. [doi: 10.2522/ptj.20090069] [Medline: 20360052]
29. Kang L, Han P, Wang J, Ma Y, Jia L, Fu L, et al. Timed Up and Go Test can predict recurrent falls: a longitudinal study of the community-dwelling elderly in China. Clin Interv Aging 2017;12:2009-2016. [doi: 10.2147/CIA.S138287] [Medline: 29238175]

30. Kojima G, Masud T, Kendrick D, Morris R, Gawler S, Treml J, et al. Does the timed up and go test predict future falls among British community-dwelling older people? Prospective cohort study nested within a randomised controlled trial. BMC Geriatr 2015 Apr 03;15(1):38 [FREE Full text] [doi: 10.1186/s12877-015-0039-7] [Medline: 25887660]
31. Garber CE, Greaney ML, Riebe D, Nigg CR, Burbank PA, Clark PG. Physical and mental health-related correlates of physical function in community dwelling older adults: a cross sectional study. BMC Geriatr 2010 Feb 03;10:6 [FREE Full text] [doi: 10.1186/1471-2318-10-6] [Medline: 20128902]
32. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Lee S, Hotta R, et al. Cognitive functioning and walking speed in older adults as predictors of limitations in self-reported instrumental activity of daily living: prospective findings from the Obu Study of Health Promotion for the Elderly. Int J Environ Res Public Health 2015 Mar;12(3):3002-3013 [FREE Full text] [doi: 10.3390/ijerph120303002] [Medline: 25768239]

Abbreviations

AUC: area under the curve
FallSA: Falls Risk Screening App
ICC: intraclass correlation
LRGS: Long-Term Research Grant Scheme—Toward Useful Ageing
NICE: National Institute for Health and Care Excellence
PPA: Physiological Profile Assessment
ROC: receiver operating characteristic
TUG: Timed Up and Go

Edited by L Buis; submitted 19.08.20; peer-reviewed by K Hsieh, B Thies; comments to author 28.11.20; revised version received 20.03.21; accepted 02.06.21; published 12.10.21.

Please cite as:
Kaur Ajit Singh D, Goh JW, Shaharudin MI, Shahar S
A Mobile App (FallSA) to Identify Fall Risk Among Malaysian Community-Dwelling Older Persons: Development and Validation Study
JMIR Mhealth Uhealth 2021;9(10):e23663
URL: https://mhealth.jmir.org/2021/10/e23663
doi:10.2196/23663
PMID:34636740

Original Paper

Six-Month Outcomes from the NEXit Junior Trial of a Text Messaging Smoking Cessation Intervention for High School Students: Randomized Controlled Trial With Bayesian Analysis

Marcus Bendtsen[1*], PhD; Preben Bendtsen[1,2*], MD, PhD; Ulrika Müssener[1*], PhD

[1]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[2]Department of Medical Specialist, Motala Hospital, Motala, Sweden

[*]all authors contributed equally

Corresponding Author:
Marcus Bendtsen, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Building 511
Linköping, 58183
Sweden
Phone: 46 13286975
Email: marcus.bendtsen@liu.se

Abstract

Background: The prevalence of daily or occasional smoking among high school students in Sweden was approximately 20% in 2019, which is problematic since lifestyle behaviors are established in adolescence and track into adulthood. The Nicotine Exit (NEXit) Junior trial was conducted in response to a lack of evidence for the effects of text message smoking cessation interventions among high school students in Sweden.

Objective: The aim of this study was to estimate the 3- and 6-month effects of a text messaging intervention among high school students in Sweden on smoking cessation outcomes.

Methods: A 2-arm, single-blind randomized controlled trial was employed to estimate the effects of the intervention on smoking cessation in comparison to treatment as usual. Participants were recruited from high schools in Sweden using advertising and promotion by school staff from January 10, 2018, to January 10, 2019. Weekly or daily smokers who were willing to make a quit attempt were eligible for inclusion. Prolonged abstinence and point prevalence of smoking cessation were measured at 3 and 6 months after randomization.

Results: Complete case analysis was possible on 57.9% (310/535) of the participants at 6 months, with no observed statistically significant effect on 5-month prolonged abstinence (odds ratio [OR] 1.27, 95% CI 0.73-2.20; *P*=.39) or 4-week smoking cessation (OR 1.42; 95% CI 0.83-2.46; *P*=.20). Sensitivity analyses using imputation yielded similar findings. Unplanned Bayesian analyses showed that the effects of the intervention were in the anticipated direction. The findings were limited by the risk of bias induced by high attrition (42.1%). The trial recruited high school students in a pragmatic setting and included both weekly and daily smokers; thus, generalization to the target population is more direct compared with findings obtained under more strict study procedures.

Conclusions: Higher than expected attrition rates to follow-up 6 months after randomization led to null hypothesis tests being underpowered; however, unplanned Bayesian analyses found that the effects of the intervention were in the anticipated direction. Future trials of smoking cessation interventions targeting high school students should aim to prepare strategies for increasing retention to mid- and long-term follow-up.

Trial Registration: IRCTN Registry ISRCTN15396225; https://www.isrctn.com/ISRCTN15396225

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-3028-2

(JMIR Mhealth Uhealth 2021;9(10):e29913) doi:10.2196/29913

KEYWORDS

smoking; cessation; text messaging; high school; randomized controlled trial; intervention; student; young adult; teenager; outcome; Bayesian; Sweden; prevalence; lifestyle; behavior

Introduction

A steady decline in smoking prevalence has been observed in Sweden over the past decade, with the most recent data indicating that 7% of the general population were daily smokers in 2018 [1]. Although this decline is promising, the prevalence of daily smoking among individuals aged 16 to 29 years was 5% in 2018, and this rate increased to 16% when including occasional smokers. Among Swedish high school students specifically, approximately 5% self-reported being everyday smokers in 2019, and this rate increased to 20% when including occasional smokers [2]. Thus, young individuals in Sweden still start smoking. This is problematic since unhealthy lifestyle behaviors, including smoking, are established in adolescence and track into adulthood [3-5]. Effective smoking cessation interventions that target adolescents are therefore important for the declining trend in smoking prevalence to continue in Sweden.

Mobile phone–based interventions could potentially have far reach among adolescents, as mobile phone ownership in this group is almost universal in Sweden. Of particular interest are text message interventions, as they rely on standard technology that is ubiquitous in all mobile phones, and they have shown promise in other populations. Three meta-analyses have concluded that text messaging interventions have a positive effect on smoking cessation: one reported a summary effect size of 0.25 (95% CI 0.13-0.38) [6], the second meta-analysis reported an overall summary odds ratio (OR) of 1.37 (95% CI 1.25-1.51) of smoking cessation in favor of text messaging interventions [7], and the third analysis similarly found that quit rates were higher among those who had access to text messaging interventions (OR 1.36, 95% CI 1.23-1.51) [8]. Thus, there exists a relatively strong body of evidence for the effects of text messaging interventions. In addition, text messaging interventions may increase access to education and support services that promote smoking cessation [7].

However, only three trials have investigated the effects of smoking cessation text messaging interventions in adolescent populations exclusively [9-11], and only two of these have been conducted in high school student populations [10,11]. Two out of the three were small-scale trials including 179 [11] and 72 [9] participants, respectively, whereas the third was a large-scale cluster randomized trial including 2638 participants [10]. Findings from the large-scale trial suggested that although the number of cigarettes smoked per day was lower among those with access to the intervention, there were no marked differences with respect to smoking abstinence.

In 2018, we conducted the Nicotine Exit Junior (NEXit Junior) trial in response to the lack of evidence for the effects of text message smoking cessation interventions among high school students. We found that a 12-week text message smoking cessation intervention had a positive effect on the 4-week point prevalence of smoking cessation 3 months after randomization among weekly and daily smoking students (OR 1.87, 95% CI 1.12-3.17, *P*=.02) [12]. This report presents the results from the 6-month follow-up of the trial. Analyses prespecified in the study protocol are presented [13], as are unplanned Bayesian analyses of primary outcomes for both the 3- and 6-month follow-ups.

Methods

Design

A 2-arm, single-blind, parallel-groups randomized controlled trial was employed to estimate the effects of the NEXit Junior intervention on smoking cessation. The trial was prospectively registered (ISRCTN15396225) and a trial protocol was published in advance of trial recruitment [13]. This report adheres to CONSORT guidelines, and a CONSORT-EHEALTH checklist has been made available with this report. The trial received ethical approval by the Regional Ethical Committee in Linköping, Sweden (Dnr 2017/388-31).

Participants

Participants were simultaneously recruited from 630 high school units in Sweden. Students were recruited through paper advertising (poster and leaflets) and digital advertising (email, school website, app), and by promotion by school staff (teachers, mentors, school health centers). Recruitment began on January 10, 2018, and ended 1 year later (January 10, 2019).

Students interested in participating in the trial sent a text message to a dedicated telephone number. An automatic response was sent back with a hyperlink to a webpage where information about the trial was presented and students were asked to leave informed consent by pressing a button. Students who consented to participate were taken to a web-based questionnaire including items for both eligibility screening and baseline assessment.

High school students who reported being weekly or daily smokers, and were willing to make a quit attempt, were eligible for the trial. We included weekly smokers for two reasons: (1) this was the criterion for our previous trial of a text messaging smoking intervention among university students [14,15], and (2) weekly smoking in adolescents is cause for concern as it may escalate to daily smoking and long-term dependence. In addition, it was expected that participants have access to a mobile phone throughout the trial period. Individuals reporting not smoking, or doing so on a monthly basis only, were excluded from the trial.

Interventions

The text message intervention was a 12-week automatic, and unguided, program consisting of a total of 121 messages. During the first 2 weeks of the program, participants received 2 to 4 messages per day, which reduced to 2 messages per day during week 3, 1 message per day during weeks 4 to 7, and on average 5 messages per week during weeks 8 to 12.

The content of the messages was based on a similar intervention targeting university students in Sweden [14,15], which was based on existing evidence-based practice [10,16-22]. The content was further developed and refined for high school students using formative methods and behavior change technique analysis [23-27]. The content of the messages focused on the health risks of smoking and behavior change advice. In particular, participants were asked to make a public declaration about their intention to quit, encouraged to ask for support, taught distraction techniques, given tips on how to cope with cravings, and given advice on how to avoid smoking triggers.

Individuals allocated to the control group were advised that they would not obtain access to the novel intervention and that they instead could use the website of the national quit line [28] or contact their high school's health service for more help. This was considered treatment as usual at the time of the trial, as this was in general what high school students who wanted help quitting were advised to do.

Outcomes

At 3 and 6 months after randomization, all participants were sent a text message with a hyperlink to a web-based follow-up questionnaire. Participants were sent two reminders 2 days apart, after which they were called by phone for follow-up (maximum of 10 attempts). The questionnaire contained items for the two primary outcomes: prolonged abstinence and point prevalence of smoking cessation.

Prolonged abstinence, following the Russel standard [29], was defined at the 3-month follow-up as not smoking more than 5 cigarettes in the past 8 weeks. At 6 months, the definition was altered to not smoking more than 5 cigarettes in the past 5 months. This outcome thus measures abstinence from the start of the 12-week smoking cessation program, allowing for a 4-week grace period.

Point prevalence of smoking cessation, a recommended outcome by the Society for Research on Nicotine and Tobacco [30], was defined as not having smoked a single cigarette in the past 4 weeks. This question measures current behavior, and thus was the same at both the 3- and 6-month follow-ups.

Sample Size

Sample size was determined by assuming that differences in cessation rates between the two groups would be similar to what was observed in a previous study of a text messaging smoking cessation intervention targeting Swedish university students [14,15]. We therefore expected a difference of approximately 10 percentage points between the two groups, which would require 195 individuals in each group to be detected with 80% power at the .05 (two-sided) significance threshold. We assumed that we would have 30% attrition at follow-up, resulting in a required total sample size of 558.

It should be noted that the expected 10 percentage point difference is much higher than what has been synthesized in meta-analyses of text messaging smoking cessation support versus minimal smoking cessation advice. The meta-analyses suggested that the differences may be in the range of 3-4 percentage points [31], which may be a more conservative choice for other trials to adopt. We used a greater difference in consideration of our previous research among university students in Sweden that yielded this difference using a similar intervention and trial design [14,15].

Randomization

After baseline assessment, eligible students were randomly allocated to the intervention or control group with equal probability. This was done on the backend computer server using Java's built-in random number generator; thus, the sequence generation was computerized and fully automatic. Allocation was thereby also fully concealed until interventions were assigned.

Research personnel were blind to allocation at the time of randomization; however, participants were informed of which group they were allocated to. Follow-up assessments were completed using online questionnaires by participants on their phones; thus, group allocation was not revealed at the time of follow-up. However, there was a potential risk of group disclosure among those who did not respond to text message prompts and subsequently were called by research personnel to collect follow-up data (see Limitations).

Statistical Methods

The analyses conformed with the prespecified analyses in the trial protocol [13]. Specifically, logistic regression models were used to compare groups at the different follow-up intervals, adjusting for baseline variables: gender, number of years smoking, average number of cigarettes smoked weekly, severity of dependence (Fagerström Test for Nicotine Dependence), and the use of snuff. Participants were analyzed in the groups to which they were randomized.

Attrition analyses were planned to investigate potential differences between those who did and did not respond to follow-up. A penalized logistic regression model (lasso) was used to identify the baseline characteristics that were potentially predictive of nonresponse. This allowed us to explore departures from the missing completely at random (MCAR) assumption underlying the primary complete case analyses. Sensitivity analyses using multiple imputation with chained equations [32] were performed to assess robustness of the findings (using predictive mean matching, with 500 imputations and 30 iterations).

In addition to the prespecified analyses, unplanned Bayesian analyses were performed. The higher than anticipated attrition rate underpowered the planned null hypothesis tests, and the Bayesian analyses were included to calculate the probability that the intervention had an effect on smoking outcomes. Using a Bayesian approach also removes the emphasis on null hypothesis testing and prespecified significance thresholds, which can be sensitive to single data points and can create counterintuitive results [33-38]. Normal priors with mean 0 and SD 1 were used for all regression coefficients to encode an a priori skepticism of effect. Inference was carried out using the Hamiltonian Monte Carlo procedure (50,000 samples, using the first half as burn-in).

All analyses were performed in R (version 3.6.1); Bayesian inference was performed using the probabilistic programming language Stan (RStan version 2.19.1).

Results

Participant Characteristics

The CONSORT flow diagram in Figure 1 describes the progress of the two groups in the NEXit Junior trial, and Table 1 presents the baseline characteristics of the two groups. A total of 621 high school students were screened for eligibility between January 10, 2018, and January 10, 2019. Among these, 86 students were excluded due to either reporting that they did not smoke on a daily or weekly basis or that they did not want to make a quit attempt, resulting in 535 randomized students. Of those randomized, 276 students were allocated to the intervention group and 259 were allocated to the control group.

Figure 1. CONSORT flow diagram of the progress through the phases of the Nicotine Exit (NEXit) Junior trial of two groups.

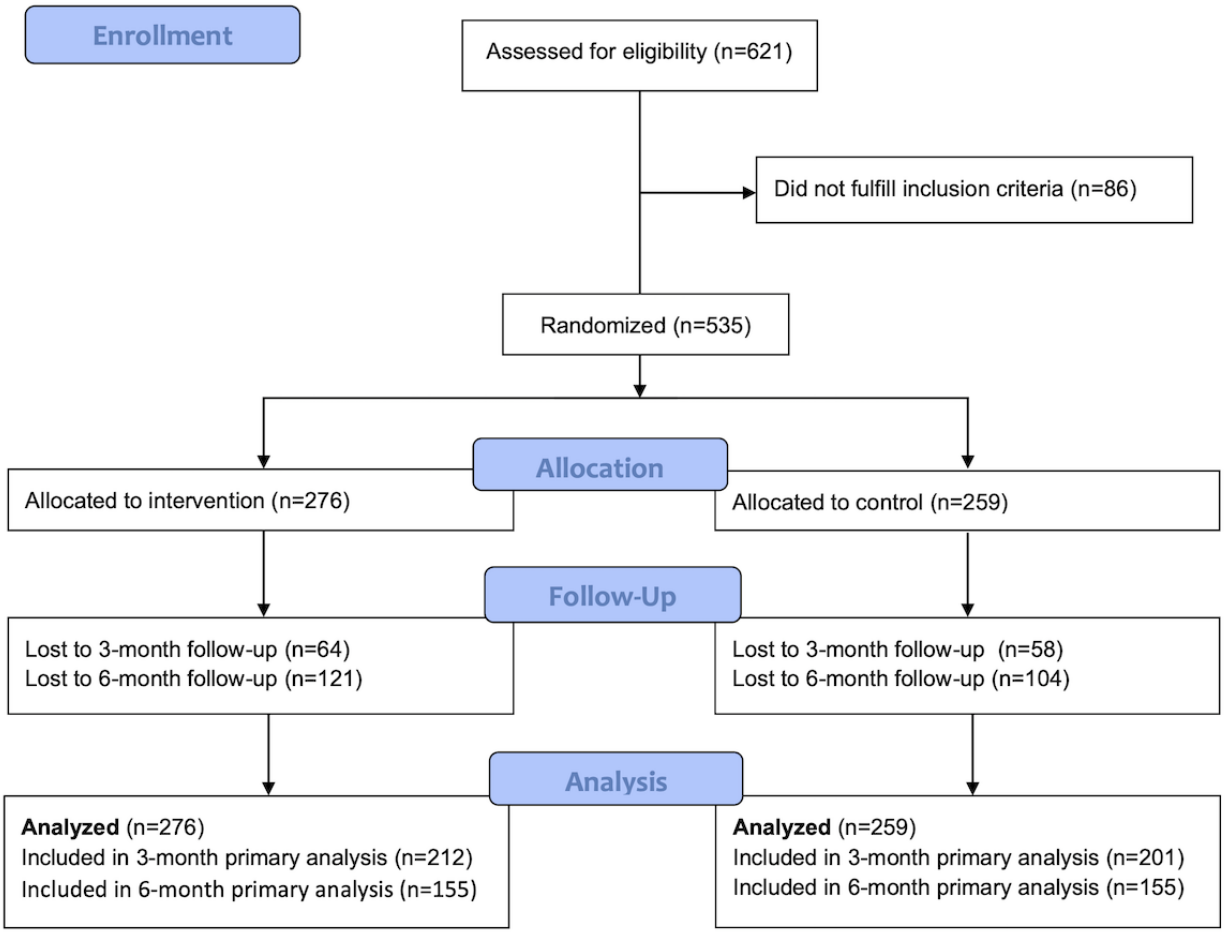


Table 1. Baseline characteristics of trial participants in the Nicotine Exit (NEXit) Junior trial.

Variables	Intervention (n=276)	Control (n=259)	*P* value
Gender, n (%)			.91[a]
Female	164 (59.4)	162 (62.5)	
Male	102 (37.0)	90 (34.7)	
Other	3 (1.1)	2 (0.8)	
Decline to answer	4 (1.4)	2 (0.8)	
Do not know	3 (1.1)	3 (1.2)	
Age (years), median (IQR)	17 (16-18)	17 (16-18)	.69[b]
Years smoking, median (IQR)	3 (2-5)	3 (2-5)	.22[b]
Cigarettes smoked per week, median (IQR)	60 (35-84)	70 (42-105)	.06[b]
Regularly using snuff, n (%)	84 (30.4)	74 (28.6)	.71[c]
Fagerström Nicotine Dependence Scale, median (IQR)	4 (2-6)	4 (3-5.5)	.26[b]
Quit attempts, median (IQR)	2 (1-4)	2 (1-4)	.85[b]
Use of nicotine replacement therapies, median (IQR)	0 (0-1)	0 (0-1)	.96[b]
Cessation counseling experience, n (%)			.21[c]
No	250 (90.6)	226 (87.3)	
Yes, previously	13 (4.7)	22 (8.5)	
Yes, currently	13 (4.7)	11 (4.2)	
Recruitment strategy, n (%)			.46[a]
Poster	71 (25.7)	80 (30.9)	
Homepage	49 (17.8)	37 (14.3)	
Student health center	45 (16.3)	44 (17.0)	
Staff	34 (12.3)	34 (13.1)	
School’s mobile app	33 (12.0)	32 (12.4)	
Friend	21 (7.6)	10 (3.9)	
Flyer	10 (3.6)	5 (1.9)	
Email	4 (1.4)	5 (1.9)	
Other	9 (3.3)	12 (4.6)	

[a]Fisher exact test.

[b]Wilcoxon rank-sum test.

[c]χ^2 test.

Primary Analysis

Complete case analysis was possible on 57.9% (310/535) of participants at 6 months and on 77.2% (413/535) of participants at 3 months. Sample estimates of prolonged abstinence and smoking cessation (primary outcomes), along with ORs, 95% CIs, and *P* values, are presented in Table 2. Results are presented for both 3- and 6-month outcomes for completeness.

The OR for 4-week point prevalence of smoking cessation, which was statistically significant at the 3-month follow-up, was no longer distinguishable from 1 at the 6-month follow-up at the conventional $P<.05$ threshold; thus, a null finding could not be ruled out. None of the other ORs was statistically significant.

Table 2. Primary outcome intention-to-treat analyses with complete cases at 3- and 6-month follow-ups.

Outcomes	Intervention,[a] n (%)	Control,[b] n (%)	OR[c] (95% CI)	*P* value
3-month outcomes				
8-week prolonged abstinence	49 (23.1)	39 (19.4)	1.21 (0.73-2.01)	.46
4-week smoking cessation	53 (25.0)	31 (15.4)	1.87 (1.12- 3.17)	.02
6-month outcomes				
5-month prolonged abstinence	41 (26.5)	32 (20.6)	1.27 (0.73-2.20	.39
4-week smoking cessation	44 (28.4)	31 (20.0)	1.42 (0.83-2.46)	.20

[a]n=212 at 3-month follow-up, n=155 at 6-month follow-up.

[b]n=201 at 3-month follow-up, n=155 at 6-month follow-up.

[c]OR: odds ratio based on logistic regression, adjusted for gender, number of years smoking, average number of cigarettes smoked weekly, severity of dependence (Fagerström Test for Nicotine Dependence), and the use of snuff.

Sensitivity Analysis

We explored potential deviations from the MCAR assumption underlying the primary analyses in Table 1. Penalized logistic regression (lasso) did not reveal any baseline characteristics that were predictive of nonresponse to 6-month outcomes, offering no evidence against the MCAR assumption.

The analyses of 6-month primary outcomes with imputed data found similar effect estimates of the intervention as the complete case analysis for 5-month prolonged abstinence (OR 1.34, 95% CI 0.81-2.20, *P*=.25) and 4-week point prevalence of smoking cessation (OR 1.51, 95% CI 0.93-2.46, *P*=.09). This suggests that the findings in Table 1 are robust to the missing data. Despite no evidence against the MCAR assumption, interpretation of the findings from the imputed values should be made with caution as the rate of attrition (42.1%) was higher than is generally advised when using multiple imputation [32].

Bayesian Analysis

Figure 2 shows histograms of the samples drawn during the Hamiltonian Monte Carlo simulations. These histograms are representative of the marginal posterior probability of the ORs contrasting the intervention and control groups. The histograms indicate which ORs are more likely than others with respect to the number of samples above or below a specific value. For instance, in the top right histogram, a strong majority of samples lie above 1, indicating a high probability that the intervention had an effect on the 4-week point prevalence of smoking cessation at the 3-month follow-up.

Figure 2. Samples from Hamiltonian Monte Carlo simulations for calculating the odds ratio of intervention vs control.

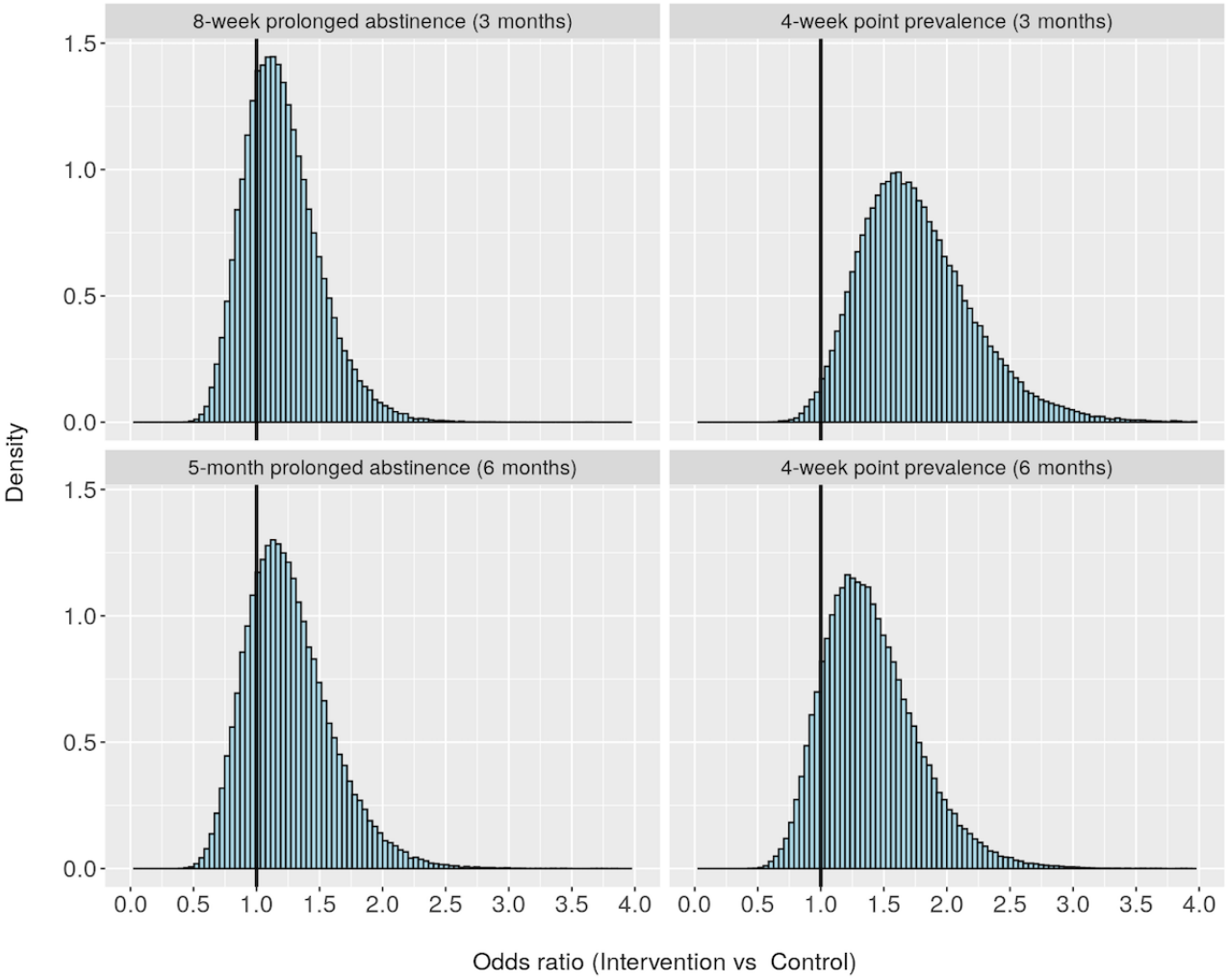


Table 3 presents the probabilities for three different OR thresholds (1, 1.25, and 1.5) for the primary outcomes. The probability that the text message intervention had an effect on prolonged abstinence (ie, OR>1) was 73.8% at 3 months and was 76.9% at 6 months. The probability that the intervention had an effect on 4-week smoking cessation was 98.4% at 3 months and was 87.5% at 6 months. In addition, the probability that the OR was greater than 1.25 for 4-week smoking cessation at 3 months was 89.6% and was 61.7% at 6 months.

Table 3. Primary outcome intention-to-treat analyses using Bayesian inference for complete cases at 3- and 6-month follow-ups.

Outcomes	Intervention,[a] n (%)	Control,[b] n (%)	Marginal posterior probability (%)		
			OR[c]>1	OR>1.25	OR>1.5
3-month outcomes					
8-week prolonged abstinence	49 (23.1)	39 (19.4)	73.8	38.7	15.0
4-week smoking cessation	53 (25.0)	31 (15.4)	98.4	89.6	69.8
6-month outcomes					
5-month prolonged abstinence	41 (26.5)	32 (20.6)	76.9	45.6	21.2
4-week smoking cessation	44 (28.4)	31 (20.0)	87.5	61.7	34.8

[a]n=212 at 3-month follow-up, n=155 at 6-month follow-up.

[b]n=201 at 3-month follow-up, n=155 at 6-month follow-up.

[c]OR: odds ratio obtained by logistic regression, adjusted for gender, number of years smoking, average number of cigarettes smoked weekly, severity of dependence (Fagerström Test for Nicotine Dependence), and the use of snuff.

Discussion

Principal Findings

The primary analyses indicated that null findings could not be ruled out when estimating the effect of a 12-week text message intervention on prolonged abstinence and point prevalence of smoking cessation among high school students 6 months after randomization. Sensitivity analyses where missing data were imputed indicated that these findings were robust. Unplanned Bayesian analyses of both 3- and 6-month outcomes indicated that the effects of the intervention were in the anticipated direction.

Among the several studies of smoking cessation text message interventions to a wide range of populations carried out over the past decade, a follow-up period of 6 months is not uncommon. A meta-analysis from 2015 summarized the evidence of nine studies at the 6-month follow-up [8], finding an overall standardized OR of 1.35 (95% CI 1.18-1.49), suggesting that there was a small detectable effect. This is in contrast to our findings, which may be severely affected by attrition and targeted a less explored population; however, this also contrasts with the findings of a large pragmatic randomized trial that did not suffer from severe attrition [39]. The pragmatic trial was performed among both treatment and nontreatment seekers in Australia, with a total of 3550 participants being recruited with high retention (86.5% for the 6-month follow-up); nevertheless, no significant effect of a text message program was found on its own or in combination with a tailored online program. Therefore, although the synthesized evidence does suggest that there are persistent effects, there are certainly some questions remaining regarding the size of these effects and their overall generalizability on health for both subpopulations and the general population.

As mentioned in the Introduction, few trials of smoking cessation text messaging interventions have been conducted targeting adolescent populations exclusively [9-11]. The findings have suggested no marked influence of text messages on smoking abstinence, which is corroborated by our findings in this trial. However, our Bayesian analyses do suggest that effect estimates are in the anticipated direction, even after having taken a skeptical a priori view on the magnitude of the effect. Moreover, although compliance is hard to measure when investigating text messaging interventions, as it is not possible to know with any certainty the degree to which participants read and comply with the messages they receive, during this trial, 21% (n=114) of participants decided that they no longer wished to receive any more messages. This number is difficult to interpret without further data, as some may have been displeased about the support received, some may have just been curious about the trial, while others may have quit smoking and not needed any more support. In any case, we believe that retaining 79% of participants to the end of a 12-week intervention should be considered a success in this population; thus, it is feasible to deliver a smoking cessation intervention via text messages to high school students. The question remains if such an intervention can be designed to also be effective in the long term. Future research should therefore carefully consider if text messaging interventions to this population are worth pursuing. If so, factorial trial designs should be employed to assess multiple components simultaneously [40,41], and Bayesian group sequential designs should be used to inform both successful and futile experiments early on [42-44].

Smoking is considered to be the second leading risk factor for disability adjusted life years globally [45]; thus, effective ways to help individuals quit smoking permanently are needed. Synthesized evidence of smoking cessation interventions targeting adolescents suggest that group counseling may be effective in the long term (risk ratio 1.35, 95% CI 1.03-1.77), whereas individual counseling, pharmacological interventions, and digital interventions did not show long-term effectiveness [4]. However, the quality of the body of evidence is low, as there are several concerns regarding risk of bias and heterogeneity. Although an intervention that encourages smoking cessation and promotes abstinence for a period of time may be considered helpful, we should be careful to offer interventions that may take time away from other smoking cessation activities. It is important to develop interventions that utilize digital platforms as an option for young individuals in Sweden, as they are digital natives and nicotine products can only be bought by people 18 years or older. Moreover, text message interventions targeting university students have been found to be more effective among those without a strong nicotine dependence [46], suggesting that acting early by offering text messaging interventions at the high school level may still be an important public health measure.

Generalizability

Recruitment to the trial was performed to closely mimic how the intervention could be disseminated to high schools in a full-scale implementation (ie, through print and digital advertisement managed by each school). There was no additional contact with participants throughout the trial period; thus, the trial closely resembles how the intervention would be used in routine care. In addition, inclusion criteria were broad, excluding only those who smoked on a monthly basis or less. The trial may therefore be seen as measuring effectiveness rather than efficacy, which strengthens the external validity of the findings. However, the limitations of this trial should be taken into account when interpreting the findings and the generalizability of the results, especially the high attrition rates.

Limitations

The most severe limitation of the 6-month findings from the NEXit Junior trial was the risk of bias induced by the high attrition rate (42.1%). Although we found no evidence of systematic missingness, and the findings were robust under imputation, high attrition increases the probability that unobserved factors may predict nonresponse, and thereby bias both the primary and sensitivity analyses. This issue has previously been raised as a concern when studying young adults [47], and future trials on this age group need to implement strategies to retain participants, potentially through incentives.

A second limitation was the lack of blinding of trial participants, which may induce performance bias, an issue prevalent in trials of health interventions as participants are often aware that they

signed up for a trial of a particular intervention. Performance bias may be the source of the measured effects rather than the interventions themselves; thus, future trials should prioritize making study information to participants consonant with allocation to intervention and control groups. Additionally, caution should be taken when assessing the present findings since the outcomes were self-reported. This does not necessarily warrant a concern of bias in and of itself; however, in combination with the nonblinding of participants, the risk of performance bias is exacerbated.

Finally, although all processes were automated and participants were treated equally within trial groups, there was a risk of detection bias due to calling nonresponders to collect follow-up data. Despite the fact that the research personnel responsible for the telephone interviews were experienced and instructed not to prompt participants to disclose allocation, some participants are expected to reveal the group to which they were randomized to, and thereby create a risk of bias. However, we judged this risk to be inferior to the increased risk of attrition bias, which would have been induced by not collecting outcome data from those not responding to initial attempts. Furthermore, the analyses revealed no differences in baseline or outcome data between early and late responders, suggesting that the risk of disclosure bias did not lead to any actual bias [12].

These limitations may all influence effect estimates by introducing bias. One way to address these sources of bias post hoc is to consider accounting for them in analyses using causal models [48]. Toward this end, future research should consider estimating the magnitude of effects induced by these sources of bias so that unbiased effects of interventions may be estimated.

Conclusions

Higher than expected attrition rates to follow-up 6 months after randomization led to null hypothesis tests being underpowered. Although null findings could not be ruled out, unplanned Bayesian analyses suggested that the effects of the intervention were in the anticipated direction both at the 3- and 6-month follow-ups. Future trials of smoking cessation interventions targeting high school students should aim to blind participants to avoid risk of performance bias, and also prepare strategies for increasing retention to mid- and long-term follow-up.

When making policy decisions to implement text messaging interventions for smoking cessation in high schools, consideration should be taken to other alternatives with stronger evidence for long-term effects. However, such decisions should also consider that text messaging interventions are relatively cheap to implement and potentially have great reach among those who may benefit. Moreover, although the effects past the 3-month follow-up period have been difficult to establish due to study limitations, the evidence put forth herein suggests that the intervention was more likely than not to have a positive effect, even in the long run. Considering the health benefits that could come from quitting smoking at such an early age, text messaging interventions could be an important part of a public health campaign in this population.

Acknowledgments

We thank all of the high school students and staff for participating in this study.

Authors' Contributions

All authors participated in defining the study. UM and PB developed the intervention, UM took the lead on recruitment, and MB designed the trial and performed all analyses. All authors contributed equally to preparation of this manuscript, and all authors have read and approved the final manuscript.

Conflicts of Interest

MB and PB own a private company that develops and distributes evidence-based lifestyle interventions to be used in health care settings. UM declares no conflict of interest.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 983 KB - mhealth_v9i10e29913_app1.pdf]

References

1. Nationella folkhälsoenkäten - Hälsa på lika villkor. Folkhälsomyndigheten. 2018. URL: https://www.folkhalsomyndigheten.se/folkhalsorapportering-statistik/om-vara-datainsamlingar/nationella-folkhalsoenkaten/ [accessed 2019-06-03]
2. Skolelevers drogvanor 2019 - CAN rapport 187. Centralförbundet för alkohol- och narkotikaupplysning (C.A.N). 2019. URL: https://www.can.se/publikationer/skolelevers-drogvanor-2019/ [accessed 2021-10-10]
3. Hall WD, Patton G, Stockings E, Weier M, Lynskey M, Morley KI, et al. Why young people's substance use matters for global health. Lancet Psychiatry 2016 Mar;3(3):265-279. [doi: 10.1016/S2215-0366(16)00013-4] [Medline: 26905482]
4. Fanshawe T, Halliwell W, Lindson N, Aveyard P, Livingstone-Banks J, Hartmann-Boyce J. Tobacco cessation interventions for young people. Cochrane Database Syst Rev 2017 Nov 17;11:CD003289 [FREE Full text] [doi: 10.1002/14651858.CD003289.pub6] [Medline: 29148565]

5. Stockings E, Hall WD, Lynskey M, Morley KI, Reavley N, Strang J, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X] [Medline: 26905481]
6. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb 17;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
7. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
8. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
9. Mason MJ, Campbell L, Way T, Keyser-Marcus L, Benotsch E, Mennis J, et al. Development and outcomes of a text messaging tobacco cessation intervention with urban adolescents. Subst Abus 2015 Oct 19;36(4):500-506. [doi: 10.1080/08897077.2014.987946] [Medline: 25551337]
10. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]
11. Shi H, Jiang X, Yu C, Zhang Y. Use of mobile phone text messaging to deliver an individualized smoking behaviour intervention in Chinese adolescents. J Telemed Telecare 2016 Jul 09;19(5):282-287. [doi: 10.1177/1357633x13495489]
12. Müssener U, Linderoth C, Thomas K, Bendtsen M. mHealth smoking cessation intervention among high school students: 3-month primary outcome findings from a randomized controlled trial. PLoS One 2020 Mar 6;15(3):e0229411 [FREE Full text] [doi: 10.1371/journal.pone.0229411] [Medline: 32142514]
13. Thomas K, Bendtsen M, Linderoth C, Müssener U. mHealth smoking cessation intervention among high-school pupils (NEXit Junior): study protocol for a randomized controlled trial. Trials 2018 Nov 22;19(1):648. [doi: 10.1186/s13063-018-3028-2]
14. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. JAMA Intern Med 2016 Mar 01;176(3):321-328 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
15. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. SMS-based smoking cessation intervention among university students: study protocol for a randomised controlled trial (NEXit trial). Trials 2015 Apr 08;16(1):140 [FREE Full text] [doi: 10.1186/s13063-015-0640-2] [Medline: 25872503]
16. 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 Jan 15;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
17. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
18. 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;378(9785):49-55. [doi: 10.1016/s0140-6736(11)60701-0]
19. Abroms LC, Whittaker R, Free C, Mendel Van Alstyne J, Schindler-Ruwisch JM. Developing and pretesting a text messaging program for health behavior change: recommended steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
20. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug 01;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
21. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 30;1(2):e17 [FREE Full text] [doi: 10.2196/mhealth.2522] [Medline: 25098502]
22. Ybarra M, Holtrop J, Prescott T, Rahbar M, Strong D. Pilot RCT results of stop my smoking USA: a text messaging-based smoking cessation program for young adults. Nicotine Tob Res 2013 Aug;15(8):1388-1399 [FREE Full text] [doi: 10.1093/ntr/nts339] [Medline: 23348969]
23. Brown KG, Gerhardt MW. Formative evaluation: an integrative practice model and case study. Pers Psychol 2002 Dec;55(4):951-983. [doi: 10.1111/j.1744-6570.2002.tb00137.x]
24. Lee A, Sandvei M, Asmussen HC, Skougaard M, Macdonald J, Zavada J, et al. The development of complex digital health solutions: formative evaluation combining different methodologies. JMIR Res Protoc 2018 Jul 16;7(7):e165 [FREE Full text] [doi: 10.2196/resprot.9521] [Medline: 30012548]

25. Fitts Willoughby J, Furberg R. Underdeveloped or underreported? Coverage of pretesting practices and recommendations for design of text message-based health behavior change interventions. J Health Commun 2015 Apr 06;20(4):472-478. [doi: 10.1080/10810730.2014.977468] [Medline: 25749250]
26. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
27. 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 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
28. Sluta-Röka-Linjen. URL: https://www.slutarokalinjen.se/ [accessed 2021-10-01]
29. West R, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction 2005 Mar;100(3):299-303. [doi: 10.1111/j.1360-0443.2004.00995.x] [Medline: 15733243]
30. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May 1;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]
31. 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 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
32. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
33. Bendtsen M. A gentle introduction to the comparison between null hypothesis testing and Bayesian analysis: reanalysis of two randomized controlled trials. J Med Internet Res 2018 Oct 24;20(10):e10873 [FREE Full text] [doi: 10.2196/10873] [Medline: 30148453]
34. Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. Nat Hum Behav 2018 Jan;2(1):4. [doi: 10.1038/s41562-017-0224-0] [Medline: 30980046]
35. Nuzzo R. Scientific method: statistical errors. Nature 2014 Feb 13;506(7487):150-152. [doi: 10.1038/506150a] [Medline: 24522584]
36. Wasserstein RL, Lazar NA. The ASA Statement on P-values: context, process, and purpose. Am Stat 2016 Jun 09;70(2):129-133. [doi: 10.1080/00031305.2016.1154108]
37. Bendtsen M. Electronic screening for alcohol use and brief intervention by email for university students: reanalysis of findings from a randomized controlled trial using a Bayesian framework. J Med Internet Res 2019 Nov 07;21(11):e14419 [FREE Full text] [doi: 10.2196/14419] [Medline: 31697242]
38. Bendtsen M. An electronic screening and brief intervention for hazardous and harmful drinking among Swedish university students: reanalysis of findings from a randomized controlled trial using a Bayesian framework. J Med Internet Res 2019 Dec 17;21(12):e14420 [FREE Full text] [doi: 10.2196/14420] [Medline: 31845903]
39. Borland R, Balmford J, Benda P. Population-level effects of automated smoking cessation help programs: a randomized controlled trial. Addiction 2013 Mar 01;108(3):618-628. [doi: 10.1111/j.1360-0443.2012.04091.x] [Medline: 22994457]
40. Baker TB, Smith SS, Bolt DM, Loh W, Mermelstein R, Fiore MC, et al. Implementing clinical research using factorial designs: a primer. Behav Ther 2017 Jul;48(4):567-580 [FREE Full text] [doi: 10.1016/j.beth.2016.12.005] [Medline: 28577591]
41. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Med Res Methodol 2003 Nov 24;3(1):26 [FREE Full text] [doi: 10.1186/1471-2288-3-26] [Medline: 14633287]
42. Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H. A practical guide to Bayesian group sequential designs. Pharm Stat 2014 Aug 24;13(1):71-80. [doi: 10.1002/pst.1593] [Medline: 24038922]
43. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006 Jan;5(1):27-36. [doi: 10.1038/nrd1927] [Medline: 16485344]
44. Bendtsen M. The P value line dance: when does the music stop? J Med Internet Res 2020 Aug 27;22(8):e21345 [FREE Full text] [doi: 10.2196/21345] [Medline: 32852275]
45. GBD 2013 Risk Factors Collaborators, Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386(10010):2287-2323. [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]
46. Bendtsen M. Heterogeneous treatment effects of a text messaging smoking cessation intervention among university students. PLoS One 2020 Mar 5;15(3):e0229637 [FREE Full text] [doi: 10.1371/journal.pone.0229637] [Medline: 32134977]
47. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
48. Bendtsen M, McCambridge J. Causal models accounted for research participation effects when estimating effects in a behavioral intervention trial. J Clin Epidemiol 2021 Aug;136:77-83 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.03.008] [Medline: 33727133]

Abbreviations

MCAR: missing completely at random
NEXit: Nicotine Exit
OR: odds ratio

Edited by G Eysenbach; submitted 25.04.21; peer-reviewed by A Pandey; comments to author 14.05.21; revised version received 14.05.21; accepted 19.09.21; published 21.10.21.

Please cite as:
Bendtsen M, Bendtsen P, Müssener U
Six-Month Outcomes from the NEXit Junior Trial of a Text Messaging Smoking Cessation Intervention for High School Students: Randomized Controlled Trial With Bayesian Analysis
JMIR Mhealth Uhealth 2021;9(10):e29913
URL: https://mhealth.jmir.org/2021/10/e29913
doi:10.2196/29913
PMID:34673532

Original Paper

Post-COVID Public Health Surveillance and Privacy Expectations in the United States: Scenario-Based Interview Study

John S Seberger[1], PhD; Sameer Patil[2], PhD

[1]College of Communication Arts & Sciences, Michigan State University, East Lansing, MI, United States

[2]School of Computing, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
John S Seberger, PhD
College of Communication Arts & Sciences
Michigan State University
404 Wilson Rd
East Lansing, MI, 48824
United States
Phone: 1 (517) 416 0743
Email: seberge1@msu.edu

Abstract

Background: Smartphone-based apps designed and deployed to mitigate the COVID-19 pandemic may become infrastructure for postpandemic public health surveillance in the United States. Through the lenses of privacy concerns and user expectations of digital pandemic mitigation techniques, we identified possible long-term sociotechnical implications of such an infrastructure.

Objective: We explored how people in the United States perceive the possible routinization of pandemic tracking apps for public health surveillance in general. Our interdisciplinary analysis focused on the interplay between privacy concerns, data practices of surveillance capitalism, and trust in health care providers. We conducted this analysis to achieve a richer understanding of the sociotechnical issues raised by the deployment and use of technology for pandemic mitigation.

Methods: We conducted scenario-based, semistructured interviews (n=19) with adults in the United States. The interviews focused on how people perceive the short- and long-term privacy concerns associated with a fictional smart thermometer app deployed to mitigate the "outbreak of a contagious disease." In order to elicit future-oriented discussions, the scenario indicated that the app would continue functioning "after the disease outbreak has dissipated." We analyzed interview transcripts using reflexive thematic analysis.

Results: In the context of pandemic mitigation technology, including app-based tracking, people perceive a core trade-off between public health and personal privacy. People tend to rationalize this trade-off by invoking the concept of "the greater good." The interplay between the trade-off and rationalization forms the core of sociotechnical issues that pandemic mitigation technologies raise. Participants routinely expected that data collected through apps related to public health would be shared with unknown third parties for the financial gain of the app makers. This expectation suggests a perceived alignment between an app-based infrastructure for public health surveillance and the broader economics of surveillance capitalism. Our results highlight unintended and unexpected sociotechnical impacts of routinizing app-based tracking on postpandemic life, which are rationalized by invoking a nebulous concept of the greater good.

Conclusions: While technologies such as app-based tracking could be useful for pandemic mitigation and preparedness, the routinization of such apps as a form of public health surveillance may have broader, unintentional sociotechnical implications for individuals and the societies in which they live. Although technology has the potential to increase the efficacy of pandemic mitigation, it exists within a broader network of sociotechnical concerns. Therefore, it is necessary to consider the long-term implications of pandemic mitigation technologies beyond the immediate needs of addressing the COVID-19 pandemic. Potential negative consequences include the erosion of patient trust in health care systems and providers, grounded in concerns about privacy violations and overly broad surveillance.

(JMIR Mhealth Uhealth 2021;9(10):e30871) doi:10.2196/30871

KEYWORDS

COVID-19; pandemic-tracking apps; privacy concerns; infrastructure; health surveillance; scenario; interview; thematic analysis

Introduction

Background

The COVID-19 pandemic has raised profound concerns about the potential effects of app-based public health surveillance. On one hand, apps for contact tracing and exposure notification — which we refer to as "pandemic tracking apps" [1] — are potentially useful for pandemic mitigation [2-5] and preparedness for future pandemics [6]. In this regard, such apps are understood to contribute to "the greater good" by helping achieve stable and acceptable levels of public health [6,7]. On the other hand, the same technologies have stoked fears of mass surveillance with an ever-increasing scope [8-12]. In this regard, the relationship between pandemic tracking apps and the long-term greater good is less clear. Even positive assessments of such apps acknowledge their inherently "creepy" nature [13], implicitly underlining a parallel between the restoration of public health and the data-driven economics of surveillance capitalism [14].

Mobile health (mHealth) research is fertile ground for improving quality of life (eg, [15-17]). The relationship between short- and long-term forms of the greater good that may be achieved through the routinization of pandemic tracking apps deserves deeper consideration in terms of the broader quality of life these apps could foster. These considerations include people's beliefs about the right to privacy in an increasingly technology-mediated world, people's trust in the health care institutions that provide care to them, and the affective outcomes of such beliefs.

While other work at the intersection of privacy and pandemic mitigation technology has (rightfully) focused on the immediacy of the pandemic [7,18,19], we approach people's perceptions about pandemic tracking apps with a focus on the long-term implications. Echoing prior work in human-computer interaction (HCI) [1], we adopted a future-oriented lens to analyze how people understand the potential ramifications of app-based public health surveillance after the pandemic. We paid particular attention to the potential effects of such surveillance on everyday sociotechnical conditions (ie, the ways in which social norms and technological capabilities mutually influence each other and thereby shape daily life) [20,21].

We begin by covering salient literature in the sections that follow. Given the disciplinary intersectionality of this work, we discuss related work from several domains.

Health, Surveillance, and Pandemic Tracking Apps

Individual health, public health, and privacy converge within the field of health surveillance. The history of health surveillance (eg, [22,23]) shows a tendency toward general surveillance [24]. Although the ultimate object of health surveillance is a specific disease, such surveillance necessarily includes the person that carries that disease. Recent work has framed the COVID-19 pandemic as an opportunity for heightened general surveillance [8,12,25]. Such framing is appropriate because of 2 characteristics of technology use at the time of the COVID-19 pandemic. First, pandemic mitigation techniques, such as quarantines and shutdowns, have increased people's daily reliance on technology [26]. Second, technological solutions to the pandemic are entangled with institutions whose survival is predicated on data-driven "dehumanization" of the user [27] and corporate initiatives that foster people's resignation to accepting privacy violations as inevitable [28]. Both characteristics are broadly aligned with the data-fueled economics of surveillance capitalism [14], in which people are implicitly operationalized as monetizable sources of data that are trafficked and profited upon through practices such as personalized advertising.

For example, in the United States, Google and Apple have worked together to create an exposure notification infrastructure [5]. The infrastructure these parties have developed is privacy-preserving, but in a short-term way. That is, the proposed and implemented privacy protections focus on limiting the identifiability of data, rather than understanding people's perceptions of the data collection in the long term. Researchers have argued that the short-term measures taken by Google and Apple may limit the value of the data collected through such an infrastructure [2]. Such an argument implies that app-based data should be squeezed for all the information it might provide. Digital pandemic interventions elsewhere in the world have leveraged similarly large-scale corporate or military platforms [25]. Such collaborations between industry, military, and the government highlight the blurry line between health surveillance and the data-driven objectification of the user [8].

Despite the murkiness of the economic, institutional, and social realms in which pandemic tracking apps function, such apps are clearly aligned with the foundational logic of health surveillance (ie, [22,29-31]) and have well-defined benefits in terms of public health. However, they raise concerns about what is termed as "surveillance creep" (ie, the tendency for the surveillance abilities of a technology used in one context to transfer to a similar technology in a different context, thereby achieving subtle forms of social control [32]). For example, in the case of the Google/Apple collaboration, surveillance creep manifests as the extension of platform-specific surveillance capabilities (eg, user tracking, geolocation) into the domain of public health. Surveillance creep in public health may have a profound impact on the postpandemic everyday lives of users [11,24]. This is particularly apparent in light of medical-technical programs that prescribe smartphones as part of mitigation routines (cf. [33]).

Broad concerns about surveillance creep are manifested in short-term concerns about end-user privacy. By short-term concerns, we mean design-oriented concerns intended to increase the use of apps without sufficiently considering the long-term implications of widespread adoption. For example, recent work has shown that mHealth apps do not typically provide sufficient information about the third parties to whom data access might be provided [34]. Similar work has identified widespread insufficiencies in mHealth privacy policies [35]. However, no work to date has focused specifically on the potential routinization of pandemic tracking apps as infrastructure for public health surveillance. Two necessary areas of inquiry regarding the impact of routinization are (1) people's beliefs about their rights to privacy in an increasingly

technology-mediated world and (2) people's perceptions of the health care systems that provide care to them.

Pandemic Tracking Apps as Potential Infrastructure

Pandemic tracking apps may form future infrastructure for public health surveillance. As such, the concept of the "hopeful monster" is relevant to the potential routinization of pandemic tracking apps. The hopeful monster is a technology that aspires to the functional invisibility of infrastructure [36,37]. In this way, the hopeful monster exists as a precursor to a successful technology, where successful technologies are those that become invisible through routinization [38].

However, the potential success of pandemic tracking apps, and therefore the likelihood of their invisibility, remains uncertain. In the short-term, pandemic tracking apps will be considered successful if they prove to be effective in controlling the current COVID-19 pandemic. In the long-term, however, it is naïve to assume that a successful, population-scale mode of health surveillance will be dismantled after the pandemic or used only for its originally intended purpose (ie, surveillance of the COVID-19 pandemic) [39]. Indeed, research is already underway to examine the applicability of pandemic tracking infrastructure for tracking other diseases including HIV/AIDS in the United States [40], the benefits of routinizing app-based surveillance in the form of smart technologies [41], and the development of international interoperability for pandemic tracking apps [42]. Recent work has further highlighted the likelihood of pandemic tracking apps being routinized as future means of pandemic preparedness and mitigation [6]. The effects that such apps have on their users will likely extend beyond the end of the current pandemic.

Technology adoption models have been employed to understand how and under what conditions individuals may adopt pandemic tracking apps [18,43,44]. Once a technology is adopted for use, the satisfaction that users feel when an adopted technology "does what it is supposed to do" fosters continued use [45], particularly in mHealth apps [44,46]. Moreover, people have been shown to approach apps that serve the greater good with less privacy wariness and greater willingness to overlook privacy concerns [6,47]. Assuming that adoption of pandemic tracking apps is effective for mitigating the spread of the COVID-19 pandemic, such apps are primed for continued use in nonpandemic circumstances and become typical means for public health surveillance [1,6].

Data collected from a sample of Amazon Mechanical Turk workers in the United States indicate a generally favorable expectation to continue using pandemic tracking apps for general health monitoring even after the COVID-19 pandemic has been controlled [1]. People demonstrating a collectivist social orientation are more likely to continue using such apps for general health monitoring. Similarly, younger users expect to continue using such apps after the pandemic [1], aligning with their higher inclination to adopt pandemic tracking apps [48].

At the same time, Seberger and Patil [1] found that privacy concerns regarding mobile apps are negatively correlated with the expectation to use pandemic tracking apps as infrastructure for general health surveillance. People who are more concerned about privacy when using mobile apps are less likely to use pandemic tracking apps for general health monitoring *after* the pandemic. However, no such relationship was found between privacy concerns and the perceived benefit or expected use of pandemic tracking apps *during* the pandemic [1]. Given the prevalence of hyperbolic discounting [49] (ie, people's general inability to judge future impacts of current privacy decisions), it is likely that the relaxation of privacy concerns caused by the immediacy of the pandemic implicitly fosters long-term routinization of pandemic tracking apps.

Apps are designed to be used, and success in meeting user needs fosters their continued use [44,45]. If pandemic tracking apps prove to be successful in the mitigation of COVID-19, then it is possible that they will form the core of a more general, app-based form of public health surveillance. As we begin transitioning beyond the pandemic, it is essential to examine how people understand the sociotechnical implications of such an app-based infrastructure for public health surveillance that extends into the future.

The Relevance of Privacy to Pandemic Tracking Apps

We consider people's privacy concerns as a lens through which we can understand their expectations of postpandemic public health infrastructure. Theoretical approaches to end-user privacy concerns vary widely across scholarly disciplines [50]. Recently, there has been an increased focus on the politics of definitions and theorizations of privacy [51]. Whether privacy is understood to be relational [52,53], normative [54-57], calculative [58-60], or affective [57,61], policies and practices derived from such approaches necessarily shape the daily expectations and practices of the worlds in which they are realized. Such expectations and practices are increasingly characterized by the economics of surveillance capitalism [14].

The pervasiveness of end-user privacy concerns that arise in relation to surveillance capitalism [14] has been demonstrated to contribute to digital resignation [28], learned helplessness [62], and security fatigue [63,64]. Digital resignation refers to a social stance toward technology in which people are resigned to the inevitability of negative effects of technology. Digital resignation is thus a form of learned helplessness, which has been demonstrated to arise in relation to negative user experiences. Similarly, security fatigue refers to the taxing effects of maintaining security and privacy over time. It is possible to extrapolate from security fatigue [63,64] to the longer-term fatigue that may arise in the context of an ever-changing ecology of devices, apps, and protocols. Further, privacy researchers often encounter the so-called privacy paradox [65]: People tend to *state* that they are concerned about privacy and *act* as though they are not. Similar to recent work [66], we consider the privacy paradox obliquely through one of its symptoms: fatalistic attitudes toward the inevitability of app-based privacy intrusions.

To frame our work, we drew on research about the experience of creepiness in relation to technology use [62,67]. More specifically, we explored the potential for creepiness in modern platforms (eg, Facebook, Google, Amazon) that are capable of increasing levels of surveillance [20,68,69]. We approached our research in terms of the varied and ad hoc groupings of

technologies used in a person's daily life (ie, assemblages [70]). We further consider 2 privacy-related phenomena: the appearance of hyperfunctional infrastructures [71] and hyperbolic scaling [66]. Hyperfunctional infrastructures work according to their intended purpose but do so in ways that render them visible through the affective conditions they bring about [71]. Relatedly, hyperbolic scaling refers to user tendencies to project the negative characteristics of 1 app or category of apps onto the whole app ecology [66].

Research Question

Future public health may well benefit from the routinization of app-based surveillance based on the infrastructure resulting from the deployment of apps to track the COVID-19 pandemic. Given the immediacy of triage and the severe impact of COVID-19 on public health, it is tempting to focus exclusively on the short-term need to stabilize public health to an acceptable level. However, improved or stable public health is only one possible outcome of app-based public health surveillance. Other outcomes play out on a longer temporal scale.

The goal of our research was to understand the interplay between perceived public health benefits of pandemic tracking apps and their potential effects on nonmedical concerns (eg, sociotechnical conditions, surveillance capitalism). To that end, we addressed one core research question: How can people's privacy concerns about pandemic tracking apps improve our understanding of the sociotechnical ramifications of a potential postpandemic technical infrastructure for public health surveillance?

We identified a need to engage in the present work for 2 reasons. First, prior work has shown that people are interested in continuing to use pandemic tracking apps as infrastructure for public health surveillance [1]. Second, people expect such continued use to occur within a broader economy of surveillance capitalism [14]. Given that technology often "moves fast and breaks things" [10,39], the academic research community should devote energy to understanding these possible outcomes before they are realized. We adopted such a stance with the hope of engaging with the broader health-related research community.

Methods

Overview

To answer the research question posed in the previous section, we conducted semistructured scenario-based interviews (n=19) during the spring and summer of 2020 with adult participants from the United States. All study procedures were approved by the Institutional Review Board of Indiana University (Protocol #1902443119). We present the research according to the Standards for Reporting Qualitative Research guidelines [72].

The pragmatic nature of our research called for bottom-up qualitative analysis [73]. Building on prior human-centered work in HCI, we engaged in inductive analysis to surface beliefs about the possible ramifications of developing infrastructure for public health surveillance on the back of pandemic tracking apps. Recent work has highlighted the value of such inquiries for understanding the user experience of mHealth devices [74]. We analyzed interview transcripts by engaging in the inductive and constructivist process of exploratory pragmatic research using reflexive thematic analysis [75].

Scenario Development

The use of scenarios is common in privacy-related research (eg, [1,66,76-78]). We developed a scenario that describes a fictional pandemic tracking application linked to a smart thermometer:

> *During the outbreak of a contagious disease, you start using a popular smart thermometer app. Apart from recording your temperature, the app allows you to input additional symptoms you might be experiencing. Based on your symptoms, you receive suggestions for actions you should take to protect yourself and others in the community from the contagious disease. The app uses a combination of Bluetooth and location data to measure exposure to the disease in communities. The app allows the data to be accessed by the authorities, doctors, and scientists so that it can be used to track the spread of the disease and enforce people's compliance with containment measures, such as quarantines. The app continues to operate in the same manner even after the disease outbreak has dissipated.*

We created the scenario in accordance with best practices for scenario-based research, such that it was open to interpretation, realistic, and did not elicit "right or wrong" responses [79]. The semantic content of the scenario was iteratively developed by both authors based on analyses of existing smart health devices (eg, thermometers, glucose monitors) that may be reasonably paired with contact tracing apps. We piloted the scenario with several individuals of diverse backgrounds, including different native languages, professions and education levels, and genders and ages. While changes based on the pilots were minimal, they helped us refine word choices and sentence structure for improved clarity and comprehension of the scenario.

The resulting scenario satisfies Meinert's [79] best practices for scenario development by being realistic: Several smart thermometer apps are available to end users and are easily discoverable through any search engine. Each of these real-world, smart thermometers is paired with a smartphone-based app that stores and analyzes the data collected using the smart thermometer. We informed participants that the fictional smart thermometer app described in the scenario was "popular." The provision of such information adheres to Meinert's [79] best practices for scenario development relative to realism and openness to interpretation.

We added information on the app's use of Bluetooth and location data to satisfy the openness to interpretation criterion of scenario development [79]. We further described data collected by the app as being available to a vague set of actors including "the authorities, doctors, and scientists." The inclusion of such information additionally satisfies the realism requirement: It is common to describe pandemic tracking apps as using location data (eg, [80]) that, according to prior work on hyperbolic scaling [66] and digital resignation [28], can be expected to be accessed by unknown third parties. Empirical work has similarly shown that mHealth apps often do not present

transparent or full accounts of third parties to whom data access may be granted [34].

Further, the scenario referenced the continued functionality of pandemic tracking apps after the pandemic has been controlled. The inclusion of this information is in line with current trends in mHealth research that examine the usefulness of pandemic tracking apps for postpandemic monitoring of health conditions (eg, [40]). We included the information in order to engage participants in a discussion of the long-term implications of digital contact tracing.

Recruitment, Participants, and Interviews

We recruited participants through ads posted on Craigslist, Reddit, Facebook, and LinkedIn. Recruitment was entirely online due to restrictions on in-person research activities during the COVID-19 pandemic. The advertisements directed interested individuals to a screening questionnaire that asked for basic demographics (eg, age, gender, employment status, duration of residence in the United States). The complete screening questionnaire is available in Multimedia Appendix 1.

We used the responses to the screening questionnaire to compose a diverse sample. Given the exploratory nature of our study, we strove for a diverse pool of interviewees to explore as much of the terrain as possible. The sample diversity helped us collect rich data that include multiple perspectives [73].

We invited the selected individuals to participate in one-on-one interviews for which participants were compensated US $10. We specifically chose the one-on-one format to be sensitive to participant comfort given the likelihood that participants would discuss sensitive information (eg, personal experiences with COVID-19, privacy concerns) [81]. All interviews took place over Zoom and lasted between 45 minutes and 60 minutes. During the interviews, participants discussed several separate scenarios, including the one described in the earlier section. We ensured that participants understood the fictional nature of the scenarios. We included several scenarios to develop an understanding of privacy concerns across a wide range of apps. Participants were free to spend as long as they wanted to answer questions related to each of the scenarios. The portions of the interview guide pertaining to the smart thermometer scenario are available in Multimedia Appendix 2.

Table 1 presents an overview of participant demographics. Of the 19 participants, 9 (48%) identified as female, 8 (42%) as male, and 2 (11%) as nonbinary. The median age of the participants was 30 years. Of the 19 participants, 5 (26%) were university students, while the remaining 12 (74%) were professionals in fields varying from web design to package handling.

Table 1. Participant demographics.

Participant ID	Gender	Age (years)	Occupation
A	Female	25	Student (law)
B	Female	26	Web designer
C	Female	27	Student (human-computer interaction [HCI])
D	Nonbinary	21	Student (history)
E	Male	26	Writer
F	Male	29	Systems analyst
G	Female	23	Student (pharmacy)
H	Male	31	Student (informatics)
I	Female	26	Business analyst
J	Male	47	Call support
K	Female	25	Student (HCI)
L	Male	30	Software quality assurance
M	Nonbinary	30	Model
N	Male	36	Database admin
O	Female	42	Court clerk
P	Male	37	Sales executive
Q	Female	36	Homemaker
R	Male	36	Package handler
S	Female	46	Admin assistant

Analysis

We analyzed the transcripts of the interviews using thematic analysis, adhering to the multistage process described by Braun and Clarke [75]. Initial engagement with the data took the form of repeated and collaborative readings of interview transcripts among 5 members of the research team. This phase culminated

in open coding of the complete corpus of interview transcripts. The researchers then condensed the codes and identified emergent themes, arriving at a group of 5 initial themes. Upon further analysis, we reduced the initial 5 to the 3 themes we present in detail in the next section.

Results

In this section, we present the results of the thematic analysis. We begin with findings surfaced through the consideration of pandemic tracking apps as part of a broader app-enabled health care system. We proceed to analyze the privacy concerns that emerge from the relationship between pandemic tracking apps and the broader ecology of smartphone apps to which they belong. We build upon this narrative by introducing and analyzing contradictory forms of the greater good that emerge from participant responses. Our findings highlight that perceptions of pandemic tracking apps indicate an expectation that such apps will form another brick in the wall of surveillance that people increasingly experience in their everyday lives.

Public Health Versus Individual Privacy: The Perceived Trade- Off

Participants saw the use of pandemic tracking apps as a potentially effective way of interfacing with the health care system. For example, the fictional smart thermometer app described in the scenario was understood to serve a triage function that indicates when an individual should seek medical attention:

> *…it's really important to resolve worries that people might have on a personal level, but then also if their fears are justified and they might have a disease or something like that. I think it's also important for them to know that so that they can receive medical care or do what they need to do.* [Participant D, nonbinary, 21 years old, student (history)]

Participant D's perception of the relationship between pandemic tracking apps and the broader health care system is generally positive. However, further analysis of this common sentiment expressed by several participants revealed a bigger picture. The entanglement of pandemic tracking app data with institutions, including but not limited to health care systems, implies the presence of unknown third parties who may gain access to user or patient data. Such privacy concerns are inherited from the broader app culture [66] and present users with difficult choices:

> *I'm feeling a bit conflicted. Part of me is like, "Okay, interesting." Another part of me is like, "Oh, this is a little bit like Big Brother."* [Participant B, female, 26 years old, web designer]

People routinely expect apps to collect data about them and to share those data with third parties [66] as part of surveillance capitalism [14]. Through Participant B's invocation of "Big Brother," it is apparent that pandemic tracking apps are not an exception to this expectation. Participants were concerned about who would have access to the data generated by pandemic tracking apps during, as well as after, the pandemic. They were similarly concerned about what those unknown third parties might do with the data:

> *I think just being aware of who's using it, where is it being used, and how they're using it is one thing. Like, what are their practices, who is using it, and what company is it being used for? Is it being used by just health workers? Is it being used to be able to track and sell information to further recommend products to us?* [Participant K, female, 25 years old, student (HCI)]

Indeed, supported by the implicit connectivity of pandemic tracking apps, each participant raised concerns about who would ultimately gain access to any data generated and collected by these apps. Participants raised similar concerns about the kinds of data that might be generated, collected, and stored. Such concerns ultimately frame the deployment of public health surveillance infrastructure built on the back of pandemic tracking apps as a "slippery slope":

> *An acceptable purpose to me would be just to gather data on the disease and how it behaves in different communities and stuff. A malicious purpose would be them using this information to figure out where you live or more information about you, such as your age. Just in any way revealing your identity would be, to me, way more on the malicious side, although it's not necessarily malicious outright. But to me, it's a slippery slope.* [Participant F, male, 29 years old, systems analyst]

Participants further demonstrated the belief that data collected by contact tracing apps could be shared for profit with advertisers or as-yet-unknown third parties:

> *They could probably give it to advertisers or other people that I probably don't want it to go into the hands of. ...I can't think of a reason why it would be bad, but it just sounds bad, you know? It sounds sketchy. I don't know what advertisers could use other than for advertising like vitamin C or other get-well things. But either way, it just seems bad.* [Participant G, female, 23 years old, student (pharmacy)]

In discussing concerns that public health apps would share user data with unknown third parties, participants referred to a category of unknown third parties that would gain access to data, but rarely identified specific actors who might belong to that category. Instead, participants expressed the belief that public health apps would work "too well," thus facilitating the trafficking of data to unknown parties:

> *…it's just knowing that the application is working and doing its thing, but potentially it's working too well and that power getting into the hands of individuals that weren't necessarily designed to have that information to begin with.* [Participant G, female, 23 years old, student (pharmacy)]

In considering the routinization of pandemic tracking apps as infrastructure for postpandemic public health surveillance, participants were faced with a difficult and contradictory set of possibilities. In particular, we observed a fundamental perceived trade-off between personal privacy and public health. This core trade-off appeared as an either/or condition: People can expect either personal privacy or better public health. Achieving both

simultaneously did not seem realistic to participants. Moreover, participants seemed to place values in a hierarchy wherein public health was superior to personal privacy:

> *I value public health and public safety over personal privacy.* [Participant A, female, 25 years old, student (law)]

Participants routinely reported such higher consideration for public health over personal privacy even when expressing deep wariness regarding existing pandemic tracking infrastructure:

> *Google and Apple have added their tracing APIs to their operating system so that governments and health authorities could start pushing out these apps that do this tracing to track down where you've been and who you've been around to understand contagious disease, disease control, and all that stuff. I understand the benefit of it. But it's highly intrusive, right? It's tracking all your data, it's tracking who you're coming in contact with, it's tracking your location. It's kind of Big Brother-ish, in a way.* [Participant G, female, 23 years old, student (pharmacy)]

This quote from Participant G demonstrates that people are aware of pandemic tracking apps as they exist in the real world but generally hold an inaccurate understanding of their functionality. Participants expected pandemic tracking apps to trade identifiable data, even though the Apple/Google collaboration in the United States goes to great lengths to maintain user anonymity. Such inaccurate understanding likely arises from the pervasiveness of apps that use the sale of access to data as a primary economic motive [14,66]. However, that which is perceived to be real is real in its consequences. As such, when understanding people's perceptions and expectations of routinized pandemic tracking apps, our focus is the way in which people believe such apps to work, regardless of whether the actual functionality matches these beliefs [82].

The perception of pandemic tracking apps as inherently privacy-intrusive constitutes a core component of the discourse of such apps. Indeed, being wary of the privacy implications of pandemic-tracking apps has real-world consequences:

> *You said that you saw someone voicing some concerns. What was that concern?* [Interviewer]
>
> *Oh, this is somebody that I personally know, and she kind of is an outlier in terms of her thinking. But she had something on there. She knew these apps, she had named some, and I did not. I have not heard of the specific names of the apps, but she had some of the apps listed, and she said that if you have any of these apps and you're thinking about using them too, please delete me from your contacts, because I will not be traced and all this kind of stuff. It was kind of really out there.*" [Participant O, female, 42 years old, court clerk]

Despite general wariness and expectations that pandemic tracking apps will violate privacy by virtue of their connectivity with unknown third parties, participants routinely justified the necessity and use of pandemic tracking apps by referring to a nebulous concept — the greater good:

> *I'm fine with [using the app] for this particular reason, just because it's for the greater good of society to know who's sick and who could potentially be sick. I think that something that could help a lot of people and save lives kind of outweighs your right to privacy in some ways, so I don't have a problem.* [Participant O, female, 42 years old, court clerk]

As we describe in the next subsection, participants routinely employed "the greater good" as a means of justifying or overlooking the perceived long-term privacy concerns raised by pandemic tracking apps. Yet, we found that this greater good is multifaceted, and the relationship between the different facets of the greater good is contradictory, highlighting pandemic tracking apps as a potential site for surveillance creep: the colonization of health by the economics of surveillance capitalism [14].

Rationalizing the Health/Privacy Trade-Off With the Greater Good

Pandemic tracking apps exist within a wider ecology of apps. Therefore, privacy concerns that emerge from user interactions with this wider ecology color perceptions and expectations of pandemic tracking apps [66]. When pandemic tracking apps are expected to align with the economic practices of surveillance capitalism (ie, monetizing user data by selling it to third parties), privacy concerns about pandemic tracking apps take center stage. Such concerns include wariness regarding unknown third parties who may gain access to the data generated, collected, and stored by pandemic tracking apps and inaccurate, but nonetheless meaningful and deep-seated, perceptions of how such apps function. The greater good is the primary means by which people rationalize and justify engaging in what they perceive as the inherently privacy-affecting decision to use pandemic tracking apps. The use of the greater good to rationalize the adoption or routinization of pandemic tracking apps can be rather sharp, demonstrating frustration with the complexity of issues raised by such apps:

> *But at the same time, yeah, actually no ... I think it's fine because it just seems to take small vitals like temperature, and if they're probably sick, it might be fine. We'll say it's fine because it's for the greater good.* [Participant G, female, 23 years old, student (pharmacy)]

This excerpt depicts a trade-off in motion. Participant G quickly reverted to the greater good as a means of rationalization. She vacillated between further expressing her concerns about the fictional smart thermometer app — "but at the same time, yea, actually no" — and finding an easier conversational way out of such expression. Participant G's use of the phrase, "We'll say it's fine," indicates a clear tension: the implicit recognition of privacy-related problems and the desire to achieve the greater good despite them.

The scenario suggested that pandemic tracking apps may be used as a foundation for postpandemic public health monitoring. Participants generally reacted negatively to this possibility:

> *It would be most useful for you to share everything or as much as you can during a pandemic, but outside of a pandemic, it starts to be kind of a gray area.* [Participant F, male, 29 years old, systems analyst]

Invocation of the greater good is not sufficient to rationalize the long-term implications of pandemic tracking apps should their use be routinized as a means of health surveillance in general. The greater good that pandemic tracking apps serve appears to have a shelf life:

> *So, during the outbreak of a contagious disease, I would think it's worth it to give up privacy and data and information. If it's an app that will really help and literally save lives, I'd be happy to do that. But I don't like the last part that it continues to operate in this way after the outbreak has dissipated.* [Participant A, female, 25 years old, student (law)]

The greater good appears to refer specifically to the mitigation and control of the pandemic, but does not necessarily extend to the routinization of pandemic tracking apps as a form of postpandemic public health surveillance:

> *Obviously when there's an outbreak of a contagious disease, everyone's focus is really just on that. Now, of course, you never know that 100%. Other people can still utilize that data for malicious reasons or their own reasons. But you would think that there's something bad going on there. Everyone's trying to help as much as possible. You're okay with it. You're okay with exposing yourself in this way. It's just the app being able to be used in the same manner after disease outbreak has dissipated … that part is more uncomfortable.* [Participant L, male, 30 years old, software quality assurance]

It is not, however, Participant L's expected or experienced discomfort that takes the center stage. Rather, it is what such discomfort implies. While our scenario intentionally primed participants to consider postpandemic routinization of pandemic tracking apps, the phrasing was ambiguous as to postpandemic functionality of such apps. We allowed participants to extrapolate from their contemporary perceptions and expectations of pandemic tracking apps in order to understand what routinization of such apps might mean.

When read in terms of people's expectation that data collected by pandemic tracking apps will be shared with third parties for financial gain, Participant L's discomfort signals the expectation that the apps will contribute to a broader culture of surveillance. Yet, participants generally accepted the likelihood of such apps being routinized for public health surveillance after the pandemic, with phrases such as, "We'll get to that when we get to that."

> *It's just the app being able to be used in the same manner after disease outbreak has dissipated…that part is more uncomfortable because it's unnecessary information being exposed when it doesn't really need to be unless the app can sort of justify it, right? If it's justified and valid, let's say let's see. Doctors and scientists and authorities. Who are you? Well, I don't know who the authorities are, but I guess it is the police. I don't know. I guess we'll get to that when we get to that.* [Participant L, male, 30 years old, software quality assurance]

Taken together, participant discussion of the greater good and their willingness to forego privacy concerns for that greater good can be troubling. Even though it is defined only implicitly, the greater good becomes a floating signifier used to justify and rationalize the perceived privacy risks of contact-tracing applications:

> *The app continues to operate in the same manner even after the disease outbreak has disappeared? Doing something for the greater good always gets my vote. So, exposing my information like this, if it's toward a greater good like this, I will tell you, again, like I sort of mentioned before, I know what's happening.* [Participant L, male, 30 years old, software quality assurance]

People understand and even expect privacy breaches as a result of app-based contact tracing. Further, they see the possibility of "Big Brother" levels of surveillance as real. The invocation of "Big Brother" refers to the expectation that contact tracing apps will result in heightened surveillance. That is, people perceive that the routinization of pandemic tracking apps for public health surveillance will lead to oversurveillance by a heterogeneous set of actors (eg, technology corporations, government bodies, health care organizations, insurance providers, third-party advertisers) variously responsible for app maintenance, data analysis, public health regulation, and health care delivery. This amounts to a future that is not accurately characterized by widespread positive affective experience despite that future's roots in the greater good:

> *What do you imagine the enforcement with compliance could mean?* [Interviewer]
>
> *Nothing that I want to imagine in real life.* [Participant M, nonbinary, 30 years old, model]

Participant M implies that public health surveillance achieved through the routinization of pandemic tracking apps would lead to outcomes too negative to even imagine. Such expectations directly call into question the short- and long-term forms of the greater good to which adoption of pandemic tracking apps contributes. On one hand, the adoption of pandemic tracking apps can contribute to the greater good of public health during a pandemic. On the other hand, the long-term adoption of such apps as infrastructure for public health surveillance contributes to widespread concerns of mass surveillance and privacy violations that cannot reasonably be described as a greater good. In fact, the privacy ramifications could arguably erode the greater good.

Discussion

Two facets of our findings are particularly worthy of discussion. First, we describe the benefits and necessity of taking a long-term, sociotechnical approach to understanding the potential routinization of mobile apps as part of the public health infrastructure. Second, we describe how user privacy concerns

constitute a productive means for identifying long-term implications of an emergent class apps used for public health surveillance.

People Expect Privacy-Related Infrastructural Problems in Public Health Surveillance

Direct consideration of people's privacy concerns about pandemic tracking apps indicates their wariness about the routinization of such apps as infrastructure for postpandemic public health surveillance. Although users are willing to justify the use of apps that raise privacy concerns with the concept of the greater good, such justification has a limited shelf life. By stating expectations that privacy concerns related to pandemic tracking apps extend to mHealth apps in general, participants highlighted the need to decouple or disentangle mHealth infrastructure from the privacy-invasive practices of surveillance capitalism.

People are wary of pandemic tracking apps because they might "work too well" (Participant G) and therefore demonstrate hyperfunctionality as identified by Seberger and Bowker [71]: conditions wherein infrastructures function within the boundaries of their designed functionality but do so in a way that gives rise to unexpected outcomes. In other words, if pandemic tracking apps achieve the form of control over the spread of the pandemic that they are designed to achieve, such success might spill over to other uses. In our scenario, we pointed out the potential for such other uses in the form of postpandemic public health surveillance that uses such apps as its infrastructure.

Participants routinely assumed that data collected via pandemic tracking apps would be shared with an identified-but-unknown category of third parties. We interpret participant concerns over third-party access to data collected by pandemic tracking apps as a form of hyperbolic scaling [66], where concerns regarding privacy violations transfer from known apps to new apps. In this logic, if App A (a known app) demonstrates problematic functions relative to end-user privacy, then *all* apps are assumed to demonstrate the same characteristics.

Stakeholders Should Take a Long-Term View of Public Health Surveillance

A pandemic often affects people's daily lives and routines in profound ways. Because the conditions of a pandemic can be all-encompassing, it can be difficult to take a long-term perspective. Yet infrastructure, like surveillance, creeps. It is learned and normalized through use [37]. The misalignment between short- and long-term concerns is compounded by the immediacy of the language used to communicate and frame the threats of the pandemic [83]. With visions of large-scale pandemic tracking already being considered (eg, [6,40-42]), it is poised to become infrastructure for public health surveillance. We suggest that it is necessary to understand the impact of public health surveillance at a temporal scale aligned more closely with the long-term one of infrastructure [84]. People's judgments related to trade-offs that unfold at the scale of the short- and long-term are notoriously problematic [85]. The severity of the pandemic and its social effects emphasize the urgency of triage to fix the most immediate problems first by focusing on the greater good of public health. Yet, given the combined power of medicine and technology to achieve social change, it is not entirely appropriate to focus solely on the immediate threats posed by the pandemic. The immediate actions that governments, corporations, and institutions take in service of pandemic mitigation will reverberate in the sociotechnical structures of the postpandemic world [86].

People's perceptions of a core trade-off between individual privacy and public health highlight a need for long-term thinking about public health surveillance. People are willing to forego privacy concerns for the greater good even though they are uncomfortable with surveillance by the technologies used to achieve that greater good. Our study therefore supports the findings of prior work on the mitigating effects of health concerns on people's privacy concerns [87]. We provide further support for findings derived from a study with UK participants who saw the greater good as a motivating factor for app use [7]. We extend the work of Williams et al [7] by adding practical nuance to the role that the greater good plays in pandemic mitigation and surfacing the potentially slippery slope toward increased and routinized public health surveillance.

The likely impacts of app-based public health surveillance extend beyond public health itself. This extension includes potential impact on people's expectations of privacy as well as colonization of mHealth by the economics of surveillance capitalism. Such colonization may lead to diminished trust between patients and health care providers. Given people's deep-seated privacy concerns, app creators, policy makers, and health institutions should engage in detailed and transparent public relations work to communicate the ways in which user privacy is protected. Such communications should describe exactly which third parties, if any, will be given access to data collected through apps used for public health surveillance and under what circumstances. While privacy concerns raised by pandemic tracking apps are likely as widespread as the pandemic itself, the steps described should be taken within existing local legal frameworks (eg, Health Insurance Portability and Accountability Act [HIPAA], General Data Protection Regulation [GDPR]). Overall, our insight highlights the need for an interdisciplinary approach to the development and deployment of app-based public health surveillance.

It is Necessary to Disentangle Public Health Surveillance From Surveillance Capitalism

People's perceptions of the alignment between public health apps and the economics of surveillance capitalism may have a detrimental effect on the relationship between patients and health care systems. Trust between patients and health care providers is foundational to effective medical care [88]. As such, the enrollment of patients as users in a sociotechnical system contextualized by potential mistrust serves neither the patient nor health care professionals. In a way, it breaches the core medical credo of "first do no harm."

Users are inherently vulnerable to the power structures instantiated by the apps that they use [66]. But, when users are also patients, they are doubly vulnerable: vulnerable to the conditional empowerment implicit in app use [66] and vulnerable to the loss of control that comes with the status of

being a patient [89] (perhaps they are also vulnerable to the dehumanization of data-driven representation [27]). If pandemic tracking apps achieve the status of infrastructure — invisibility through successful deployment and adoption [37,38] — they may do so at the risk of solidifying the vulnerability-inducing conditions of surveillance capitalism within a medical context [14]. Such solidification would directly contrast with the notion of care central to medical practice. In this light, a broader, socially oriented approach to the medical credo, "first do no harm," would benefit end users of pandemic tracking apps.

We contend that people's expectation of, and resignation to, being used may foreshadow the colonization of public health surveillance by the data-hungry and profit-driven mechanics of surveillance capitalism. We further contend that this is as much a public health issue as the maintenance of more narrowly defined bodily health. We therefore suggest the development of public health policies that account for the long-term sociotechnical effects of public health surveillance. Such policies, developed within the legal frameworks of specific regions, would necessarily view patients not solely with the medical gaze, but through a more sensitive lens that considers end-user empowerment, trust in digital systems, and the emerging broader relationship between health care and surveillance capitalism.

We further suggest that the activation of infrastructure for public health surveillance be legislatively limited to times of scientifically defined public health crises. It would be unwise to discard infrastructure that's already been built. The Google and Apple collaboration in the United States, for example, likely constitutes a useful resource for pandemic preparedness. However, such infrastructure does not necessarily present an ethical or beneficial means of general public health surveillance, given the profound privacy concerns about pandemic tracking apps. The routinization of pandemic tracking apps as infrastructure for public health surveillance has the potential for profound negative impact on the trust that should characterize the relationship between patients and health care providers. We found that people believe that health surveillance apps will operate according to the economics of surveillance capitalism, wherein access to data is sold for profit, typically without much regard for user privacy concerns. When people believe that health surveillance apps align with surveillance capitalism, it is unreasonable to expect them to trust the system that profits from the sale of their data. Such a dynamic sets up the potential for such apps to diminish patient trust in the health care systems that ostensibly care for them.

Given the dangers present in the perceived alignment between app-based public health surveillance and surveillance capitalism, we suggest that the activation of public health surveillance be limited to populations that meet specific criteria. Such criteria might include population density, smartphone saturation, and rate of disease transmission. In this way, an app for public health surveillance might be utilized to the greatest effect, while minimizing the potential for surveillance creep. Beyond meeting well-defined medical standards, such population-specific deployment should be overseen by multidisciplinary teams capable of assessing the sociotechnical and surveillance-related impact on populations who are already, by virtue of an infectious agent, fundamentally at risk.

Limitations and Future Work

We collected data during a specific period from a relatively small group of participants. Since the COVID-19 pandemic is still ongoing, additional studies are needed to investigate how the trajectory of the pandemic might influence these matters. Public health and privacy concerns are inherently cultural. Further studies involving participants from other populations would add greater context to our findings. In particular, it is possible that our findings may be specific only to the United States. Many social norms and expectations in the United States differ from those in other countries. In the United States, perceived economic relationships between government agencies, health care providers, and for-profit corporations (eg, app makers, health insurance companies) create a structure in which people may expect their privacy to be breached. Such relationships qualify people's trust (or lack thereof) in government that may be different than that in other regions across the world. Nonetheless, our findings highlight the need for further work regarding the role of trust in institutions and the possible success of app-based public health surveillance after the pandemic has ended. Further work is required to understand perceptions and expectations of public health apps in other cultural contexts.

Given the exploratory nature of this work, we did not consider ethnicity as a factor in our sample construction. However, when choosing interviewees from a pool of eligible participants, we did strive for diversity in ages, genders, professions, and educational backgrounds. Prior work [7] has not analyzed the effect of ethnicity either nor can a qualitative study of limited sample size such as ours yield robust findings relative to such a sensitive topic. That said, we recognize the profound importance of ethnicity in matters pertaining to the greater good and identify this as an important area for further research.

Conclusion

Given that pandemic tracking apps have been promoted as a useful tool for pandemic mitigation, it is likely that such apps will form the backbone of postpandemic infrastructure for public health surveillance. While it is clearly important to understand the short-term, pandemic-specific privacy concerns that arise from such apps, we contend that it is equally important to understand their long-term sociotechnical implications. People frame the use of pandemic tracking apps in terms of achieving a greater good: acceptable levels of public health. However, such framing is contextualized by a core trade-off with individual privacy. The trade-off reveals a deep-seated tension in the use of the greater good rationalization: the perception that pandemic tracking apps implicitly align with the data-hungry economics of surveillance capitalism. To make matters worse, people are resigned to viewing such alignment as inevitable, which may harm their trust in the institutions that provide health care. We highlight the relevance of analyzing privacy concerns when considering the potential routinization of apps for public health surveillance and call for the multidisciplinary application of medically influenced ethics in

the design, development, deployment, and data use of mHealth apps designated for everyday use.

Acknowledgments

The research described in this paper was undertaken while the authors were affiliated with Indiana University Bloomington. We thank the participants of the study for their time and insight. We acknowledge Marissel Llavore and Nicholas Nye Wyant for their contributions to data collection. We are grateful to the anonymous reviewers for their feedback that helped improve this paper. This research was supported by a grant from Indiana University Office of the Vice President for Research. Open access to this article is funded in part by the Indiana University Open Access Fund. Contents of the paper are the work of the authors and do not necessarily reflect the views of the sponsors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screening questionnaire.
[DOCX File , 15 KB - mhealth_v9i10e30871_app1.docx]

Multimedia Appendix 2
Scenario.
[DOCX File , 18 KB - mhealth_v9i10e30871_app2.docx]

References

1. Seberger JS, Patil S. Us and Them (and It): Social Orientation, Privacy Concerns, and Expected Use of Pandemic-Tracking Apps in the United States. 2021 Presented at: Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445485]
2. Seto E, Challa P, Ware P. Adoption of COVID-19 Contact Tracing Apps: A Balance Between Privacy and Effectiveness. J Med Internet Res 2021 Mar 04;23(3):e25726 [FREE Full text] [doi: 10.2196/25726] [Medline: 33617459]
3. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
4. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
5. Virtual Workshop on Privacy Aspects of Contact Tracing. ProperData. 2020. URL: https://properdata.eng.uci.edu/events/privacy-contact-tracing/ [accessed 2021-09-23]
6. Utz C, Becker S, Schnitzler T, Farke FM, Herbert F, Schaewitz L, et al. Apps Against the Spread: Privacy Implications and User Acceptance of COVID-19-Related Smartphone Apps on Three Continents. Cornell University. 2021 Feb 01. URL: https://arxiv.org/abs/2010.14245 [accessed 2021-09-23]
7. Williams S, Armitage C, Tampe T, Dienes K. Public attitudes towards COVID-19 contact tracing apps: A UK-based focus group study. Health Expect 2021 Apr;24(2):377-385 [FREE Full text] [doi: 10.1111/hex.13179] [Medline: 33434404]
8. Desai BC. Pandemic and big tech. 2020 Presented at: 24th Symposium on International Database Engineering & Applications; August 12-14, 2020; Seoul, Republic of Korea. [doi: 10.1145/3410566.3410585]
9. French MA. Woven of War-Time Fabrics: The globalization of public health surveillance. Surveillance & Society 2009 Feb 27;6(2):101-115. [doi: 10.24908/ss.v6i2.3251]
10. COVID-19 apps pose serious human rights risks: recommendations for governments considering technology in addressing pandemic tech. Human Rights Watch. 2020 May 13. URL: https://www.hrw.org/news/2020/05/13/covid-19-apps-pose-serious-human-rights-risks [accessed 2021-09-23]
11. Leslie D. Tackling COVID-19 through responsible AI innovation: Five steps in the right direction. Harvard Data Science Review 2020. [doi: 10.1162/99608f92.4bb9d7a7]
12. Kitchin R. Civil liberties or public health, or civil liberties or public health? Using surveillance technologies to tackle the spread of COVID-19. Space and Polity 2020 Jun 03;24(3):362-381. [doi: 10.1080/13562576.2020.1770587]
13. Bell G. We need mass surveillance to fight COVID-19 - but it doesn't have to be creepy. MIT Technology Review. 2020 Apr 12. URL: https://www.technologyreview.com/2020/04/12/999186/covid-19-contact-tracing-surveillance-data-privacy-anonymity/ [accessed 2021-09-23]
14. Zuboff S. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. London, England: Profile Books; 2019.

15. Stoyanov SR, Zelenko O, Staneva A, Kavanagh DJ, Smith C, Sade G, et al. Development of the Niggle App for Supporting Young People on Their Dynamic Journey to Well-being: Co-design and Qualitative Research Study. JMIR Mhealth Uhealth 2021 Apr 20;9(4):e21085 [FREE Full text] [doi: 10.2196/21085] [Medline: 33877050]
16. Shalaby R, Vuong W, Hrabok M, Gusnowski A, Mrklas K, Li D, et al. Gender Differences in Satisfaction With a Text Messaging Program (Text4Hope) and Anticipated Receptivity to Technology-Based Health Support During the COVID-19 Pandemic: Cross-sectional Survey Study. JMIR Mhealth Uhealth 2021 Apr 15;9(4):e24184 [FREE Full text] [doi: 10.2196/24184] [Medline: 33750738]
17. MacIsaac A, Mushquash AR, Mohammed S, Grassia E, Smith S, Wekerle C. Adverse Childhood Experiences and Building Resilience With the JoyPop App: Evaluation Study. JMIR Mhealth Uhealth 2021 Jan 04;9(1):e25087 [FREE Full text] [doi: 10.2196/25087] [Medline: 33393908]
18. Walrave M, Waeterloos C, Ponnet K. Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
19. Liu C, Graham R. Making sense of algorithms: Relational perception of contact tracing and risk assessment during COVID-19. Big Data & Society 2021 Feb 17;8(1):205395172199521. [doi: 10.1177/2053951721995218]
20. Tene O, Polonetsky J. A theory of creepy: technology, privacy and shifting social norms. Yale JL & Tech 2013;16:59.
21. Berg M, Aarts J, van der Lei J. ICT in Health Care: Sociotechnical Approaches. Methods Inf Med 2018 Feb 08;42(04):297-301. [doi: 10.1055/s-0038-1634221]
22. Langmuir AD. The Surveillance of Communicable Diseases of National Importance. N Engl J Med 1963 Jan 24;268(4):182-192. [doi: 10.1056/nejm196301242680405]
23. Langmuir AD. Communicable Disease Surveillance. Proceedings of the Royal Society of Medicine 2016 Sep;64(6):681-684. [doi: 10.1177/003591577106400646]
24. French M, Monahan T. Disease Surveillance: How Might Surveillance Studies Address COVID-19? Surveillance & Society 2020 Mar 16;18(1):1-11. [doi: 10.24908/ss.v18i1.13985]
25. Calvo RA, Deterding S, Ryan RM. Health surveillance during covid-19 pandemic. BMJ 2020 Apr 06;369:m1373. [doi: 10.1136/bmj.m1373] [Medline: 32253180]
26. Goldkind L, LaMendola W, Taylor-Beswick A. Tackling COVID-19 is a crucible for privacy. Journal of Technology in Human Services 2020 May 02;38(2):89-90. [doi: 10.1080/15228835.2020.1757559]
27. Chancellor S, Baumer EPS, De Choudhury M. Who is the "Human" in Human-Centered Machine Learning. Proc ACM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-32. [doi: 10.1145/3359249]
28. Draper NA, Turow J. The corporate cultivation of digital resignation. New Media & Society 2019 Mar 08;21(8):1824-1839. [doi: 10.1177/1461444819833331]
29. Foege WH, Hogan RC, Newton LH. Surveillance projects for selected diseases. Int J Epidemiol 1976 Mar;5(1):29-37. [doi: 10.1093/ije/5.1.29] [Medline: 944166]
30. Lee LM, Thacker SB, Centers for Disease ControlPrevention (CDC). The cornerstone of public health practice: public health surveillance, 1961-2011. MMWR Suppl 2011 Oct 07;60(4):15-21. [Medline: 21976162]
31. Lee LM. An Ethics for Public Health Surveillance. Am J Bioeth 2020 Oct 18;20(10):61-63. [doi: 10.1080/15265161.2020.1806382] [Medline: 33016826]
32. Marx GT. Undercover: police surveillance in America. Oakland, CA: University of California Press; 1988:2.
33. Kazevman G, Mercado M, Hulme J, Somers A. Prescribing Phones to Address Health Equity Needs in the COVID-19 Era: The PHONE-CONNECT Program. J Med Internet Res 2021 Apr 06;23(4):e23914 [FREE Full text] [doi: 10.2196/23914] [Medline: 33760753]
34. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the Privacy of mHealth Apps for Self-Tracking: Heuristic Evaluation Approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
35. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Assessment of the Fairness of Privacy Policies of Mobile Health Apps: Scale Development and Evaluation in Cancer Apps. JMIR Mhealth Uhealth 2020 Jul 28;8(7):e17134 [FREE Full text] [doi: 10.2196/17134] [Medline: 32720913]
36. Law J. A Sociology of monsters: Essays on power, technology, and domination. London, England: Routledge; 1991.
37. Star SL, Ruhleder K. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research 1996 Mar;7(1):111-134. [doi: 10.1287/isre.7.1.111]
38. Weiser M. The computer for the 21st century. SIGMOBILE Mob. Comput. Commun. Rev 1999 Jul 01;3(3):3-11. [doi: 10.1145/329124.329126]
39. Khosla R. Technology, Health, and Human Rights: A Cautionary Tale for the Post-Pandemic World. Health Hum Rights 2020 Dec;22(2):63-66 [FREE Full text] [Medline: 33390694]
40. Nosyk B, Armstrong W, Del Rio C. Contact Tracing for COVID-19: An Opportunity to Reduce Health Disparities and End the Human Immunodeficiency Virus/AIDS Epidemic in the United States. Clin Infect Dis 2020 Nov 19;71(16):2259-2261 [FREE Full text] [doi: 10.1093/cid/ciaa501] [Medline: 32339245]

41. Vedaei SS, Fotovvat A, Mohebbian MR, Rahman GME, Wahid KA, Babyn P, et al. COVID-SAFE: An IoT-Based System for Automated Health Monitoring and Surveillance in Post-Pandemic Life. IEEE Access 2020;8:188538-188551. [doi: 10.1109/access.2020.3030194]
42. Du L, Raposo VL, Wang M. COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e23194. [doi: 10.2196/23194]
43. Walrave M, Waeterloos C, Ponnet K. Ready or Not for Contact Tracing? Investigating the Adoption Intention of COVID-19 Contact-Tracing Technology Using an Extended Unified Theory of Acceptance and Use of Technology Model. Cyberpsychol Behav Soc Netw 2021 Jun;24(6):377-383. [doi: 10.1089/cyber.2020.0483] [Medline: 33017171]
44. Woldeyohannes HO, Ngwenyama OK. Factors Influencing Acceptance and Continued Use of mHealth Apps. In: Nah FH, Tan CH, editors. HCI in Business, Government and Organizations. Interacting with Information Systems. HCIBGO 2017. Lecture Notes in Computer Science, vol 10293. Cham, Switzerland: Springer; 2017:239-256.
45. Chaouali W. Once a user, always a user: Enablers and inhibitors of continuance intention of mobile social networking sites. Telematics and Informatics 2016 Nov;33(4):1022-1033. [doi: 10.1016/j.tele.2016.03.006]
46. 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]
47. Baig K, Mohamed R, Theus AL, Chiasson S. "I'm hoping they're an ethical company that won't do anything that I'll regret": Users Perceptions of At-home DNA Testing Companies. 2020 Presented at: Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376800]
48. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
49. Acquisti A, Grossklags J. Losses, gains, and hyperbolic discounting: an experimental approach to information security attitudes and behavior in 2nd Annual Workshop on Economics and Information Security-WEIS;3(Berkeley, CA)?27UC Berkeley 2003. 2003 Presented at: 2nd Annual Workshop on “Economics and Information Security”; March 18, 2003; Berkeley, CA.
50. Igo SE. The known citizen: A history of privacy in modern America. Cambridge, MA: Harvard University Press; 2018.
51. McDonald N, Forte A. The politics of privacy theories: moving from norms to vulnerabilities. 2020 Presented at: Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376167]
52. Gove WR, Altman I. The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding. Contemporary Sociology 1978 Sep;7(5):638. [doi: 10.2307/2065073]
53. Bannerman S. Relational privacy and the networked governance of the self. Information, Communication & Society 2018 May 29;22(14):2187-2202. [doi: 10.1080/1369118x.2018.1478982]
54. Petronio S. Boundaries of privacy: Dialectics of disclosure. Albany, NY: SUNY Press; 2002.
55. Nissenbaum H. Privacy in context: Technology, policy, and the integrity of social life. Redwood City, CA: Stanford University Press; 2009.
56. Crabtree A, Tolmie P, Knight W. Repacking 'Privacy' for a Networked World. Comput Support Coop Work 2017 May 29;26(4):453-488 [FREE Full text] [doi: 10.1007/s10606-017-9276-y] [Medline: 32025101]
57. Nippert-Eng CE. Islands of privacy. Chicago, IL: University of Chicago Press; 2010:0226586537.
58. Margulis ST. Privacy as a social issue and behavioral concept. Journal of Social Issues 2003;59:261. [doi: 10.1111/1540-4560.00063]
59. Solove DJ. A Taxonomy of Privacy. University of Pennsylvania Law Review 2006;154(3):477-564. [doi: 10.2307/40041279]
60. Solove DJ. Understanding privacy. Cambridge, MA: Harvard University Press; 2008.
61. Stark L. The emotional context of information privacy. The Information Society 2015 Dec 22;32(1):14-27. [doi: 10.1080/01972243.2015.1107167]
62. Shklovski I, Mainwaring SD, Skuladottir HH, Borgthorsson H. Leakiness and creepiness in app space: perceptions of privacy and mobile app use. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Ontario, Canada p. 2347-2356. [doi: 10.1145/2556288.2557421]
63. Stanton B, Theofanos MF, Prettyman SS, Furman S. Security Fatigue. IT Prof 2016 Sep;18(5):26-32. [doi: 10.1109/mitp.2016.84]
64. Furnell S, Thomson K. Recognising and addressing ‘security fatigue’. Computer Fraud & Security 2009 Nov;2009(11):7-11. [doi: 10.1016/s1361-3723(09)70139-3]
65. Norberg PA, Horne DR, Horne DA. The privacy paradox: personal information disclosure intentions versus behaviors. The Journal of Consumer Affairs 2007;41:126. [doi: 10.1111/j.1745-6606.2006.00070.x]
66. Seberger JS, Llavore M, Wyant NN, Shklovski I, Patil S. Empowering Resignation: There’s an App for That. 2021 Presented at: Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445293]
67. Paasonen S. As Networks Fail. Television & New Media 2014 Oct 08;16(8):701-716. [doi: 10.1177/1527476414552906]
68. Cumbley R, Church P. Is “Big Data” creepy? Computer Law & Security Review 2013 Oct;29(5):601-609. [doi: 10.1016/j.clsr.2013.07.007]

69. West E. Amazon: Surveillance as a Service. SS 2019 Mar 31;17(1/2):27-33. [doi: 10.24908/ss.v17i1/2.13008]
70. Haggerty K, Ericson R. The surveillant assemblage. Br J Sociol 2000 Dec;51(4):605-622. [doi: 10.1080/00071310020015280] [Medline: 11140886]
71. Seberger JS, Bowker GC. Humanistic infrastructure studies: hyper-functionality and the experience of the absurd. Information, Communication & Society 2020 Feb 21;24(12):1712-1727. [doi: 10.1080/1369118x.2020.1726985]
72. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
73. Levitt HM, Motulsky SL, Wertz FJ, Morrow SL, Ponterotto JG. Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. Qualitative Psychology 2017 Feb;4(1):2-22. [doi: 10.1037/qup0000082]
74. Dick S, O'Connor Y, Thompson MJ, O'Donoghue J, Hardy V, Wu TJ, et al. Considerations for Improved Mobile Health Evaluation: Retrospective Qualitative Investigation. JMIR Mhealth Uhealth 2020 Jan 22;8(1):e12424 [FREE Full text] [doi: 10.2196/12424] [Medline: 32012085]
75. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
76. Ackerman MS, Cranor LF, Reagle J. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. 1999 Presented at: 1st ACM Conference on Electronic Commerce; November 3-5, 1999; Denver, CO. [doi: 10.1145/336992.336995]
77. Hine C. Privacy in the Marketplace. The Information Society 1998 Nov;14(4):253-262. [doi: 10.1080/019722498128700]
78. Martin K, Shilton K. Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. The Information Society 2016 Apr 13;32(3):200-216. [doi: 10.1080/01972243.2016.1153012]
79. Meinert S. Field manual - Scenario building. European Trade Union Institute. 2014. URL: https://www.etui.org/publications/guides/field-manual-scenario-building [accessed 2021-09-23]
80. Kaptchuk G, Goldstein DG, Hargittai E, Hofman J, Redmiles EM. How good is good enough for COVID19 apps? The influence of benefits, accuracy, and privacy on willingness to adopt. Cornell University. 2020 May 18. URL: https://arxiv.org/abs/2005.04343 [accessed 2021-09-23]
81. Allmark P, Boote J, Chambers E, Clarke A, McDonnell A, Thompson A, et al. Ethical Issues in the Use of In-Depth Interviews: Literature Review and Discussion. Research Ethics 2009 Jun 01;5(2):48-54. [doi: 10.1177/174701610900500203]
82. Wash R, Radar E. Influencing mental models of security: a research agenda. 2011 Presented at: New Security Paradigms Workshop; September 12-15, 2011; Marin County, CA. [doi: 10.1145/2073276.2073283]
83. Halpern SD, Truog RD, Miller FG. Cognitive Bias and Public Health Policy During the COVID-19 Pandemic. JAMA 2020 Jul 28;324(4):337-338. [doi: 10.1001/jama.2020.11623] [Medline: 32597963]
84. Bowker GC. Temporality. Society for Cultural Anthropology. 2015 Sep 24. URL: https://culanth.org/fieldsights/temporality [accessed 2021-09-23]
85. Acquisti A, Grossklags J. Privacy and rationality in individual decision making. IEEE Secur. Privacy Mag 2005 Jan;3(1):26-33. [doi: 10.1109/MSP.2005.22]
86. Latour B. Imaginer les gestes-barrieres contre le retour a la production d'avant-crise. AOC. 2020 Mar 30. URL: https://aoc.media/opinion/2020/03/29/imaginer-les-gestes-barrieres-contre-le-retour-a-la-production-davant-crise/ [accessed 2021-09-23]
87. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Information & Management 2018 Jun;55(4):482-493. [doi: 10.1016/j.im.2017.11.003] [Medline: 30872122]
88. Clark CC. Trust in medicine. J Med Philos 2002 Feb 1;27(1):11-29. [doi: 10.1076/jmep.27.1.11.2975] [Medline: 11961684]
89. Irurita VF. The problem of patient vulnerability. Collegian 1999 Jan;6(1):10-15. [doi: 10.1016/s1322-7696(08)60310-8]

Abbreviations

GDPR: General Data Protection Regulation
HCI: human-computer interaction
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health

Edited by L Buis; submitted 01.06.21; peer-reviewed by M Lotto, M Holter; comments to author 28.06.21; revised version received 19.07.21; accepted 05.08.21; published 05.10.21.

Please cite as:
Seberger JS, Patil S
Post-COVID Public Health Surveillance and Privacy Expectations in the United States: Scenario-Based Interview Study
JMIR Mhealth Uhealth 2021;9(10):e30871
URL: https://mhealth.jmir.org/2021/10/e30871
doi:10.2196/30871
PMID:34519667

Original Paper

The Functionality of Mobile Apps for Anxiety: Systematic Search and Analysis of Engagement and Tailoring Features

Andreas Balaskas[1], BSc, MSc; Stephen M Schueller[2,3], PhD; Anna L Cox[4], PhD; Gavin Doherty[1], DPhil

[1]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

[2]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States

[3]Department of Informatics, University of California, Irvine, Irvine, CA, United States

[4]UCL Interaction Centre, University College London, London, United Kingdom

Corresponding Author:
Gavin Doherty, DPhil
School of Computer Science and Statistics
Trinity College Dublin
College Green
Dublin
Ireland
Phone: 353 01 8963858
Email: Gavin.Doherty@tcd.ie

Abstract

Background: A range of mobile apps for anxiety have been developed in response to the high prevalence of anxiety disorders. Although the number of publicly available apps for anxiety is increasing, attrition rates among mobile apps are high. These apps must be engaging and relevant to end users to be effective; thus, engagement features and the ability to tailor delivery to the needs of individual users are key. However, our understanding of the functionality of these apps concerning engagement and tailoring features is limited.

Objective: The aim of this study is to review how cognitive behavioral elements are delivered by anxiety apps and their functionalities to support user engagement and tailoring based on user needs.

Methods: A systematic search for anxiety apps described as being based on cognitive behavioral therapy (CBT) was conducted on Android and iPhone marketplaces. Apps were included if they mentioned the use of CBT for anxiety-related disorders. We identified 597 apps, of which 36 met the inclusion criteria and were reviewed through direct use.

Results: Cognitive behavioral apps for anxiety incorporate a variety of functionalities, offer several engagement features, and integrate low-intensity CBT exercises. However, the provision of features to support engagement is highly uneven, and support is provided only for low-intensity CBT treatment. Cognitive behavioral elements combine various modalities to deliver intervention content and support the interactive delivery of these elements. Options for personalization are limited and restricted to goal selection upon beginning use or based on self-monitoring entries. Apps do not appear to provide individualized content to users based on their input.

Conclusions: Engagement and tailoring features can be significantly expanded in existing apps, which make limited use of social features and clinical support and do not use sophisticated features such as personalization based on sensor data. To guide the evolution of these interventions, further research is needed to explore the effectiveness of different types of engagement features and approaches to tailoring therapeutic content.

(JMIR Mhealth Uhealth 2021;9(10):e26712) doi:10.2196/26712

KEYWORDS

mental health; cognitive behavioral therapy; mobile apps; anxiety; stress; mHealth; mobile phone

Introduction

Background

Anxiety disorders are the most common type of mental health problem. Several studies have shown that cognitive behavioral therapy (CBT) is an effective treatment for anxiety disorders [1-4]. CBT focuses on a person's cognitive processes (thoughts, images, beliefs, and attitudes) and behavior to recognize and address negative thinking patterns and beliefs. Depending on the specific anxiety disorder, different CBT techniques are weighted differently during therapy [3]. Unfortunately, only a minority of individuals with an anxiety disorder have access to treatment [5-7].

Internet adoption and advances in technology have led to an increase in research on internet-delivered CBT to make evidence-based psychotherapy more accessible and cost-effective [8,9]. Internet-delivered CBT is delivered via desktop computers, laptops, or tablets to help patients build core CBT knowledge and skills while reducing reliance on traditional face-to-face sessions [8]. Mobile apps provide a promising avenue for increasing access to mental health interventions. Apps can be used to deliver a range of intervention strategies, provide information about mental health, and enable real-time communication with health care professionals. Several studies have examined the effectiveness of delivering interventions using such apps [10,11]. These apps are likely to become increasingly sophisticated, and advances in technology have opened up possibilities for the delivery of *just-in-time adaptive interventions* that aim to provide the most beneficial interventions based on data collected from sensors or provided by users [12]. This creates opportunities for the delivery of deeply personalized interventions [13].

Although the number of mental health apps available is increasing, there are high attrition rates among mobile apps [14-16]. High attrition rates may be attributed to a lack of knowledge regarding the translation of treatment elements in CBT into engaging digital elements, the loss of a therapist-client relationship, the lack of individualized treatment, and the omission of important therapeutic components [8,17,18]. Hence, we need to understand the content in mental health apps from the viewpoint of not only evidence-based strategies but also of engagement elements and other features relevant to digital delivery.

Reviewing Mental Health Apps

The increased availability of mental health apps has motivated researchers to create guidelines for assessing such apps [19-22]. In addition, several papers have recently reviewed apps targeting a variety of mental health conditions, including depression [23-26], bipolar disorder [27], and anxiety disorders [28-33]. Previous reviews targeting anxiety disorders have aimed to assess the extent to which mobile apps are grounded in theory [30,31] or evaluate the extent of expert involvement in development [28,29,32]. Several reviews have also categorized different types of mobile apps [25,28] and identified the range and frequency of treatment elements available [34]. The results from these reviews have indicated that mobile apps for anxiety do not provide information on the source of their content [28] and often lack the involvement of health care professionals in their development [29]. Most apps are inconsistent with evidence-based treatment [30-32] and lack published studies on effectiveness [28]. The examination of sensors used to collect data or provide support based on collected data has received little attention among these studies. Only 1 app review that we are aware of has investigated the use of sensors—in a review of apps targeting children and adolescents with anxiety disorders [31].

A review that examined common treatment elements within depression and anxiety apps found that core treatment elements (eg, exposure and restructuring) were rarely included [34]. Looking beyond anxiety, CBT apps for depression offer an eclectic mix of features, including many that are not evidence-based. The apps offer limited CBT features—with their presence or absence not linked with expert involvement in app creation—and lack elements used in high-intensity interventions [23]. A study found that the utility of self-help CBT or behavioral activation apps is questionable, and their usability is highly variable; furthermore, apps are rarely accompanied by privacy policies [26]. Most apps targeting depression provide multiple functions [24,25], provide an interactive interface, and use text as the main type of media [25]. Similarly, the content of apps targeting bipolar disorder is not in line with practice guidelines, and most apps lack citations and privacy policies [27]. Consequently, the case for recommendation of available apps to users is not as strong as it could be.

The design of usable and effective mental health apps is a key challenge and depends on an understanding of what works, for whom, and under which circumstances. This requires a greater understanding of what these apps actually provide and what strategies they employ to engage and deliver treatment to users. Thus, there is a need to examine both the provision of features that encourage regular use and tailoring of content delivery to provide interaction that is both engaging and effective for particular users. As the concept of engagement research is broad and depends on the context that will be examined [35], we define engagement features as functionality that encourages regular use, makes app content more appealing, and in general, helps users to stay engaged with therapy or the app itself. Tailoring in the context of technology-based interventions refers to the adjustment of technology-delivered self-help programs to suit the user's needs, characteristics, and comorbidities of case formulation [36,37]. Within the context of anxiety apps, this relates primarily to the tailoring of content to specific user groups and needs.

Within the academic literature, the use of functionality based on sensors is seen as a key strategy for delivering more effective content in future systems [12,38-41]; therefore, any use of such features is of interest.

Objectives

The objectives of this study are to explore the functionality of publicly available mobile apps for anxiety that integrate CBT, with a focus on content delivery, including engagement and tailoring features. Although various studies have explored the evidence base of and evidence-based content in anxiety apps

[8,34,42], to our knowledge, no review of apps has been conducted exploring content delivery in apps for anxiety disorders.

Methods

Overview

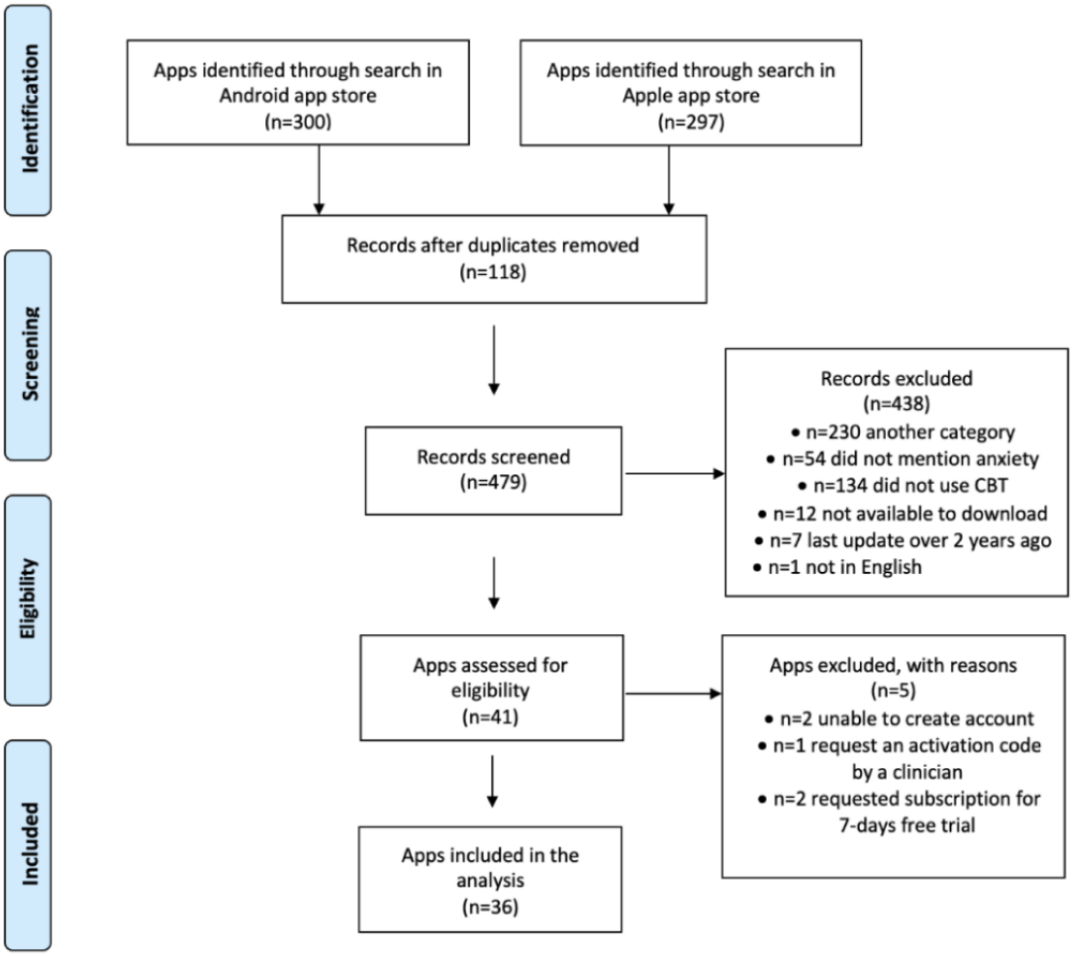
Apps were identified through a systematic search of the two most widely used platforms: Android's Google Play Store and Apple's iOS App Store [16]. Initial searches were conducted in February 2020 to explore the inclusion criteria and analysis methodology. The corpus presented here is based on a final search conducted in June 2020. The search was based on the following keywords related to anxiety disorders: *anxiety*, *stress*, *worry*, *phobia*, and *panic*. In addition, we searched for the following keyword related to CBT: *cognitive behavioural*. Separate searches for each of these keywords were conducted on each app store and applied across all categories of apps.

Studies suggest that only a small number of users look beyond the first 10 ranked apps [43]. In addition, the first five apps for a given search term are the most downloaded [43]. Therefore, as a conservative approximation of those visible to potential users, only the first 50 results were included from each unique search. Search results for each of these keywords were automatically downloaded using scripts from the US version of Google Play and Apple's App Store [44,45]. Recorded information included app name, description, price, developer name, average rating, and the number of user ratings.

Selection Criteria

We included apps that met the following criteria: (1) the app was in the health and fitness or medical category of the app store, (2) mentioned the use of CBT in the title or store description when we searched for anxiety-related keywords, (3) mentioned anxiety-related disorders when we searched for "cognitive behavioural," (4) were currently available to download, and (5) were available in English. Apps were excluded if (6) they did not mention the treatment or management of anxiety-related symptoms in their description pages and (7) were last updated more than two years ago. We downloaded apps that were free or offered in-app purchases to have the broadest appeal because studies suggest that free apps are more likely to be downloaded than paid apps [46]. We coded both the free and premium features of these apps. Two reviewers reviewed a pilot sample of apps to clarify the inclusion criteria before proceeding. One reviewer independently applied the inclusion criteria. The study selection process is illustrated in Figure 1.

Figure 1. App selection process. CBT: cognitive behavioral therapy.



Data Extraction and Coding

Apps meeting the inclusion criteria were downloaded onto either a Samsung Galaxy S9 Plus (Android version 9) or an iPhone 8 Plus (iOS version 14.0). The American Psychiatric Association model for assessing mobile mental health apps was used as a starting point [19]. Descriptive characteristics related to the following features were extracted from the marketplaces: general background information (store, price, target audience, popularity, privacy and safety, accessibility, claimed scientific underpinning, and medical disclaimer). We assessed accessibility by manually checking each app to identify options relevant to visual, auditory, or motor impairment or explicitly termed accessibility options within the apps.

Engagement features were extracted using a classification from a review by Stawarz et al [23] on the functionality of CBT apps for depression. They examined the functionality of CBT apps for depression by recording both therapeutic and engagement features from their description pages. We recorded features that encourage regular use, make app content more appealing, and in general, help users to stay engaged with therapy or the app itself. In addition, we recorded CBT therapeutic features; 2 researchers independently indicated whether a given feature type represented a CBT component used in the treatment of anxiety, similar to other efforts used to characterize evidence-based app content [26,34,47]. In cases of disagreement, discussions between the authors were conducted until full agreement was established before proceeding. See Textbox 1 for definitions of how components were operationalized for study coders.

Textbox 1. Evidence-based treatment components within study apps.

Psychoeducation

- Education about anxiety (definition, description of cycle of reinforcement, description of symptoms)
- Education on the cognitive behavioral model

Self-monitoring

- Monitoring cognition
- Monitoring emotions or symptoms
- Monitoring behaviors

Cognitive techniques

- Identifying thoughts
- Cognitive restructuring

Behavioral techniques

- Behavioral activation
- Behavioral experimentation

Relaxation skills

- Mindfulness exercises
- Progressive muscle relaxation
- Breathing exercises

A sample of apps was reviewed to identify functionality features; 10 apps were then open-coded to create the initial codebook. The codebook was refined following discussion with all authors, and all the apps were coded to the refined scheme. In cases of disagreement, discussions between the authors were conducted until full agreement was established before proceeding.

We coded general characteristics for each app, including target audience, popularity, privacy and safety, accessibility, claimed scientific underpinnings, and medical disclaimers. We coded six main types of functionality and recorded the engagement features—screening, self-monitoring, visualization of data entries, gamification, social features and support (immediate support and crisis support), and features used to deliver CBT treatment elements. We also recorded CBT treatment components, including psychoeducation, cognitive techniques, behavioral techniques, and relaxation skills.

In addition, we recorded tailoring features that aim to personalize treatment, for example, by delivering content targeted to individual user needs or allowing users to customize content based on their needs. We distinguished personalization and customization based on the nature of user involvement. Personalization is used to tailor the user's experience based on their previous behaviors, whereas customization is initiated by users and allows the modification of app features based on their preferences. The codes covered five main types of tailoring functionality: interface customization, treatment-oriented customization, content tailoring offered immediately after installation (app-driven tailoring), tailoring based on

self-monitoring entries (mood-driven personalization), and customization of push notifications. We also recorded (1) the use of sensors and (2) the provision of support based on sensor data. Textbox 2 lists the features identified in the apps in a hierarchical manner. See Textbox 3 for the classification of functionality and engagement features.

Textbox 2. Hierarchical organization of the results section.

General characteristics

- Target audience
- Popularity
- Privacy and safety
- Accessibility
- Claimed scientific underpinning
- Medical disclaimer

Functionality

- Screening
- Self-monitoring
- Visualization of data entries
- Gamification
- Social features and support
  - Immediate support
  - Crisis support
- Features used for delivery of cognitive behavioral therapy treatment elements
  - Psychoeducation
  - Cognitive techniques
  - Behavioral techniques
  - Relaxation skills

Tailoring features

- Interface customization
- Treatment-oriented customization
- App-driven tailoring
- Mood-driven personalization
- Customization of push notifications

Textbox 3. Functionality types and engagement features.

Functionality and engagement features

- Screening
- Self-monitoring
- Data visualization: graphs and charts, reports supporting graphs and charts
- Gamification: games and gamification
- Tailoring: customization options, notifications, and reminders
- Social features and support: ability to share data with others, peer support, and ability to contact a therapist
- Features used for delivery of cognitive behavioral therapy treatment elements: chat with a bot, treatment program format, ability to add pictures and videos, audio content, video content, and question and answer interface

Results

General Characteristics

Of the 479 unique apps screened, 36 met the inclusion criteria (Figure 1), in a range similar to that found in other reviews [27,32]; 11 were available on the Android platform, 4 on the Apple Store, and 21 on both platforms. One app offered an Apple smartwatch version. The apps spanned two categories: health and fitness (26/36, 72%) and medical (10/36, 28%). None of the apps focused on a specific anxiety disorder. In total, 14 apps were free to download (14/36, 39%), and the rest (22/36, 61%) were free with in-app purchases. In-app purchases were offered mostly as monthly or yearly subscriptions (21/22, 95%). One of these apps offered the option of lifetime subscription. Most apps identified through the app stores were last updated in the previous 6 months (27/36, 75%), and only nine apps were last updated before that period.

Target Audience

Both marketplaces provide formal age classification. The majority of Android apps in the sample were classified as being suitable for children ≥3 years (31/32, 97%), and only 1 app recommended parental guidance. Most apps (15/25, 60%) in the Apple Store were classified as being suitable for adolescents ≥12 years, 32% (8/25) for children ≥4 years, and 8% (2/25) for adolescents ≥17 years. In addition, 1 app recommended use for ages between 11 and 19 years and allowed use by younger children with the support of a carer, although it was classified as being suitable for children ≥3 years (Android) and 4 years (Apple).

Popularity

Most apps provided a rating score in the marketplace (33/36, 92%). The rating for most apps (out of 5 stars) was above 4.0 (26/33, 79%). Multimedia Appendix 1 provides information on app ratings. Two apps in the Android marketplace received an editor's choice award and another received a standout well-being app award.

Privacy and Safety

Privacy policies were available, either in the app or as a link from the app store description for most apps (34/36, 94%). Two apps lacked a privacy policy, which means no protection for personal information or safeguards against misuse of mental health data. The privacy policy for 1 app was not in English. Of all the apps with a privacy policy, an account or password creation was mandatory for 9 (out of 35, 26%) and optional for 1 (out of 35, 3%). A total of 15 apps provided the option to set up a personal identification number (13/15, 87%) or biometric authentication (2/15, 13%). The setup for password protection was offered in the premium version for one app. The remaining apps did not provide any security features to restrict access to the data.

Accessibility

Most of the apps required an internet connection to function (24/36, 67%), and a small number (3/36, 8%) provided reduced functionality without an internet connection, which may disadvantage those without a reliable connection. Accessibility options for those with impaired vision or other disabilities were offered by only 2 apps (out of 36, 5%). One of these apps allowed the text size to be changed. The other app (Happify) offered a variety of options, including compatibility with assistive technology (such as voice assistants), high-contrast mode for those with low vision or color blindness, accessibility warnings for activities that require visual or audio interactions, font resize support for low-vision users, and the option to disable animations.

Claimed Scientific Underpinning

All the apps claimed in their description page to be designed based on validated psychological treatments. A total of 20 apps were designed to provide techniques based on CBT (20/36, 56%). The remainder integrated CBT techniques combined with other psychological treatment approaches, including positive psychology (5/36, 14%), acceptance and commitment therapy (4/36, 11%), and dialectical behavior therapy (3/36, 8%).

Medical Disclaimer

Most apps provided a medical disclaimer, indicating that the app was not a replacement for clinical treatment (21/36, 58%). A total of 15 apps did not provide any disclaimer on the marketplace or app's website. In addition, 10 apps made the disclaimer easy to find and read by presenting it on the description page of the app on the marketplace (8/10, 80%) or when downloading the app (2/10, 20%). The remainder presented the disclaimer in their terms of use (4/11, 36%), on the app menu (5/11, 45%), or in the *frequently asked questions* section (2/11, 18%).

Functionality Analysis

Functionality Types

The following section discusses the functionality of the identified apps, with a focus on engagement and tailoring features. For all apps, we recorded details of features to support user engagement with therapy or the app itself offered in the free (Table 1) and premium versions (Table 2).

Table 1. Engagement features.

App name	Visualization	Gamification	Customization	Social	Chatbot
Bloom	✓[a]		✓		
CBT Companion	✓		✓	✓	
CBT diary	✓		✓	✓	
CBT Journey	✓			✓	
CBT MH	✓				
CBT diary	✓			✓	
CBT Tools	✓	✓	✓		
ClearFear	✓	✓	✓		
CBT	✓				
CBT (2)	✓				
De-stressMe	✓	✓			
ezeCBT	✓		✓	✓	
FearTools	✓				
FreeCBT	✓		✓		
Happify	✓	✓	✓	✓	✓
Innerhour	✓	✓	✓		✓
Life	✓		✓		✓
Mindease	✓		✓		
Mindshift	✓		✓	✓	
Moodfit	✓		✓	✓	
Moodnotes	✓		✓	✓	
Moodpath	✓		✓	✓	✓
MoodSpace	✓		✓	✓	
Panic Pit Stop					
Pocketcoach	✓				✓
Reflectly	✓		✓	✓	
Sanvello	✓	✓	✓	✓	✓
Stress & Anxiety Companion	✓		✓		
Thoughts	✓				
UpLift	✓	✓	✓		
What’s up	✓		✓	✓	
Woebot	✓		✓		✓
WorryKit	✓				
WorryTree	✓		✓	✓	
Wysa	✓		✓		
Youper	✓		✓	✓	

[a]Feature present.

Table 2. Engagement features available in the premium version of apps.

App name	Therapist	Program	Data visualization	Reports	Data sharing	Support	Other
Bloom		✓[a]	✓	✓			
CBT diary			✓		✓	✓	✓
CBT Companion		✓		✓	✓	✓	✓
Happify		✓	✓	✓			
Innerhour	✓	✓		✓		✓	
Mindshift	✓	✓	✓	✓			
Moodnotes			✓				
Moodpath		✓	✓			✓	
Pocketcoach	✓	✓	✓			✓	
Reflectly			✓				
Sanvello		✓	✓				✓
UpLift		✓					
What's up							✓
WorryTree				✓			✓
Wysa	✓	✓				✓	
Youper							✓

[a]Feature present.

Screening

In total, 12 (12/36, 33%) apps offered functionality to screen for a variety of psychological disorders using questionnaires. The screening was user-initiated for five apps. The remaining apps provided screening when downloading the app (5/12, 42%) or during app use (2/12, 17%). The purpose of the screening was to help users track and manage their progress and provide insights. In addition, 1 app offered screening to train the chatbot to learn which intervention strategies would be most relevant for each user.

Self-monitoring

Of the 36 apps, 22 (61%) offered functionality entailing tracking feelings (8/22, 36%), mood (11/22, 50%), emotions (1/22, 5%), or mood and anxiety levels (2/22, 9%). A diverse range of designs was used for self-monitoring. Common modalities included the use of emoticons or tags for different feelings, an avatar that changes based on interaction with it, and scales used to rate the intensity of different emotions. Colors were used to indicate the intensity of feelings and text to support the meaning of different emoticons. A total of 13 apps supported the entry of additional information related to the selected feeling, such as situations (4/13, 31%), factors (5/13, 38%), thoughts (3/13, 23%), or journal entries (1/13, 8%). The number of options for the selection of feelings differed among apps, with the majority presenting from 5 to 7 different feelings. Apps varied in how often users can track their mood and how mood tracking is presented. Tracking was unlimited and user-initiated in 20 apps. One app allowed 2 different types of tracking: daily tracking of worry level presented on the home page of the app, and mood tracking, which was triggered every time the user opened the chatbot feature of the app. All of the apps supported momentary tracking, with some allowing retrospective completion.

Visualization of Data Entries

A total of 30 apps provided ways to reflect on the data collected through the app. The apps offered a reflection on mood tracking data (20/30, 67%) or other kinds of data collected through the app (22/30, 73%), such as data from intervention tools. Color and emoticons were the most common elements used to display data on graphs. One app used cards showing a feeling and the factors related to that specific feeling. Other types of data included past entries in intervention tools (13/22, 59%), chatbot conversations (2/22, 9%), and frequency of app use (7/22, 32%). Customization options included the choice of the time range to display data (6/17, 35%), graph type (2/17, 12%), and selection of different variables (eg, mood vs sleep; 1/17, 6%). In addition, 3 apps provided weekly or monthly reports, including charts.

Gamification

Gamification techniques were integrated in 7 of the apps, including level upgrades (2/7, 29%), points (3/7, 43%), and badges (5/7, 71%) based on points earned from practicing different activities. One app mentioned that the reason for integrating points was to encourage regular use of the different activities. In addition, 1 app included a game to make negative thoughts concrete by knocking out negative feelings presented as cartoons.

Social Features and Support

All the apps were designed to function without professional guidance. Four of the apps offered the opportunity to involve health care experts by supporting access to counseling sessions either through the app (2/4, 50%) or over the internet with links

to external websites in the premium version (2/4, 50%). In total, 15 apps allowed users to either share their data directly through the app with a therapist (2/15, 13%) or export and download data (13/15, 87%). Interestingly, 2 of the apps allowed for data sharing with wearable devices regarding mindfulness. Peer support was provided in five apps by integrating discussion and chat groups for various topics related to mental health. One of these apps required users to download another app for community features.

Immediate Support

Eight of the apps integrated a feature to provide additional momentary support. Seven of these apps provided a feature to access different intervention strategies to manage their mood and anxiety. A feature in one of the apps connected the user directly with an available professional through WhatsApp upon paying a "nominal" fee.

Crisis Support

In total, 14 apps offered in-app support via a feature that provided links to external support services and hotlines. One of the apps offered additional support by integrating a crisis feature that, apart from providing information on hotlines, created a safety plan and integrated a *grounding* technique for panic management. This feature presented different types of exercises (breathing, mindfulness, and physical exercises) through a chatbot. In another app, a chatbot presented resources on the screen when the user indicated a crisis by typing a specific word (ie, SOS). However, this feature was presented only when the user first used the app, potentially making it difficult to remember.

Features Used for Delivery of CBT Treatment Elements

Delivery Format

Six apps used the format of one or more treatment programs, each comprising a number of modules. Four of these apps provided a single treatment program, whereas the rest provided access to multiple treatment programs at any time. Users had to complete each module in the program to access the next module. An alternative delivery format, used by 5 apps, was to use a chatbot to deliver intervention strategies; in 2 cases, the chatbot was used to deliver intervention strategies based on self-monitoring entries.

One app that integrated a chatbot provided intervention strategies only in the premium version of the app. Table 3 lists the CBT elements identified in the apps.

Table 3. Cognitive behavioral therapy evidence-based elements available in the apps.

App name	Psychoeducation	Self-monitoring	Cognitive	Behavioral	Relaxation skills
Bloom		✓[a]			
CBT Companion	✓	✓	✓•[b]	✓•	✓
CBT diary		✓			
CBT Journey			✓		
CBT MH		✓			✓
CBT diary		✓	✓		
CBT Tools	✓	✓	✓		
ClearFear	✓	✓			✓
CBT	✓		•	•	
CBT (2)	✓				
De-stressMe	✓	✓			✓
ezeCBT	✓		✓		
FearTools		✓	✓	✓	✓
FreeCBT			✓		
Happify					✓
Innerhour	✓	✓			✓
Life	✓	✓		✓	
Mindease		✓	✓		✓
Mindshift	✓	✓	✓	✓	✓
Moodfit	✓	✓	✓		✓
Moodnotes		✓•			
Moodpath	✓	✓			✓•
MoodSpace			✓		✓
Panic Pit Stop	✓				
Pocketcoach	✓	✓	•		✓•
Reflectly		✓			
Sanvello	✓	✓	✓•		✓
Stress & Anxiety Companion	✓		✓		✓•
Thoughts	✓		✓		
UpLift	✓	✓			✓
What’s up	✓	✓			
Woebot	✓	✓	✓	✓	✓
WorryKit			✓		✓
WorryTree				•	
Wysa		✓	✓		✓
Youper		✓	•		•

[a]Checkmark denotes an element offered in the free version of the app.

[b]Bullet symbol denotes additional elements offered in the premium version of the app.

Psychoeducation

Most apps delivered psychoeducation material (21/36, 58%) using a distinct feature integrated into the apps (16/21, 76%). Modalities for the delivery of psychoeducation included the use of text (11/21, 52%) separated into chunks (6/21, 29%) accompanied by illustrations (3/21, 14%). One of these apps additionally offered video and audio delivery for

psychoeducation. Two apps delivered psychoeducation using a question-and-answer interface. Another 2 offered psychoeducation through a chatbot using multiple-choice topic selection or through users' interaction with the app (eg, after check-in assessments, to explain different intervention strategies).

Cognitive Techniques

Cognitive techniques were supported by 56% (20/36) of apps and were mainly delivered through structured exercises, providing suggestions and prompts based on previously entered data. Many cognitive techniques, such as cognitive restructuring, were implemented through textual representations of thoughts, lists of thinking traps, and cognitive distortions. One app allowed audio entries as an alternative. Some apps supported a greater degree of interactivity; for example, 1 app supported identifying negative thoughts by dragging a finger over words, and a window would appear with a selection of thinking traps. Two of the apps used a chatbot to deliver a cognitive exercise aimed at reducing the burden on users by delivering content based on data collected during app use. More specifically, one of the apps remembers and presents a *rethink list* with unhelpful thoughts previously provided by the user. The user can remove or add thoughts from the list. The other app remembers and presents the most frequent distortions identified by the user in the past, allowing them to decide what to work on.

Behavioral Techniques

Behavioral techniques supported (7/36, 19%) included behavioral activation, exposure, and action planning. For example, 2 apps supported behavioral activation, focusing on positive rewarding activities. In one app (Woebot), users, after identifying negative thoughts, could schedule an activity and rate their feelings after activity completion. The other app used a chatbot, issued text instructions, and suggested activities to be completed based on the setting (home or outside).

Relaxation Skills

Most apps supported relaxation skills such as breathing (14/36, 39%), relaxation (10/36, 28%), and mindfulness exercises (17/36, 47%). Relaxation and mindfulness exercises were provided using audio tracks or text instructions and illustrations. In 22% (8/36) of apps, breathing exercises were delivered using a breathing indicator that visually represented a full breath. For example, a circle expands as users inhale and contracts as they exhale. Interestingly, one of the apps monitored heart-rate variability through the application of a finger to the phone camera to detect changes in detected *stress* during the breathing exercise.

Tailoring: Customization and Personalization

Customization and Personalization Types

Table 4 shows the personalization offered across three types. We characterized customization as interface customization, treatment-oriented customization, and customization of data visualization.

Table 4. Tailoring of mobile apps.

App name	App-driven	Mood-driven	Notifications
Bloom	✓[a]		✓
CBT Companion			✓
CBT diary			✓
CBT Journey			✓
CBT MH			✓
CBT diary			✓
CBT Tools			✓
ClearFear	✓		
CBT			
CBT (2)			✓
De-stressMe			✓
ezeCBT			
FearTools			✓
FreeCBT			✓
Happify	✓		✓
Innerhour	✓	✓	✓
Life			
Mindease		✓	✓
Mindshift		✓	✓
Moodfit		✓	✓
Moodnotes			✓
Moodpath			✓
MoodSpace			
Panic Pit Stop			
Pocketcoach	✓	✓	✓
Reflectly			✓
Sanvello	✓	✓	✓
Stress & Anxiety Companion			✓
Thoughts			
UpLift	✓		✓
What's up			
Woebot		✓	✓
WorryKit			✓
WorryTree		✓	✓
Wysa	✓	✓	✓
Youper	✓		✓

[a]Personalization offered.

Interface Customization

In total, 24 apps allowed users to customize the app either at the beginning (14/24, 58%) or through a menu (24/24, 100%). Customization options covered user profiles (nicknames, avatars, and profile photo), user interface appearance (themes, display options, animations, and language), and technical features covering notifications and location tracking.

Treatment-Oriented Customization

Three apps allowed the customization of different treatment elements from the settings menu, such as adding, deactivating,

or changing the position of emotion management (1/24, 4%), hiding navigation arrows on a diary entry (1/24, 4%), enabling voice dictation for a thought diary (1/24, 4%), adding a therapist's number (1/24, 4%), and stopping sharing data with a professional (1/24, 4%).

App-Driven Personalization

A total of 16 apps provided onboarding screens to educate users about the functions and benefits of each app. As part of this onboarding, nine apps allowed users to select the challenges or goals they wanted to work on (8/9, 89%) or request information about things that calm the user and phone numbers to be presented for emergency support (1/9, 11%). App-driven tailoring of app content was compulsory in 4 apps with no option to skip that step. Two apps did not provide information on the purpose of therapeutic tailoring.

Mood-Driven Personalization

Eleven apps suggested intervention strategies based on users' self-monitoring data. The suggested intervention strategies were either randomized (3/11, 27%) or the same set of strategies (8/11, 73%) were presented to the users each time they tracked their mood, although in 1 chatbot-based intervention, the last activity practiced would also be suggested. Four of these apps presented intervention strategies only when low mood levels were indicated. One app allowed access to intervention strategies only in the premium version.

Customization of Notifications

Notifications are used to prompt users to interact with apps. Nineteen apps integrated notifications and prompted access to the intervention content. Three apps offered notifications for tracking only in the premium version. Notifications were provided at fixed times by the user (8/19, 42%), fixed times by the app (4/19, 21%), randomly (3/19, 16%), or in combination (4/19, 21%). Users could customize notification timing for different purposes, such as tracking (16/19, 84%), accessing intervention strategies (7/19, 37%), or accessing other app features (7/19, 37%). Two apps allowed the customization of the reminder message.

Discussion

General Findings

Despite increased scholarly interest in the review of apps for a variety of mental health conditions [23-33], this is the first study to examine the delivery of content in CBT apps for anxiety, with a focus on engagement and tailoring features.

The reviewed apps targeted a variety of anxiety conditions and were not designed to tackle a specific anxiety disorder (eg, general anxiety disorder), which may limit the degree to which treatment could be tailored to individual users. There is also confusion around the appropriate age for these apps. Most apps were in the health and fitness category. However, some apps were advertised in the medical category without providing a medical disclaimer. These findings align with the results of a recent review that explored the functionality of top-rated apps for depression [24].

Three of the apps received awards from the marketplace (Google Play Store) based on the quality of their design and overall functionality. However, there are no clear guidelines on how an app's content quality is assessed.

Accessibility and Access

The results showed that accessibility was rarely considered when designing apps. The lack of consideration of accessibility during the design process potentially excludes a variety of users. To address accessibility issues, researchers have begun developing guidelines to gauge mobile app accessibility [48]. In addition, many apps cannot function or offer reduced functionality when an internet connection is not available. Several engagement features or full access to features are offered only in the premium version. Hence, valuable functionality is offered as a premium feature, and issues of access and cost should be addressed at the public health level.

Engagement With Apps and Therapeutic Content

Overall, our results indicate that self-monitoring and visualization were relatively well explored by the apps in the sample in comparison with other feature types. Screening was provided using multiple-choice questions without involving any engagement feature for that purpose, which suggests an assumption by designers that screening itself could potentially be engaging. Several types of self-monitoring functionality are used to make the process more appealing and easier to use by minimizing the completion time. A minority of apps made use of basic gamification techniques such as points, badges, and level upgrades, in line with those found in a recent review of gamification in mental health [49].

Human Support

Human support is a powerful mechanism for increasing engagement and enabling the tailoring of treatment. The most common engagement feature related to human support was the ability to share data through the app. A minority of apps provided some form of peer support feature, and for understandable reasons, the ability to contact a therapist was not offered in the free version of the apps. The therapeutic alliance is an important part of CBT treatment, and developing a relationship between a health care professional and a user could enhance the regular use of apps [50-52]. Further research is required to reveal the amount and type of clinician support needed to optimize the effectiveness of CBT for specific groups [8].

Delivery of CBT Elements

CBT elements use various features, such as audio and video, illustrations, and interactive screens, to increase engagement. By using multimedia elements and interactive exercises, we can enrich the therapy experience and support learning [39]. In particular, behavioral techniques can be enhanced through sensing.

The CBT apps identified in this review were designed to support low-intensity CBT exercises. Although a structured treatment program was offered by a minority of apps, studies have shown that mobile apps do not provide a full course of CBT but are limited to specific techniques [53]. Limited support for clinician

integration reduces the opportunity to use such apps in higher-intensity CBT.

Customization and Tailoring

Although the classifications in the existing literature on tailoring health interventions focus on the delivery of tailored messages [54,55], this study explored tailoring more broadly in the context of anxiety apps that adjust content to suit different groups and user needs. The majority of apps allowed the customization of the interface. However, customization options for CBT elements were lacking, with only 3 apps supporting this. The results showed that app-driven therapeutic tailoring is restricted to push notifications and goal selection, although CBT involves the individualization of treatment.

Many apps offered tailoring based on self-monitoring data by delivering intervention strategies to users. The results indicated that either the same set of strategies is presented after reporting a particular score in self-monitoring or intervention strategies are presented at random. Thus, there are opportunities for more interactive and integrated approaches that incorporate both user preferences and mood-based recommendations**.**

Sensors and Just-in-Time Intervention

Only one app used physiological sensing. Two of the apps provided helplines and resource information based on location data. This contrasts with the academic literature, in which much recent research concerns the vision of *just-in-time adaptive interventions* that allow the provision of intervention content when users are in need of support [12,56] and the interest in exploiting machine learning to optimize treatment delivery within mental health [57].

Improving Delivery Within CBT Apps for Anxiety

Engagement with apps may be increased if we target mobile apps to specific types of anxiety disorders and user needs to enhance the degree of tailored content to individual user needs. Greater interactivity, for example, through increased user agency or better use of gamification, could enhance engagement [35]. Human support offered through the apps should be enhanced to increase engagement with low-intensity CBT apps and explore mechanisms for use in higher-intensity treatment. Designers should consider accessibility issues and offline functionality to increase the potential reach of these systems to wider populations.

Evidence-Based Content and Effectiveness of Mobile Apps

The results showed that all apps claim in their description page to be designed based on validated psychological treatments. Previous research has shown that most apps are inconsistent with evidence-based treatments [30-32]. In addition, an app's consistency with evidence-based treatment elements does not guarantee the efficacy and effectiveness of treatment elements delivered in an app-based format. Randomized controlled trials are considered to provide the most reliable evidence for the effectiveness of an intervention. However, these studies test the effectiveness of an intervention as a whole and do not examine the mechanisms that lead to improvement in one's mental health state. Further research is thus required to determine the most effective functionality for delivering evidence-based treatment content. We hope that the breakdown of app functionality in this study will be useful as a starting point for such efforts. In addition to effectiveness, future work could also examine proximal outcomes, such as the impact of different features and evidence-based components on engagement.

Limitations

The search was based on US app stores and might not be representative of all anxiety apps available in the global market. This study focuses on apps integrating CBT elements and thus may have excluded potentially useful apps for people with anxiety. This review focused on investigating engagement features and tailoring of content; therefore, we did not evaluate the quality of content in the apps. We believe that by exploring the functionality of the apps for tailored and engaging delivery, our work complements existing research and can help inform future design efforts. Our results reflect consumer mobile apps available in the marketplace rather than the current state-of-the-art in research.

Conclusions

Although apps integrate a range of functionalities and engagement features, the provision of these features is highly uneven. Self-monitoring and visualization are relatively well explored, with social features and human support rarely integrated. Our results show that, within currently available apps, accessibility is neglected, and tailoring options are limited. Furthermore, consumer apps do not currently take advantage of the technological capabilities of smartphones to deliver *just-in-time* interventions at opportune moments [12]. Future research should explore strategies for tailoring therapeutic content and involving clinicians to facilitate user engagement.

Acknowledgments

This work was supported, in part, by Science Foundation Ireland grant 17/RC-PHD/3488 to the ADAPT Centre at Trinity College Dublin. The ADAPT Centre is funded by Science Foundation Ireland and is cofunded by the European Regional Development Fund through grant 13/RC/2106_P2.

Conflicts of Interest

SMS received payment for consulting from Otsuka Pharmaceutical, and serves on the Scientific Advisory Board for Headspace, for which he receives compensation. He receives research funding from One Mind for the operation and management of One Mind PsyberGuide. GD is a cofounder of SilverCloud Health Ltd and has a minority shareholding in the company.

Multimedia Appendix 1
Spreadsheet of main coding elements.
[XLSX File (Microsoft Excel File), 2549 KB - mhealth_v9i10e26712_app1.xlsx]

References

1. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
2. Kaczkurkin AN, Foa EB. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. Dialogues Clin Neurosci 2015 Sep;17(3):337-346 [FREE Full text] [Medline: 26487814]
3. Otte C. Cognitive behavioral therapy in anxiety disorders: current state of the evidence. Dialogues Clin Neurosci 2011;13(4):413-421 [FREE Full text] [Medline: 22275847]
4. DiMauro J, Domingues J, Fernandez G, Tolin DF. Long-term effectiveness of CBT for anxiety disorders in an adult outpatient clinic sample: a follow-up study. Behav Res Ther 2013 Feb;51(2):82-86. [doi: 10.1016/j.brat.2012.10.003] [Medline: 23262115]
5. Young AS, Klap R, Shoai R, Wells KB. Persistent depression and anxiety in the United States: prevalence and quality of care. Psychiatr Serv 2008 Dec;59(12):1391-1398 [FREE Full text] [doi: 10.1176/appi.ps.59.12.1391] [Medline: 19033165]
6. Alonso J, Angermeyer M, Bernert S, Bruffaerts R, Brugha TS, Bryson H, ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project. Use of mental health services in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004 Jun;109(420):47-54. [doi: 10.1111/j.1600-0047.2004.00330.x] [Medline: 15128387]
7. Collins KA, Westra HA, Dozois DJ, Burns DD. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. Clin Psychol Rev 2004 Sep;24(5):583-616. [doi: 10.1016/j.cpr.2004.06.001] [Medline: 15325746]
8. Wright JH, Mishkind M. Computer-assisted CBT and mobile apps for depression: assessment and integration into clinical care. Focus (Am Psychiatr Publ) 2020 Apr;18(2):162-168 [FREE Full text] [doi: 10.1176/appi.focus.20190044] [Medline: 33162853]
9. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided internet-based CBT for common mental disorders. J Contemp Psychother Cut Edg Mod Dev Psychother 2013 May 21;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
10. Loo Gee B, Griffiths K, Gulliver A. Effectiveness of mobile technologies delivering Ecological Momentary Interventions for stress and anxiety: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):221-229 [FREE Full text] [doi: 10.1093/jamia/ocv043] [Medline: 25997643]
11. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
12. Nahum-Shani I, Smith S, Spring B, 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]
13. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depress Anxiety 2017 Jun;34(6):540-545. [doi: 10.1002/da.22649] [Medline: 28494123]
14. Becker S, Kribben A, Meister S, Diamantidis CJ, Unger N, Mitchell A. User profiles of a smartphone application to support drug adherence--experiences from the iNephro project. PLoS One 2013 Oct 23;8(10):e78547 [FREE Full text] [doi: 10.1371/journal.pone.0078547] [Medline: 24194946]
15. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
16. Boulos M, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
17. Colby KM, Gould RL, Aronson G. Some pros and cons of computer-assisted psychotherapy. J Nerv Ment Dis 1989 Feb;177(2):105-108. [doi: 10.1097/00005053-198902000-00007] [Medline: 2915214]
18. Fernández-Álvarez J, Díaz-García A, González-Robles A, Baños R, García-Palacios A, Botella C. Dropping out of a transdiagnostic online intervention: a qualitative analysis of client's experiences. Internet Interv 2017 Sep 22;10:29-38 [FREE Full text] [doi: 10.1016/j.invent.2017.09.001] [Medline: 30135750]

19. Torous J, Luo J, Chan S. Mental health apps: what to tell patients. Curr Psychiatry 2018 Mar;17(3):21-25.
20. 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]
21. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
22. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
23. 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]
24. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321 [FREE Full text] [doi: 10.2196/15321] [Medline: 32012079]
25. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
26. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016 May 2;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
27. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
28. Alyami M, Giri B, Alyami H, Sundram F. Social anxiety apps: a systematic review and assessment of app descriptors across mobile store platforms. Evid Based Ment Health 2017 Aug 30;20(3):65-70. [doi: 10.1136/eb-2017-102664] [Medline: 28666986]
29. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun 15;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
30. Kertz SJ, MacLaren Kelly J, Stevens KT, Schrock M, Danitz SB. A review of free iPhone applications designed to target anxiety and worry. J Technol Behav Sci 2017 Jan 11;2(2):61-70. [doi: 10.1007/s41347-016-0006-y]
31. Bry LJ, Chou T, Miguel E, Comer JS. Consumer smartphone apps marketed for child and adolescent anxiety: a systematic review and content analysis. Behav Ther 2018 Mar;49(2):249-261 [FREE Full text] [doi: 10.1016/j.beth.2017.07.008] [Medline: 29530263]
32. Van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
33. Rodriguez-Paras C, Tippey K, Brown E, Sasangohar F, Creech S, Kum H, et al. Posttraumatic stress disorder and mobile health: app investigation and scoping literature review. JMIR Mhealth Uhealth 2017 Oct 26;5(10):e156. [doi: 10.2196/mhealth.7318] [Medline: 29074470]
34. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
35. Doherty K, Doherty G. Engagement in HCI: conception, theory and measurement. ACM Comput Surv 2019 Jan 23;51(5):1-39. [doi: 10.1145/3234149]
36. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthan E, et al. Individually-tailored, internet-based treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2011 Jan;49(1):18-24. [doi: 10.1016/j.brat.2010.10.002] [Medline: 21047620]
37. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
38. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. JMIR Serious Games 2018 May 10;6(2):e9 [FREE Full text] [doi: 10.2196/games.8902] [Medline: 29748165]
39. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
40. 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]

41. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
42. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
43. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
44. Node.js scraper to get data from Google Play. GitHub. URL: https://github.com/facundoolano/google-play-scraper [accessed 2018-02-08]
45. Scrape data from the itunes app store. GitHub. URL: https://github.com/facundoolano/app-store-scraper [accessed 2018-02-08]
46. Petsas T, Papadogiannakis A, Polychronakis M, Markatos E, Karagiannis T. Rise of the planet of the apps: a systematic study of the mobile app ecosystem. In: Proceedings of the 2013 conference on Internet measurement conference. 2013 Presented at: IMC '13: Proceedings of the 2013 conference on Internet measurement conference; Oct 23-25, 2013; Barcelona, Spain. [doi: 10.1145/2504730.2504749]
47. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: a meta-analytic review of intervention components. Depress Anxiety 2019 Mar;36(3):213-224. [doi: 10.1002/da.22860] [Medline: 30450811]
48. Ballantyne M, Jha A, Jacobsen A, Hawker J, El-Glaly Y. Study of accessibility guidelines of mobile applications. In: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia. 2018 Presented at: 17th International Conference on Mobile and Ubiquitous Multimedia; Nov 25-28, 2018; Cairo, Egypt. [doi: 10.1145/3282894.3282921]
49. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
50. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
51. Henson P, Peck P, Torous J. Considering the therapeutic alliance in digital mental health interventions. Harv Rev Psychiatry 2019;27(4):268-273. [doi: 10.1097/HRP.0000000000000224] [Medline: 30907764]
52. Tremain H, McEnery C, Fletcher K, Murray G. The therapeutic alliance in digital mental health interventions for serious mental illnesses: narrative review. JMIR Ment Health 2020 Aug 07;7(8):e17204 [FREE Full text] [doi: 10.2196/17204] [Medline: 32763881]
53. Wright JH, Mishkind M, Eells TD, Chan SR. Computer-assisted cognitive-behavior therapy and mobile apps for depression and anxiety. Curr Psychiatry Rep 2019 Jun 27;21(7):62. [doi: 10.1007/s11920-019-1031-2] [Medline: 31250242]
54. Noar SM, Grant Harrington N, Van Stee SK, Shemanski Aldrich R. Tailored health communication to change lifestyle behaviors. Am J Lifestyle Med 2010 Nov 19;5(2):112-122. [doi: 10.1177/1559827610387255]
55. Lustria M. Message tailoring in health and risk messaging. In: Oxford Research Encyclopedia of Communication. Oxford, UK: Oxford University Press; 2017.
56. Balaskas A, Schueller SM, Cox AL, Doherty G. Ecological momentary interventions for mental health: a scoping review. PLoS One 2021 Mar 11;16(3):e0248152 [FREE Full text] [doi: 10.1371/journal.pone.0248152] [Medline: 33705457]
57. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans Comput Hum Interact 2020 Oct;27(5):1-53. [doi: 10.1145/3398069]

Abbreviations

CBT: cognitive behavioral therapy

Edited by L Buis; submitted 23.12.20; peer-reviewed by C Linehan, D Di Matteo, CY Lin; comments to author 12.02.21; revised version received 02.04.21; accepted 15.07.21; published 06.10.21.

Please cite as:
Balaskas A, Schueller SM, Cox AL, Doherty G
The Functionality of Mobile Apps for Anxiety: Systematic Search and Analysis of Engagement and Tailoring Features
JMIR Mhealth Uhealth 2021;9(10):e26712
URL: https://mhealth.jmir.org/2021/10/e26712
doi:10.2196/26712
PMID:34612833

Original Paper

Characteristics and Quality of Mobile Apps Containing Prenatal Genetic Testing Information: Systematic App Store Search and Assessment

Ko-Lin Wu[1], BSc; Rebeca Alegria[2], BSc; Jazzlyn Gonzalez[3], BSc; Harrison Hu[4]; Haocen Wang[2], PhD; Robin Page[5], PhD; Patricia Robbins-Furman[6], CGC, MPH; Ping Ma[7], PhD; Tung-Sung Tseng[8], DrPH; Lei-Shih Chen[2], PhD

[1]Graduate School of Biomedical Sciences, University of North Texas Health Science Center, Fort Worth, TX, United States

[2]Department of Health and Kinesiology, Texas A&M University, College Station, TX, United States

[3]College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, TX, United States

[4]College of Science, Texas A&M University, College Station, TX, United States

[5]College of Nursing, Texas A&M University, College Station, TX, United States

[6]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States

[7]Department of Health Promotion & Community Health Sciences, School of Public Health, Texas A&M University, College Station, TX, United States

[8]Behavioral and Community Health Sciences, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA, United States

Corresponding Author:
Lei-Shih Chen, PhD
Department of Health and Kinesiology
Texas A&M University
4243 TAMU
College Station, TX, 77840
United States
Phone: 1 979 862 2912
Email: lacechen@tamu.edu

Abstract

Background: Prenatal genetic testing is an essential part of routine prenatal care. Yet, obstetricians often lack the time to provide comprehensive prenatal genetic testing education to their patients. Pregnant women lack prenatal genetic testing knowledge, which may hinder informed decision-making during their pregnancies. Due to the rapid growth of technology, mobile apps are a potentially valuable educational tool through which pregnant women can learn about prenatal genetic testing and improve the quality of their communication with obstetricians. The characteristics, quality, and number of available apps containing prenatal genetic testing information are, however, unknown.

Objective: This study aims to conduct a firstreview to identify, evaluate, and summarize currently available mobile apps that contain prenatal genetic testing information using a systematic approach.

Methods: We searched both the Apple App Store and Google Play for mobile apps containing prenatal genetic testing information. The quality of apps was assessed based on the criteria adopted from two commonly used and validated mobile app scoring systems, including the Mobile Application Rating Scale (MARS) and the APPLICATIONS evaluation criteria.

Results: A total of 64 mobile apps were identified. Of these, only 2 apps were developed for a specific prenatal genetic test. All others were either pregnancy-related (61/64, 95%) or genetics-related (1/64, 2%) apps that provided prenatal genetic testing information. The majority of the apps (49/64, 77%) were developed by commercial companies. The mean quality assessment score of the included apps was 13.5 (SD 2.9), which was equal to the average of possible theoretical score. Overall, the main weaknesses of mobile apps in this review included the limited number of prenatal genetic tests mentioned; incomprehensiveness of testing information; unreliable and missing information sources; absence of developmental testing with users (not evidence based); high level of readability; and the lack of visual information, customization, and a text search field.

Conclusions: Our findings suggest that the quality of mobile apps with prenatal genetic testing information must be improved and that pregnant women should be cautious when using these apps for prenatal genetic testing information. Obstetricians should carefully examine mobile apps before referring any of them to their patients for use as an educational tool. Both improving the

quality of existing mobile apps, and developing new, evidence-based, high-quality mobile apps targeting all prenatal genetic tests should be the focus of mobile app developers going forward.

(JMIR Mhealth Uhealth 2021;9(10):e30404) doi:10.2196/30404

KEYWORDS

mobile applications; prenatal genetic testing; pregnancy; review; evaluation

Introduction

Prenatal genetic testing, a set of genetic tests used to detect potential fetal disease risk, is an essential part of routine prenatal care [1,2]. Information provided by these tests allows pregnant women to make informed decisions about their pregnancy, including preparation for affected births, early management of infants with genetic disorders, and termination of affected pregnancies [1,3-5]. Although prenatal genetic testing is important for pregnant women, obstetricians often have insufficient time to provide comprehensive education about prenatal genetic testing in prenatal care [6,7]. Studies have found that pregnant women lack prenatal genetic testing knowledge, which may hinder informed decision-making during their pregnancies [4,8,9].

Mobile apps have been used to assist patients in accessing health information, facilitate engagement with their physicians, and strengthen patient-provider communication and relationships [10-12]. Lay people also perceive health information provided by apps to be accurate and trustworthy [13]. Among apps with medical information, pregnancy-related apps are the most common [14,15]. Nevertheless, the characteristics, quality, and number of available apps containing prenatal genetic testing information are unknown.

To fill this knowledge gap, we used a systematic approach to identify existing apps containing prenatal genetic testing information and summarize their characteristics. We then adapted existing app evaluation tools [16-18] to evaluate the quality of these apps. We believe our findings will have the potential to help obstetricians become familiar with prenatal genetic testing apps and make recommendations to their patients. Our results can assist researchers and app developers to recognize the strengths and limitations of existing apps that contain information pertaining to prenatal genetic testing, as well as to design apps that better meet the needs of pregnant women.

Methods

Overview

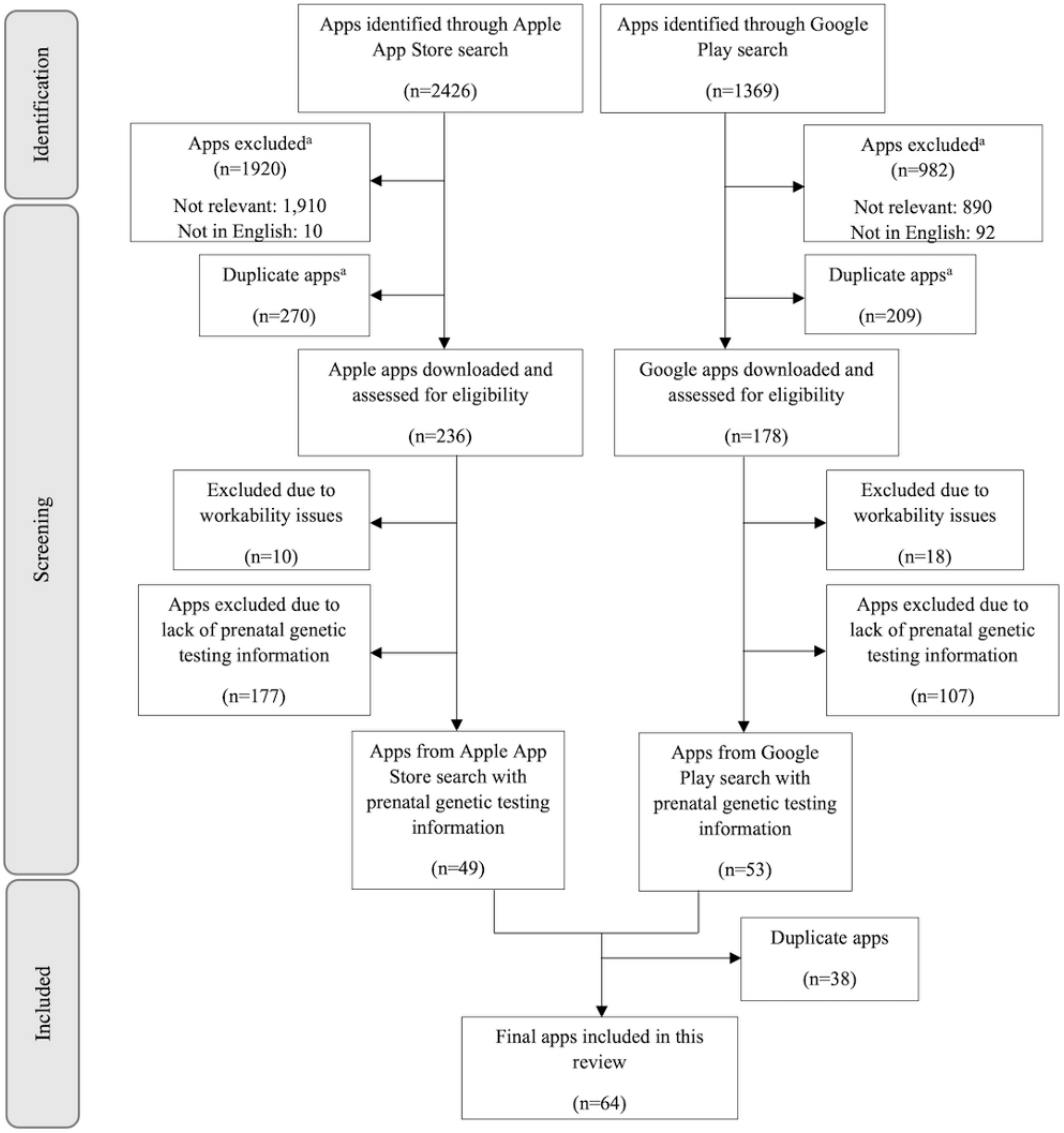
Because this study did not involve risk to human subjects, it was exempted from institutional review board oversight at Texas A&M University. A prenatal genetic counselor at a large medical center and a certified health education specialist who is an expert in prenatal genetic education guided, oversaw, and reviewed the research process and findings.

App Search

Figure 1 summarizes the search, screening, and selection procedures for the eligible prenatal genetic testing apps. Specifically, this process included four steps. First, 3 members of the research team (KW, RA, and JG) conducted a search on both the Apple App Store (iOS) and Google Play (Android) for apps with prenatal genetic testing information, between January 28, 2020, and March 23, 2020. The following search terms were used: “pregnancy,” “prenatal,” “pregnant,” “genetic testing,” “genetic screening,” “pregnancy screening,” “prenatal screening,” “prenatal screening test,” “prenatal diagnostic test,” “carrier testing,” “expanded carrier screening,” “amniocentesis,” “chorionic villus sampling,” “non-invasive prenatal testing,” “non-invasive prenatal screening,” “cell free DNA screening,” “cell free DNA testing,” “MaterniT21,” “Harmony,” “Panorama,” “QNatal,” “percutaneous umbilical blood sampling,” “cordocentesis,” “nuchal translucency scan,” “nuchal translucency ultrasound,” “quad screen,” “quadruple screen,” “quadruple test,” “first-trimester screening,” “second-trimester screening,” “modified sequential screening,” “sequential screening,” “alpha-fetoprotein screening,” “maternal serum screening,” “prenatal genetic screening test,” “prenatal genetic screening,” and “prenatal genetic testing”.

A total of 3795 apps were identified. In the second step, we used very few exclusion criteria for the initial screening in order to maximize the number of potential apps selected. A total of 2902 apps were excluded using the following exclusion criteria: (1) apps that were not relevant to prenatal genetic testing, pregnancy, and/or genetics, and (2) apps that were not in English. In the third step, we excluded apps that were duplicated within the Apple App Store (iOS) and in Google Play (Android). In the fourth step, a total of 414 apps were downloaded and thoroughly assessed. Among these, 28 apps were excluded because of workability issues, such as apps that could not be successfully downloaded or opened, and/or those that needed a specific user account from a hospital or clinic for access. Moreover, 284 apps were excluded due to the lack of prenatal genetic testing information, and 38 apps were excluded because of their duplication in both the Apple App Store and Google Play. Thus, our final sample consisted of 64 apps.

Figure 1. The selection process of mobile apps containing prenatal genetic testing information. [a]The apps were excluded by reading the app titles and descriptions.



Data Extraction

The entire research team assessed, extracted, and summarized detailed information from the included 64 apps (Multimedia Appendix 1). Extracted information included app name, operating system (ie, iOS or Android), app description, types of prenatal genetic testing mentioned in those apps, prenatal genetic test procedures mentioned therein, information on the timing of prenatal genetic tests, reliability or accuracy of prenatal genetic tests included, interpretation of the results of those prenatal genetic tests, specific disorders mentioned, prenatal genetic testing information citations, customer ratings, and cost. Moreover, according to the Mobile Application Rating Scale (MARS) app classification [18], we also extracted the app developer information (ie, details of the commercial company, hospital, or nongovernmental organization). Finally, based on the content of the apps, we utilized the Flesch-Kincaid grade level assessment to determine the readability level of the apps selected for further analysis.

App Quality Assessment

To evaluate the quality of the selected apps, we developed the App Quality Assessment Scoring System (AQASS) by adapting two common app evaluation tools—MARS [18] and the "APPLICATIONS" evaluation criteria [16,17]. The AQASS was used to assess 16 items in each app, including the number of prenatal genetic tests mentioned, comprehensiveness, information quality, quality of information sources, evidence base, readability, visual prenatal genetic testing information, customization, interactivity, text search field, ease of navigation, app layout, visual appeal, interdevice compatibility, connectivity, and price. Table 1 lists the detailed criteria for each item in AQASS. Possible AQASS scores ranged from 0 to 27, with a higher score representing higher app quality. Following the reliability checking procedure outlined in previous research [19,20], 2 raters on the team (RA and HH) used the AQASS criteria to independently score each app, after which interrater reliability was assessed using the Gwet AC1 calculation [21]. Findings suggest a strong interrater reliability between the two raters with a coefficient of 0.86 [22].

Table 1. App Quality Assessment Scoring System (AQASS) ratings of mobile apps (N=64) containing prenatal genetic testing information.

Criterion and description	AQASS score	Frequency distribution (%)
Number of prenatal genetic tests mentioned		
Using the general term "(prenatal) genetic testing" only	0	3
1-5 prenatal genetic tests	1	67
6-10 prenatal genetic tests	2	30
All 10 prenatal genetic tests	3	0
Comprehensiveness of information		
Only mentioned the name of the prenatal genetic test	0	2
Having information in 1 or 2 of the following 5 categories: test procedures, timing of the tests, reliability or accuracy of the tests, interpretation of test results, and specific diseases or conditions to be tested for	1	25
Having information in 3 or 4 of the following 5 categories: test procedures, timing of the tests, reliability or accuracy of the tests, interpretation of test results, and specific diseases or conditions to be tested for	2	38
Having information in all of the following categories: test procedures, timing of the tests, reliability or accuracy of the tests, interpretation of test results, and specific diseases or conditions to be tested for	3	36
Quality of information		
Information too general to assess quality	0	3
Most information is incorrect	1	2
Some incorrect and/or out-of-date information	2	25
Correct and up-to-date information	3	70
Reliable information sources		
No reference identified	0	75
Questionable/unreliable source(s)	1	3
Source from hospitals, specific obstetricians-gynecologists, or peer reviewed articles	2	13
Source from authoritative organizations, such as the American College of Obstetricians and Gynecologists	3	9
Evidence base		
App had not been trialed or tested	0	98
App had been tested (eg, acceptability, usability, and satisfaction ratings) and had positive outcomes in studies that were not randomized control trials (RCTs), and there was no contradictory evidence.	1	2
App had been trialed and outcome tested in 1-2 RCTs indicating positive results	2	0
App had been trialed and outcome tested in >3 high quality RCTs indicating positive results	3	0
Readability[a]		
Not applicable[b]	0	3
>6th grade	1	95
≤6th grade	2	2
Visual information[c]		
Absent	0	92
Present	1	8
Customization[d]		
Absent	0	100
Present	1	0

Criterion and description	AQASS score	Frequency distribution (%)
Interactivity[e]		
Absent	0	31
Present	1	69
Text search function		
Absent	0	61
Present	1	39
Navigation ease		
Hard to navigate	0	14
Easy to navigate	1	86
Layout of the app		
Inappropriate, unclear, and/or some options difficult to select, locate, find, or read	0	3
Appropriate, clear, and able to select, locate, find, or read items	1	97
Visual appeal		
Unprofessional or unappealing	0	2
Professional and appealing	1	98
Interdevice compatibility		
iPhone or Android phone	0	41
iPhone and Android phone	1	59
Connectivity		
Internet not required	0	58
Internet required	1	42
Price		
Paid	0	0
Free	1	100

[a]Readability was assessed by Flesch-Kincaid grade level assessment. According to Weiss [23], mobile health apps are recommended to have a reading level of sixth grade or below for general public use.

[b]App did not have enough text and/or complete sentences to allow for the calculation of Flesch-Kincaid grade level for readability.

[c]Visual information was defined as the presence in the app of photos, images, tables, figures, and videos to demonstrate pregame genetic tests

[d]Customization referred to a function that allows users to input prenatal genetic test results, testing dates, and other information in settings or preferences of the mobile app.

[e]Interactivity allowed users to provide feedback to the developers, offer reminders on the timing of prenatal genetic testing, have a chatroom function, and allow sharing information onto the social media.

Results

Among 64 evaluated apps, only 2 (3.1%) apps were specifically developed for prenatal genetic testing. NIPT Insights, developed by Five Minutes Ltd., introduced noninvasive prenatal testing (NIPT) and compared it with first-trimester screening, triple/quad screening, anatomy ultrasound, chorionic villus sampling (CVS), and amniocentesis. Similarly, cfDNA Predictive Value Calculator (Perinatal Quality Foundation) offered links to a number of scientific articles regarding NIPT, and it provided users with the ability to calculate both positive and negative predictive values of NIPT. All other apps that provided prenatal genetic testing information were either pregnancy-related (61/64, 95%) or genetics (1/64, 2%) apps.

Table 2 summarizes the key characteristics of those 64 apps. Specifically, all 64 apps could be downloaded at no cost. Of these, 5 (8%) offered an optional upgrade, which were downloaded and assessed. However, upgrading did not provide any further prenatal genetic testing information. The majority of the apps (49/64, 77%) were created by commercial companies, and over half of the apps (38/64, 59%) were compatible with both the Android and iOS operating systems. Only 52 (81%) apps had a customer rating. With a theoretical range of 1 to 5 stars, the mean rating was 4.5 (SD 0.6; range 2.0-5.0).

Table 2. Descriptive statistics of the mobile apps (N=64) containing prenatal genetic testing information.

Characteristic and category	Value, n (%)
Type of developer	
App created by commercial companies	49 (77)
App created by hospitals	11 (17)
App created by non-governmental organizations	3 (5)
Apps created by the government (ie, California Department of Public Health)	1 (2)
Operating system	
Android only	15 (23)
iOS only	11 (17)
Android and iOS	38 (59)
Type of prenatal genetic tests mentioned	
Amniocentesis	49 (77)
CVS[a]	49 (77)
First trimester screening (NT[b] screening)	46 (72)
Triple/quad screening	44 (69)
Cell-free DNA testing or NIPT[c]	35 (55)
Anatomy ultrasound	26 (41)
Cordocentesis	10 (16)
Expanded carrier screening	8 (13)
First trimester screening (blood test)	8 (13)
Genetic screening (prenatal) in general	7 (11)
Carrier screening	6 (9)
Prenatal genetic test procedures mentioned	
Amniocentesis	37 (58)
CVS	36 (56)
First trimester screening (NT screening)	34 (53)
Cell-free DNA testing or NIPT	25 (39)
Triple/quad screening	17 (27)
Cordocentesis	9 (14)
Anatomy ultrasound	5 (8)
Carrier screening	4 (6)
First trimester screening (blood test)	3 (5)
Expanded carrier screening	1 (2)
Not reported	17 (27)
Timing of prenatal genetic tests mentioned	
CVS	45 (70)
Amniocentesis	44 (69)
First trimester screening (NT screening)	40 (63)
Triple/quad screening	38 (59)
Cell-free DNA testing or NIPT	31 (48)
Anatomy ultrasound	23 (36)
Cordocentesis	9 (14)

Characteristic and category	Value, n (%)
First trimester screening (blood test)	8 (13)
Carrier screening	3 (5)
Expanded carrier screening	1 (2)
Genetic screening (prenatal) in general	1 (2)
Not reported	5 (8)
Reliability or accuracy of prenatal genetic tests mentioned	
Cell-free DNA testing or NIPT	23 (36)
Amniocentesis	21 (33)
First trimester screening (NT screening)	13 (20)
CVS	12 (19)
Triple/quad screening	12 (19)
Cordocentesis	2 (3)
Carrier screening	1 (2)
Expanded carrier screening	1 (2)
First trimester screening (blood test)	1 (2)
Anatomy ultrasound	1 (2)
Not reported	33 (52)
Interpretation of potential results of prenatal genetic tests	
Triple/quad screening	19 (30)
Cell-free DNA testing or NIPT	13 (20)
Amniocentesis	12 (19)
First trimester screening (NT screening)	9 (14)
Expanded carrier screening	8 (13)
CVS	8 (13)
Carrier screening	3 (5)
Genetic screening (prenatal) in general	1 (2)
Not reported	36 (56)
Specific disorder(s) mentioned	
Trisomy 21 (Down syndrome)	53 (83)
Trisomy 18 (Edward syndrome)	33 (52)
Cystic fibrosis	29 (45)
Trisomy 13 (Patau syndrome)	18 (28)
Sickle cell anemia	14 (22)
Tay-Sachs disease	11 (17)
Spinal muscular atrophy	8 (13)
Neural tube defect (general)	3 (5)
Spina bifida	15 (23)
Anencephaly	6 (9)
Thalassemia	3 (5)
Fragile X syndrome	2 (3)
Huntington disease	2 (3)
Muscular dystrophy	2 (3)
Klinefelter syndrome	1 (2)

Characteristic and category	Value, n (%)
Prader-Willi syndrome	1 (2)
Trisomy 16	1 (2)
Turner syndrome	1 (2)
Not reported	9 (14)
Citation for prenatal genetic testing information	
No	48 (75)
Yes (reliable source)	14 (22)
Yes (non-reliable source)	2 (3)
Cost[d]: $0	64 (100)
Customer rating[e], mean (SD); range	4.5 (0.6); 2.0-5.0

[a]CVS: chorionic villus sampling.

[b]NT: nuchal translucency.

[c]NIPT: noninvasive prenatal testing.

[d]Five apps had an optional upgrade fee but upgrading did not affect the amount of prenatal genetic testing information available to the user.

[e]Only 52 of the 64 apps (81.3%) had customer ratings.

A total of 10 different types of prenatal genetic tests were mentioned across all apps, including carrier screening, expanded carrier screening, first-trimester screening (blood test), first-trimester screening (nuchal translucency screening), cell-free DNA testing or NIPT, CVS, amniocentesis, triple/quad screening, anatomy ultrasound, and cordocentesis/percutaneous umbilical blood sampling. Amniocentesis and CVS were most common types of prenatal genetic tests referred to in apps (49/64, 77% for both), whereas carrier screening (6/64, 9%) was the least mentioned.

All apps specifically referred to 17 genetic disorders tested for by prenatal genetic testing; these included testing for the following disorders: trisomy 21 (53/64, 83%), trisomy 18 (33/64, 52%), cystic fibrosis (29/64, 45%), trisomy 13 (18/64, 28%), spina bifida (15/64, 23%), sickle cell anemia (14/64, 22%), Tay-Sachs disease (11/64, 17%), spinal muscular atrophy (8/64, 13%), anencephaly (6/64, 9%), neural tube defect as a general term (3/64, 5%), thalassemia (3/64, 5%), Fragile X syndrome (2/64, 3%), Huntington disease (2/64, 3%), muscular dystrophy (2/64, 3%), Klinefelter syndrome (1/64, 2%), Prader-Willi syndrome (1/64, 2%), trisomy 16 (1/64, 2%), and Turner syndrome (1/64, 2%). Nevertheless, 14.1% (9/64) of the apps did not mention any genetic disorders that could be identified by prenatal genetic testing. Information that was missing in the reviewed apps included context about the interpretation of test results (36/64, 56%), the reliability or the accuracy of prenatal genetic tests (33/64, 52%), testing procedures (17/64, 27%), and testing time (5/64, 8%). In terms of references and citations found in these apps, only 14 of 64 apps (22%) included reliable sources (eg, the Mayo Clinic, the American Pregnancy Association, and the American College of Obstetricians and Gynecologists [ACOG]) to support their prenatal genetic testing information.

Table 1 summarizes the AQASS ratings for the quality of the 64 apps we evaluated. With a possible score range of 0 to 27, the mean score for all apps was 13.5 (SD 2.89; range 5-18). In particular, a majority of the apps were found to be visually appealing (63/64, 98%), have appropriate and clear layouts (62/64, 97%), easy to navigate (55/64, 86%), contain interactivity functions (44/64, 69%), and require no internet to access once downloaded (37/64, 58%). Nevertheless, none of the apps included all types of prenatal genetic tests. Most apps (43/64, 67%) reported only 1 to 5 types of prenatal genetic tests. Only 23 (36%) of all the evaluated apps addressed procedures, timing, reliability or accuracy, and interpretation of the results of the prenatal genetic tests. Although 45 (70%) of the apps presented correct and up-to-date information, merely 8 apps (13%) provided reliable information sources from hospitals, specific obstetricians-gynecologists, and/or peer reviewed articles, or referenced authoritative organizations such as the ACOG (6/64, 9%). Furthermore, a large percentage of the apps had a reading level higher than the sixth grade (61/64, 95%), did not include any visual information for prenatal genetic testing (eg, photos, images, tables, figures, and videos) (59/64, 92%), and lacked a text search field (39/64, 61%). None of the apps had a customization function to allow for the tailoring of the prenatal genetic information to the patients' circumstances (eg, inputting gestational weeks as well as prenatal genetic testing dates and results in settings or preferences). Nearly all apps (63/64, 98%) had not been tested by research studies to determine their quality and efficacy.

Discussion

Principal Findings

It is important for pregnant women to understand prenatal genetic testing in order to make informed decisions about whether or not to utilize these tests during their pregnancy [4]. Apps are a potential educational tool to help pregnant women understand such complex information [24]. To our knowledge, our study is the first systematic review of apps containing prenatal genetic testing information. We found 64 apps currently available that provide information about prenatal genetic testing.

Nevertheless, none of those apps presented comprehensive information about all prenatal genetic tests.

Our results showed that commercial companies had developed the majority of the 64 apps with prenatal genetic testing information we found. We also noticed that approximately one-third of all evaluated apps contained incorrect and/or out-of-date information. It is recommended that developers in commercial companies collaborate with a team of health care professionals, including obstetricians, geneticists, and genetic counselors to design apps with accurate and up-to-date information. Moreover, due to the rapid advancement of genomic technologies, information and guidelines regarding prenatal genetic testing may evolve periodically [25]. Thus, developers should review their apps constantly and update information as needed.

The average AQASS score of the 64 apps containing prenatal genetic testing information for our final sample was 13.5, which was equal to the mean score of the possible theoretical AQASS score (range 0-27). This low score was in line with a previous review that suggested that the quality of genetics and genomics apps that need improvement [19]. The low scores of the apps we evaluated was mainly due to the limited number of prenatal genetic tests mentioned, incomprehensiveness of testing information, unreliable and missing information sources; absence of developmental and clinical testing with users; high readability levels; and the lack of visual information presented, customization options, and a text search field.

Unfortunately, the regulation of health apps has not been well established in the United States. Although the Food and Drug Administration (FDA) has published a policy regulating medical health apps, only mobile apps classified as medical devices are currently under federal oversight and require the FDA approval before they are released to the public [26,27]. Because the FDA does not regulate health apps focused on providing health information [26,27], the often-low quality we found in these apps that include information of prenatal genetic testing can cause pregnant women to be exposed to incomplete, unreliable information regarding prenatal genetic testing. As such, comprehensive regulatory standards and guidelines for these apps are urgently required.

Clinical and Research Implications

Our study has important clinical and research implications. First, due to the lack of high-quality and comprehensive apps that provide prenatal genetic testing information, we caution obstetricians to be prepared to address the questions raised by patients who may have previously obtained incorrect and incomplete information from those apps. Second, obstetricians should also be aware of the limitations of the apps with prenatal genetic testing information when communicating with their patients. Moreover, establishing policies to regulate apps that include prenatal genetic testing information is warranted. Finally, from both research and clinical standpoints, the creation of a high-quality app specifically designed for prenatal genetic testing that has been evaluated using a randomized control trial to test its efficacy in improving pregnant women's prenatal genetic testing decision-making is recommended.

Strengths and Limitations

There are several limitations to this study. As we only reviewed apps with information on prenatal genetic testing in English, there may be apps developed in languages other than English that were not included in this review. Similarly, our app search was conducted in the United States. Future researchers may consider changing platform and search settings to include apps available in countries outside the United States to allow for a broader search. Furthermore, as a team, we carefully conducted the app search together and asked 3 additional researchers to independently check our search results. Some apps, however, may have been overlooked. As the app search was concluded on March 23, 2020, new apps with prenatal genetic testing information may have become available after that date.

Despite these limitations, this study has a number of strengths. It is a unique study and the first of its kind attempting to identify apps containing prenatal genetic testing information. After screening a total of 3795 apps on the Apple App Store and Google Play platforms, we identified 64 apps. In addition, we also carefully examined each app to present a thorough summary of the information they contained. Beyond reporting the characteristics of the identified apps, we developed the AQASS to capture a detailed overview of the quality of the apps we reviewed, as most apps were not created with the primary purpose of conveying prenatal genetic testing to pregnant women. Our AQASS was based on commonly used and validated app scoring systems (ie, MARS and APPLICATIONS) [16-18]. We adopted both these systems because several items of MARS (eg, "Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?") and APPLICATIONS (eg, "advertisements" and "subjective presentation") were not applicable to the evaluation of the apps that provide prenatal genetic testing information in our study. Conversely, some items from either MARS (eg, "Evidence base: Has the app been trialed/tested?") or APPLICATIONS (eg, "connectivity" and "navigation ease") were found to be suitable for the evaluation of apps that include information on prenatal genetic testing. Beyond the use of AQASS for our study, it can also be adopted to assess other apps that were not originally designed to primarily address the specific topic under evaluation. It should be noted that the reliability and validity of AQASS also need to be further examined in the future.

Conclusions

This study is the first to extensively search apps containing information on prenatal genetic testing in both the Apple App Store and Google Play, summarize their characteristics, and assess their quality. We identified 64 available apps containing information about prenatal genetic testing and found that the quality of those apps needs to be improved. Additionally, none of the apps we evaluated were specifically designed to introduce all prenatal genetic tests. As such, pregnant women should be cautious when using these apps for prenatal genetic testing information. Obstetricians should carefully examine apps before any recommendation is made for their use as an educational tool. Improving the quality of existing apps and developing new, evidence-based, high-quality apps with a targeted focus

on prenatal genetic testing is strongly recommended in the future.

Acknowledgments

The authors thank Ms Madeline Zubrod and Sahiba Trehan for their contributions to data checking and preparing the summary. This study was funded by a Global Institute for Hispanic Health Seed Grant from the Texas A&M University System and by the Driscoll Children's Hospital (grant 0004459).

Authors' Contributions

KW, RA, JG, and LC developed the initial research question. LC led the study design. KW, RA, and JG conducted the app search and screening. The content of the reviewed apps was extracted and evaluated by KW, RA, JG, HH, HW, RP, PR, and LC. RA and HH independently rated the apps. All authors contributed to the data interpretation. KW, RA, and JG drafted the initial manuscript. All authors reviewed, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the reviewed mobile apps.
[DOCX File , 103 KB - mhealth_v9i10e30404_app1.docx]

References

1. Practice Bulletin No. 162: Prenatal Diagnostic Testing for Genetic Disorders. Obstet Gynecol 2016 May;127(5):e108-e122. [doi: 10.1097/AOG.0000000000001405] [Medline: 26938573]
2. American College of Obstetricians and Gynecologists' Committee on Practice Bulletins—Obstetrics, Committee on Genetics, Society for Maternal-Fetal Medicine. Screening for Fetal Chromosomal Abnormalities: ACOG Practice Bulletin, Number 226. Obstet Gynecol 2020 Oct;136(4):e48-e69. [doi: 10.1097/AOG.0000000000004084] [Medline: 32804883]
3. Dukhovny S, Norton ME. What are the goals of prenatal genetic testing? Semin Perinatol 2018 Aug;42(5):270-274. [doi: 10.1053/j.semperi.2018.07.002] [Medline: 30195989]
4. Sharma MN, Gayathri R, Priya VV. Awareness of prenatal genetic screening test among pregnant mothers. Drug Discov Today 2019;11(9):2309-2312.
5. Chen L, Xu L, Huang T, Dhar SU. Autism genetic testing: a qualitative study of awareness, attitudes, and experiences among parents of children with autism spectrum disorders. Genet Med 2013 Apr;15(4):274-281. [doi: 10.1038/gim.2012.145] [Medline: 23288207]
6. Gammon BL, Kraft SA, Michie M, Allyse M. "I think we've got too many tests!": Prenatal providers' reflections on ethical and clinical challenges in the practice integration of cell-free DNA screening. Ethics Med Public Health 2016;2(3):334-342 [FREE Full text] [doi: 10.1016/j.jemep.2016.07.006] [Medline: 28180146]
7. Nagle C, Lewis S, Meiser B, Gunn J, Halliday J, Bell R. Exploring general practitioners' experience of informing women about prenatal screening tests for foetal abnormalities: a qualitative focus group study. BMC Health Serv Res 2008 May 28;8:114 [FREE Full text] [doi: 10.1186/1472-6963-8-114] [Medline: 18507850]
8. Seven M, Akyüz A, Eroglu K, Daack-Hirsch S, Skirton H. Women's knowledge and use of prenatal screening tests. J Clin Nurs 2017 Jul;26(13-14):1869-1877. [doi: 10.1111/jocn.13494] [Medline: 27487389]
9. Smith SK, Cai A, Wong M, Sousa MS, Peate M, Welsh A, et al. Improving women's knowledge about prenatal screening in the era of non-invasive prenatal testing for Down syndrome - development and acceptability of a low literacy decision aid. BMC Pregnancy Childbirth 2018 Dec 17;18(1):499 [FREE Full text] [doi: 10.1186/s12884-018-2135-0] [Medline: 30558569]
10. Ling C, Harnish D, Shehab R. Educational Apps: Using Mobile Applications to Enhance Student Learning of Statistical Concepts. Hum. Factors Man 2014 Aug 28;24(5):532-543. [doi: 10.1002/hfm.20550]
11. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: A systematic, narrative review. Patient Educ Couns 2019 Jan 31;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
12. Kratzke C, Cox C. Smartphone technology and apps: Rapidly changing health promotion. Int Electron J Health Educ 2012;15(1):72-82.
13. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
14. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy apps freely available in the Google Play Store. Health Promot J Austr 2020;31(3):340-342. [doi: 10.1002/hpja.270] [Medline: 31225924]

15. Tripp N, Hainey K, Liu A, Poulton A, Peek M, Kim J, et al. An emerging model of maternity care: smartphone, midwife, doctor? Women Birth 2014 Mar;27(1):64-67. [doi: 10.1016/j.wombi.2013.11.001] [Medline: 24295598]
16. Chyjek K, Farag S, Chen KT. Rating Pregnancy Wheel Applications Using the APPLICATIONS Scoring System. Obstet Gynecol 2015 Jun;125(6):1478-1483. [doi: 10.1097/AOG.0000000000000842] [Medline: 26000520]
17. Moglia ML, Nguyen HV, Chyjek K, Chen KT, Castaño PM. mobile evaluation of smartphone menstrual cycle tracking applications using an adapted APPLICATIONS scoring system. Obstet Gynecol 2016 Jun;127(6):1153-1160. [doi: 10.1097/AOG.0000000000001444] [Medline: 27159760]
18. 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]
19. Talwar D, Yeh Y, Chen W, Chen L. Characteristics and quality of genetics and genomics mobile apps: a systematic review. Eur J Hum Genet 2019 Jun;27(6):833-840 [FREE Full text] [doi: 10.1038/s41431-019-0360-2] [Medline: 30809045]
20. Talwar D, Chen W, Yeh Y, Foster M, Al-Shagrawi S, Chen L. Characteristics and evaluation outcomes of genomics curricula for health professional students: a systematic literature review. Genet Med 2019 Aug;21(8):1675-1682. [doi: 10.1038/s41436-018-0386-9] [Medline: 30563987]
21. Gwet K. Handbook of inter-rater reliability, 4th Edition: the definitive guide to measuring the extent of agreement among raters. Gaithersburg: Advanced Analytics, LLC; 2014.
22. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
23. Weiss B. Health literacy: a manual for clinicians.: American Medical Association Foundation and American Medical Association; 2003. URL: http://lib.ncfh.org/pdfs/6617.pdf [accessed 2021-04-27]
24. Page RL, Murphey C, Aras Y, Chen L, Loftin R. Pregnant Hispanic women's views and knowledge of prenatal genetic testing. J Genet Couns 2021 Jan 26;30(3):838-848. [doi: 10.1002/jgc4.1383] [Medline: 33496987]
25. Farwell Hagman KD, Lamb Thrush D, Freeze S, Dorsainville DL, Eichmeyer J, Banks KC, Genomic Technologies Special Interest Group of the National Society of Genetic Counselors. Facing the challenge of genetic counselors' need for rapid continuing education about genomic technologies. J Genet Couns 2020 Oct;29(5):838-848. [doi: 10.1002/jgc4.1213] [Medline: 31916674]
26. Yasini M, Marchand G. health applications, in the absence of an authentic regulation, does the usability score correlate with a better medical reliability? Stud Health Technol Inform 2015;216:127-131. [Medline: 26262024]
27. U.S. Food and Drug Administration. Policy for device software functions and mobile medical applications: Guidance for industry and food and drug administration staff. 2019. URL: https://www.fda.gov/media/80958/download [accessed 2021-04-27]

Abbreviations

ACOG: American College of Obstetricians and Gynecologists
AQASS: App Quality Assessment Scoring System
CVS: chorionic villus sampling
FDA: Food and Drug Administration
MARS: Mobile Application Rating Scale

Edited by L Buis; submitted 12.05.21; peer-reviewed by D Talwar, B Chaudhry; comments to author 22.06.21; revised version received 19.07.21; accepted 23.08.21; published 14.10.21.

Please cite as:
Wu KL, Alegria R, Gonzalez J, Hu H, Wang H, Page R, Robbins-Furman P, Ma P, Tseng TS, Chen LS
Characteristics and Quality of Mobile Apps Containing Prenatal Genetic Testing Information: Systematic App Store Search and Assessment
JMIR Mhealth Uhealth 2021;9(10):e30404
URL: https://mhealth.jmir.org/2021/10/e30404
doi:10.2196/30404
PMID:34647898

Original Paper

Stakeholders' Perceptions of Benefits of and Barriers to Using Video-Observed Treatment for Monitoring Patients With Tuberculosis in Uganda: Exploratory Qualitative Study

Juliet Nabbuye Sekandi[1,2], MS, DrPH, MD; Vicent Kasiita[3], BA, MA; Nicole Amara Onuoha[2], MPH, MD; Sarah Zalwango[4,5], MPH, MD; Damalie Nakkonde[4], BPH; David Kaawa-Mafigiri[6], MPH, PhD; Julius Turinawe[4], PA; Robert Kakaire[2], BA, MPH, DrPH; Paula Davis-Olwell[1,2], BS, MA, PhD; Lynn Atuyambe[4], BA, MPH, PhD; Esther Buregyeya[4], MPhil, MD, PhD

[1]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA, United States

[2]Global Health Institute, College of Public Health, University of Georgia, Athens, GA, United States

[3]Infectious Disease Institute, Kampala, Uganda

[4]School of Public Health, Makerere University, Kampala, Uganda

[5]Kampala Capital City Authority, Kampala, Uganda

[6]School of Social Sciences and Social Work, Makerere University, Kampala, Uganda

Corresponding Author:
Juliet Nabbuye Sekandi, MS, DrPH, MD
Department of Epidemiology and Biostatistics
College of Public Health
University of Georgia
100 Foster Road
Athens, GA, 30602
United States
Phone: 1 17063387993
Email: jsekandi@uga.edu

Abstract

Background: *Nonadherence* to treatment remains a barrier to tuberculosis (TB) control. Directly observed therapy (DOT) is the standard for monitoring adherence to TB treatment worldwide, but its implementation is challenging, especially in resource-limited settings. DOT is labor-intensive and inconvenient to both patients and health care workers. Video DOT (VDOT) is a novel patient-centered alternative that uses mobile technology to *observe* patients taking medication remotely. However, the perceptions and acceptability of potential end users have not been evaluated in Africa.

Objective: This study explores stakeholders' acceptability of, as well as perceptions of potential benefits of and barriers to, using VDOT to inform a pilot study for monitoring patients with TB in urban Uganda.

Methods: An exploratory, qualitative, cross-sectional study with an exit survey was conducted in Kampala, Uganda, from April to May 2018. We conducted 5 focus group discussions, each comprising 6 participants. Groups included patients with TB (n=2 groups; male and female), health care providers (n=1), caregivers (n=1), and community DOT volunteer workers (n=1). The questions that captured perceived benefits and barriers were guided by domains adopted from the Technology Acceptance Model. These included perceived usefulness, ease of use, and intent to use technology. Eligible participants were aged ≥18 years and provided written informed consent. For patients with TB, we included only those who had completed at least 2 months of treatment to minimize the likelihood of infection. A purposive sample of patients, caregivers, health care providers, and community DOT workers was recruited at 4 TB clinics in Kampala. Trained interviewers conducted unstructured interviews that were audio-recorded, transcribed, and analyzed using inductive content analysis to generate emerging themes.

Results: The average age of participants was 34.5 (SD 10.7) years. VDOT was acceptable to most participants on a scale of 1 to 10. Of the participants, 70% (21/30) perceived it as highly acceptable, with scores ≥8, whereas 30% (9/30) scored between 5 and 7. Emergent themes on perceived benefits of VDOT were facilitation of easy adherence monitoring, timely follow-up on missed doses, patient-provider communication, and saving time and money because of minimal travel to meet in person. Perceived barriers included limited technology usability skills, inadequate cellular connectivity, internet access, availability of electricity,

cost of the smartphone, and use of the internet. Some female patients raised concerns about the disruption of their domestic work routines to record videos. The impact of VDOT on privacy and confidentiality emerged as both a perceived benefit and barrier.

Conclusions: VDOT was acceptable and perceived as beneficial by most study participants, despite potential technical and cost barriers. Mixed perceptions emerged about the impact of VDOT on privacy and confidentiality. Future efforts should focus on training users, ensuring adequate technical infrastructure, assuring privacy, and performing comparative cost analyses in the local context.

(JMIR Mhealth Uhealth 2021;9(10):e27131) doi:10.2196/27131

KEYWORDS

tuberculosis; adherence; mHealth; video directly observed therapy; Uganda; mobile phone

Introduction

Nonadherence to treatment of tuberculosis (TB) remains a significant challenge to meeting the World Health Organization's End TB Strategy to reduce deaths by 90% and incidence by 80% [1]. Approximately 10 million new TB cases occur annually, and 1.5 million die from the disease worldwide [2]. An estimated 33% to 50% of patients who start treatment are nonadherent to their prescribed medication regimens, particularly in low- and middle-income countries (LMICs) [3]. Nonadherence to medication can result in the emergence of drug resistance, prolonged infectiousness, treatment failure, and relapse [4-6]. Poor adherence is particularly common in LMICs, where TB rates are high and resources for health care delivery are limited [7].

The Uganda National Tuberculosis Prevalence Survey reported a high TB case rate of 253 per 100,000 people in 2015 [8]. The Uganda National TB Program uses the recommended standard directly observed therapy (DOT) but implements a mixture of facility-based DOT and community-based DOT (CB-DOT) [9]. In practice, CB-DOT is predominantly used in urban settings such as Kampala city. CB-DOT typically involves a trained community DOT volunteer designated by the TB program or a treatment supporter (a family member, friend, or neighbor) as selected by the patient who is responsible for watching the daily intake of each medication dose [5,10,11]. Previous studies conducted in Uganda have shown the effectiveness of CB-DOT in rural settings, but mixed findings have been reported in urban settings [10,12,13]. Proper implementation and sustainability of standard DOT has been limited because of a lack of funding and its heavy reliance on volunteers [14]. Other factors such as a severe shortage of health workers, high cost of transportation, and the inconvenience of the need for face-to-face provider-patient contact largely impede the feasibility of DOT [7,12,13]. In the end, many patients take medications on their own and then self-report pill ingestion. The lack of an objective method to validate self-reported adherence highlights the need to explore alternative methods of medication monitoring.

The updated guidelines for the treatment of drug-susceptible TB issued by the World Health Organization in 2017 recommended the use of digital adherence technologies such as video DOT (VDOT) as an alternative method to monitor adherence [15]. Digital technologies have been shown to overcome the common systemic barriers to TB treatment delivery [16-19]. The VDOT process involves using a smartphone app to record and send daily medication intake videos to a secure computer system. The submitted videos are then accessed by health care providers treating TB for remote observation [20]. Although VDOT has been generally shown to be feasible, acceptable, and effective when evaluated in high-income countries and LMICs [20-25], low acceptability has also been reported among some users in LMICs [24]. In Uganda, a growing body of evidence on the acceptability of digital adherence monitoring abounds mostly in HIV-infected populations on treatment [26-29] but is limited in populations with TB [30]. Moreover, there are no published studies on the acceptability and perceptions of patients with TB and other stakeholders related to the use of VDOT in Uganda. This exploratory qualitative study was conducted to inform a pilot feasibility study of VDOT for adherence monitoring and support in Uganda. The findings in this paper add to the limited evidence needed to inform future implementation and scale-up of digital adherence technologies in Uganda and other LMICs.

Methods

Ethical Review

All participants provided written informed consent (including permission to audio-record the sessions for transcription and coding purposes) in English or Luganda as they preferred and were informed of their freedom to withdraw at any time during the interview. The institutional review boards approved the study at the University of Georgia, Office of Research (STUDY00004974), and Makerere University Higher Degrees, Research and Ethics Committee, in Uganda (Protocol #562). All participants were offered an equivalent of US $5.00 in Uganda Shillings at the end of the interviews as a refund for travel expenses and compensation for their time.

Theoretical Framework

There is considerable evidence suggesting that theories of behavior change are useful for informing the effective design, uptake, and adoption of new health interventions by health care workers and patients for the management of TB [31]. Generally, before changes in behavior occur, intended users have to perceive it as valuable to their lives and accept the intervention. We adopted some constructs from the Technology Acceptance Model, which is an information systems framework for understanding how users accept and eventually use new technology [32]. The model posits that attitudes, perceived usefulness, and ease of use of technology predict the intention to use the technology, which subsequently correlates with actual use [32]. The strength of the Technology Acceptance Model is

that it can be used to explore views from a broad spectrum of users, including patients and health care providers.

Setting and Population

The study was conducted in Kampala, the capital city of Uganda, from April to May 2018. This work was accomplished through collaborative efforts between researchers at the Makerere University School of Public Health, the University of Georgia, and the Uganda National TB Program staff. Kampala district has the highest TB burden in Uganda, with nearly 25% of annually reported TB cases occurring in the metropolitan area of the capital city [9]. In Kampala, care for patients with active TB is delivered through public and designated private clinics supervised by the Kampala Capital City Authority, with an oversight from the Uganda National TB Program. Diagnosis and treatment services for TB are provided free of charge to all patients.

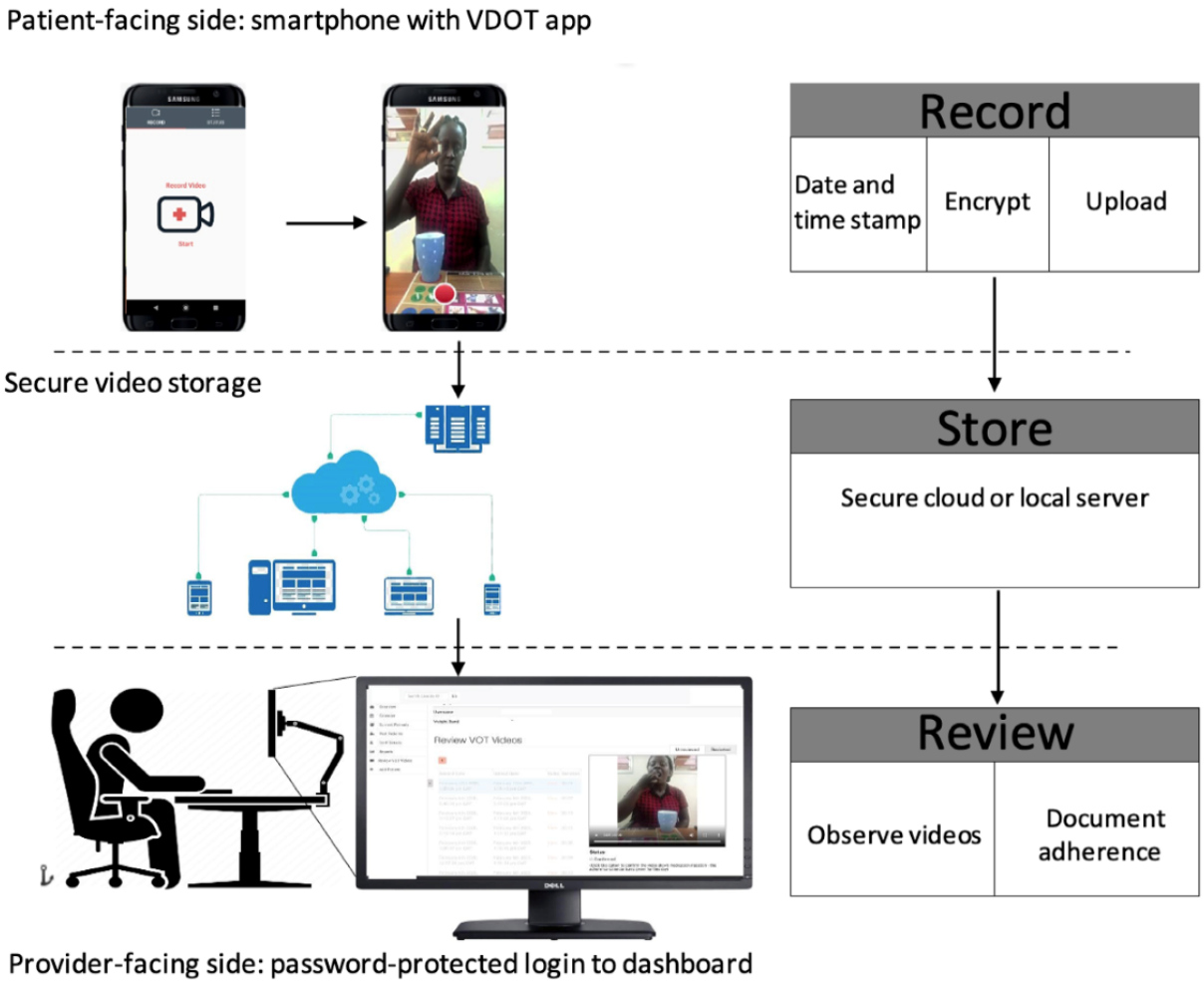
Study Design

We conducted a cross-sectional qualitative study of 5 focus group discussions involving a total of 30 participants. The focus groups were composed of a minimum of 6 participants, as recommended for qualitative studies [33,34]. Two groups comprised patients with TB stratified by sex, 1 group had caregivers, 1 group had TB health care providers (nurses and clinicians), and 1 group had trained DOT community volunteer workers. The caregiver category was broadly defined as any person (a spouse, other family member, or a friend) who regularly provided supportive care, such as accompanying the patient during the TB clinic visits. A DOT community worker was a volunteer who was previously trained by the National TB Program and designated to support DOT in the community for patients attending a specific TB clinic. The patient focus groups were stratified by sex to create more homogeneous groups to foster an environment for free discussion. Owing to cultural norms in the Ugandan settings, men tend to dominate conversations and women tend to be less inclined to speak in the presence of men. Therefore, sex stratification was important whenever possible, and the focus groups of health care providers and DOT community workers were not stratified by sex, as the job positions and number of workers were somewhat fixed. We used a hypothetical scenario to describe the step-by-step process of VDOT, as none of the participants had prior experience with it. We then gathered participants' perceptions of the potential benefits and barriers.

Detailed Description of VDOT

The hypothetical scenario involved the following: a description coupled with a demonstration of how VDOT works were offered to the participants before the start of interviews in the focus group discussions. The VDOT system, as shown in Figure 1, has 3 main components: a patient-facing side, which is composed of a smartphone app for recording and submitting videos; a secure Health Insurance Portability and Accountability Act–compliant cloud server that stores the encrypted videos; and a provider-facing, computer-based log-in system with a dashboard where submitted medication videos are watched and adherence is confirmed. Daily dosing can be monitored, and individual or aggregated reports can be generated within the system. The system also sends automatic text reminders to patients' phone numbers. A second reminder was sent 8 hours later if the medication video was not received in the system. An example of these messages is "It's time to take your pills and send a video. Taking your pills will help you to get cured." A translated version of the message is also sent in a local dialect (Luganda), which is predominantly used in the Kampala region.

Figure 1. Schematic of the asynchronous video directly observed therapy system for monitoring tuberculosis treatment.



Participant Recruitment and Enrollment

A purposive sample of adult male and female patients with TB with their caregivers was invited to participate in the focus groups. Participants were approached face to face with the help of a clinic nurse at the TB clinic under the National TB Program. Patients with TB were eligible if they were aged ≥18 years and were receiving treatment for at least 2 months under the usual in-person DOT to ensure that they were no longer infectious. Consenting participants were recruited from 4 public TB clinics. Community DOT volunteers were also selected according to the participating TB clinics to which they were attached. All the participants who were invited agreed to participate in the study.

Data Collection and Focus Group Procedures

Focus group discussions were conducted in Luganda (a commonly spoken local dialect) for patients, caregivers, and community DOT workers and in English for the health care providers. The moderator (RT), a male qualitative interview expert, reviewed the purpose of the study and the agenda for the meeting. He then described the VDOT process and demonstrated how it works using a smartphone and the app. Participants were shown the special features of the app, password access to the app to ensure privacy, and how the videos were recorded and sent to the cloud system. In addition, the provider-facing dashboard and what happens during a video review session were shown on the computer. RT introduced the interview topics and posed questions, followed by probing. The assistant moderator (RK), a female social worker with vast experience in qualitative data collection, asked follow-up questions, audio-recorded the discussion, and captured nonverbal expressions to enrich the collected data. Interview guides were pretested and revised, and the final versions were used for the data collection process (see an example of the guide in Multimedia Appendix 1). RK and other trained research team members also administered a brief exit survey on demographics, phone ownership and use, and perceived acceptability rating. Focus group discussion sessions were conducted at a selected clinic over the weekends to minimize disruption of routine health care activities and ensure that nonparticipants were not present. Each session lasted between 75 and 125 minutes.

Qualitative Data Processing and Analysis

To ensure data integrity, RT transcribed the data verbatim from the audio-recordings in an iterative process within 48 hours of completing the interviews. During transcription, RK reviewed the transcripts for accuracy by playing back the audio-recordings while reading the transcripts. In case of a discrepancy, the transcripts were edited to match the audio-recordings. The transcripts were then imported into ATLAS.ti for thematic content analysis using an inductive approach. Two authors (RT

and RK) independently analyzed 3 phases: data immersion, coding, and coding sort. In the first phase, the research team led by JS and EB reviewed the transcripts several times to achieve familiarity and identify emerging issues. In the second phase, the authors flagged relevant transcripts with appropriate descriptive words (codes). In the third phase, the reviewers met and harmonized the independent codes. The harmonized codes were then combined to form categories, and the categories were combined to form emergent themes and subthemes. A third independent reviewer (VK) with vast qualitative analysis expertise verified the codes and themes. The reporting of the study results was guided by the Consolidated Criteria for Reporting Qualitative Studies guidelines [35]. Data from the exit survey were summarized as frequencies, means, SDs, and percentages. Summary statistics of the acceptability rating scores on a scale of 1 to 10 are presented as means and SDs (Table 1). For this study, the scores were further categorized into 3 groups to reflect levels of acceptability as low (1-4), moderate (5-7), or high (8-10).

Table 1. Baseline characteristics and technology experience of focus group participants in Kampala, Uganda, 2018 (N=30).

Demographic characteristics	Stakeholder category					
	Female patients with TB[a] (n=6)	Male patients with TB (n=6)	Health care providers treating TB (n=6)	DOT[b] community workers (n=6)	Caregivers (n=6)	All participants[c]
Age, mean (range)	28.7 (20-37)	29 (24-35)	39.3 (30-57)	34.7 (25-62)	40.7 (28-54)	34.5 (10.7; 20-62)
Highest level of education, n						
Primary	3	3	0	2	0	8 (27)
Secondary	3	1	0	0	2	6 (20)
Tertiary or university	0	2	6	4	4	16 (53)
Cell phone ownership, n						
Basic feature phone only	4	3	0	2	3	12 (40)
Smartphone only	2	2	4	3	3	14 (47)
Both regular and smartphone	0	1	2	1	0	4 (13)
Cell phone experience, n						
Uses cell phone camera regularly—yes	2	3	6	4	4	19 (63)
Takes selfies—yes	1	2	4	3	3	13 (43)
Uses phone to take videos—yes	1	3	5	5	4	18 (60)
Sends photographs or videos via phone—yes	1	4	6	5	4	20 (67)
Uses phone for internet access regularly—yes	1	4	6	4	4	19 (63)
Uses WhatsApp or Facebook on phone—yes	1	5	6	5	5	22 (73)
Level of perceived acceptability for VDOT[d] on a scale of 1 to 10, n						
Value, mean (range)	N/A[e]	N/A	N/A	N/A	N/A	8.23 (1.87; 5-10)
1-4 (low)	0	0	0	0	0	0 (0)
5-7 (moderate)	1	2	2	3	1	9 (30)
8-10 (high)	5	4	4	3	5	21 (70)

[a]TB: tuberculosis.

[b]DOT: directly observed therapy.

[c]Values in the column *All participants* are presented as mean (SD; range) for mean age, and as n (%) for the remaining characteristics.

[d]VDOT: video directly observed therapy.

[e]N/A: not applicable.

Results

Characteristics of Participants

A total of 5 focus groups were conducted with a total of 30 participants, with an average age of 34.5 (SD 10.7) years and an age range of 20 to 62 years. All participants owned a cell phone, but more health care providers owned smartphones compared with other categories. Overall, experience with cell phone use was modest, with at least 60% (18/30) using their phones to take photographs, record videos, or use the internet. However, patients reported relatively lower use of these phone features of interest compared with health care providers. Participants expressed mixed views about acceptability, with ratings varying somewhat across categories. The mean score for acceptability was 8.2 (SD 1.87), with a range of 5 to 10 on a scale of 1 to 10. About 70% (21/30) of participants rated VDOT as high, with a score of >8, whereas 30% (9/30) perceived it as moderately acceptable. Detailed information on the baseline characteristics of the participants is provided in Table 1.

Overview of Results

The results from the focus group discussions are broadly categorized into perceived benefits and barriers. Within each category, we present themes and subthemes that emerged across the 4 categories of participants, including patients, caregivers, community DOT workers, and health care providers.

Perceived Benefits of VDOT

Three themes and interrelated subthemes emerged from discussions on perceived benefits. These included easy monitoring and support of adherence, enhancing patient-provider communication, and saving time and money. The subthemes presented under ease of monitoring reflect an aspect of this theme. These perceived benefits of using technology for monitoring TB medication adherence could indicate the degree of acceptability of VDOT.

Easy Monitoring of Medication Adherence

Participants generally perceived that using VDOT would make it easier to monitor medication adherence. Community DOT workers and TB health care providers revealed that sometimes patients may falsely report taking their medications under the standard *DOT* method. For context, the TB health care providers in the focus group acknowledged that the Uganda TB program faces practical challenges, such as limited personnel, which makes it difficult to administer proper in-person DOT. Ideally, DOT should involve consistent observation of the patient's daily swallowing of medications by a health worker. However, in practice, most patients are seldom supported by designated community DOT volunteer workers.

Subtheme A: Objective Evidence of Adherence

Patients often take medicine on their own (self-administered treatment) and then self-report adherence at monthly clinic visits. In such situations, the health workers believed that VDOT would be more reliable because it provides video evidence of swallowing the medicines instead of relying on patients' self-reports:

> *I think VDOT is more evidence-based, you find that you'll be seeing what the patient is doing not like DOT where you wait for the patients to tell you what they did.* [TB health provider #2]

> *But this video will be able to show you that this person has taken drugs, because you have to place the pills on the tongue to show us that you have really taken drugs and take water, then you put a second pill and take some water...* [Community DOT worker #5]

Subtheme B: Timely Follow-up of Missed Doses

Patients perceived VDOT to be beneficial for the early identification of missed doses and follow-up to solve problems. Patients thought the technology would help a health worker to find out what is going on with a patient before the next visit and also plan to intervene early if a patient is not doing well:

> *When using video, it helps the health worker to follow you up, it becomes easy for the health worker to follow you up like yesterday you were not able to take your drugs then he can even remind you that your time has passed and you should have taken your drugs. At times, the health workers think we don't take the drugs.* [Patient #5, female]

> *The health worker will be able to follow up the patient and she will be able to know how you are doing at times the health workers think that we don't take the drugs, they know some of us don't take the drugs and when I tell him or her that's what he takes. But on the video [VDOT] the health worker will be on the right track.* [Patient #4, female]

Subtheme C: VDOT Is Convenient and Efficient

Participants described VDOT as a convenient and efficient strategy for observing many patients within a limited time while being seated at one place compared with a community DOT worker who moves from place to place to observe medication intake:

> *I think the VDOT may be more effective like p#4 [another participant] said that you may be having 50 patients to attend to or follow up but I am not going to be around to watch all these people when they are taking their drugs in one day,...[with VDOT] you will sit on your computer in one place and monitor all the people and you will be able to follow up people by calling only those whose videos are missing...* [Community DOT worker #3]

> *You know we left analogue ["old fashioned ways"] and we are now in digital ["modern ways"], and everything is digitalized so this method seems to be more improved than the analogue method where you have to monitor someone by visiting that person, this one will help the health workers to easily monitor a patient when the patient is in his/her room and the health worker is in his office without having to wait for the patient to come...* [Caregiver #5]

Improved Patient-Provider Communication

Although VDOT would take away the in-person contact between patients and health workers provided by DOT, some participants believed that VDOT would increase provider-patient communication. Patients were pleased to learn that the asynchronous VDOT sessions would have no time limitations, instead these sessions would allow them to share their treatment-related problems. Health workers agreed that by eliminating the time constraints, VDOT would enable them to focus more on patients' needs and concerns, thereby fostering good patient-provider relationships and improving medication adherence:

> *[Patient] talking when taking drugs creates a good atmosphere between the patient and the health worker. With time, the patient will develop trust in the health worker because of that routine interaction. This will encourage him or her to continue taking medications.* [TB health care provider #3]
>
> *This video is going to help the patient to air out his/her complaints, not complaints but how the patient is progressing with the treatment, or the side effects the patient is getting out of these drugs like my legs paralyzes, I vomit, I feel pain in such and such a place, so this person will have got a platform to air out his or her problems early than waiting to see the health worker [at the clinic or home visit]...* [Community DOT worker #6]

VDOT Saves Money and Time

Both health workers and patients believed that VDOT could potentially save both time and money. Health workers pointed out that patients' visits at home and workplaces for DOT are costly, especially when visiting multiple patients daily. They thought that the cost of internet data might be much lower than the transportation costs. Participants also believed that time spent in transit to patients' homes and workplaces, especially with traffic jams, could be used in ways that would be more beneficial to the patients:

> *It [VDOT] is straightforward; everywhere you are, any time you want. [However] this health worker who has to come to you may lack time, she/he may have other things to do. Also [with VDOT], the health worker will save on transport. Where she has used UGX 5000 [on transportation], she will use UGX 400 on the cost of internet data.* [Patient #4, female]
>
> *With a video in practice, I will not have to waste my time visiting many patients or even telling the DOT volunteers to schedule visits with many patients. We shall only visit those who do not send their videos and use time saved to do other work like the paper work at facilities.* [TB health care provider #1]

Perceived Barriers to VDOT

Although VDOT was generally perceived to have several benefits by the respondents, they highlighted some potential barriers to its use. A total of 4 themes and related subthemes emerged prominently from the discussion. These included concerns about limited technology skills among users, inadequate technical infrastructure to support VDOT use, costs related to using the technology, and disruption of domestic routines for female patients. Subthemes that reflect aspects of inadequate infrastructure are also presented. Perceived barriers to VDOT are likely to reduce the acceptability and speed of adoption during large-scale implementation. The detailed findings and related quotes are provided later.

Limited Technology Usability Skills

Participants were concerned that patients with limited technological skills would have a difficult time using VDOT successfully. Although all of the participants owned cell phones, very few patients owned smartphones and were not very familiar with features such as video functions or mobile apps. Several participants described the difficulties they faced while trying to operate a smartphone:

> *Because of my age, I might have challenges in taking a good video, because I cannot operate my phone [smartphone] very well or I cannot focus it [the camera] very well so what I do I just go to someone to ask that person to help me to do what I want to do with my phone, so smart phones are complicated for some people...* [TB health care provider #4]
>
> *The challenge I see is that like my brother here P5 [fellow participant in group] who said that he doesn't know how to use a smart phone, he will find it hard to record a video and send it even when he is given MBs.* [Patient #4, male]

Inadequate Technical Infrastructure

Subtheme A: Poor Cellular Network Connectivity and Internet Access

Some community DOT workers and health care providers expressed concerns about the unstable cellular network coverage in some areas. They worried about the disruption of the video recording, submission process, and review by the health workers. They also stated that an unstable network could result in more internet data being used up, thereby increasing the operational costs of using VDOT:

> *If there is no network or internet, you [a health provider] will not be able to receive that video...If there is no network where this person [patient] is, he [patient] will not be able to send the video, and you [health provider] will not get the video that day, or the following day.* [Community DOT worker #2]
>
> *Some areas have poor internet network, and at times there is no network at all...because of poor network, where you would have used five hundred shillings worth of internet M.B.s, you find yourself rather using two thousand shillings worth.* [Community DOT worker #6]

Subtheme B: Unstable Electricity

Some participants also had concerns about unstable electricity supply that might prevent some patients with TB from charging the smartphone battery, interrupt daily video recordings and uploads, and disrupt adherence monitoring:

> *The challenge...[is] power may go off when you have not charged the phone and yet it is time to take the drugs, power is off, but you want the health worker to see you so you have nothing to do.* [Patient #2, female]

> *Sometimes we run out of battery power, then you have a power blackout even going on for two days, and the phone is off. So, I think that could be one of the challenges...the phone is not charged...also without power you do not have enough light to take your video.* [TB health care provider #2]

Recurrent Costs Associated With Using VDOT

Participants in each group expressed concerns about the costs associated with using VDOT. Most patients stated that although they would be willing to record videos of themselves while taking medications, the recurrent costs of accessing the internet would be a major barrier. There were concerns that for some low-income patients, daily recording of videos might require them to choose between purchasing food and internet bundles:

> *There is a situation when it is time for taking drugs, and you have to take a video. At that moment you may be having only 500UGX on your table. You will not use that money to buy internet M.B.s when you don't have money to buy something to eat...so you will find yourself missing one of the two.* [Patient #5, male]

> *With the data everyone is suffering with this infection differently at times you have no strength and you can't afford to go and work so you find that even data will be another challenge because at times I take a week without working when I feel so weak and it comes to 8.00am when I don't want to get out of the bed.* [Patient #1, female]

VDOT May Disrupt Domestic Work Routines

Some female participants were concerned about how VDOT might interfere with their domestic routines, for example, cleaning, cooking, and caring for the family. These duties are mostly performed by women in the Ugandan context, and they often fill up the entire day:

> *In the morning, we usually have no time. You are preparing yourself or preparing children for school. Now you are telling me that I [should] get the phone [to record a video]? The time is little.* [Patient #3, female]

VDOT as a Potential Benefit and Barrier

Mixed perceptions of confidentiality, privacy, and other closely related issues emerged prominently in the discussions across all participant categories. Although some respondents believed that VDOT could potentially enhance privacy, others expressed fear that their privacy and confidentiality could be breached. Both ends of the spectrum are presented as subthemes to show the perceived benefits of and threats to privacy or confidentiality.

Subtheme A: Benefits of Preserving Privacy

Participants believed that the VDOT method would allow patients to take their medications discreetly, unlike DOT where community members witness health workers visiting their neighbors daily for medication intake. Patients revealed that they would be motivated to use VDOT if their TB status would be kept private, and videos would not be witnessed by non–health care staff:

> *I don't want health workers at my home; everyone will know that I have TB, which will scare me from taking my drugs. So, the big thing VDOT will save us [from] is the issue of health workers following us in the community; there is no privacy because you will come with a car which has the hospital words, it will expose me, so for this one [VDOT] I will be in my room and I will take a video and send it to the health worker and that will be all.* [Patient #3, male]

> *With the DOT method, they [DOT workers] come to my home. If they do not find you, they will ask a neighbor where you work. So you will just see them coming to your work. Or they will find you in the market and people will be wondering why they are looking for you...but with VDOT you will do your things in a private way.* [Caregiver #4]

Subtheme B: Perceived Threats to Privacy

Unintended disclosure of the TB disease status to close family members or workmates was cited as one of the concerns that could result from the use of VDOT. Some participants feared that videos sent to a health worker could be shared or accessed by other people, and others worried that the process of recording a video might attract unwanted attention from a passerby in the vicinity. Some patients also expressed perceptions of stigma as a result of unintended disclosure of the TB disease status:

> *The video we would use it but now like us who are married I had already talked about it, I am not married fully because my husband comes from the other end he has another home [polygamous situation] I have never disclosed to him that I have TB he may get that video there he gets to know and its bad.* [Patient #2, female]

> *My fear is taking a video and it is shared with someone else...when people realize that you have TB they try to dissociate from you. People are scared of TB, even me. I also protect myself with the people I work with, so it does not require people to know that you have TB.* [Patient #1, male]

Subtheme C: Breach in Confidentiality and Security of Videos

Patients, health workers, and community workers expressed concerns and fears about the confidentiality of information in the videos once received on the health system side. Multiple participants asked about who would have access to the videos and whether the health workers could be trusted to protect the videos. Some patients feared that uploaded videos might be

shared on social media sites or be used for advertisements without their permission:

> *I would be concerned about that because in most cases people misuse social media, so I may send you my video taking medications, and the video may be misused and sent for advertisement or something like that.* [Patient #5, male]

Subtheme D: Lack of a Private Place to Record Videos

Some participants raised another threat to privacy stemming from the lack of a convenient place to record videos. They noted that some patients might be uncomfortable or embarrassed to reveal their living or work environment to health care providers through video recordings. This may deter patients from recording and uploading VDOT videos:

> *People will even fear to record videos in such environment [their homes]...they don't want you [the health worker] to know where they stay.* [Community DOT worker #1]

> *It would be difficult for our clients because most of them work in the market, [or in] taxis where there so many people. So getting a private place to take his medicine and taking the video will be very difficult.* [TB health care provider #2]

> *When you are at [the] work place, there are many people. As you know the setup of the town, we are many people working from [a] shed. For you [a patient] to get out the [TB] drugs then [record] the video because you are sending it to your health care worker, it would not be good.* [Patient #5, female]

Discussion

Principal Findings

In this study, we explored perceived acceptability and benefits of and barriers to using VDOT among patients with TB, their caregivers, health care providers, and community DOT volunteer workers in Uganda. Although we used a hypothetical scenario of VDOT to interview participants who had no prior experience, we gathered very rich information about perceived benefits and barriers. We found that VDOT was perceived as acceptable by most participants, despite some differences between patients and providers. Prominent benefits related to the ease of patient monitoring; timely follow-up of missed doses; and enhancement of provider-patient communication, efficiency, and cost savings. The perceived barriers cited were limited technology skills; suboptimal technical infrastructure; costs of smartphones; internet access; and disruption of routine domestic work, especially for female patients. There were mixed perceptions about the impact of VDOT on privacy and confidentiality, particularly with unintended disease closure, potentially resulting in stigma. Similar perceptions of structural and privacy- and patient-related barriers to VDOT implementation have been previously reported [24,36,37]. For example, some studies have reported that video recordings and repeated text reminders are perceived as being overly intrusive [37,38]. In contrast, a study in India reported that patients with TB perceived VDOT as being more private than DOT [24]. The mixed findings underscore the need for a critical evaluation of potential barriers within local contexts. To our knowledge, this is the first study to examine the perspectives of key stakeholders in the use of VDOT in Uganda. More studies are needed to contribute to the sparse evidence base on the topic.

The perceived benefits of VDOT highlight aspects that can be enhanced to increase acceptability among users. The ease of monitoring, convenience, access to objective evidence of dosing, and facilitation of timely follow-up have also been reported in previous studies [23,25,37,39]. Health care providers and patients perceived VDOT as a cost- and time-saving approach in the long run compared with costs that are likely to be incurred with frequent travel when using in-person DOT. Similar findings were reported in a study of health care providers in the private sector in urban Vietnam [39]. In addition, 4 studies quantified the time spent on the treatment observation process and found a significantly larger saving in time and money compared with the usual DOT [21,36,40,41]. VDOT was also perceived to enhance health care provider–patient communication, which in turn affects patient engagement with their treatment [42]. The consistency in these positive aspects makes VDOT a promising patient-centered approach for TB disease management.

The perceived barriers to using VDOT were mostly related to technology usability skills, given the limited experience with smartphones and apps. The technology usability barrier is likely to affect mostly older patients and those without formal education. Similar findings have been reported in VDOT studies conducted in Vietnam, Cambodia, and South India [23,37,39]. In our experience with the VDOT pilot study, intensive training with clear instructions to patients helped them gain the required skills. Inadequate network connectivity and electricity are structural barriers that are more challenging to overcome at the patient or programmatic level. Failure to charge a dead phone battery is one of the most common reasons for missing videos in the VDOT system [43]. The use of solar power banks for charging smartphones might be a short-term solution where electricity is unstable. The absolute costs of smartphones with internet data plans have also been documented as potential barriers in other studies [23,41]. The cost of smartphones on the global market is gradually decreasing, thereby boosting ownership, even in LMICs [44]. Our VDOT pilot study in Kampala showed that 70% of the participants owned a smartphone [43]. However, smartphones may remain unaffordable for some patients with TB. One possible way to minimize inequitable access to digital interventions is to set up a loaner system where patients can borrow and return their phones on completion of treatment. We used a similar system in our VDOT pilot study in Uganda, which worked well [43]. Regarding the cost of the internet, preloading the phones with prepaid internet could be a good option, but it raises other concerns of misusing it for other personal activities. Other creative ways can be explored to address the cost of internet access to support patients' video submissions. For example, public-private partnerships between the National TB Program and telecommunication companies might tap into resources designated for corporate social responsibility from the business side. Gender roles emerged as a potential issue, especially for women, but have been rarely cited [38], warranting the need

for more research. Despite these perceived barriers, patients and health workers believed that VDOT would be more flexible, convenient, and patient-friendly if the main barriers are addressed. Cost and cost-effectiveness studies are needed to inform the implementation and scale-up of VDOT [44].

Privacy and confidentiality issues with the use of VDOT were raised as perceived benefits and potential threats. Previous VDOT studies conducted in India, the United States, and Vietnam have also reported similar mixed concerns with privacy [24,25,39]. The importance of preserving confidentiality, guarding against unintended disease disclosure, and stigma also emerged as prominent themes in studies on acceptance of mobile health (mHealth) interventions among people living with HIV in rural and urban Uganda [26,30,45]. Indeed, the use of any mHealth interventions in monitoring treatment should be ethically comparable with the standard of care [38]. Specifically, for VDOT, the major privacy concerns raised have been addressed by several security features such as a unique personal identification number to facilitate secure log-in into the phone app, video encryption, secure cloud server, and password-protected log-in to the health system dashboard for the health care provider. However, to minimize feelings of mistrust, providers need to reassure patients that caution should be taken to prevent intentional breaches. It is also important to understand the complex cultural perspectives related to perceptions of privacy and cater to them accordingly. For example, some patients in South India perceived repeated adherence text message reminders as intrusive [37]. Finally, patients should also be encouraged to be active players in the process of protecting their own information, for example, by finding a private place to record videos and not sharing their personal identification numbers. DiStefano and Schmidt [38] proposed a valuable framework to guide ethical planning, implementation, and evaluation when using mHealth interventions. The collective goal is to minimize stigmatization and preserve patient autonomy [38].

Broadly, digital adherence technology studies on VDOT, electronic pillbox or Medication Event Reminder Monitor, and 99DOTS performed in LMICs have shown promising acceptability patterns among TB populations [22,37,43,46-48]. General lessons can be learned and applied across interventions and populations. For example, high acceptance in recent studies of VDOT was related to perceptions of convenience, ease of use, and perception of better privacy, whereas lower acceptance was mostly tied to technology skills, app glitches, and cellular connectivity challenges [23,24,36]. In India, a qualitative study evaluating differences in acceptability of 99DOTS, a low-cost, phone call–based strategy for reporting doses, found that high acceptance was related to improved patient-provider communication and the convenience of reduced clinic visits, among others [48]. Similar to our study, low acceptance was related to concerns about cell phone access, technology literacy, and poor cellular connectivity. Health care providers specifically expressed concerns about inadequate training in the use of the technology, changes in workload, and a lack of needed 99DOTS supplies. Although the health care providers in our VDOT study only had a hypothetical exposure, it is possible that similar challenges could arise. In a qualitative study using the electronic pillbox in patients with TB in Vietnam, the technology was perceived to be useful. However, the study participants pointed out that it would be most beneficial as a medication reminder for older patients. The device was less acceptable for people who worked outside of their homes, as they thought that the device was inconvenient to carry around [47]. Concerns about stigma, disease disclosure, or other privacy issues and costs related to technology seem to be crosscutting but to varying degrees from one study to another. Further research is required to expand our understanding on this key area.

Strengths and Limitations

This study has several strengths and limitations that should be considered when interpreting the findings. To our knowledge, this exploratory qualitative study is among the first to document the perspectives of TB stakeholders in relation to VDOT in Uganda. It provided helpful insights into potential barriers and benefits that informed the design of the quantitative pilot study that subsequently followed [43]. One limitation of this study was that we used hypothetical scenarios of VDOT; therefore, respondents had no real-life experience using the technology intervention. This could have resulted in increased socially desirable responses and perhaps limited the ability of the respondents to envision potential facilitators and barriers comprehensively. Despite this limitation, respondents identified some critical issues such as cost of internet, the smartphone, and privacy or confidentiality concerns. The sample of respondents was limited to urban Uganda; therefore, findings may not necessarily be generalizable to all TB stakeholders in this local setting. The small number of focus group discussions could have fallen short of reaching the point of saturation. Overall, the findings of this study build on the sparse evidence on the acceptability of digital adherence technologies and their use in TB management in LMICs. More robust research is still needed to explore various aspects of acceptability, feasibility, and ethical issues related to VDOT. Other future research priorities have been highlighted as accuracy, clinical effectiveness, and cost-effectiveness of digital technologies [44].

Future Implications

In the real world, successful implementation of VDOT will require both patients and health care providers to be equipped and willing to use the system. Users must have an adequate level of skills to operate the smartphone, mobile app, and technology system on the back end for health care providers. The baseline characteristics showed that the cell phone technology experience was higher among health care providers than among patients. This will require intensive training and perhaps run-in periods needed to allow users to gain a baseline functional level of skills at the beginning of a VDOT monitoring program. It is plausible that as individuals use VDOT, their level of comfort and usability skills will increase over time [49]. Special considerations must be made for subgroups such as older adults who may need more time to acquire new technology skills [50]. Any modifiable barriers must be addressed to ensure high uptake and sustained engagement with VDOT. Context- and culture-specific barriers for patients and health care providers must be evaluated before deploying the technology.

The influence of digital technology use on gender roles emerged as a new aspect with little or no information from previous studies; therefore, further research is needed, especially in the African context.

Conclusions

VDOT was relatively acceptable and perceived as beneficial by most study participants despite the potential technical and cost barriers. There were mixed perceptions about privacy and confidentiality issues related to the use of VDOT. Although some participants thought it would increase patients' autonomy and privacy, others indicated fears about unintended disclosure of one's disease status that could lead to stigma. Future efforts should focus on training users, ensuring adequate technical infrastructure, assurance of privacy, and comparative cost analysis studies in the local context.

Acknowledgments

This work was supported by the University of Georgia Research Foundation, through the Office of the Vice President for Research and the Global Collaboration Research Grant. The authors would like to thank Ronald Tenywa and Robinah Komuhendo for their technical assistance in conducting and transcribing the focus group interviews; Julius Turinawe for his assistance in data collection; and Joan Namatovu, Nicholas Kawuki, and Jonah Lubega for providing the administrative and technological support for the video directly observed therapy (VDOT) pilot study at the Lubaga tuberculosis (TB) clinic. In addition, the authors would like to thank Dr Christopher Whalen and Dr Noah Kiwanuka for their input during the study development and providing a physical space for the research team. We thank all the participants and the TB clinics for providing access to the respondents and Dr Kelly Collins, Dr Richard Garfein, and Ms Lisa Skow of SureAdhere Mobile Technology for the technical support and training of the research team on the use of VDOT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient focus group interview guide.
[DOCX File , 29 KB - mhealth_v9i10e27131_app1.docx]

References

1. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new end TB strategy. Lancet 2015 May;385(9979):1799-1801. [doi: 10.1016/s0140-6736(15)60570-0]
2. Global tuberculosis report 2020. World Health Organization. 2020 Oct 14. URL: https://www.who.int/publications/i/item/9789240013131 [accessed 2021-09-23]
3. Moonan PK, Quitugua TN, Pogoda JM, Woo G, Drewyer G, Sahbazian B, et al. Does directly observed therapy (DOT) reduce drug resistant tuberculosis? BMC Public Health 2011 Jan 07;11(1):19 [FREE Full text] [doi: 10.1186/1471-2458-11-19] [Medline: 21214913]
4. Whalen CC, Nsubuga P, Okwera A, Johnson JL, Hom DL, Michael NL, et al. Impact of pulmonary tuberculosis on survival of HIV-infected adults: a prospective epidemiologic study in Uganda. AIDS 2000 Jun 16;14(9):1219-1228 [FREE Full text] [doi: 10.1097/00002030-200006160-00020] [Medline: 10894287]
5. Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2015:3343-3387. [doi: 10.1002/14651858.cd003343.pub4]
6. Toczek A, Cox H, du Cros P, Cooke G, Ford N. Strategies for reducing treatment default in drug-resistant tuberculosis: systematic review and meta-analysis. Int J Tuberc Lung Dis 2013 Mar 01;17(3):299-307. [doi: 10.5588/ijtld.12.0537] [Medline: 23211716]
7. Keugoung B, Fouelifack FY, Fotsing R, Macq J, Meli J, Criel B. A systematic review of missed opportunities for improving tuberculosis and HIV/AIDS control in Sub-saharan Africa: what is still missed by health experts? Pan Afr Med J 2014;18:320 [FREE Full text] [doi: 10.11604/pamj.2014.18.320.4066] [Medline: 25478041]
8. The Uganda National Tuberculosis Prevalence Survey, 2014-2015 Survey Report. The Uganda Ministry of Health. 2015 Jul. URL: https://www.health.go.ug/cause/the-uganda-national-tuberculosis-prevalence-survey-2014-2015-survey-report/ [accessed 2021-09-23]
9. Ministry of Health Manual of the National Tuberculosis and Leprosy Programme. The Uganda Ministry of Health. 2010. URL: http://library.health.go.ug/publications/tuberculosis/ministry-health-manual-national-tuberculosis-and-leprosy-programme [accessed 2021-09-23]
10. Mafigiri DK, McGrath JW, Whalen CC. Task shifting for tuberculosis control: a qualitative study of community-based directly observed therapy in urban Uganda. Glob Public Health 2012 Mar;7(3):270-284 [FREE Full text] [doi: 10.1080/17441692.2011.552067] [Medline: 21360381]

11. Gabriel AP, Mercado CP. Evaluation of task shifting in community-based DOTS program as an effective control strategy for tuberculosis. Scient World J 2011;11:2178-2186 [FREE Full text] [doi: 10.1100/2011/984321] [Medline: 22125465]
12. Okello D, Floyd K, Adatu F, Odeke R, Gargioni G. Cost and cost-effectiveness of community-based care for tuberculosis patients in rural Uganda. Int J Tuberc Lung Dis 2003 Sep;7(9 Suppl 1):72-79. [Medline: 12971657]
13. Adatu F, Odeke R, Mugenyi M, Gargioni G, McCray E, Schneider E, et al. Implementation of the DOTS strategy for tuberculosis control in rural Kiboga District, Uganda, offering patients the option of treatment supervision in the community, 1998-1999. Int J Tuberc Lung Dis 2003 Sep;7(9 Suppl 1):63-71. [Medline: 12971656]
14. Izudi J, Tamwesigire IK, Bajunirwe F. Explaining the successes and failures of tuberculosis treatment programs; a tale of two regions in rural eastern Uganda. BMC Health Serv Res 2019 Dec 19;19(1):979 [FREE Full text] [doi: 10.1186/s12913-019-4834-2] [Medline: 31856817]
15. Guidelines for treatment of drug-susceptible tuberculosis and patient care (2017 update). World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/255052/9789241550000-eng.pdf [accessed 2021-09-23]
16. Mirsaeidi M, Farshidpour M, Banks-Tripp D, Hashmi S, Kujoth C, Schraufnagel D. Video directly observed therapy for treatment of tuberculosis is patient-oriented and cost-effective. Eur Respir J 2015 Sep 18;46(3):871-874 [FREE Full text] [doi: 10.1183/09031936.00011015] [Medline: 25792632]
17. Story A, Garfein RS, Hayward A, Rusovich V, Dadu A, Soltan V, et al. Monitoring therapy compliance of tuberculosis patients by using video-enabled electronic devices. Emerg Infect Dis 2016 Mar;22(3):538-540 [FREE Full text] [doi: 10.3201/eid2203.151620] [Medline: 26891363]
18. Garfein RS, Liu L, Cuevas-Mota J, Collins K, Muñoz F, Catanzaro DG, et al. Tuberculosis treatment monitoring by video directly observed therapy in 5 health districts, California, USA. Emerg Infect Dis 2018 Oct;24(10):1806-1815 [FREE Full text] [doi: 10.3201/eid2410.180459] [Medline: 30226154]
19. Ngwatu BK, Nsengiyumva NP, Oxlade O, Mappin-Kasirer B, Nguyen NL, Jaramillo E, Collaborative Group on the Impact of Digital Technologies on TB. The impact of digital health technologies on tuberculosis treatment: a systematic review. Eur Respir J 2018 Jan 11;51(1):1701596 [FREE Full text] [doi: 10.1183/13993003.01596-2017] [Medline: 29326332]
20. Garfein RS, Collins K, Muñoz F, Moser K, Cerecer-Callu P, Raab F, et al. Feasibility of tuberculosis treatment monitoring by video directly observed therapy: a binational pilot study. Int J Tuberc Lung Dis 2015 Sep 01;19(9):1057-1064 [FREE Full text] [doi: 10.5588/ijtld.14.0923] [Medline: 26260824]
21. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar;393(10177):1216-1224. [doi: 10.1016/s0140-6736(18)32993-3]
22. Nguyen TA, Pham MT, Nguyen TL, Nguyen VN, Pham DC, Nguyen BH, et al. Video Directly Observed Therapy to support adherence with treatment for tuberculosis in Vietnam: a prospective cohort study. Int J Infect Dis 2017 Dec;65:85-89 [FREE Full text] [doi: 10.1016/j.ijid.2017.09.029] [Medline: 29030137]
23. Rabinovich L, Molton JS, Ooi WT, Paton NI, Batra S, Yoong J. Perceptions and acceptability of digital interventions among tuberculosis patients in Cambodia: qualitative study of video-based directly observed therapy. J Med Internet Res 2020 Jul 27;22(7):e16856 [FREE Full text] [doi: 10.2196/16856] [Medline: 32716309]
24. Holzman SB, Atre S, Sahasrabudhe T, Ambike S, Jagtap D, Sayyad Y, et al. Use of smartphone-based Video Directly Observed Therapy (vDOT) in tuberculosis care: single-arm, prospective feasibility study. JMIR Form Res 2019 Aug 27;3(3):e13411 [FREE Full text] [doi: 10.2196/13411] [Medline: 31456581]
25. Zúñiga ML, Collins K, Muñoz F, Moser K, Rangel G, Cuevas-Mota J, et al. A qualitative study exploring stakeholder perceptions of video directly observed therapy for monitoring tuberculosis treatment in the US-Mexico border region. J Mobile Technol Med 2016 Jul;5(2):12-23. [doi: 10.7309/jmtm.5.2.3]
26. Campbell JI, Aturinda I, Mwesigwa E, Burns B, Santorino D, Haberer JE, et al. The Technology Acceptance Model for Resource-Limited Settings (TAM-RLS): a novel framework for mobile health interventions targeted to low-literacy end-users in resource-limited settings. AIDS Behav 2017 Nov 18;21(11):3129-3140 [FREE Full text] [doi: 10.1007/s10461-017-1765-y] [Medline: 28421356]
27. Campbell JI, Eyal N, Musiimenta A, Burns B, Natukunda S, Musinguzi N, et al. Ugandan study participants experience electronic monitoring of antiretroviral therapy adherence as welcomed pressure to adhere. AIDS Behav 2018 Oct 20;22(10):3363-3372 [FREE Full text] [doi: 10.1007/s10461-018-2200-8] [Medline: 29926301]
28. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
29. Musiimenta A, Atukunda EC, Tumuhimbise W, Pisarski EE, Tam M, Wyatt MA, et al. Acceptability and feasibility of real-time antiretroviral therapy adherence interventions in rural Uganda: mixed-method pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 May 17;6(5):e122 [FREE Full text] [doi: 10.2196/mhealth.9031] [Medline: 29773527]
30. Musiimenta A, Tumuhimbise W, Mugaba AT, Muzoora C, Armstrong-Hough M, Bangsberg D, et al. Digital monitoring technologies could enhance tuberculosis medication adherence in Uganda: mixed methods study. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100119 [FREE Full text] [doi: 10.1016/j.jctube.2019.100119] [Medline: 31788561]

31. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. PLoS Med 2007 Jul 24;4(7):e238 [FREE Full text] [doi: 10.1371/journal.pmed.0040238] [Medline: 17676945]
32. 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]
33. Carlsen B, Glenton C. What about N? A methodological study of sample-size reporting in focus group studies. BMC Med Res Methodol 2011 Mar 11;11(1):26 [FREE Full text] [doi: 10.1186/1471-2288-11-26] [Medline: 21396104]
34. Kitzinger J. Qualitative research. Introducing focus groups. Br Med J 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
35. Tong A, 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 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
36. Guo X, Yang Y, Takiff HE, Zhu M, Ma J, Zhong T, et al. A comprehensive app that improves tuberculosis treatment management through video-observed therapy: usability study. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e17658 [FREE Full text] [doi: 10.2196/17658] [Medline: 32735222]
37. Kumar AA, De Costa A, Das A, Srinivasa G, D'Souza G, Rodrigues R. Mobile health for tuberculosis management in south India: is video-based directly observed treatment an acceptable alternative? JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11687 [FREE Full text] [doi: 10.2196/11687] [Medline: 30942696]
38. DiStefano MJ, Schmidt H. mHealth for tuberculosis treatment adherence: a framework to guide ethical planning, implementation, and evaluation. Glob Health Sci Pract 2016 Jun 27;4(2):211-221. [doi: 10.9745/ghsp-d-16-00018]
39. Nguyen LH, Tran PT, Dam TA, Forse RJ, Codlin AJ, Huynh HB, et al. Assessing private provider perceptions and the acceptability of video observed treatment technology for tuberculosis treatment adherence in three cities across Viet Nam. PLoS One 2021 May 7;16(5):e0250644 [FREE Full text] [doi: 10.1371/journal.pone.0250644] [Medline: 33961645]
40. Ravenscroft L, Kettle S, Persian R, Ruda S, Severin L, Doltu S, et al. Video-observed therapy and medication adherence for tuberculosis patients: randomised controlled trial in Moldova. Eur Respir J 2020 Aug 07;56(2):2000493. [doi: 10.1183/13993003.00493-2020] [Medline: 32381495]
41. Holzman SB, Zenilman A, Shah M. Advancing patient-centered care in tuberculosis management: a mixed-methods appraisal of video directly observed therapy. Open Forum Infect Dis 2018 Apr;5(4):ofy046 [FREE Full text] [doi: 10.1093/ofid/ofy046] [Medline: 29732378]
42. Garfein RS, Doshi RP. Synchronous and asynchronous video observed therapy (VOT) for tuberculosis treatment adherence monitoring and support. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100098 [FREE Full text] [doi: 10.1016/j.jctube.2019.100098] [Medline: 31867442]
43. Sekandi JN, Buregyeya E, Zalwango S, Dobbin KK, Atuyambe L, Nakkonde D, et al. Video directly observed therapy for supporting and monitoring adherence to tuberculosis treatment in Uganda: a pilot cohort study. ERJ Open Res 2020 Jan 06;6(1):00175-2019 [FREE Full text] [doi: 10.1183/23120541.00175-2019] [Medline: 32280670]
44. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018 Oct 11;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
45. Twimukye A, Naggirinya A, Parkes-Ratanshi R, Kasirye R, Kiragga A, Castelnuovo B, et al. Acceptability of a mobile phone support tool (Call for Life Uganda) for promoting adherence to antiretroviral therapy among young adults in a randomized controlled trial: exploratory qualitative study. JMIR Mhealth Uhealth 2021 Jun 14;9(6):e17418 [FREE Full text] [doi: 10.2196/17418] [Medline: 34121665]
46. Sinkou H, Hurevich H, Rusovich V, Zhylevich L, Falzon D, de Colombani P, et al. Video-observed treatment for tuberculosis patients in Belarus: findings from the first programmatic experience. Eur Respir J 2017 Mar 22;49(3):1602049 [FREE Full text] [doi: 10.1183/13993003.02049-2016] [Medline: 28331042]
47. Drabarek D, Anh NT, Nhung NV, Hoa NB, Fox GJ, Bernays S. Implementation of medication event reminder monitors among patients diagnosed with drug susceptible tuberculosis in rural Viet Nam: a qualitative study. PLoS One 2019 Jul 22;14(7):e0219891 [FREE Full text] [doi: 10.1371/journal.pone.0219891] [Medline: 31329610]
48. Thomas BE, Kumar JV, Onongaya C, Bhatt SN, Galivanche A, Periyasamy M, et al. Explaining differences in the acceptability of 99DOTS, a cell phone-based strategy for monitoring adherence to tuberculosis medications: qualitative study of patients and health care providers. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16634 [FREE Full text] [doi: 10.2196/16634] [Medline: 32735220]
49. Do D, Garfein RS, Cuevas-Mota J, Collins K, Liu L. Change in patient comfort using mobile phones following the use of an app to monitor tuberculosis treatment adherence: longitudinal study. JMIR Mhealth Uhealth 2019 Feb 01;7(2):e11638 [FREE Full text] [doi: 10.2196/11638] [Medline: 30707103]
50. Westerman SJ, Davies DR. Acquisition and application of new technology skills: the influence of age. Occup Med (Lond) 2000 Sep 01;50(7):478-482. [doi: 10.1093/occmed/50.7.478] [Medline: 11198671]

Abbreviations

CB-DOT: community-based directly observed therapy
DOT: directly observed therapy
LMIC: low- and middle-income country
mHealth: mobile health
TB: tuberculosis
VDOT: video directly observed therapy

Edited by L Buis, G Eysenbach; submitted 13.01.21; peer-reviewed by R Garfein, M DiStefano, R Subbaraman, B Thies; comments to author 18.02.21; revised version received 15.08.21; accepted 07.09.21; published 27.10.21.

Please cite as:
Sekandi JN, Kasiita V, Onuoha NA, Zalwango S, Nakkonde D, Kaawa-Mafigiri D, Turinawe J, Kakaire R, Davis-Olwell P, Atuyambe L, Buregyeya E
Stakeholders' Perceptions of Benefits of and Barriers to Using Video-Observed Treatment for Monitoring Patients With Tuberculosis in Uganda: Exploratory Qualitative Study
JMIR Mhealth Uhealth 2021;9(10):e27131
URL: https://mhealth.jmir.org/2021/10/e27131
doi:10.2196/27131
PMID:34704961

Original Paper

The Effect of a WeChat-Based Tertiary A-Level Hospital Intervention on Medication Adherence and Risk Factor Control in Patients With Stable Coronary Artery Disease: Multicenter Prospective Study

Boqun Shi[1], MD; Xi Liu[1], MD; Qiuting Dong[1], MD; Yuxiu Yang[1], MD; Zhongxing Cai[1], MD; Haoyu Wang[1], MD; Dong Yin[1], MD; Hongjian Wang[1], MD; Kefei Dou[1*], MD; Weihua Song[1*], MD

Cardiometabolic Medicine Center, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Weihua Song, MD
Cardiometabolic Medicine Center, Fuwai Hospital
National Center for Cardiovascular Diseases
Chinese Academy of Medical Sciences and Peking Union Medical College
No 167 North Lishi Road, Xicheng District
Beijing, 100037
China
Phone: 86 1088396863
Email: songweihua@fuwai.com

Abstract

Background: In China, ischemic heart disease is the main cause of mortality. Having cardiac rehabilitation and a secondary prevention program in place is a class IA recommendation for individuals with coronary artery disease. WeChat-based interventions seem to be feasible and efficient for the follow-up and management of chronic diseases.

Objective: This study aims to evaluate the effectiveness of a tertiary A-level hospital, WeChat-based telemedicine intervention in comparison with conventional community hospital follow-up on medication adherence and risk factor control in individuals with stable coronary artery disease.

Methods: In this multicenter prospective study, 1424 patients with stable coronary artery disease in Beijing, China, were consecutively enrolled between September 2018 and September 2019 from the Fuwai Hospital and 4 community hospitals. At 1-, 3-, 6-, and 12-month follow-up, participants received healthy lifestyle recommendations and medication advice. Subsequently, the control group attended an offline outpatient clinic at 4 separate community hospitals. The intervention group had follow-up visits through WeChat-based telemedicine management. The main end point was medication adherence, which was defined as participant compliance in taking all 4 cardioprotective medications that would improve the patient's outcome (therapies included antiplatelet therapy, β-blockers, statins, and angiotensin-converting-enzyme inhibitors or angiotensin-receptor blockers). Multivariable generalized estimating equations were used to compare the primary and secondary outcomes between the 2 groups and to calculate the relative risk (RR) at 12 months. Propensity score matching and inverse probability of treatment weighting were performed as sensitivity analyses, and propensity scores were calculated using a multivariable logistic regression model.

Results: At 1 year, 88% (565/642) of patients in the intervention group and 91.8% (518/564) of patients in the control group had successful follow-up data. We matched 257 pairs of patients between the intervention and control groups. There was no obvious advantage in medication adherence with the 4 cardioprotective drugs in the intervention group (172/565, 30.4%, vs 142/518, 27.4%; RR 0.99, 95% CI 0.97-1.02; *P*=.65). The intervention measures improved smoking cessation (44/565, 7.8%, vs 118/518, 22.8%; RR 0.48, 95% CI 0.44-0.53; *P*<.001) and alcohol restriction (33/565, 5.8%, vs 91/518, 17.6%; RR 0.47, 95% CI 0.42-0.54; *P*<.001).

Conclusions: The tertiary A-level hospital, WeChat-based intervention did not improve adherence to the 4 cardioprotective medications compared with the traditional method. Tertiary A-level hospital, WeChat-based interventions have a positive effect

on improving lifestyle, such as quitting drinking and smoking, in patients with stable coronary artery disease and can be tried as a supplement to community hospital follow-up.

Trial Registration: ClinicalTrials.gov NCT04795505; https://clinicaltrials.gov/ct2/show/NCT04795505

(JMIR Mhealth Uhealth 2021;9(10):e32548) doi:10.2196/32548

KEYWORDS

WeChat; telemedicine; coronary artery disease; medication adherence; mobile phone

Introduction

Background

In China, the main cause of mortality is a cardiac condition known as ischemic heart disease [1]. According to current recommendations, having cardiac rehabilitation and a secondary prevention program in place is a class IA recommendation for individuals with coronary artery disease (CAD) [2-4]; however, there is a large gap between clinical practice and guideline recommendations. The cost of treating cardio-cerebrovascular illnesses in China was Chinese ¥540.64 billion (US $83.90 billion) in 2017. More than 80% of the costs of cardio-cerebrovascular diseases in China were incurred in hospitals and over 70% of the costs incurred in inpatient care. These allocations were unreasonable, and the primary medical and health facilities accounted for less than 12% of the costs [5]. To reduce the economic burden of cardiovascular illnesses in China, efforts have concentrated on improving the quality of treatment for acute myocardial infarctions (MIs) and percutaneous coronary interventions (PCIs) [6]. Since the development of a clinical performance quality control system for adults with acute ST-elevation MI, significant improvements have been achieved in China regarding the prescription of medications during hospitalization, and these medications are evidence-based [7]. However, approximately half of the patients with acute MI in China do not have good compliance in taking their medications after discharge, which substantially increases morbidity and mortality [8-10]. It is difficult for patients to be hospitalized in tertiary A-level hospitals in China, and many patients do not return for follow-up after discharge due to the patient perspective of treatment being much more important than prevention. The low-density lipoprotein cholesterol (LDL-C) goals were not met by a statistically significant percentage (74.5%) of individuals with a high risk of arteriosclerotic cardiovascular disease [11]. First- and second-level preventive care needs to be improved to increase patient compliance and to change modifiable risk factors [12,13].

In response to this phenomenon, facilities and agencies are trying to engage patients, change behaviors, and help to control the risk factors. Traditional patient education methods include in-office patient counseling, health seminars, follow-up via telephone calls, text messages or emails, etc. Traditional teaching methods had no effect on fatal or nonfatal MI, total revascularization, or hospitalization, according to a Cochrane comprehensive study [14], and innovative strategies are required for routine clinical use.

Tencent introduced WeChat (Chinese version: Wei Xin), a free social networking app, in January 21, 2011, to offer instant messaging services across all platforms. It not only offers basic text, voice, photo and video sharing, web-based payment, and news subscription services but also provides integration with intelligent hardware, such as smart bracelets, blood pressure (BP) monitors, and body fat scales. WeChat now has over one billion active users, making it the most popular social networking site on the planet. After considering its extensive population coverage, strong peripheral features, and seamless integration into everyday life [15], many hospitals have introduced web-based follow-up measures based on WeChat to strengthen secondary prevention measures and risk factor interventions and to improve the drug compliance of patients. WeChat-based interventions seem to be feasible and efficient for the follow-up and management of chronic diseases. A review by Chen et al [16] discussed the following reasons why WeChat might be useful in chronic illness management: (1) it provides continuous health services. Hospitals or community health centers might develop distinct WeChat groups or official WeChat accounts based on the categories of chronic illnesses. (2) WeChat can help patients change their unhealthy lifestyle by constantly sending patient education materials to them. (3) A WeChat-based follow-up approach can improve physician-patient relationships by delivering personalized health advice and enhancing user engagement. (4) Doctors can spread their successful experiences and measures widely and quickly to many patients through group messages.

Objectives

To our knowledge, no studies have compared WeChat web-based interventions with traditional community hospital follow-ups [17,18] This study evaluates the benefits of a tertiary A-level hospital WeChat-based telemedicine in comparison with a conventional community hospital follow-up on medication adherence and risk factor control in individuals with stable CAD.

Methods

Study Design

A secondary prevention telemedicine program based on the WeChat platform provided by a tertiary A-level hospital was assessed in this 2-arm, parallel multicenter prospective study. It was one of the Prevention and Control Projects of the Major Chronic Noninfectious Disease (grant 2018YFC1315600), which was supported by the Ministry of Science and Technology of China. The National Center for Cardiovascular Diseases and the Fuwai Hospital led the study design, follow-up, data collection, and analysis of this study. Trial development and reporting were in accordance with the Strengthening the

Reporting of Observational Studies in Epidemiology Statement. We registered this study on ClinicalTrials.gov (NCT04795505).

At the initial trial visit, all participants signed a written informed consent form, and the study adhered to the principles of the Declaration of Helsinki. The primary ethical committee of the National Center for Cardiovascular Diseases approved the research protocol.

Recruitment

In this multicenter prospective study, 1424 patients with stable CAD in Beijing, China, were consecutively enrolled between September 2018 and September 2019 from the Fuwai Hospital and 4 community hospitals. The inclusion criteria were as follows: participants were required to be aged at least 18 years and to have a diagnosis of stable CAD according to the guidelines [19,20]. All participants underwent coronary computed tomography angiography or coronary angiography. Patients who could potentially participate in the research were checked as outpatients and given a form to return with their information. Participants in the intervention group were required to own a smartphone with an active WeChat account and to have the ability to communicate fluently in Chinese with the cardiac rehabilitation team via WeChat. Participants in the control group were eligible to participate if they were registered in one of the 4 community hospitals. Participants were excluded if they refused to provide signed informed consent or had a life expectancy of less than a year because of comorbidities. Participants were assigned to either the intervention or control group at their own discretion. Demographic information and reasons for study withdrawal were recorded for each participant during the entire study period.

Interventions

At the 1-, 3-, 6-, and 12-month follow-ups, participants received healthy lifestyle recommendations and medication advice. Subsequently, the control group went to an offline outpatient clinic at 4 separate community hospitals. The control group received conventional outpatient cardiology care, including formal cardiac rehabilitation and secondary preventive measures whereas the intervention group had follow-up visits through WeChat-based telemedicine management (Figure 1). Participants in this group were trained on how to interact with the WeChat official account (Figure 2). Each appointment included inquiries, evaluations, and comments. A questionnaire (Multimedia Appendix 1) was administered remotely before formal WeChat-based follow-up. The questionnaire included symptoms and adverse events, control of risk factors, basic physical examination and auxiliary examination, and medication status. Participants can answer the above questions by voice, text, or picture. The results of the questionnaire are only for improving the efficiency of information collection, and doctors will further confirm the authenticity of information during follow-up visits on WeChat. On the basis of the above preliminary data, the researcher appointed time to further communicate with the subjects on WeChat and took intervention measures such as adjusting the treatment plan, strengthening the control of risk factors, and improving the lifestyle. During every consultation, the participant's medication adherence and risk factor modification status were evaluated, and the participant was given personalized feedback, encouragement, and suggestions. In our study, risk factor modification included improving cholesterol management, quitting smoking and drinking, monitoring BP, and maintaining a healthy weight. At the end of each visit, the participant received an evaluation report (Multimedia Appendix 2), highlighting areas for improvement. The researchers focused on outcomes where the participant did not perform well at the previous follow-up to trigger a virtuous circle and help them achieve optimal cardiovascular health. The official WeChat account also had other functions, such as regularly sending health education materials, physician-patient communication, and medical appointments. Participants were provided with a variety of teaching materials on coronary heart disease that had been evaluated by cardiologists and that they may read whenever and wherever they wished (Figure 3). Relevant information was updated and sent weekly. If participants had questions, they could always ask the physician in the form of text and pictures via WeChat (Figure 4). Doctors could see participants' questions in the backstage (Figure 5) and answer them on a mobile phone (Figure 6). Participants would be contacted by phone call if their condition changed or they cannot be contacted by WeChat, or the investigator deemed it necessary. The cardiac rehabilitation team received uniform training before first contact with the participant to minimize the heterogeneity of the interventions.

Figure 1. Overview of the WeChat-based telemedicine intervention. HbA_{1c}: glycated hemoglobin A_{1c}; LDL-C: low-density lipoprotein cholesterol.

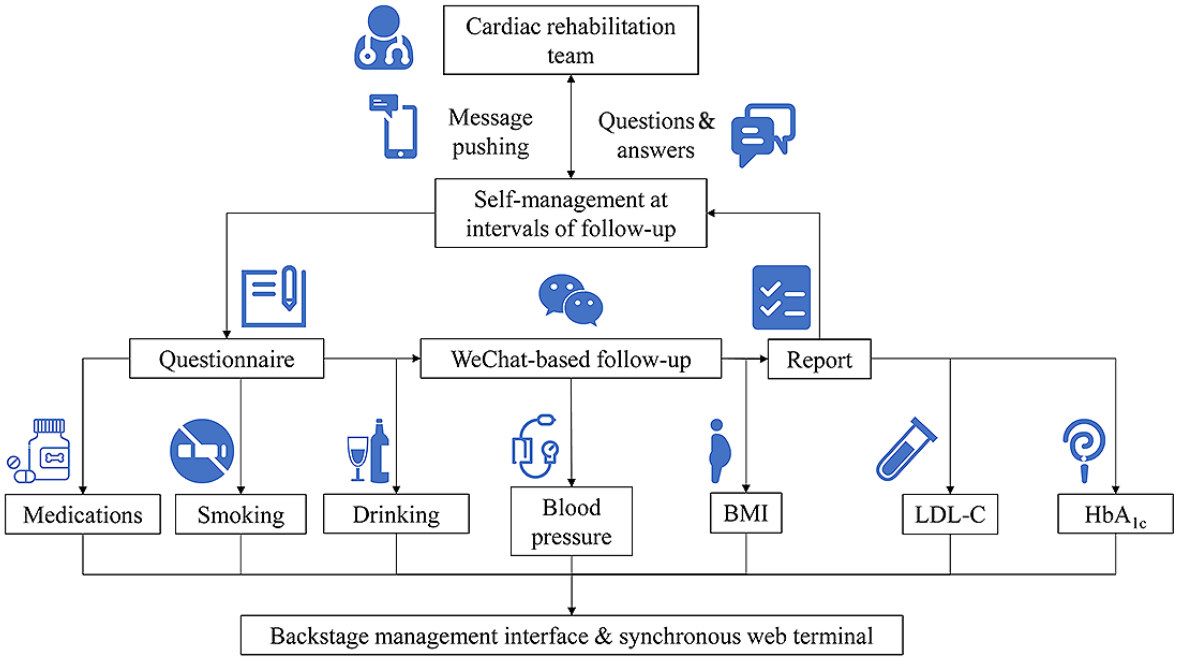


Figure 2. Screenshots of the user registration and binding interface.

Figure 3. Screenshots of educational materials related to coronary heart disease in the patient terminal.

Figure 4. Screenshot of the initiation of a question in the patient terminal.

Figure 5. Backstage management interface of the WeChat-based secondary prevention program.

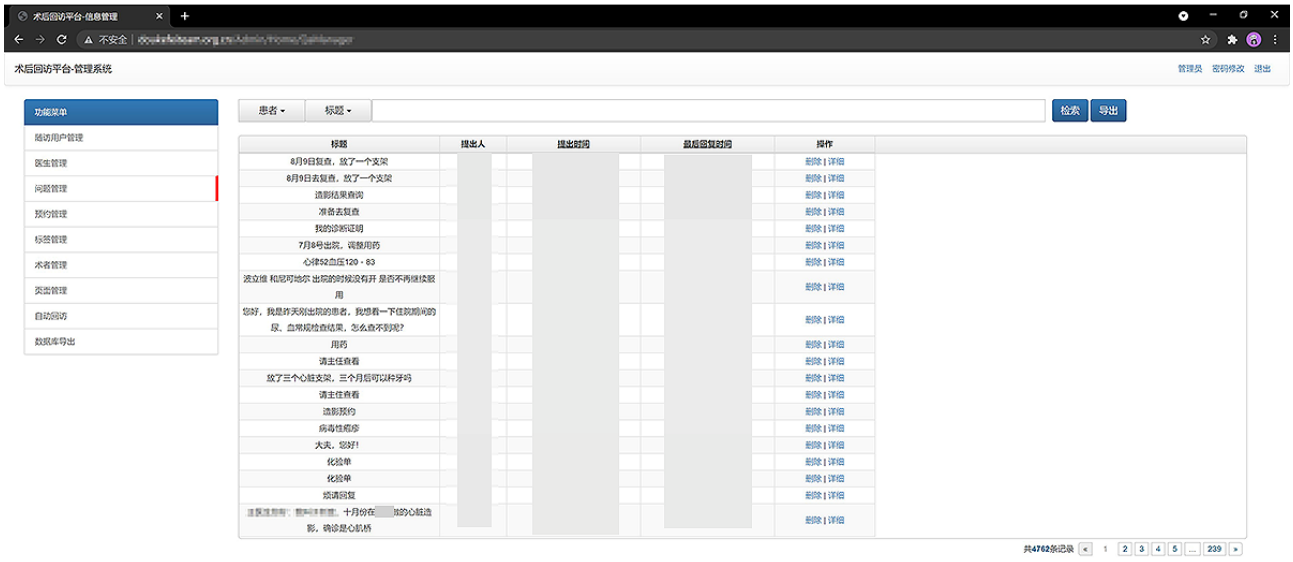


Figure 6. Screenshots of the answering of questions in the physician terminal.

Outcome Measures and Data Collection

Participant characteristics included age, gender, alcohol consumption, and cigarette smoking, clinical data (systolic BP, diastolic BP, BMI, LDL-C, glycated hemoglobin A_{1c} [HbA_{1c}], and ejection fraction), past medical history (hypertension, dyslipidemia, CAD family history, previous PCI, previous coronary artery bypass grafting, previous MI, previous ischemic stroke, diabetes mellitus [DM], peripheral vascular disease, chronic kidney disease, and heart failure), and medications (antiplatelets, β-blockers, statins, and angiotensin-converting-enzyme inhibitor or angiotensin-receptor blocker [ACEI/ARB]).

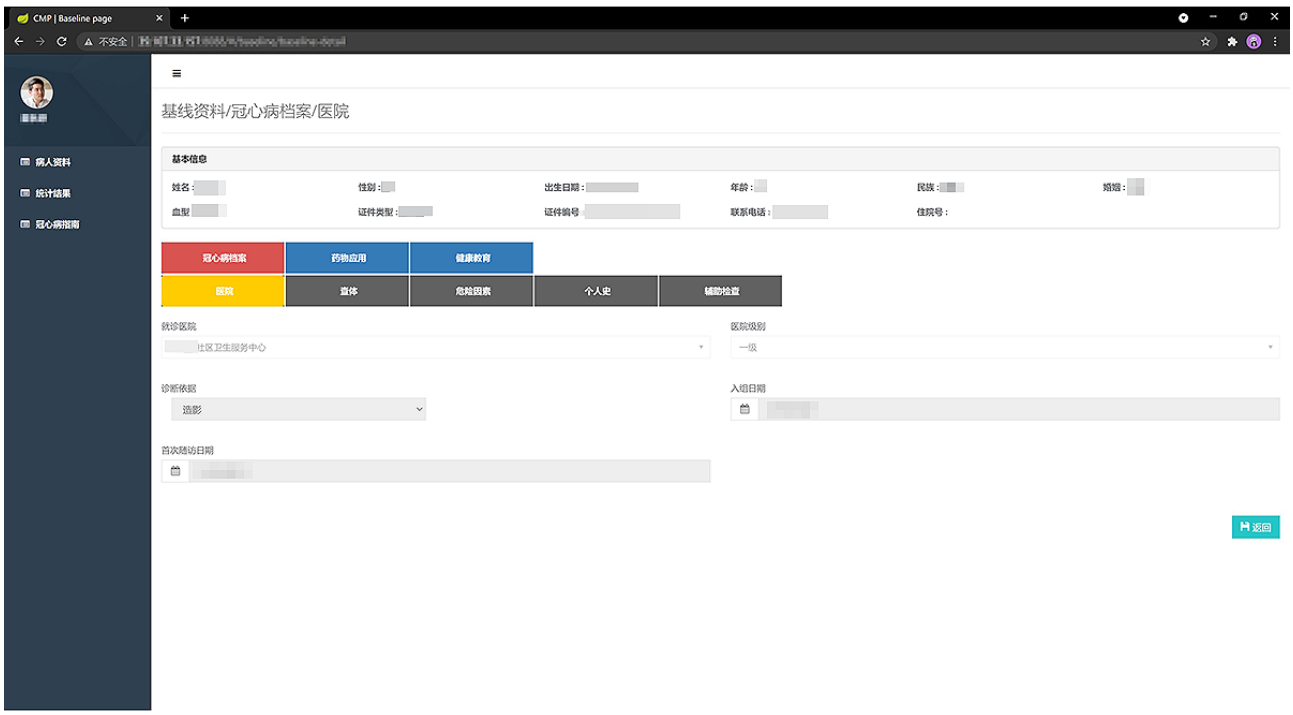
The main endpoint was medication adherence, which was defined as the participants compliance in taking all 4 cardioprotective medications that would improve their outcome (therapies included antiplatelet therapy, β-blockers, statins, and ACEI/ARB). Participants were considered to be taking all 4 of the cardiovascular protective medications if they were taking all 4 medications at the time of the follow-up and had no more than 10% of the days without medication. If the participant can provide details of the prescription, the investigator will calculate the medication status based on the prescription. The secondary outcomes included control of hypertension, current smoking, current alcohol consumption, $18.5 \leq BMI < 25.0$ kg/m^2, LDL-C<1.8 mmol/L and HbA_{1c}<7%. BP less than 140/90 mm Hg in individuals was considered to have good control of hypertension for the purposes of this research. These target values are based on guidelines for the diagnosis and treatment of stable CAD [20]. The whole data set included 4 parts: baseline characteristics and the 1-, 3-, 6-, and 1-year follow-up characteristics. All the baseline characteristics were extracted from the participants' medical records.

Researchers performed face-to-face interviews with participants in the control group during the first visit, as well as at the 1-month, 3-month, 6-month, and 1-year follow-up visits. At the follow-up, body height and body weight were measured by the physicians. Two BP readings were taken by using an electronic BP monitor with the participant sitting in a chair with back support after 10 minutes of rest, and the average was considered as the final reading. Behavioral changes in drinking and smoking status and adherence to secondary prevention medications were self-reported by the participants. We evaluated 4 medications that were commonly prescribed to patients with stable CAD; specifically, each participant was asked about their current medications during each follow-up visit.

Follow-up data were collected for the intervention group via our official WeChat account. To minimize the impact of the discrepancy between BP recorded at home and BP measured at the clinic, all members of this group were requested to report their height, weight, and BP using conventional procedures at a nearby clinic. Behavioral changes in drinking and smoking status and adherence to medications were collected using self-reported questionnaires. Blood samples in both groups for LDL-C and HbA_{1c} levels were analyzed in the respective laboratories using standard procedures.

An electronic data capture system (Figure 7) was used to gather and handle all the data. To enter and analyze the data, researchers needed to be given appropriate permissions, and all the researchers were unable to access the database until they underwent data safety training.

Figure 7. Synchronous data-capture system of the WeChat-based secondary prevention program.



Statistical Analysis

Categorical variables were described using frequencies and percentages, and continuous variables using means with SDs or medians with IQRs. The baseline characteristics of participants were compared across groups using the chi-square or Fisher exact test for categorical variables and Student *t* tests (2-tailed) or the Wilcoxon rank-sum test for continuous data. Multivariable generalized estimating equations (GEEs) were used to compare the primary and secondary outcomes between the 2 groups and to calculate the relative risk (RR) at 12 months. Propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) were performed as sensitivity analyses, and propensity scores were calculated using a multivariable logistic regression model. These variables, including gender, age, current smoking, current alcohol consumption, hypertension, dyslipidemia, CAD family history, previous coronary artery bypass grafting, previous PCI, previous MI, ischemic stroke, DM, peripheral vascular disease, chronic kidney disease, heart failure, BP, BMI, LDL-C, HbA_{1c}, antiplatelet medications, β-blocker use, statin use, and ACEI/ARB, were chosen as covariates because the differences in the baseline characteristics reached statistical significance (P<.10) or were associated with the outcome. The PSM process was based on the nearest neighbor matching algorithm without replacement under a 0.02 caliper at a 1:1 ratio, yielding 257 participants in the intervention group and 257 participants in the control group. IPTW was performed using the same propensity score as previously estimated. A standardized mean difference of <0.2 indicated an acceptable balance after matching or weighting. We used this set of tests to account for baseline variables and draw conclusions about the effect of telemedicine intervention on the results at the individual participant level. Furthermore, comparisons of the primary endpoint between the 2 groups were made based on the prespecified baseline characteristics including gender, age, control of hypertension, current smoker, current drinker, BMI, LDL-C, and HbA_{1c} subgroups. The interaction between treatment effects and subgroups was evaluated using the multivariable GEE models. The analysis was performed in the whole population and adjusted for baseline factors including gender, age, control of hypertension, current smoker, current drinker, BMI, LDL-C, and HbA_{1c}. On the basis of previous studies [8,18], it is estimated that the proportion of the control group in this study who were persistent with taking the 4 cardiovascular protective drugs at the 1-year follow-up was approximately 30%, whereas the proportion of the intervention group was estimated to be 40%. We calculated according to a 90% power (2-sided α=.05) and considering a 10% participant loss to follow-up, a total of 1060 participants needed to be enrolled in this study. The ratio between the intervention and control groups was 1:1, and 530 participants were included in the 2 groups.

A 2-tailed P value <.05 was considered statistically significant. All the statistical analyses were performed using STATA 16.0 (Stata Corp) and R 4.0.2 (R Foundation for Statistical Computing). The missing values are filled in by the average of the 10 multiple interpolations. None of the variables had missing values of >5%. Missing values varied from 0.1% (BP) to 3.6% (HbA_{1c}).

Results

Baseline Characteristics

Table 1 summarizes the unadjusted baseline characteristics.

Table 1. Baseline characteristics (N=1206).

Variables	Total	Intervention (n=642)	Control (n=564)	*P* value	SMD[a]
Demographics					
Male, n (%)	875 (72.6)	488 (76)	387 (68.6)	.005	0.166
Age (years), mean (SD)	64.83 (10.59)	61.27 (10.20)	68.89 (9.52)	<.001	0.772
Age ≥65 years, n (%)	637 (52.8)	253 (39.4)	384 (68.1)	<.001	0.601
Current smoker, n (%)	308 (25.5)	171 (26.6)	137 (24.3)	.39	0.054
Current drinker, n (%)	192 (15.9)	93 (14.5)	99 (17.6)	.17	0.084
Past medical history, n (%)					
Hypertension	856 (71)	409 (63.7)	447 (79.3)	<.001	0.35
Dyslipidemia	907 (75.2)	564 (87.9)	343 (60.8)	<.001	0.651
CAD[b] family history	120 (10)	60 (9.3)	60 (10.6)	.52	0.043
Previous CABG[c]	59 (4.9)	20 (3.1)	39 (6.9)	.004	0.175
Previous PCI[d]	304 (25.2)	205 (31.9)	99 (17.6)	<.001	0.338
Previous myocardial infarction	172 (14.3)	99 (15.4)	73 (12.9)	.25	0.071
Previous ischemic stroke	85 (7)	53 (8.3)	32 (5.7)	.10	0.102
Diabetes mellitus	464 (38.5)	205 (31.9)	259 (45.9)	<.001	0.29
Peripheral vascular disease	40 (3.3)	26 (4)	14 (2.5)	.18	0.088
Chronic kidney disease	11 (0.9)	2 (0.3)	9 (1.6)	.04	0.132
Heart failure	34 (2.8)	30 (4.7)	4 (0.7)	<.001	0.247
Clinical data					
Good control of hypertension, n (%)	860 (71.3)	369 (57.5)	491 (87.1)	<.001	0.7
Systolic BP[e] (mm Hg), mean (SD)	131.65 (16.33)	135.10 (18.67)	127.73 (12.04)	<.001	0.47
Diastolic BP (mm Hg), mean (SD)	77.06 (10.38)	78.06 (11.92)	75.93 (8.16)	<.001	0.209
BMI (kg/m^2), mean (SD)	25.61 (3.19)	25.74 (3.21)	25.45 (3.17)	.12	0.091
18.5≤BMI<25.0 kg/m^2, n (%)	516 (42.8)	276 (43)	240 (42.6)	.92	0.009
LDL-C[f] (mmol/L), mean (SD)	2.32 (0.80)	2.34 (0.84)	2.29 (0.75)	.34	0.055
LDL-C<1.8 mmol/L, n (%)	328 (27.2)	172 (26.8)	156 (27.7)	.79	0.02
HbA_{1c}[g] (%), mean (SD)	6.50 (1.18)	6.55 (1.23)	6.44 (1.11)	.11	0.094
HbA_{1c}<7%, n (%)	903 (74.9)	475 (74)	428 (75.9)	.49	0.044
Medications, n (%)					
Medications adherence	374 (31)	211 (32.9)	163 (28.9)	.16	0.086
Antiplatelet	1182 (98)	632 (98.4)	550 (97.5)	.35	0.066
β-blocker	867 (71.9)	514 (80.1)	353 (62.6)	<.001	0.394
Statin	1144 (94.9)	633 (98.6)	511 (90.6)	<.001	0.359
ACEI/ARB[h]	565 (46.8)	277 (43.1)	288 (51.1)	.007	0.159

[a]SMD: standardized mean difference.

[b]CAD: coronary artery disease.

[c]CABG: coronary artery bypass grafting.

[d]PCI: percutaneous coronary intervention.

[e]BP: blood pressure.

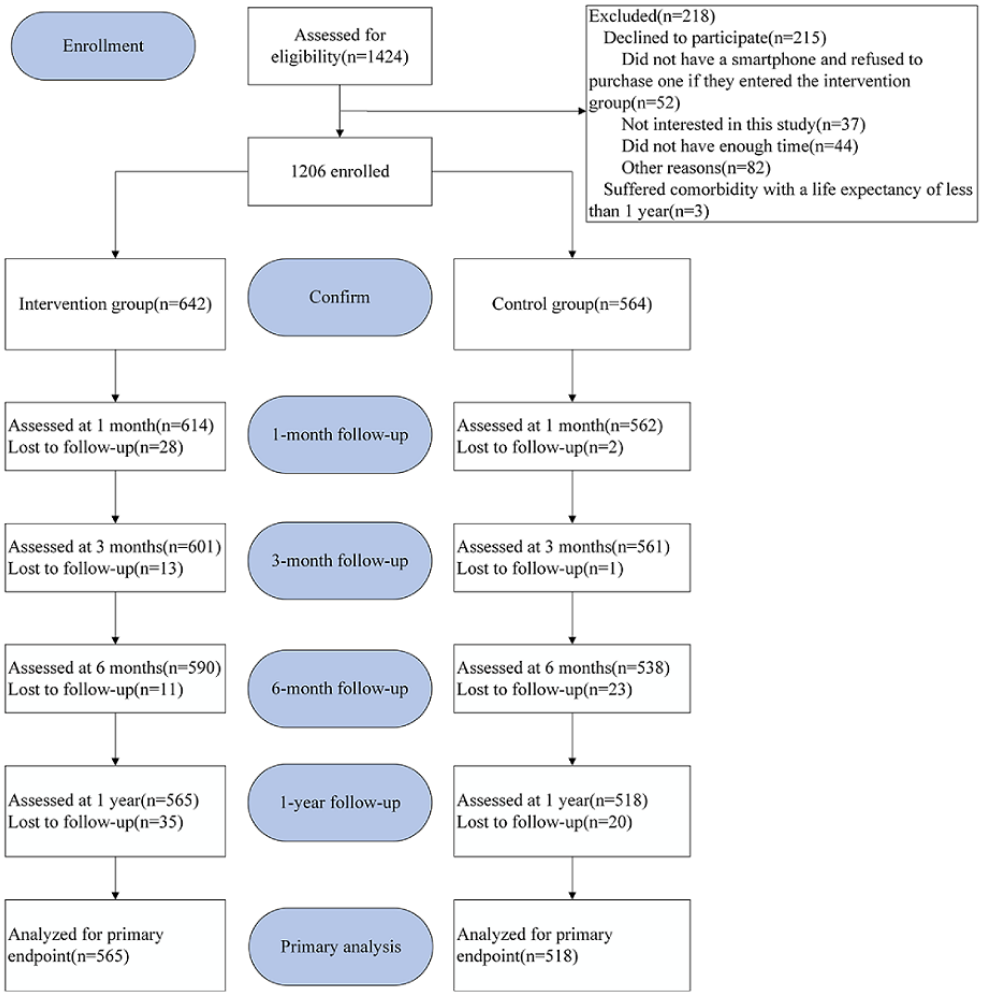
[f]LDL-C: low-density lipoprotein cholesterol.

[g]HbA_{1c}: glycated hemoglobin A_{1c}.

[h]ACEI/ARB: angiotensin-converting-enzyme inhibitor or angiotensin-receptor blocker.

In this study, 1424 participants were identified between September 2018 and September 2019. After screening the participants based on the exclusion criteria, 1206 participants were analyzed in this study. At 1 year, 88% (565/642) of participants in the intervention group and 91.8% (518/564) of participants in the control group had successful follow-up data (Figure 8). The loss to follow-up rate was lower in the control group (77/642, 12%, vs 46/564, 8.2%; *P*=.03), and 84.1% (475/565) of participants in the intervention group were followed up via the WeChat platform. In summary, participants in the intervention group were more likely to be male (488/642, 76%, vs 387/564, 68.6%; *P*<.001) and younger (61.27 vs 68.89; *P*<.001). The intervention group showed a reduced prevalence of comorbidities such as hypertension (409/642, 63.7%, vs 447/564, 79.3%; *P*<.001), DM (205/642, 31.9%, vs 259/564, 45.9%; *P*<.001), and chronic kidney disease (2/642, 0.3%, vs 9/564, 1.6%; *P*<.001) when compared with the control group. Regarding clinical data, the intervention group had worse BP control (369/642, 57.5%, vs 491/564, 87.1%; *P*<.001) than the control group; however, heart failure was more common in the intervention group (30/642, 4.7%, vs 4/564, 0.7%; *P*<.001) as was dyslipidemia (564/642, 87.9%, vs 343/564, 60.8%; *P*<.001). Regarding medication adherence with the 4 cardioprotective drugs, participants in the intervention group more frequently received β-blockers (514/642, 80.1%, vs 163/564, 62.6%; *P*<.001) and statins (633/642, 98.6%, vs 163/564, 90.6%; *P*<.001) and less frequently received ACEI/ARB (277/642, 43.1%, vs 288/564, 51.1%; *P*=.007). There were no statistically significant differences between the 2 groups with regard to current smoker, current drinker, previous PCI, previous MI, previous ischemic stroke, medication adherence, BMI, LDL-C, or HbA_{1c}. Overall, the on-target proportions of BP, BMI, LDL-C, and HbA_{1c} were 71.31% (860/1206), 42.79% (516/1206), 27.2% (328/1206), and 74.88% (903/1206), respectively, and 54.7% (254/464) of patients with known diabetes had HbA_{1c}≥7%. Regarding unhealthy lifestyles, the proportions of smokers and drinkers were 25.54% (308/1206) and 15.92% (192/1206), respectively. The prevalence of the 4 cardiovascular drugs at the beginning was 31.01% (374/1206, 95% CI 28.4%-33.6%). Among them, the proportion of antiplatelet drugs (98.01%, 1182/1206) and statins (94.86%, 1144/1206) was higher, whereas the proportion of β-blockers (71.89%, 867/1206) and ACEI/ARBs (46.85%, 565/1206) was lower. Among participants treated with statins, 73.6% (842/1144) did not achieve the goal LDL-C level of 1.8 mmol/L. The reasons for participants' loss of follow-up included not being able to keep in touch (82.1%, 101/123) and participants requesting withdrawal from the study (17.9%, 22/123). Clinical demographics of follow-up and lost to follow-up participants are shown in Multimedia Appendix 3. Compared with participants with regular follow-up, participants who were lost to follow-up had a higher proportion of hypertension (757/1083, 69.9%, vs 99/123, 80.5%; *P*=.02), diabetes (403/1083, 37.21%, vs 61/123, 49.6%; *P*=.01), a lower proportion of dyslipidemia (825/1083, 76.18%, vs 82/123, 66.7%; *P*=.03), and better medication adherence (324/1083, 29.92%, vs 50/123, 40.7%; *P*=.02).

Figure 8. Participant flow diagram.



Sensitivity Analyses Using PSM and IPTW

We matched 257 pairs of participants between the intervention group and the control group using PSM. To avoid decreasing the sample size and weakening the statistical power, we also performed IPTW using the same covariates in the PSM. After matching and weighing, almost all covariates were well-balanced, except for age (Multimedia Appendix 4). Detailed baseline characteristics and standard mean differences after PSM and IPTW are depicted in Multimedia Appendix 5.

Primary and Secondary Outcome Analyses

Table 2 presents an overview of the primary and secondary outcomes at the 1-year follow-up (comparison within groups).

Table 2. Primary and secondary outcomes at the 1-year follow-up (comparison within groups).

Outcomes	Intervention			Control		
	Baseline, n (%)	1 year, n (%)	*P* value	Baseline, n (%)	1 year, n (%)	*P* value
Primary outcome						
Medication adherence	211 (32.9)	172 (30.4)	.38	163 (28.9)	142 (27.4)	.63
Secondary outcomes						
Antiplatelet	632 (98.4)	540 (95.6)	.005	550 (97.5)	494 (95.4)	.08
β-blocker	514 (80.1)	441 (78.1)	.42	353 (62.6)	329 (63.5)	.80
Statin	633 (98.6)	532 (94.2)	<.001	511 (90.6)	479 (92.5)	.32
ACEI/ARB[a]	277 (43.1)	227 (40.2)	.31	288 (51.1)	258 (49.8)	.73
Current smoker	171 (26.6)	44 (7.8)	<.001	137 (24.3)	118 (22.8)	.61
Current drinker	93 (14.5)	33 (5.8)	<.001	99 (17.6)	91 (17.6)	.99
Good control of hypertension	369 (57.5)	416 (73.6)	<.001	491 (87.1)	486 (93.8)	<.001
18.5≤BMI<25.0 kg/m^2	276 (43)	237 (41.9)	.74	240 (42.6)	226 (43.6)	.77
LDL-C[b]<1.8 mmol/L	172 (26.8)	198 (35)	.002	156 (27.7)	280 (54.1)	<.001
HbA_{1c}[c]<7%	475 (74)	439 (77.7)	.16	428 (75.9)	484 (93.4)	<.001

[a]ACEI/ARB: angiotensin-converting-enzyme inhibitor or angiotensin-receptor blocker.

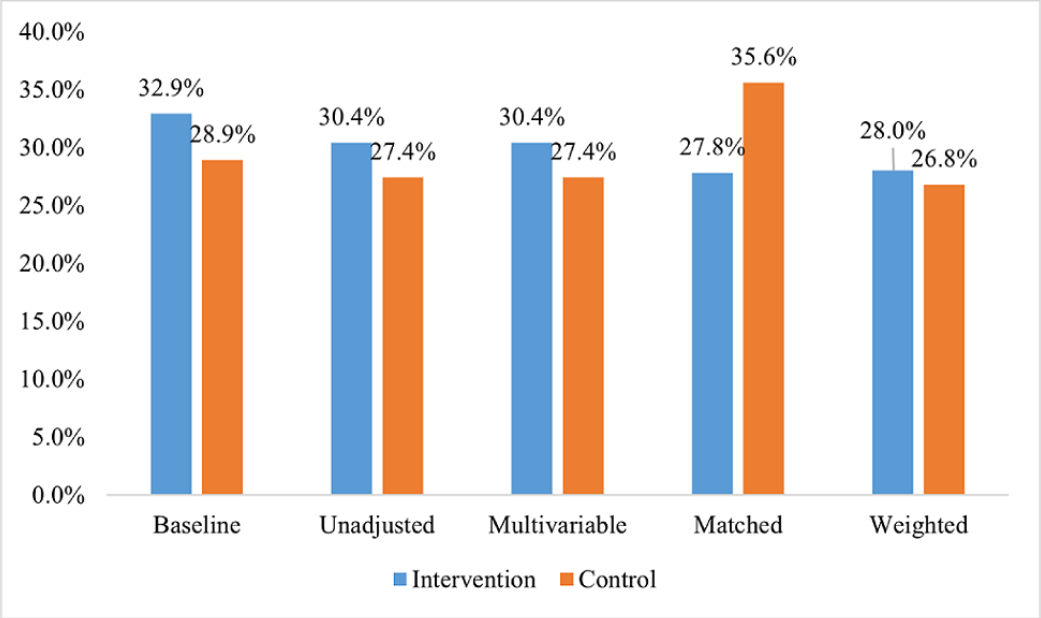
[b]LDL-C: low-density lipoprotein cholesterol.

[c]HbA_{1c}: glycated hemoglobin A_{1c}.

Compared with the previous year, there was no significant difference in the drug adherence with the 4 cardioprotective medications in either the intervention or the control group (172/565, 30.4%, vs 211/642, 32.9%, *P*=.38; 142/518, 27.4%, vs 163/564, 28.9%, *P*=.63). Compared with the previous year, an increased prevalence of good hypertension management was observed among the intervention group (416/565, 73.6%, vs 369/642, 57.5%; *P*<.001), an LDL-C on target (198/565, 35%, vs 172/642, 26.8%; *P*<.001) and a reduction in the proportion of current smokers (44/565, 7.8%, vs 171/642, 26.6%; *P*<.001) and drinkers (33/565, 5.8%, vs 93/642, 14.5%; *P*<.001). After the 1-year follow-up, the proportion of medication adherence to antiplatelet treatment (540/565, 95.6%, vs 632/642, 98.4%; *P*=.005) and statins (532/565, 94.2%, vs 633/642, 98.6%; *P*<.001) decreased. In the control group, participants achieved a better BP level (486/518, 93.8%, vs 491/564, 87.1%; *P*<.001), improved lipid levels (280/518, 54.1%, vs 156/564, 27.7%; *P*<.001) and improved control of blood glucose (484/518, 93.4%, vs 428/564, 75.9%; *P*<.001) at the 1-year follow-up.

Multimedia Appendix 6 presents 1-year primary and secondary outcomes (intervention vs control). Figure 9 depicts proportions of medical adherence to the 4 cardioprotective drugs in the intervention group and control group by different statistical methods. Compared with the routine follow-up in community hospitals, there was no obvious advantage in the medication adherence with the 4 cardioprotective drugs in the intervention group (172/565, 30.4%, vs 142/518, 27.4%; RR 0.99, 95% CI 0.97-1.02; *P*=.65). The mean difference of medications adherence between the intervention and control groups is 3% (95% CI 0.2%-11.5%). The intervention measures improved the smoking cessation (44/565, 7.8%, vs 118/518, 22.8%; RR 0.48, 95% CI 0.44-0.53; *P*<.001), alcohol restriction (33/565, 5.8%, vs 91/518, 17.6%; RR 0.47, 95% CI 0.42-0.54; *P*<.001). The control group was superior to the intervention group in medication adherence in regard to ACEI/ARBs (227/565, 40.2%, vs 258/518, 49.8%; RR 0.98, 95% CI 0.96-0.99; *P*<.001), BMI (237/565, 41.9%, vs 226/518, 43.6%; RR 0.95, 95% CI 0.93-0.97; *P*<.001), LDL-C (198/565, 35%, vs 280/518, 54.1%; RR 0.79, 95% CI 0.73-0.84; *P*<.001), and blood glucose (439/565, 77.7%, vs 484/518, 93.4%; RR 0.95, 95% CI 0.94-0.97; *P*<.001) targets. All of these results were still significant after the multivariable analysis using GEE, PSM, and IPTW.

Figure 9. Proportions of medical adherence to 4 cardioprotective drugs in the intervention group and control group.

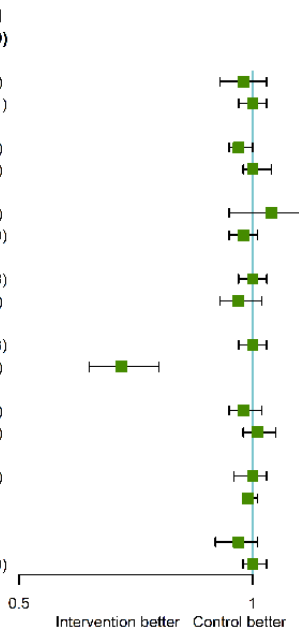


Subgroup Analysis

The intervention and control groups were divided into 16 subgroups according to gender, age, control of hypertension, current smoker, current drinker, BMI, LDL-C, and HbA_{1c} (Figure 10). No significant difference in medication adherence between the 2 groups was consistent across all subgroups, and no significant interaction was observed. A trend of increased medication adherence in the intervention group was observed in the current drinker subgroup (8/33, 24%, vs 34/91, 37%; RR 0.72, 95% CI 0.65-0.80). However, the subgroup analysis did not indicate any significant interactions between medication adherence and stratification variables.

Figure 10. Subgroup analysis of primary outcome. Values are n (%) for categorical variables. The interaction between treatment effect and subgroups was evaluated using multivariable generalized estimating equations models. The analysis was performed in the whole population and adjusted for baseline factors including sex, age, control of hypertension, current smoker, current drinker, BMI, low-density lipoprotein cholesterol, and glycated hemoglobin. HbA_{1c}: glycated hemoglobin A_{1c}; LDL-C: low-density lipoprotein cholesterol. RR: relative risk.

Subgroup	No. of patients	Intervention	Control	RR (95%CI)	*P* values	*P* values for interaction
Overall	**1083**	**211 (32.9)**	**163 (28.9)**	**0.99 (0.97-1.02)**	**0.648**	
Gender						**0.493**
Female	290	36 (26.7)	40 (25.8)	0.98 (0.93-1.03)	0.419	
Male	793	136 (31.6)	102 (28.1)	1.00 (0.97-1.03)	0.942	
Age						**0.097**
<65 yrs	512	104 (29.8)	48 (29.5)	0.97 (0.95-1.00)	0.044	
>=65 yrs	571	68 (31.5)	94 (26.5)	1.00 (0.98-1.04)	0.578	
Good control of hypertension						**0.409**
No	181	54 (36.2)	11 (34.4)	1.04 (0.95-1.15)	0.41	
Yes	902	118 (28.4)	131 (27.0)	0.98 (0.95-1.01)	0.249	
Current smoker						**0.091**
No	921	159 (30.5)	111 (27.8)	1.00 (0.97-1.03)	0.99	
Yes	162	13 (29.6)	31 (26.3)	0.97 (0.93-1.02)	0.263	
Current drinker						**0.261**
No	959	164 (30.8)	108 (25.3)	1.00 (0.97-1.03)	0.962	
Yes	124	8 (24.2)	34 (37.4)	0.72 (0.65-0.80)	<.001	
18.5<=BMI<25.0kg/m2						**0.275**
No	620	110 (35.5)	92 (31.5)	0.98 (0.95-1.02)	0.346	
Yes	463	62 (26.2)	50 (22.1)	1.01 (0.98-1.05)	0.551	
LDL-C<1.8mmol/L						**0.861**
No	605	69 (34.9)	77 (27.5)	1.00 (0.96-1.03)	0.834	
Yes	478	38 (30.2)	7 (20.6)	0.99 (0.98-1.01)	0.452	
HbA_{1c} <7.0mmol/L						**0.129**
No	160	38 (30.2)	7 (20.6)	0.97 (0.92-1.01)	0.157	
Yes	923	134 (30.5)	135 (27.9)	1.00 (0.98-1.03)	0.695	



No study-related adverse events were recorded during this trial.

Discussion

This is a prospective study, which directly compares web-based tertiary A-level hospital, WeChat-based secondary prevention

with a traditional community, hospital-based cardiac rehabilitation program in patients with stable CAD.

Principal Findings

Most patients with CAD had poor management of cardiovascular risk factors, and the percentage of participants who met the prescribed BP, BMI, LDL-C, and HbA_{1c} goals recommended were 71.31% (860/1206), 42.79% (516/1206), 27.2% (328/1206), and 74.88% (903/1206), respectively. Current smoking and alcohol consumption accounted for 25.54% (308/1206) and 15.92% (192/1206) of the total participants, respectively. A total of 31.01% (374/1206) of the participants had good medication adherence for the use of the 4 cardioprotective drugs. Cardioprotective medications included antiplatelet therapy, statins, β-blockers, and ACEIs/ARBs in 98.01% (1182/1206), 94.86% (1144/1206), 71.89% (867/1206), and 46.85% (565/1206) of patients, respectively. The current situation that uses secondary prevention is not ideal, especially regarding the participants' blood lipids and blood glucose control. More than 73.60% (842/1144) of the participants in our study were taking statins to reduce their blood lipids, but their LDL-C levels remained high, suggesting that they needed more rigorous cholesterol treatment. A possible explanation is that the initial dose of the drug was too low, and the dose was not adjusted as soon as the treatment began, or failure to strengthen cholesterol management in a timely manner, such as the addition of ezetimibe and proprotein convertase subtilisin/kexin type 9 inhibitors. Our findings are in line with those of previous studies conducted in Europe, China, the United States, and other areas of the globe. Previous results from the European Action on Secondary and Primary Prevention by Intervention to Reduce Events V [21-24], Dyslipidemia International Study [25], Dyslipidemia International Study-China [26], the Prospective Urban Rural Epidemiology study [12], the prospective observational longitudinal registry of patients with stable CAD (CLARIFY [Prospective Observational Longitudinal Registry of Patients With Stable Coronary Artery Disease]) study [27], Report on Cardiovascular Health and Diseases Burden in China 2020 [28] confirmed that the current situation of secondary prevention in patients with coronary heart disease is concerning.

The most significant result of this study is that a WeChat-based intervention provided by a tertiary A-level hospital had no obvious advantage in improving patient adherence with the 4 cardioprotective medications compared with the traditional method. This finding is in contrast to a previous study [18], which suggested that the WeChat intervention can improve the medication adherence of patients. However, our study did not record the reasons for discontinuation of medication, and it is impossible to determine whether patients stopped medication because of a change in their disease or because of poor patient compliance. The WeChat remote intervention leads to better lifestyle improvements, including abstinence from smoking and alcohol consumption. The possible mechanism for the improved smoking cessation and alcohol use in participants in the intervention group includes several aspects, including how the relevant content is presented to the participants. All the messages sent out by our official WeChat account include smoking cessation and alcohol use content. Second, the participants were asked about smoking and alcohol consumption at each follow-up. Risk factors, including BP, LDL-C, blood glucose, and BMI, were more controlled by the traditional community follow-up, which may be explained by the fact that community hospital doctors were trained before the trial to improve their clinical skills.

Limitations

This study had some limitations. First, this was an observational study. Bias may be introduced by variations in the baseline features of the 2 groups. Second, our study did not record the use of all antihypertensive drugs used in patients, so it was impossible to specifically analyze whether the reason for the better control of hypertension in participants in the control group was due to better compliance with antihypertensive drugs or an improved lifestyle. Third, because of the short follow-up period, it was difficult to assess the long-term consequences. In addition, some factors that might affect patient adherence, such as income and education, were not recorded in our study. Finally, participants were not asked to participate in designing messages and the intervention so that the interventions we offered may not fully meet their needs. Participants may be asked to contribute to the design of interventions in future trials.

Comparison With Prior Work

According to a literature review by Farsi et al [29], multidimensional health care, which includes the integration of health care with social media and other kinds of communication, has been shown to be very effective. A systematic review by Indraratna et al [30] included 26 randomized controlled trials (n=6713). In patients with heart failure, mobile phone technologies were associated with lower hospitalization rates, and in patients with hypertension, mobile phone technologies significantly reduced the systolic BP [30]. A systematic review of 9 randomized controlled studies evaluated by Hamilton et al [31] confirmed that participants had high rates of participation, acceptance, use, and adherence to mobile health (mHealth). In addition, the health care provided by mHealth is just as effective as a traditional central health care and significantly improves the quality of life [31]. A few recent studies have demonstrated the feasibility and effectiveness of WeChat in chronic disease management, such as in hypertension and CAD. In a 30-day follow-up, Ni et al [17] discovered that the experimental group's medication nonadherence score dropped more. Participants in that group were given an mHealth intervention created by combining 2 apps: *WeChat* and *BB Reminder*. The medication nonadherence score, heart rate, systolic BP, and diastolic BP were all included as outcome variables in this study. The remaining 3 outcomes were not examined owing to the short follow-up period (30 days) and small sample size of the study [17]. Dorje et al [18] created the Smartphone and Social Media–Based Cardiac Rehabilitation and Secondary Prevention in China program, which is a smartphone-based cardiac rehabilitation and secondary prevention program provided through the social media platform WeChat. The participants were monitored for a year in this randomized controlled study, which included 312 participants. The Smartphone and Social Media–Based Cardiac Rehabilitation and Secondary Prevention in China group improved substantially more in the 6-minute

walk distance at 2 and 6 months than in the control group [18]. Participants in this group had better secondary outcomes, including knowledge of CAD total score, systolic BP, lipid profile, and cardioprotective drug compliance. However, they found no differences between groups in other secondary outcomes, such as current smoker, BMI, waist-to-hip ratio, psychosocial status, and quality of life. In addition, participants in the control group did not receive formal cardiac rehabilitation and secondary prevention, which may have led to an overestimation of the WeChat effect. BP control in the target range is an important strategy for secondary prevention of CAD. Several reports have shown that WeChat-based interventions are associated with a better control of hypertension. An investigation by Li et al [32] involving 464 patients with hypertension found that after 6 months of using WeChat for self-care, BP control was better in the intervention group. They built separate group chats according to the different risk factors and developed a punch in an innovative system to promote healthy behaviors. In a study by Xiao et al [33], participants reported feeling more willing to use and satisfied when using the WeChat platform for routine BP monitoring. Chang et al [34] examined participants' experiences of physician-patient communication and peer interaction in a social media–based (WeChat) weight management program. The interactive nature of social media mitigates the practice of social support and social comparison and creates new forms of supervision [34]. However, such communication in a public group carries a potential privacy risk. In a study by Chen et al [35], 80 people were randomly allocated to 1 of 2 groups: intervention and control. The intervention group was given the entire Chinese smoking cessation plan, which was based on applicable guidelines. The features included projects that were used in a specific intervention program to help users plan and record good protocols to promote quitting smoking, promote smoking cessation games, provide information on smoking hazards, help users overcome impulse behaviors, evaluate the level of nicotine dependence and standardized lung health tests, and provide a social platform that encourages social support among users. A total of 25% (10/40) of the intervention participants and 5% (2/40) of the control participants (RR 5, 95% CI 1.2-21.4; *P*=.03) had biochemically validated cessation at 6 weeks. It has been suggested that using the WeChat platform for smoking cessation is a novel and acceptable intervention for smoking cessation. Zhang et al [36] found in their study that WeChat self-monitoring tended to increase the medication compliance of patients with ischemic stroke. However, owing to the study's limited sample size, no significant conclusions could be drawn [36]. The study by Li et al [37] confirmed that the videoconference follow-up based on WeChat has better effectiveness, reliability, and higher user satisfaction and trust than the traditional telephone follow-up. The results of this study are consistent with those of our study, which verifies the feasibility of WeChat as a new method of long-term follow-up.

According to this study, WeChat-based telemedicine is particularly effective for lifestyle interventions. Owing to the COVID-19 crisis, it has been inconvenient and even impossible for patients with chronic diseases to receive outpatient follow-up. Remote follow-up can be used as an effective medical treatment.

We identified other problems in this study. First, participants in the intervention group often want consultation for their comorbidities, and our cardiac rehabilitation team may not be able to provide detailed explanations for their comorbidities, resulting in participants needing to go to the hospital. In the future, this problem can be solved by integrating a chronic disease management team to manage all comorbidities of participants. Second, some participants with persistent chest pain and who were suspected of having an MI still submitted a consultation through the WeChat platform rather than calling or going to the emergency center, which may lead to a delay in revascularization. Therefore, patient education may need to be enhanced when using these platforms, and it is important to inform the patient to be transported to the emergency department for additional care in case of life-threatening situations.

Conclusions

Despite the prevalent use of cardioprotective medications, many patients with CAD fail to achieve ideal control of cardiovascular risk factors, as recommended by the guidelines. After initial treatment, the patient's target should be monitored. The treatment regimen should be adjusted in time, and lifestyle interventions should be strengthened to try to control the risk factors and reach the target as soon as possible. Tertiary A-level hospital, WeChat-based intervention did not improve adherence to the 4 cardioprotective medications compared with the traditional method. Traditional community hospital follow-up was superior to WeChat remote follow-up in risk factor control, including BP, LDL-C, blood glucose, and BMI. The tertiary A-level hospital, WeChat-based intervention has a positive effect on improving lifestyle, such as quitting drinking and smoking, in patients with stable CAD and can be tried as a supplement to community hospital follow-up. Additional research on social media interventions aimed specifically at improving the lifestyle of patients with CAD is necessary.

Acknowledgments

This study was one of the Prevention and Control Projects of the Major Chronic Noninfectious Disease (grant 2018YFC1315600), which was supported by the Ministry of Science and Technology of China. They thank the community hospital doctors for their contributions to this study.

Authors' Contributions

All authors contributed to the study design, data collection, data analysis, and manuscript writing. KD and WS contributed equally to the study and they are cocorresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire that was sent remotely before the formal WeChat-based follow-up.
[PNG File , 350 KB - mhealth_v9i10e32548_app1.png]

Multimedia Appendix 2
An example of a report that was sent remotely after the formal WeChat-based follow-up.
[PNG File , 609 KB - mhealth_v9i10e32548_app2.png]

Multimedia Appendix 3
Clinical demographics of follow-up and lost to follow-up participants.
[XLSX File (Microsoft Excel File), 13 KB - mhealth_v9i10e32548_app3.xlsx]

Multimedia Appendix 4
Standardized mean difference in the unadjusted, matched, and weighted cohorts.
[PNG File , 28 KB - mhealth_v9i10e32548_app4.png]

Multimedia Appendix 5
Baseline characteristics of the propensity score-matched and weighted cohorts.
[XLSX File (Microsoft Excel File), 15 KB - mhealth_v9i10e32548_app5.xlsx]

Multimedia Appendix 6
Primary and secondary outcomes at the 1-year follow-up (intervention vs control).
[DOCX File , 18 KB - mhealth_v9i10e32548_app6.docx]

Multimedia Appendix 7
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1166 KB - mhealth_v9i10e32548_app7.pdf]

References

1. Zhou M, Wang H, Zeng X, Yin P, Zhu J, Chen W, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 Sep 28;394(10204):1145-1158 [FREE Full text] [doi: 10.1016/S0140-6736(19)30427-1] [Medline: 31248666]
2. Task Force on Chinese Guidelines for the Prevention of Cardiovascular Diseases(2017), Editorial Board of Chinese Journal of Cardiology. [Chinese guidelines for the prevention of cardiovascular diseases (2017)]. Zhonghua Xin Xue Guan Bing Za Zhi 2018 Jan 24;46(1):10-25. [doi: 10.3760/cma.j.issn.0253-3758.2018.01.004] [Medline: 29374933]
3. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. J Am Coll Cardiol 2019 Sep 10;74(10):177-232 [FREE Full text] [doi: 10.1016/j.jacc.2019.03.010] [Medline: 30894318]
4. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention and Rehabilitation (EACPR). Eur Heart J 2016 Aug 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
5. Yuhui Z, Peipei C, Quan W. Study on accounting and analysis of curative expenditure on cardio-cerebrovascular diseases in China. Chin Circ J 2020;35(9):859-865 [FREE Full text]
6. Li J, Li X, Wang Q, Hu S, Wang Y, Masoudi FA, China PEACE Collaborative Group. ST-segment elevation myocardial infarction in China from 2001 to 2011 (the China PEACE-Retrospective Acute Myocardial Infarction Study): a retrospective analysis of hospital data. Lancet 2015 Jan 31;385(9966):441-451 [FREE Full text] [doi: 10.1016/S0140-6736(14)60921-1] [Medline: 24969506]
7. Working Group on Coronary Artery Disease, National Center for Cardiovascular Quality Improvement (NCCQI). Clinical performance and quality measures for adults with acute ST-elevation myocardial infarction in China. China National Center for Cardiovascular Diseases 2020 Apr;35(4):313-325. [doi: 10.3969/j.issn.1000-3614.2020.04.001]

8. Shang P, Liu GG, Zheng X, Ho PM, Hu S, Li J, et al. Association between medication adherence and 1-year major cardiovascular adverse events after acute myocardial infarction in China. J Am Heart Assoc 2019 May 07;8(9):e011793 [FREE Full text] [doi: 10.1161/JAHA.118.011793] [Medline: 31057004]
9. Ni Z, Dardas L, Wu B, Shaw R. Cardioprotective medication adherence among patients with coronary heart disease in China: a systematic review. Heart Asia 2019 Jun 24;11(2):e011173 [FREE Full text] [doi: 10.1136/heartasia-2018-011173] [Medline: 31297162]
10. Kones R, Rumana U, Morales-Salinas A. Confronting the most challenging risk factor: non-adherence. Lancet 2019 Jan 12;393(10167):105-106. [doi: 10.1016/S0140-6736(18)33079-4] [Medline: 30522920]
11. Zhang M, Deng Q, Wang L, Huang Z, Zhou M, Li Y, et al. Prevalence of dyslipidemia and achievement of low-density lipoprotein cholesterol targets in Chinese adults: a nationally representative survey of 163,641 adults. Int J Cardiol 2018 Jun 01;260:196-203. [doi: 10.1016/j.ijcard.2017.12.069] [Medline: 29622441]
12. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. Lancet 2020 Mar 07;395(10226):795-808 [FREE Full text] [doi: 10.1016/S0140-6736(19)32008-2] [Medline: 31492503]
13. Greenland P, Fuster V. Cardiovascular risk factor control for all. J Am Med Assoc 2017 Jul 11;318(2):130-131. [doi: 10.1001/jama.2017.7648] [Medline: 28697239]
14. Anderson L, Brown JP, Clark AM, Dalal H, Rossau HK, Bridges C, et al. Patient education in the management of coronary heart disease. Cochrane Database Syst Rev 2017 Jun 28;6(6):CD008895 [FREE Full text] [doi: 10.1002/14651858.CD008895.pub3] [Medline: 28658719]
15. Science on WeChat. Nat Methods 2020 Sep;17(9):863. [doi: 10.1038/s41592-020-0954-1] [Medline: 32873983]
16. Chen X, Zhou X, Li H, Li J, Jiang H. The value of WeChat application in chronic diseases management in China. Comput Methods Programs Biomed 2020 Nov;196:105710 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105710] [Medline: 32858284]
17. Ni Z, Liu C, Wu B, Yang Q, Douglas C, Shaw RJ. An mHealth intervention to improve medication adherence among patients with coronary heart disease in China: development of an intervention. Int J Nurs Sci 2018 Sep 8;5(4):322-330 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.003] [Medline: 31406843]
18. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019 Nov;1(7):363-374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
19. Task Force Members, Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, ESC Committee for Practice Guidelines, Document Reviewers, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. Eur Heart J 2013 Oct;34(38):2949-3003. [doi: 10.1093/eurheartj/eht296] [Medline: 23996286]
20. Section of Interventional Cardiology of Chinese Society of Cardiology, Section of Atherosclerosis and Coronary Artery Disease of Chinese Society of Cardiology, Specialty Committee on Prevention and Treatment of Thrombosis of Chinese College of Cardiovascular Physicians. [Guideline on the diagnosis and treatment of stable coronary artery disease]. Zhonghua Xin Xue Guan Bing Za Zhi 2018 Sep 24;46(9):680-694. [doi: 10.3760/cma.j.issn.0253-3758.2018.09.004] [Medline: 30293374]
21. Kotseva K, De Backer G, De Bacquer D, Rydén L, Hoes A, Grobbee D, EUROASPIRE Investigators*. Lifestyle and impact on cardiovascular risk factor control in coronary patients across 27 countries: results from the European Society of Cardiology ESC-EORP EUROASPIRE V registry. Eur J Prev Cardiol 2019 May;26(8):824-835. [doi: 10.1177/2047487318825350] [Medline: 30739508]
22. Jennings CS, Kotseva K, Bassett P, Adamska A, Wood D, ASPIRE-3-PREVENT Investigators. ASPIRE-3-PREVENT: a cross-sectional survey of preventive care after a coronary event across the UK. Open Heart 2020 Apr;7(1):e001196 [FREE Full text] [doi: 10.1136/openhrt-2019-001196] [Medline: 32354740]
23. Kotseva K, De Backer G, De Bacquer D, Rydén L, Hoes A, Grobbee D, et al. Primary prevention efforts are poorly developed in people at high cardiovascular risk: a report from the European Society of Cardiology EURObservational Research Programme EUROASPIRE V survey in 16 European countries. Eur J Prev Cardiol 2020 Mar 20:2047487320908698 [FREE Full text] [doi: 10.1177/2047487320908698] [Medline: 32195597]
24. Blom DJ, Santos RD, Daclin V, Mercier F, Ruiz AJ, Danchin N, ICLPS study group. The challenge of multiple cardiovascular risk factor control outside Western Europe: findings from the International ChoLesterol management Practice Study. Eur J Prev Cardiol 2020 Sep;27(13):1403-1411 [FREE Full text] [doi: 10.1177/2047487319871735] [Medline: 31533447]
25. Gitt AK, Drexel H, Feely J, Ferrières J, Gonzalez-Juanatey JR, Thomsen KK, DYSIS Investigators. Persistent lipid abnormalities in statin-treated patients and predictors of LDL-cholesterol goal achievement in clinical practice in Europe and Canada. Eur J Prev Cardiol 2012 Apr;19(2):221-230. [doi: 10.1177/1741826711400545] [Medline: 21450578]
26. Wang F, Ye P, Hu D, Min Y, Zhao S, Wang Y, DYSIS-China Study Investigators. Lipid-lowering therapy and lipid goal attainment in patients with metabolic syndrome in China: subgroup analysis of the Dyslipidemia International Study-China (DYSIS-China). Atherosclerosis 2014 Nov;237(1):99-105. [doi: 10.1016/j.atherosclerosis.2014.08.023] [Medline: 25238215]

27. Ferrari R, Ford I, Greenlaw N, Tardif J, Tendera M, Abergel H, CLARIFY Registry Investigators. Geographical variations in the prevalence and management of cardiovascular risk factors in outpatients with CAD: data from the contemporary CLARIFY registry. Eur J Prev Cardiol 2015 Aug;22(8):1056-1065. [doi: 10.1177/2047487314547652] [Medline: 25147344]
28. The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on cardiovascular health and diseases burden in China: an updated summary of 2020. Chin Circ J 2021 Jun;36(6):521-545 [FREE Full text]
29. Farsi D. Social media and health care, Part I: literature review of social media use by health care providers. J Med Internet Res 2021 Apr 05;23(4):e23205 [FREE Full text] [doi: 10.2196/23205] [Medline: 33664014]
30. Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e16695 [FREE Full text] [doi: 10.2196/16695] [Medline: 32628615]
31. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018 Feb 07;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]
32. Li X, Li T, Chen J, Xie X, An X, Lv Y, et al. A wechat-based self-management intervention for community middle-aged and elderly adults with hypertension in Guangzhou, China: a cluster-randomized controlled trial. Int J Environ Res Public Health 2019 Oct 23;16(21):4058 [FREE Full text] [doi: 10.3390/ijerph16214058] [Medline: 31652688]
33. Xiao M, Lei X, Zhang F, Sun Z, Harris VC, Tang X, et al. Home blood pressure monitoring by a mobile-based model in Chongqing, China: a feasibility study. Int J Environ Res Public Health 2019 Sep 10;16(18):3325 [FREE Full text] [doi: 10.3390/ijerph16183325] [Medline: 31509950]
34. Chang L, Chattopadhyay K, Li J, Xu M, Li L. Interplay of support, comparison, and surveillance in social media weight management interventions: qualitative study. JMIR Mhealth Uhealth 2021 Mar 01;9(3):e19239 [FREE Full text] [doi: 10.2196/19239] [Medline: 33646130]
35. Chen J, Ho E, Jiang Y, Whittaker R, Yang T, Bullen C. Mobile social network-based smoking cessation intervention for Chinese male smokers: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e17522 [FREE Full text] [doi: 10.2196/17522] [Medline: 33095184]
36. Zhang Y, Fan D, Ji H, Qiao S, Li X. Treatment adherence and secondary prevention of ischemic stroke among discharged patients using mobile phone- and WeChat-based improvement services: cohort study. JMIR Mhealth Uhealth 2020 Apr 15;8(4):e16496 [FREE Full text] [doi: 10.2196/16496] [Medline: 32293574]
37. Li L, Huang J, Wu J, Jiang C, Chen S, Xie G, et al. A mobile health app for the collection of functional outcomes after inpatient stroke rehabilitation: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 May 13;8(5):e17219 [FREE Full text] [doi: 10.2196/17219] [Medline: 32401221]

Abbreviations

ACEI/ARB: angiotensin-converting-enzyme inhibitor or angiotensin-receptor blocker
BP: blood pressure
CAD: coronary artery disease
CLARIFY: Prospective Observational Longitudinal Registry of Patients With Stable Coronary Artery Disease
DM: diabetes mellitus
GEE: generalized estimating equation
HbA_{1c}: glycated hemoglobin A_{1c}
IPTW: inverse probability of treatment weighting
LDL-C: low-density lipoprotein cholesterol
MI: myocardial infarction
mHealth: mobile health
PCI: percutaneous coronary intervention
PSM: propensity score matching
RR: relative risk

Edited by G Eysenbach; submitted 01.08.21; peer-reviewed by JJ Mira, L Li; comments to author 26.08.21; revised version received 07.09.21; accepted 26.09.21; published 27.10.21.

Please cite as:
Shi B, Liu X, Dong Q, Yang Y, Cai Z, Wang H, Yin D, Wang H, Dou K, Song W
The Effect of a WeChat-Based Tertiary A-Level Hospital Intervention on Medication Adherence and Risk Factor Control in Patients With Stable Coronary Artery Disease: Multicenter Prospective Study
JMIR Mhealth Uhealth 2021;9(10):e32548
URL: https://mhealth.jmir.org/2021/10/e32548
doi:10.2196/32548
PMID:34569467

Original Paper

Complete and Resilient Documentation for Operational Medical Environments Leveraging Mobile Hands-free Technology in a Systems Approach: Experimental Study

MinJae Woo[1], PhD; Prabodh Mishra[2], MS; Ju Lin[2], MS; Snigdhaswin Kar[2], MS; Nicholas Deas[3], BS; Caleb Linduff[2], MS; Sufeng Niu[4], PhD; Yuzhe Yang[5], MS; Jerome McClendon[6], PhD; D Hudson Smith[7], PhD; Stephen L Shelton[8], MD; Christopher E Gainey[8], MD; William C Gerard[8], MD; Melissa C Smith[2], PhD; Sarah F Griffin[9], MPH, PhD; Ronald W Gimbel[9], PhD; Kuang-Ching Wang[9], PhD

[1]School of Data Science and Analytics, Kennesaw State University, Kennesaw, GA, United States

[2]Department of Electrical and Computing Engineering, Clemson University, Clemson, SC, United States

[3]School of Computing, Clemson University, Clemson, SC, United States

[4]Linkedin Inc, Mountain View, CA, United States

[5]NetApp, Sunnyvale, CA, United States

[6]Department of Automotive Engineering, Clemson University, Clemson, SC, United States

[7]Watt Family Innovation Center, Clemson University, Clemson, SC, United States

[8]Department of Emergency Medical Services, Prisma Health Richland Hospital, Columbia, SC, United States

[9]Department of Public Health Sciences, Clemson University, Clemson, SC, United States

Corresponding Author:
Ronald W Gimbel, PhD
Department of Public Health Sciences
Clemson University
501 Edwards Hall
Clemson, SC, 29634
United States
Phone: 1 864 656 1969
Email: rgimbel@clemson.edu

Abstract

Background: Prehospitalization documentation is a challenging task and prone to loss of information, as paramedics operate under disruptive environments requiring their constant attention to the patients.

Objective: The aim of this study is to develop a mobile platform for hands-free prehospitalization documentation to assist first responders in operational medical environments by aggregating all existing solutions for noise resiliency and domain adaptation.

Methods: The platform was built to extract meaningful medical information from the real-time audio streaming at the point of injury and transmit complete documentation to a field hospital prior to patient arrival. To this end, the state-of-the-art automatic speech recognition (ASR) solutions with the following modular improvements were thoroughly explored: noise-resilient ASR, multi-style training, customized lexicon, and speech enhancement. The development of the platform was strictly guided by qualitative research and simulation-based evaluation to address the relevant challenges through progressive improvements at every process step of the end-to-end solution. The primary performance metrics included medical word error rate (WER) in machine-transcribed text output and an F1 score calculated by comparing the autogenerated documentation to manual documentation by physicians.

Results: The total number of 15,139 individual words necessary for completing the documentation were identified from all conversations that occurred during the physician-supervised simulation drills. The baseline model presented a suboptimal performance with a WER of 69.85% and an F1 score of 0.611. The noise-resilient ASR, multi-style training, and customized lexicon improved the overall performance; the finalized platform achieved a medical WER of 33.3% and an F1 score of 0.81 when compared to manual documentation. The speech enhancement degraded performance with medical WER increased from 33.3% to 46.33% and the corresponding F1 score decreased from 0.81 to 0.78. All changes in performance were statistically significant (P<.001).

Conclusions: This study presented a fully functional mobile platform for hands-free prehospitalization documentation in operational medical environments and lessons learned from its implementation.

(JMIR Mhealth Uhealth 2021;9(10):e32301) doi:10.2196/32301

KEYWORDS

emergency medical services; prehospital documentation; speech recognition software; natural language processing; military medicine; documentation; development; challenge; paramedic; disruption; attention; medical information; audio; speech recognition; qualitative; simulation

Introduction

Prehospitalization information provided by the first responders is often essential to subsequent treatment efforts including the accurate assessment of a patient, medical diagnosis, and rationale for treatment decisions in the emergency care settings. A patient record documented in the field promotes a continuum of care, playing a vital clinical role in the subsequent treatment of patients in emergency rooms, trauma centers, or other receiving facilities. Complete and effective documentation of prehospitalization care informs clinicians and staff of presenting vitals and symptoms, the initial assessment of the condition, attempted prehospitalization interventions, and observed response to the interventions [1-3]. Failure to report initial findings and interventions in the field may result in clinical errors such as inadvertent overdose due to duplicate administration of the same medication by paramedic and emergency department physicians [4-6]. However, prehospitalization documentation is a challenging task and prone to loss of information, as paramedics operate under urgent and disruptive environments requiring their constant attention to the patient [6-9].

The US military has demonstrated an ongoing interest in potential technological approaches that enable efficient prehospitalization documentation at the point of injury in advance of a patient's arrival to a field hospital [10-14]. Adequate prehospitalization documentation plays a critical role in ensuring casualties' maximal chance of survival in the operational environments [15-19]. In the past, the United States Army Medical Research and Development Command has successfully deployed a PDA-based mobile platform that enabled efficient data entry to the electronic patient record and transmission of patient information through a wireless network [12,13]. A new challenge has arisen from the PDA-based design interrupting the flow of care when entering electronic health record (EHR) data using keyboards or a stylus. The loss of time for direct patient care is often prohibitive in emergency environments as medical personnel have to continuously conduct hands-on interventions for patients to save their life and limb.

Given the necessity of seeking solutions that will not degrade the clinical workflow, technology solutions using automatic speech recognition (ASR) have been explored for hands-free clinical documentation [20-23]. A mobile platform based on ASR technologies has the potential to enable hands-free documentation by extracting medical information from the incoming audio stream without hand-operated input devices. However, the technologies have not yet proven to be reliable in noise-intensive real-world environments in the context of emergency medicine. In military operations, the environment often involves high levels of noise from factors such as blasts, gunshots, and aircraft. It has been well-documented that the performance of contemporary ASR systems is degraded by heavy background noise, leading to more word errors in speech recognition output [24-29]. Moreover, the noise in ASR audio input may result in specific types of word errors in the output text interfering with the documentation when extracting relevant medical information. The existing publicly and commercially available ASR models are optimized for the daily conversation and thus may perform poorly when applied to domain-specific clinical speech [30,31].

ASR consists of multiple components to convert input audio to output text. There are componentwise interventions known to address the listed challenges at a single component level. Recent studies demonstrated acoustic signal processing algorithms that offer improved resilience of ASR to background noise [27,32]. Some studies improved the noise resilience by implementing speech enhancement algorithms for noise filtering in input audio [33,34], while others trained ASR for various noise patterns to improve its robustness against noise [35,36]. There are well-established methods to establish a customized lexicon for a domain of interest so that ASR could better detect domain-specific terms [37-39]. Some research demonstrated solutions to effectively extracting medical information from clinical text containing both semantic and syntactic errors [40,41].

Despite a number of available component-level interventions, it remains unknown how a combination of all these interventions simultaneously affects the overall performance of hands-free prehospitalization documentation in a noise-intensive operational environment. A technology approach encompassing all possible improvements at every process step of the end-to-end solution has the potential to make a substantial contribution to addressing similar challenges in the daily emergency and prehospital clinical practice.

In this paper, we describe the design of our mobile platform for hands-free documentation in the operational medical environment and lessons learned from its use in a simulated environment. The purpose of the study is to perform a systematic evaluation of improvement opportunities for the platform by aggregating and assessing all possible component-level solutions at every process step. The platform was built to extract meaningful medical information from the real-time audio streaming and generate complete documentation before a patient arrives at a simulated field hospital. To this end, the state-of-the-art ASR solutions with relevant component interventions for modular improvement were thoroughly

explored. Development of our platform was guided by qualitative research and structured evaluation to identify and address the relevant challenges through modular improvement at every process step of the end-to-end solution. Physician-supervised clinical simulation drills were conducted for the precise assessment of the system performance in the emergency settings.

Methods

This research was approved by the Institutional Review Boards of Clemson University (Clemson, South Carolina) and Palmetto Health System (Columbia, South Carolina), with secondary review and approval by the US Army Medical Research and Material Command (Ft Detrick, Maryland).

Qualitative Analysis for Platform Design and Clinical Simulation

Presimulation focus groups and follow-on simulation drill observations were used to assess medical workflow, scope medical information communicated, user requirements during operation, documentation needs, and overall design of platform. Six focus groups were held with 26 individuals across three categories of employment including emergency medical services, transport nurses, and emergency department physicians (Table 1). Focus groups were conducted by trained facilitators using a semistructured interview guide organized to facilitate a workflow discussion of tasks, communication, and documentation strategies as they approach an emergency, the transition to active treatment, and then transition the patient to the next care team. A total of 21 simulation drills were observed over 3 days. Observers monitored their interaction between equipment, verbal communication, and nonverbal communication as they approached the scene, provided active treatment, and transitioned the patient to the next phase of care. Short debriefing interviews were conducted after each drill to gather participant feedback on the process. A postsimulation focus group was also conducted with participants after each day of drills. Data were documented through detailed notes provided on the observation forms and from the two postsimulation drill focus groups (Multimedia Appendix 1).

Table 1. Focus group participants.

	Participants, n (%)	Male, n (%)	Female, n (%)	Experience <10 years, n (%)	Experience 10-20 years, n (%)	Experience >20 years, n (%)	Experience unknown, n (%)
Emergency medical services	6 (23)	5(83)	1 (17)	1 (17)	1 (17)	3 (50)	1 (17)
Transport nurses	11 (42)	8 (73)	3 (27)	1 (9)	2 (18)	8 (73)	0 (0)
Physicians	9 (35)	1 (11)	8 (89)	3 (33)	2 (22)	4 (44)	0 (0)
Total	26 (100)	14 (54)	12 (46)	5 (19)	5 (19)	15 (58)	1 (4)

Hardware Architecture Design

The overall system architecture design consisted of three major platforms: field mobile platform, field hospital platform, and headquarter back-end platform (Figure 1). The field mobile platform operated on a GoPro camera (video capture), microphone, onboard storage (SDXC memory card), and a mobile form factor graphics processing unit (GPU) system (NVIDIA Jetson TX2; Figure 2). The field mobile platform operated on a 7.4V 7000 mAh LiPo battery, which provided continuous power to the platform for up to 8 hours. The transmission between the field device and the field hospital platform was realized through a closed secure network with multiple Linksys Velop WHW0303 routers under Wi-Fi Protected Access II (WPA2) encryption. The field hospital platform operated on a laptop computer where the received information from the field platform was displayed and converted to Fast Healthcare Interoperability Resources–based data types for improved interoperability with EHR platforms. Dell Poweredge R620 equipped with Cerner Sandbox was deployed as a virtual EHR server throughout the project. The headquarter platform operated on a NVIDIA DGX1 with 8 x NVIDIA Tesla 32GB V100 GPUs and 2 x 20-Core 2.20 GHz Intel Xeon E5-2698v4 central processing units. The ASR training was performed on DGX1 from the headquarter platform, and the output model was downloaded to TX2 in the field platform.

Figure 1. Documentation platform architecture design. EHR: electronic health record; FHIR: Fast Healthcare Interoperability Resources; GPU: graphics processing unit; HQ: headquarters.

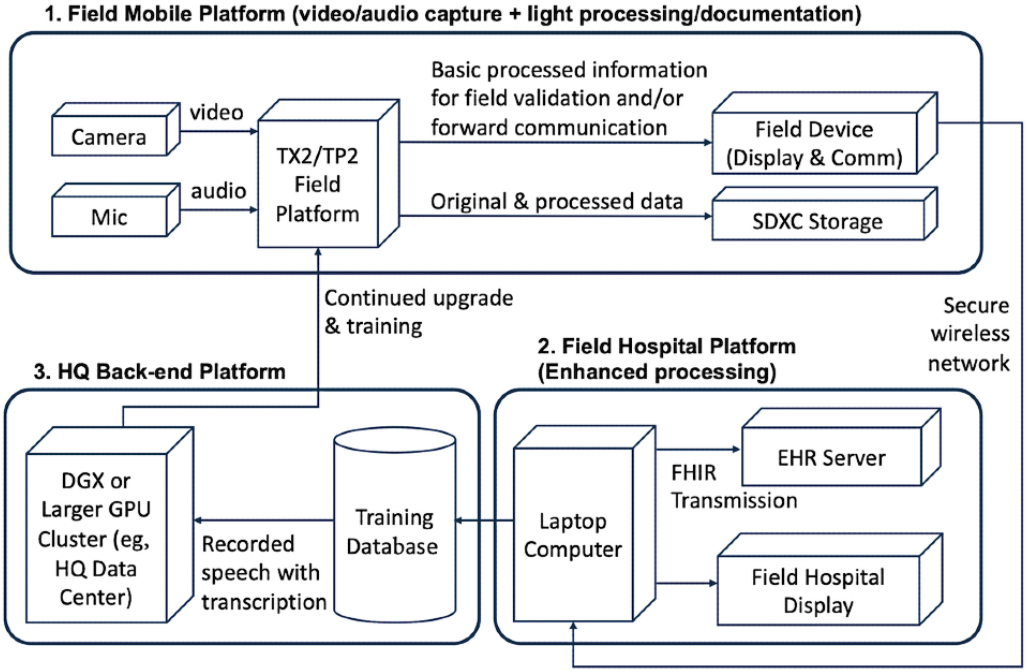
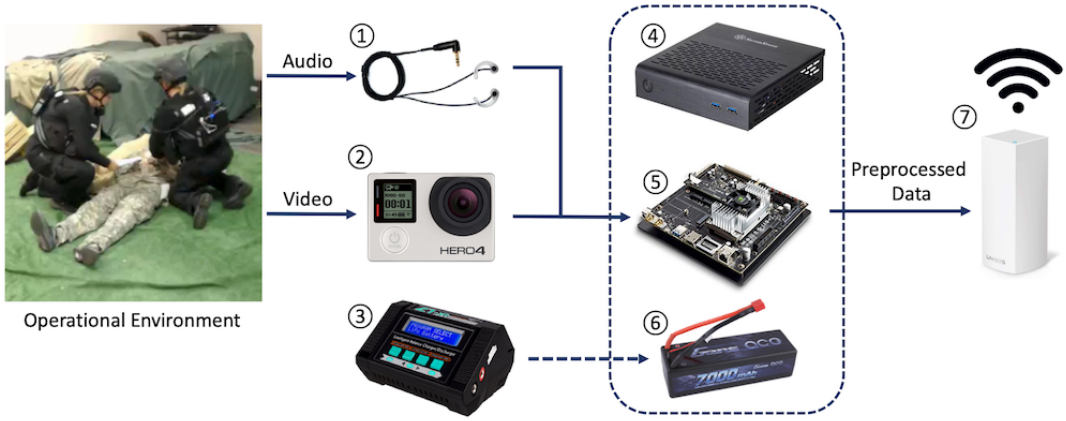


Figure 2. Overview of hardware specifications for field platform.



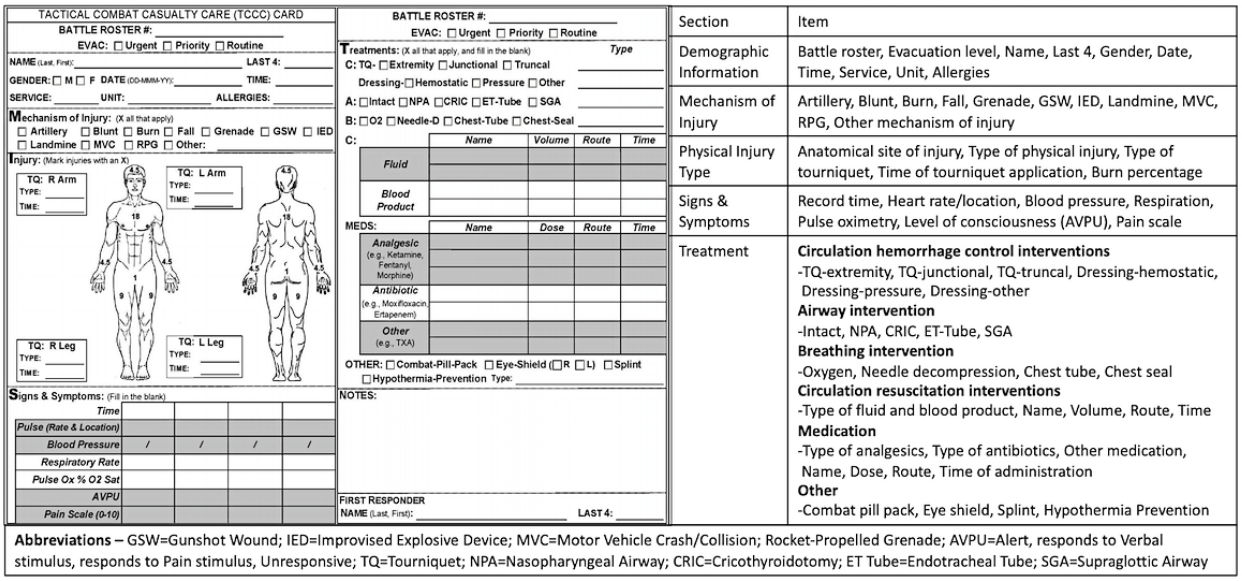
Device	Product	Function
①	SP-TFB-2 Sound Professionals In-ear Binaural Microphones	Audio streaming capture for automatic speech recognition
②	GoPro HERO4 v05.00	Video streaming capture & real-time monitoring
③	EV-Peak C1-XR Charger	LiPo battery charger for field platform
④	SilverStone ITX Case PT13B-120	Enclosure for the field platform
⑤	NVIDIA Jetson TX2	Preprocessing of captured data
⑥	Gens LiPo Battery Pack 7000mAh 14.8V	Continuous power supply up to 6 hours
⑦	Linksys Velop WHW0303 Router	Secure network connection between field and field hospital

Software Architecture Design

The field mobile platform was designed to perform a basic preprocessing of audio captured at the point of injury for the hands-free prehospitalization documentation. The captured audio was converted into a transcript through the ASR module. The Tactical Combat Casualty Care (TCCC) card was selected as the standard format for prehospitalization documentation

throughout the study; it has been well-documented that the complete TCCC documentation results in a higher casualty survival rate [11,15]. The ASR output was analyzed to generate bookmarks for the captured video for immediate retrieval of video footages relevant to injuries of interest and to fill out a TCCC card for patients (Figure 3). The captured audio was first passed on to the voice activity detection module, which decides whether the given input is a human voice or not. Next, the audio containing the human voice was processed by a speech enhancement module to emphasize the human voice and minimize background noise (Figure 4). Upon the audio preprocessing, the acoustic model generates the initial transcriptions, which then are corrected and improved through the language model. The language model was designed to infer each word based on its context by using a probability distribution over sequences of words. During the postanalysis, the transcribed text was processed by a post natural language processing module to generate bookmarks for the point of injuries and preliminary documentation of injuries on the TCCC card.

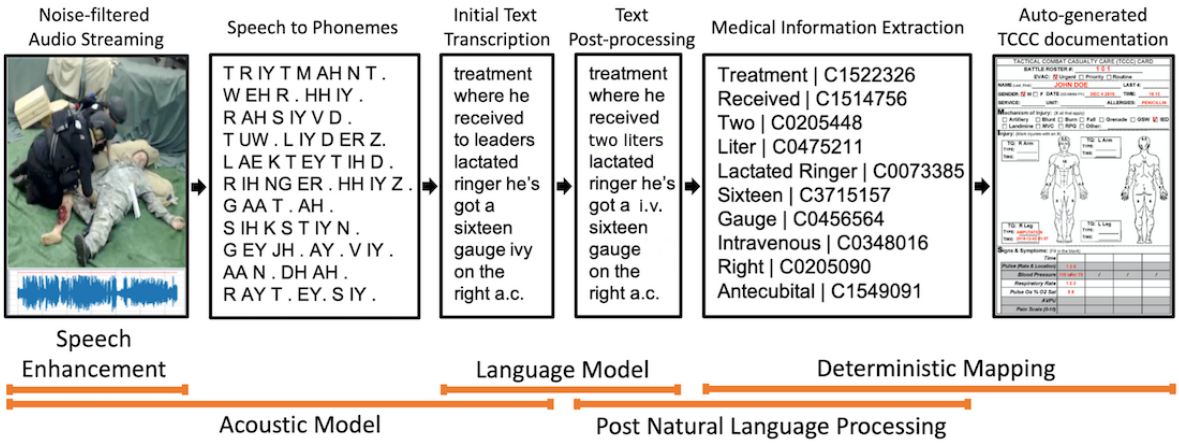
Figure 3. Overview of Tactical Combat Casualty Care card.



Section	Item
Demographic Information	Battle roster, Evacuation level, Name, Last 4, Gender, Date, Time, Service, Unit, Allergies
Mechanism of Injury	Artillery, Blunt, Burn, Fall, Grenade, GSW, IED, Landmine, MVC, RPG, Other mechanism of injury
Physical Injury Type	Anatomical site of injury, Type of physical injury, Type of tourniquet, Time of tourniquet application, Burn percentage
Signs & Symptoms	Record time, Heart rate/location, Blood pressure, Respiration, Pulse oximetry, Level of consciousness (AVPU), Pain scale
Treatment	**Circulation hemorrhage control interventions** -TQ-extremity, TQ-junctional, TQ-truncal, Dressing-hemostatic, Dressing-pressure, Dressing-other **Airway intervention** -Intact, NPA, CRIC, ET-Tube, SGA **Breathing intervention** -Oxygen, Needle decompression, Chest tube, Chest seal **Circulation resuscitation interventions** -Type of fluid and blood product, Name, Volume, Route, Time **Medication** -Type of analgesics, Type of antibiotics, Other medication, Name, Dose, Route, Time of administration **Other** -Combat pill pack, Eye shield, Splint, Hypothermia Prevention

Abbreviations – GSW=Gunshot Wound; IED=Improvised Explosive Device; MVC=Motor Vehicle Crash/Collision; Rocket-Propelled Grenade; AVPU=Alert, responds to Verbal stimulus, responds to Pain stimulus, Unresponsive; TQ=Tourniquet; NPA=Nasopharyngeal Airway; CRIC=Cricothyroidotomy; ET Tube=Endotracheal Tube; SGA=Supraglottic Airway

Figure 4. Data processing workflow for hands-free medical documentation. TCCC: Tactical Combat Casualty Care.



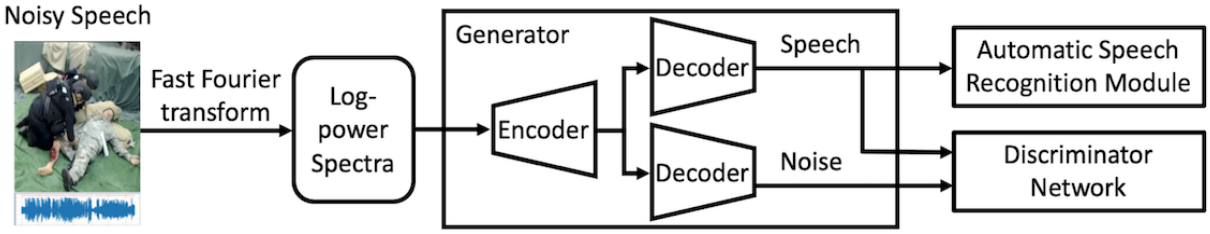
Modular Improvement for Noise Resilience

The selection of each componentwise intervention was guided by relevant literature and a series of preliminary experiments (Multimedia Appendix 2). A hybrid deep neural network model was used to achieve noise-resilient ASR with its performance comparable to that of the current state of the art. For the implementation of the ASR module, an open-source speech recognition platform, Kaldi, was used for the training of the selected models. A Gaussian mixture model–hidden Markov model was first trained to obtain senones (ie, tied triphone states). Next, the corresponding aligned frames were used for training time delay neural network (TDNN) [42]. The TDNN structure includes an input layer, 11 TDNN layers, and one linear output layer with each TDNN layer set to have 1536 nodes [43]. All weights and biases were discriminatively trained by optimizing the cross-entropy between the target probability and the actual SoftMax output with the backpropagation algorithm [44]. The initial training data consisted of the Switchboard data set (260 hours) and the Common Voice data set (500 hours). Parallel training of the TDNNs using up to 8 NVIDIA Tesla 32GB V100 GPUs was done on the training data with 6 epochs.

The speech enhancement module was deployed based on Speech Enhancement Generative Adversarial Network (SEGAN), which enabled the rapid enhancement process without the need for

explicit assumptions about the raw data and generalizability to various speakers and noise types [45]. The module was trained using noisy data sets generated by mixing clean data sets with battlefield noise. The original SEGAN has been further improved through log-power spectra-based operation and forked generative adversarial network (ForkGAN) structure to extract both speech and noise information (Figure 5). The ForkGAN architecture operated directly on spectral domain features instead of on raw audio with aims to learn a mapping from the log-power spectra feature input to its feature output, which has demonstrated to outperform other well-known GAN-based speech enhancement techniques [33].

Figure 5. Overview of the generative adversarial network–based speech enhancement architecture.



Multi-style training was adopted for additional noise resilience in the operational environment. In specific, ASR was trained with a noisy audio data set containing various types of battlefield noise. A total of 17 battlefield noise files were collected from Signal Processing Information Base [46]. These noises included different types of guns, helicopters, tanks, jets with different speeds, speech babble, and white noise. Additionally, the following other continuous noise types were randomly selected and added to the original training data sets: helicopters, armored vehicles, and tanks. Continuous signal-to-noise ratio values from 0 dB to 20 dB were used to signify different noise power levels. The noisy training data set was created in addition to the original training data sets.

Modular Improvement for Medical Information Extraction

Our initial investigations showed the original language model was unable to detect medical and military terms used by the medical professionals during the simulation drills. The primary cause of the failure was that these terms were not present in the dictionary that was created from the original ASR training data. To address the issue, a new customized lexicon was trained from medical and military terms used in battlefield-related injuries and medical evacuation. First, the relevant medical and military fields were identified in the TCCC card, the most predominant documentation template of battlefield injuries. Using the Carnegie Mellon University Sphinx Knowledge Base Tool, a dictionary with these domain-specific words and their corresponding phonemes was generated to update the existing language model [47]. The original dictionary and language models were merged with their corresponding new versions, and then the new merged dictionary was compiled to acquire the new lexicon. The Stanford Research Institute Language Modeling toolkit was used to combine the merged language model and dictionary to generate the new grammar model [48]. The new lexicon, new grammar model, and the existing hidden Markov model context-dependency lexicon grammar (HCLG) graph used for the baseline ASR model were combined to construct the updated HCLG graph.

Although all of the aforementioned methods focused on the accurate transcription of conversation between patients and medics, additional processing extracting medical information from the machine-transcribed unstructured text was necessary for completing TCCC documentation. MetaMap is a key tool developed by the National Library of Medicine that has been widely used in biomedical information retrieval and data mining applications to obtain Unified Medical Language System Concept Unique Identifiers (CUIs) with corresponding textual descriptions [49]. The post natural language processing module used MetaMap 2018 for medical information extraction. The following semantic type mappings were configured for the implementation: anatomical abnormality, anatomical structure, antibiotics, body substance, body location, body part, clinical drug, drug delivery device, diagnostic procedure, disease, finding, medical device, quantitative and qualitative concepts, sign, temporal concept, and therapeutic procedure. To prevent excessive false-positive issues [50], a number of sample clinical notes on gunshot, explosion, and head trauma were manually crafted and inputted to MetaMap for identification of the potential CUIs of interest and the corresponding entry location within the TCCC documentation. To clarify, the module was designed for a closed domain application by discarding concepts that are not in the preidentified CUI list. Lastly, the extracted information was automatically entered into the appropriate TCCC sections through a predetermined mapping.

Clinical Simulation

A total of three clinical simulation drills were conducted in 2017-2019 at Palmetto Health Simulation Center in Columbia, South Carolina. Each physician-supervised drill simulated a typical rescue mission in the medical operational environment. The scope of the simulation spanned from the battlefield to the field hospital, and thus, only the field and field hospital platforms were deployed during the drills. Three common battlefield injury types were used for the clinical scenarios: gunshot wound, amputation due to explosion, head trauma [10,51]. Throughout the drills, all emergency medical care providers taking a role as a medic were wearing the field mobile platform described in Figure 2. The scenarios were loosely scripted by suggesting general descriptions and numbers for vital signs. The medics were allowed to improvise in their verbal reports. The participants acting as patients were also allowed to improvise their responses to medics based on the general description of scenarios.

Each simulation drill started in a room simulating the landscape of field and sky. Various types of battlefield noises were

simulated in the room using multichannel high-output speakers. The medics treated patients as they would on a real battlefield during the first encounter. After the initial treatment, the patients were escorted to the flight paramedics waiting at the next meeting point. The patients were then transported to the next room simulating inside of a medical helicopter. Likewise, helicopter noises were simulated in the room using multichannel high-output speakers. After a certain amount of flight time, patients were then transported to an outdoor space where a field hospital had been set up. The patients received the basic examinations at the field hospital, which concluded one simulation drill. The same three clinical scenarios (gunshot wound, amputation, head trauma) were used for each simulation drill in turn. A total of 27 complete patient cases spanning from field to field hospital were simulated and collected, resulting in a total of 5.05 hours of audio recordings. The maximum noise level of 89 decibels was maintained for gunshot and helicopter noise when measured from the patient's position.

Overall Performance Evaluation and Statistical Analysis

For qualitative evaluation, thematic analysis using a hybrid inductive and deductive approach was completed in Atlas.Ti 8 (Scientific Software Development GmbH) [52,53]. The analysis process began by reviewing focus group transcripts and observation notes using an open coding format to identify various ways participants described their experiences during different stages within the emergency. This was followed by a round of deductive coding focused on communication strategies and device interaction through the ABCDs of Emergency Care. A final round of axial coding produced four thematic areas. All coding was conducted by one member of the research team.

A standard measure to evaluate ASR performance, word error rate (WER), was used to verify whether the acoustic and language models achieved performance comparable to the current state of the art. However, it was suboptimal to include all conversations captured throughout the drills measure since the main goal of our platform was adequate documentation of injuries rather than transcribing daily conversations. Thus, the primary evaluation measures relied on medical WER and referred to WER for only the sentences from medically oriented speech. For example, sentences from nonmedical conversations between the medics were not considered when evaluating the medical WER. The WER was calculated by comparing machine-transcribed text output and text transcribed by human medical transcriptionists who listened to the audio recording of all simulations. Another primary performance measure was based on the completeness of captured clinical information in the autogenerated TCCC documentation. The captured clinical information was assessed using the F1 score calculated by comparing the autogenerated documentation to the manual documentation by physicians. The cost-effect analysis to identify opportunities for modular improvement was based on how much more clinical information could be captured after each componentwise intervention. McNemar test with Bonferroni correction was used to detect the statistical significance of the improvement effect with respect to medical WER. A total of 4 settings with different combinations of modular improvements were tested using the selected measures. Additionally, one setting based on a commercial ASR solution was assessed using the same performance measures. Dragon Medical Practice Edition 4 (DMPE 4) software (Nuance Communications) is one of the predominant speech recognition solutions that assist clinicians with hands-free voice-dictated documentation in clinics. A setting with its ASR powered by DMPE 4 was compared with the settings with the different modular improvements (Multimedia Appendix 3).

Results

Qualitative Study Findings

Four thematic areas include communication methods, communication content, device interaction, and information use (Table 2). Communication methods varied across workflow phases, provider type, and care setting. For example, several focus group participants described frequently using verbal and nonverbal communication strategies with their partner while providing care, and those with military experience discussed this even further. Participants also described situations that they labeled complex communication, whereby they are communicating with and about different patients at one time. This was most frequently discussed as a battlefield experience more so than a transport or field hospital phenomena. Although communication content could vary greatly depending on the workflow phase, the content was remarkably similar within each phase, regardless of the provider type.

Focus group and simulation drill participants' feedback emphasized the need for device flexibility and for the person wearing it to have control. They also encouraged the design team to make the device strong, durable, and lightweight. Simulation participants recommended that users would have to be trained to use the device and to talk aloud during care so that the device can capture what is being done. Finally, participants shared that short notes and recording that could replace charting would increase user perception of value and thus motivation to use. Physician providers noted that short notes or videos or photos of the injury or emergency site transmitted before patient arrival could be helpful.

Table 2. Overview of qualitative study findings.

Theme	Antecedent	Behavior during interaction	Context	Delegation
Communication methods	• Use of mnemonics • Verbal	• More nonverbal • To patient • To partner	• Often chaotic • Can dictate if verbal or nonverbal	• Must be charted/recorded • Very different process at each phase
Communication content	• Roles tasks	• Only what is necessary • If not safe, very little verbal communication • Conversely sometimes lots of content at same time—chaotic	• Dictates depth/detail • Sound an issue for some settings	• Preference for who provides hands-off by provider type • Content is same at each phase of delegation
Device interaction	• Ability to turn on and off prior to hot zone	• Cannot get in the way	• Flexible locations for different types of providers—helmets, chest, shoulder, etc	• When/how to turn off device
Information use	• Planning and preparation	• N/A[a]	• N/A	• Help next team

[a]N/A: not applicable.

Modular Improvement With Componentwise Interventions

The total number of 15,139 words necessary for completing TCCC documentation were identified through transcription from audio recordings collected from all simulation drills. The field mobile platform equipped with baseline ASR achieved a medical WER of 69.9% with 10,582 word errors of 15,139 words (Table 3). Multi-style training incorporating both clean and noise-injected training data sets improved medical WER by a 26.9% decrease in the error rate from 69.9% to 43.0%. The updated language models further reduced medical WER to 33.3%. Although the multi-style training and updated language model decreased the medical WER, deployment of the speech enhancement module increased the error rate to 46.3%. All increases and decreases in the medical WER with the componentwise intervention were statistically significant. The participating physicians identified a total of 768 unique CUIs relevant to the TCCC documentation of gunshot wounds, amputations, and head trauma on the battlefield. The field mobile platform equipped with baseline ASR achieved an F1 score of 0.61. Upon the deployment of the multi-style training, the F1 score increased by 0.11 to 0.72. The updated language models further improved the score to 0.81. However, the score decreased to 0.78 with the deployment of the speech enhancement module.

Among all the componentwise interventions, the combination of multi-style training and an updated language model resulted in the most improvement in medical WER; the error rate was reduced by 36.6% when compared to the baseline model. For specific examples of improvement made by the updated language model, see Table 4. The autogenerated TCCC documentation from our best model (baseline + multi-style training + new language model) achieved an F1 score of 0.81 with 559 true positives, 119 false positives, and 137 false negatives.

Table 3. Automated transcription and documentation performance by different settings.

Setting	ASR[a] transcription output		Automated TCCC[b] documentation		
	Medical word error rate (%)	*P* value[c]	Precision	Recall	F_1 score
Setting 1: baseline	69.9	N/A[d]	0.484	0.828	0.611
Setting 2: baseline + multi-style training	43.0	Setting 1 vs 2: <.001	0.634	0.824	0.717
Setting 3: baseline + multi-style training + updated language model	33.3	Setting 2 vs 3: <.001	0.803	0.824	0.813
Setting 4: baseline + multi-style training + updated language model + speech enhancement	46.3	Setting 3 vs 4: <.001	0.747	0.819	0.781

[a]ASR: automatic speech recognition.

[b]TCCC: Tactical Combat Casualty Care.

[c]McNemar test with Bonferroni correction was used to calculate the statistical significance.

[d]N/A: not applicable.

Table 4. Example of domain-specific word correction with updated language model.

Original language model	Updated language model
air movement *by literally*[a]	air movement *bilaterally*
exit as a *poster here*	exit as a *posterior*
patient is *take kid nick*	patient is *tachypneic*
take *nor vast* for hypertension	take *norvasc* for hypertension
with *pebble radio balls*	with *palpable radial pulses*
active *orchard real* bleeding	active *arterial* bleeding
ten planting numbering preparation	*tympanic membrane perforation*
michael grams offend a nil	*micrograms of fentanyl*
full toxins ninety eight percent	*pulse oximetry* ninety eight percent
soldier triggered *naive do*	soldier triggered *an I.E.D.*

[a]Italics indicate the change between models.

Discussion

Performance and Lessons Learned

The previous studies on the extraction of medical information from the human-written clinical text have reported F1 scores ranging from 0.757 to 0.872, depending on a target entity to be recognized [53,54]. Our platform achieved the comparable F1 score of 0.81, despite the multiple challenges posed by errors that are attributed to the machine transcription under noise-intensive operational environments. Our experience deploying the mobile platform has given us four lessons that may be useful in the development of other similar platforms for speech to patient record applications.

Lesson 1: Closed Domain Strategy

The observation made by the focus group identified considerable similarity between all patient transportation processes regardless of injury types. For example, all medical personnel described a similar set of information that is expected to share as they transition the patient from one setting to the next. The identified similarity between the processes enabled the labor-intensive closed domain solutions for the post natural language processing without concern for resource constraints (eg, physician time). In our experience, both language model and medical information extraction could be further improved through rule-based or manual tasks such as observation-driven lexicon updates and preidentification of relevant CUIs for reducing false positives. The qualitative study to identify the similarity may provide the basis for cost-effect analysis to examine the feasibility of similar closed domain strategies.

Lesson 2: User Training

The importance of user training was pointed out during the focus group study. Accordingly, users were trained to turn on and off the system whenever appropriate, which could prevent the potential false positives incurred by nonclinic conversation. Next, the users were also trained to repeat the information whenever possible. It was observed that, if the same information is repeatedly spoken by a user, the system has a higher chance for complete documentation by properly capturing the information at least once, resulting in the improved high F1 score despite a relatively high WER. We have learned that the proper user training may result in performance improvement as significant as state-of-the-art componentwise interventions.

Lesson 3: Impact of Speech Recognition

The performance of the ASR module had a direct impact on the quality of the autogenerated documentation in our speech-to-patient-record application. It was observed that improvement in medical WER after each componentwise intervention is likely to improve autogenerated documentation quality evaluated by the F1 score. As expected, more medical word errors in ASR-transcribed text interfered with the post–natural language processing to extract medical information for documentation. A preliminary observation on the autogenerated documentation revealed that missing words in ASR output and incorrect negation due to word errors were the major causes of false negatives and positives, respectively.

Lesson 4: Context of Componentwise Intervention

To some extent, our mobile platform resembles a personal artificial intelligence assistant platform on the commercial market, as it listens to its user and executes desired actions (ie, documentation). Although our platform could deploy the same types of componentwise interventions known to be effective for the commercial platforms, not all interventions were effective in our application. In the context of everyday life, the personal assistance platform can benefit from speech enhancement that emphasizes the speech of the primary speaker (eg, owner of device) while suppressing the speech of secondary speakers. However, in the context of medicine, the same speech enhancement module may cause a higher medical WER by filtering out the patient's response to doctors or speech from other care providers attempting to deliver information to the primary speaker. Our experience of the performance degradation reveals a necessity for more context-sensitive training for speech enhancement modules to enhance speech from both primary and secondary speakers in the emergency care settings.

Our field mobile platform used only verbal communications for the documentation. As documented in qualitative study findings, information extraction from nonverbal communication

along with the verbal communication is essential to reducing the loss of information. Future research may incorporate the existing computer vision solutions to examine if additional information can be extracted from nonverbal communication for more resilient documentation. In response to lesson 2, future studies are warranted to perform a hypothesis-driven study to assess the effect of user training on the resilience of documentation. Lastly, our platform was designed for the closed domain application exclusively for the three most common injury types on the battlefield. Although our study demonstrated that the closed domain strategy can be developed to significantly improve speech recognition performance for the target medical conditions, future speech-to-text and medical information extraction modules may explore to expand the platform design for more variety of medical conditions.

To the best of our knowledge, this was the first attempt to create a fully functional platform for hands-free prehospitalization documentation in operational medical environments. Our application contributes to the body of existing knowledge for the development and assessment of platforms to enable hands-free clinical documentation in real-world noisy environments. The development of our platform was strictly guided by domain experts and a series of structured evaluations to examine modular improvement at every process step of the end-to-end solution. The lessons learned suggest potential refinements in the future endeavors to develop other similar platforms for speech-to-patient-record application.

Acknowledgments

This study was supported by a grant from the US Army Medical Research and Materiel Command, Fort Detrick, Maryland (award W81XWH-17-C-0238).

The authors acknowledge and appreciate the logistics and regulatory support of Karen Edwards, MS, of the Department of Public Health Sciences, Clemson University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete and Resilient Documentation simulation drill observation form.
[PNG File , 406 KB - mhealth_v9i10e32301_app1.png]

Multimedia Appendix 2
Preliminary experiments using a standardized testing data set to guide the selection of componentwise interventions for modular improvement.
[PNG File , 210 KB - mhealth_v9i10e32301_app2.png]

Multimedia Appendix 3
Comparison of performance between the presented platform and commercial clinical voice recognition software.
[PNG File , 311 KB - mhealth_v9i10e32301_app3.png]

References

1. Joyce SM, Dutkowski KL, Hynes T. Efficacy of an EMS quality improvement program in improving documentation and performance. Prehosp Emerg Care 1997;1(3):140-144. [doi: 10.1080/10903129708958807] [Medline: 9709356]
2. Staff T, Søvik S. A retrospective quality assessment of pre-hospital emergency medical documentation in motor vehicle accidents in south-eastern Norway. Scand J Trauma Resusc Emerg Med 2011 Mar 31;19:20 [FREE Full text] [doi: 10.1186/1757-7241-19-20] [Medline: 21453536]
3. Tønsager K, Krüger AJ, Ringdal KG, Rehn M, P-EMS Template Collaborating Group. Template for documenting and reporting data in physician-staffed pre-hospital services: a consensus-based update. Scand J Trauma Resusc Emerg Med 2020 Apr 03;28(1):25 [FREE Full text] [doi: 10.1186/s13049-020-0716-1] [Medline: 32245496]
4. Starmer AJ, Spector ND, Srivastava R, West DC, Rosenbluth G, Allen AD, I-PASS Study Group. Changes in medical errors after implementation of a handoff program. N Engl J Med 2014 Nov 06;371(19):1803-1812. [doi: 10.1056/NEJMsa1405556] [Medline: 25372088]
5. Laudermilch DJ, Schiff MA, Nathens AB, Rosengart MR. Lack of emergency medical services documentation is associated with poor patient outcomes: a validation of audit filters for prehospital trauma care. J Am Coll Surg 2010 Feb;210(2):220-227 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2009.10.008] [Medline: 20113943]
6. Meisel ZF, Shea JA, Peacock NJ, Dickinson ET, Paciotti B, Bhatia R, et al. Optimizing the patient handoff between emergency medical services and the emergency department. Ann Emerg Med 2015 Mar;65(3):310-317.e1. [doi: 10.1016/j.annemergmed.2014.07.003] [Medline: 25109535]

7. Butler F. Tactical combat casualty care: combining good medicine with good tactics. J Trauma 2003 May;54(5 Suppl):S2-S3. [doi: 10.1097/01.TA.0000064504.55800.73] [Medline: 12768093]
8. Carter AJE, Davis KA, Evans LV, Cone DC. Information loss in emergency medical services handover of trauma patients. Prehosp Emerg Care 2009;13(3):280-285. [doi: 10.1080/10903120802706260] [Medline: 19499462]
9. Evans SM, Murray A, Patrick I, Fitzgerald M, Smith S, Andrianopoulos N, et al. Assessing clinical handover between paramedics and the trauma team. Injury 2010 May;41(5):460-464. [doi: 10.1016/j.injury.2009.07.065] [Medline: 19735917]
10. Seeley BE. The ability of military health systems applications to coordinate combat casualty care. Urol Nurs 2013;33(2):73-78. [Medline: 23734552]
11. Robinson JB, Smith MP, Gross KR, Sauer SW, Geracci JJ, Day CD, et al. Battlefield documentation of tactical combat casualty care in Afghanistan. US Army Med Dep J 2016(2-16):87-94. [Medline: 27215873]
12. Bukhres O, Morton S, Mossman M. Mobile computing architecture for a battlefield environment. CODAS 1996:130-139.
13. Morris TJ, Pajak J, Havlik F, Kenyon J, Calcagni D. Battlefield Medical Information System-Tactical (BMIST): the application of mobile computing technologies to support health surveillance in the Department of Defense. Telemed J E Health 2006 Aug;12(4):409-416. [doi: 10.1089/tmj.2006.12.409] [Medline: 16942412]
14. Therien S, Nesbitt M, Duran-Stanton A, Gerhardt R. Prehospital medical documentation in the Joint Theater Trauma Registry: a retrospective study. J Trauma 2011 Jul;71(1 Suppl):S103-S108. [doi: 10.1097/TA.0b013e3182218fd7] [Medline: 21795867]
15. Butler FK. Two decades of saving lives on the battlefield: tactical combat casualty care turns 20. Mil Med 2017 Mar;182(3):e1563-e1568. [doi: 10.7205/MILMED-D-16-00214] [Medline: 28290925]
16. Butler JF, Blackbourne L. Battlefield trauma care then and now: a decade of Tactical Combat Casualty Care. J Trauma Acute Care Surg 2012 Dec;73(6 Suppl 5):S395-S402. [doi: 10.1097/TA.0b013e3182754850] [Medline: 23192061]
17. Kotwal R, Montgomery H, Mechler K. A prehospital trauma registry for tactical combat casualty care. US Army Med Dep J 2011:15-17. [Medline: 21607902]
18. Elster EA, Butler FK, Rasmussen TE. Implications of combat casualty care for mass casualty events. JAMA 2013 Aug 07;310(5):475-476. [doi: 10.1001/jama.2013.167481] [Medline: 23925612]
19. Mabry RL. Challenges to improving combat casualty survivability on the battlefield. Joint Force Q 2015 Jan:78-84.
20. Hodgson T, Magrabi F, Coiera E. Efficiency and safety of speech recognition for documentation in the electronic health record. J Am Med Inform Assoc 2017 Nov 01;24(6):1127-1133 [FREE Full text] [doi: 10.1093/jamia/ocx073] [Medline: 29016971]
21. Goss FR, Zhou L, Weiner SG. Incidence of speech recognition errors in the emergency department. Int J Med Inform 2016 Sep;93:70-73 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.05.005] [Medline: 27435949]
22. Blackley S, Huynh J, Wang L, Korach Z, Zhou L. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. J Am Med Inform Assoc 2019 Apr 01;26(4):324-338 [FREE Full text] [doi: 10.1093/jamia/ocy179] [Medline: 30753666]
23. Vogel M, Kaisers W, Wassmuth R, Mayatepek E. Analysis of documentation speed using web-based medical speech recognition technology: randomized controlled trial. J Med Internet Res 2015 Nov 03;17(11):e247 [FREE Full text] [doi: 10.2196/jmir.5072] [Medline: 26531850]
24. Strand O, Holter T, Egeberg A, Stensby S. On the feasibility of ASR in extreme noise using the PARAT earplug communication terminal. 2003 Presented at: 2003 IEEE Workshop on Automatic Speech Recognition and Understanding; November 30-December 4, 2003; St. Thomas, VI. [doi: 10.1109/asru.2003.1318460]
25. Noyes JM, Baber C, Leggatt AP. Automatic speech recognition, noise and workload. Proc Hum Factors Ergonomics Soc Annu Meeting 2016 Nov 05;44(22):762-765. [doi: 10.1177/154193120004402269]
26. Rasch R, Kott A, Forbus KD. AI on the battlefield: an experimental exploration. 2002 Presented at: The Eighteenth National Conference on Artificial Intelligence; July 28-August 1, 2002; Edmonton, AB p. 906-912.
27. Yu D, Li J. Recent progresses in deep learning based acoustic models. IEEE/CAA J Automatica Sinica 2017;4(3):396-409. [doi: 10.1109/jas.2017.7510508]
28. Khademian M, Homayounpour MM. Monaural multi-talker speech recognition using factorial speech processing models. Speech Commun 2018 Apr;98:1-16. [doi: 10.1016/j.specom.2018.01.007]
29. Hermansky H, Sharma S. Temporal patterns (TRAPs) in ASR of noisy speech. 1999 Presented at: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing; March 15-19, 1999; Phoenix, AZ p. 289-292. [doi: 10.1109/icassp.1999.758119]
30. Kodish-Wachs J, Agassi E, Kenny P, Overhage JM. A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. AMIA Annu Symp Proc 2018;2018:683-689 [FREE Full text] [Medline: 30815110]
31. Latif S, Qadir J, Qayyum A, Usama M, Younis S. Speech technology for healthcare: opportunities, challenges, and state of the art. IEEE Rev Biomed Eng 2021;14:342-356. [doi: 10.1109/RBME.2020.3006860] [Medline: 32746367]
32. Tan T, Qian Y, Hu H, Zhou Y, Ding W, Yu K. Adaptive very deep convolutional residual network for noise robust speech recognition. IEEE/ACM Trans Audio Speech Lang Processing 2018 Aug;26(8):1393-1405. [doi: 10.1109/taslp.2018.2825432]

33. Lin J, Niu S, Wei Z. Speech enhancement using forked generative adversarial networks with spectral subtraction. Interspeech 2021:195-209. [doi: 10.21437/interspeech.2019-2954]
34. Donahue C, Li B, Prabhavalkar R. Exploring speech enhancement with generative adversarial networks for robust speech recognition. 2018 Presented at: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing; April 15-20, 2018; Calgary, AB p. 5024-5028. [doi: 10.1109/icassp.2018.8462581]
35. Xu H, Dalsgaard P, Tan Z, Lindberg B. Noise condition-dependent training based on noise classification and SNR estimation. IEEE Trans Audio Speech Lang Processing 2007 Nov;15(8):2431-2443. [doi: 10.1109/tasl.2007.906188]
36. Gemmeke JF, Cranen B, Remes U. Sparse imputation for large vocabulary noise robust ASR. Computer Speech Lang 2011 Apr;25(2):462-479. [doi: 10.1016/j.csl.2010.06.004]
37. Cucu H, Buzo A, Besacier L, Burileanu C. SMT-based ASR domain adaptation methods for under-resourced languages: application to Romanian. Speech Commun 2014 Jan;56:195-212. [doi: 10.1016/j.specom.2013.05.003]
38. Edwards E, Salloum W, Finley GP, Fone J, Cardiff G, Miller M, et al. Medical speech recognition: reaching parity with humans. In: Karpov A, Potapova R, Mporas I, editors. Speech and Computer: 19th International Conference, SPECOM 2017, Hatfield, UK, September 12-16, 2017, Proceedings. Cham: Springer; 2017:512-524.
39. Anantaram C, Sangroya A, Rawat M, Chhabra A. Repairing ASR output by artificial development and ontology based learning. 2018 Presented at: Twenty-Seventh International Joint Conference on Artificial Intelligence; July 13-19, 2018; Stockholm, Sweden p. 5799-5801. [doi: 10.24963/ijcai.2018/842]
40. Kim M, Xu Y, Zaiane O, Goebel R. Patient information extraction in noisy tele-health texts. 2013 Presented at: 2013 IEEE International Conference on Bioinformatics and Biomedicine; December 18-21, 2013; Shanghai, China p. 326-329. [doi: 10.1109/bibm.2013.6732511]
41. Selvaraj SP, Konam S. Medication regimen extraction from medical conversations. In: Shaben-Nejad A, Michalowski M, Buckeridge DL, editors. Explainable AI in Healthcare and Medicine. Cham: Springer; 2021:195-209.
42. Peddinti V, Povey D, Khudanpur S. A time delay neural network architecture for efficient modeling of long temporal contexts. Interspeech 2015:3214-3218. [doi: 10.21437/interspeech.2015-647]
43. Povey D, Ghoshal A, Boulianne G, Burget L, Glembek O, Goel N, et al. The Kaldi speech recognition toolkit. 2011 Presented at: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding; December 11-15, 2011; Waikoloa, HI.
44. Chen X, Eversole A, Li G, Yu D, Seide F. Pipelined back-propagation for context-dependent deep neural networks. 2012 Presented at: Thirteenth Annual Conference of the International Speech Communication Association; September 9-13, 2012; Portland, OR. [doi: 10.21437/interspeech.2012-7]
45. Pascual S, Bonafonte A, Serrà J. SEGAN: speech enhancement generative adversarial network. arXiv Preprint posted online on March 28, 2017.
46. Johnson D, Shami P. The Signal Processing Information Base. IEEE Signal Processing Magazine 1993 Oct;10(4):36-42. [doi: 10.1109/79.248556]
47. Chowdhury SS, Talukdar A, Mahmud A, Rahman T. Domain specific intelligent personal assistant with bilingual voice command processing. 2018 Presented at: TENCON 2018 - 2018 IEEE Region 10 Conference; October 28-31, 2018; Jeju Island, Korea. [doi: 10.1109/tencon.2018.8650203]
48. Upadhyaya P, Farooq O, Abidi MR, Varshney YV. Continuous Hindi speech recognition model based on Kaldi ASR toolkit. 2017 Presented at: 2017 International Conference on Wireless Communications, Signal Processing and Networking; March 22-24, 2017; Chennai, India p. 786-789. [doi: 10.1109/wispnet.2017.8299868]
49. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
50. Reátegui R, Ratté S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes. BMC Med Inform Decis Mak 2018 Sep 14;18(Suppl 3):74 [FREE Full text] [doi: 10.1186/s12911-018-0654-2] [Medline: 30255810]
51. Shackelford SA, Fowler M, Schultz K, Summers A, Galvagno SM, Gross KR, et al. Prehospital pain medication use by U.S. Forces in Afghanistan. Mil Med 2015 Mar;180(3):304-309. [doi: 10.7205/MILMED-D-14-00257] [Medline: 25735021]
52. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
53. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qualitative Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
54. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]

Abbreviations

ASR: automatic speech recognition
CUI: Concept Unique Identifier
DMPE 4: Dragon Medical Practice Edition 4
EHR: electronic health record
ForkGAN: forked generative adversarial network

GPU: graphics processing unit
HCLG: hidden Markov model context-dependency lexicon grammar
SEGAN: Speech Enhancement Generative Adversarial Network
TCCC: Tactical Combat Casualty Care
TDNN: training time delay neural network
WER: word error rate
WPA2: Wi-Fi Protected Access II

Edited by G Eysenbach; submitted 23.07.21; peer-reviewed by P Sengstack, M Elbattah; comments to author 14.08.21; revised version received 03.09.21; accepted 16.09.21; published 12.10.21.

Please cite as:
Woo M, Mishra P, Lin J, Kar S, Deas N, Linduff C, Niu S, Yang Y, McClendon J, Smith DH, Shelton SL, Gainey CE, Gerard WC, Smith MC, Griffin SF, Gimbel RW, Wang KC
Complete and Resilient Documentation for Operational Medical Environments Leveraging Mobile Hands-free Technology in a Systems Approach: Experimental Study
JMIR Mhealth Uhealth 2021;9(10):e32301
URL: https://mhealth.jmir.org/2021/10/e32301
doi:10.2196/32301
PMID:34636729

Review

Identification and Evaluation of Methodologies to Assess the Quality of Mobile Health Apps in High-, Low-, and Middle-Income Countries: Rapid Review

Fionn Woulfe[1]; Kayode Philip Fadahunsi[2], MBBS, MPH; Simon Smith[1], BSc, MSc; Griphin Baxter Chirambo[3], BSc, MSc, PhD; Emma Larsson[4], BSc(Hons), MBBS; Patrick Henn[5], Bch, BAO, MB, MSc, MATLHE; Mala Mawkin[6], BSc(Hons), MBBS; John O' Donoghue[7], BSc, MSc, PhD, Cert (ICL)

[1]School of Medicine, University College Cork, Cork, Ireland

[2]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[3]Faculty of Health Sciences, Mzuzu University, Mzuzu, Malawi

[4]Ashford and St Peter's Hospitals NHS Trust, Chertsey, United Kingdom

[5]Assert Centre, College of Medicine & Health, University College Cork, Cork, Ireland

[6]School of Medicine, Imperial College London, London, United Kingdom

[7]Malawi eHealth Research Centre, University College Cork, College Road, Cork, Ireland

Corresponding Author:
Fionn Woulfe
School of Medicine
University College Cork
College Road
Cork, T12 K8AF
Ireland
Phone: 353 21 490 3000
Email: fionnwoulfe@gmail.com

Abstract

Background: In recent years, there has been rapid growth in the availability and use of mobile health (mHealth) apps around the world. A consensus regarding an accepted standard to assess the quality of such apps has yet to be reached. A factor that exacerbates the challenge of mHealth app quality assessment is variations in the interpretation of quality and its subdimensions. Consequently, it has become increasingly difficult for health care professionals worldwide to distinguish apps of high quality from those of lower quality. This exposes both patients and health care professionals to unnecessary risks. Despite progress, limited understanding of the contributions of researchers in low- and middle-income countries (LMICs) exists on this topic. Furthermore, the applicability of quality assessment methodologies in LMIC settings remains relatively unexplored.

Objective: This rapid review aims to identify current methodologies in the literature to assess the quality of mHealth apps, understand what aspects of quality these methodologies address, determine what input has been made by authors from LMICs, and examine the applicability of such methodologies in LMICs.

Methods: This review was registered with PROSPERO (International Prospective Register of Systematic Reviews). A search of PubMed, EMBASE, Web of Science, and Scopus was performed for papers related to mHealth app quality assessment methodologies, which were published in English between 2005 and 2020. By taking a rapid review approach, a thematic and descriptive analysis of the papers was performed.

Results: Electronic database searches identified 841 papers. After the screening process, 52 papers remained for inclusion. Of the 52 papers, 5 (10%) proposed novel methodologies that could be used to evaluate mHealth apps of diverse medical areas of interest, 8 (15%) proposed methodologies that could be used to assess apps concerned with a specific medical focus, and 39 (75%) used methodologies developed by other published authors to evaluate the quality of various groups of mHealth apps. The authors in 6% (3/52) of papers were solely affiliated to institutes in LMICs. A further 15% (8/52) of papers had at least one coauthor affiliated to an institute in an LMIC.

Conclusions: Quality assessment of mHealth apps is complex in nature and at times subjective. Despite growing research on this topic, to date, an all-encompassing appropriate means for evaluating the quality of mHealth apps does not exist. There has

been engagement with authors affiliated to institutes across LMICs; however, limited consideration of current generic methodologies for application in LMIC settings has been identified.

Trial Registration: PROSPERO CRD42020205149; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=205149

(JMIR Mhealth Uhealth 2021;9(10):e28384) doi:10.2196/28384

KEYWORDS

mHealth app; health app; mobile health; health website; quality; quality assessment; methodology; high-income country; low-income country; middle-income country; LMIC; mobile phone

Introduction

Background

Mobile health (mHealth) apps can be defined as software "incorporated into smartphones to improve health outcome, health research, and health care services" [1]. In 2017, >325,000 mHealth apps were available for download [2]. These apps can enhance health promotion and disease prevention, resulting in improved patient outcomes and economic savings [3,4].

In 2020, 35% of US health care consumers used mHealth apps compared with just 16% in 2014 [5]. Access to and use of these apps is also increasing in many low- and middle-income countries (LMICs) [6]. In 2015, there were >7 billion mobile telephone subscriptions worldwide, 70% of which were in LMICs [7,8]. Furthermore, 95% of the global population resides in an area covered by mobile cellular networks, with 84% of people having access to mobile broadband networks [9]. Such widespread use and access to smartphones has helped incorporate mHealth solutions into health care systems within LMICs [10].

Since the introduction of mHealth in the late 2000s, apps have facilitated improvements in disease management, reductions in health care costs and boosted service efficiency [3,4,11]. Despite the growing popularity of mHealth apps, research has also identified the potential risks associated with their use. Regardless of location, quality of content and software functionality are areas of concern in mHealth apps [12], as are data privacy and security [10,13]. For successful implementation of mHealth in LMIC settings, additional factors such as user-prospective and technical factors should also be considered [10].

At present, there is no comprehensive, universally available methodology to assess the quality of mHealth apps [14]. In addition, the existing five-star rating scales available within app stores provide subjective indications of quality, which are often unreliable [15]. Given the paucity of current methodologies, unreliability associated with star ratings, and the ever-expanding mHealth app market, the challenge for health care professionals to identify high-quality apps is becoming increasingly difficult.

A factor that exacerbates the conundrum of quality assessment is indeed the word quality itself. Quality can be considered an umbrella term encompassing many dimensions, depending on its context. Hence, disparities exist in the depth and focus of its definition. The Institute of Medicine defines quality in health care broadly as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" [16]. The International Organization for Standardization adopts a more expansive approach and defines quality as the "degree to which a set of inherent characteristics of an object fulfills requirements" [17]. Nouri et al [18] proposed a broad classification model to address quality in relation to the mHealth app evaluation. Within this model, criteria and subcriteria are outlined for consideration when evaluating the quality of mHealth apps.

Perhaps it is this hierarchical, multifaceted nomenclature that has rendered it difficult to unify on a standard of quality when discussing mHealth apps. Various approaches have been taken to help identify higher quality apps. In the United Kingdom, the National Health Service app library provides a collection of mHealth apps of approved quality [19]. The Federal Institute for Drugs and Medical devices in Germany is set to examine the quality of apps with a view to doctors ultimately being able to "prescribe health care apps to patients" [20].

Efforts are also being made for mHealth app evaluation methods in LMIC settings [21]. However, despite the rapidly increasing market access, significant developments have yet to occur. Given the variability of socioeconomics across the globe, additional parameters in methodologies for mHealth app evaluation in emerging economics may be required.

Objectives

The primary aim of this rapid review is to identify current methodologies in the literature to assess the quality of mHealth apps. Second, it aims to determine what aspects of quality these methodologies consider. Third, it aims to examine global research input on this topic since 2005. Finally, this review examines the applicability of such methodologies in LMIC settings.

Methods

Study Design

Rapid reviews draw upon traditional systematic review processes to accelerate and streamline research while preserving the rigor and quality of review methodology [22]. Given the aforementioned research aims and objectives, a rapid review approach was deemed appropriate.

The broad principles of scoping review methodology, as defined by Arksey and O'Malley [23], were followed to formulate the research question and identify relevant studies for selection. A concept-centric approach was taken for the charting procedure in line with the advice given by Webster and Watson [24] for writing literature reviews in the field of information sciences.

A standard protocol was followed in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [25].

Search Strategy

A systematic search strategy was developed and applied across four databases: PubMed, EMBASE, Scopus, and Web of Science. These databases have strong scientific and medical focuses. This combination of databases was chosen in an effort to guarantee adequate and efficient coverage of relevant papers [26].

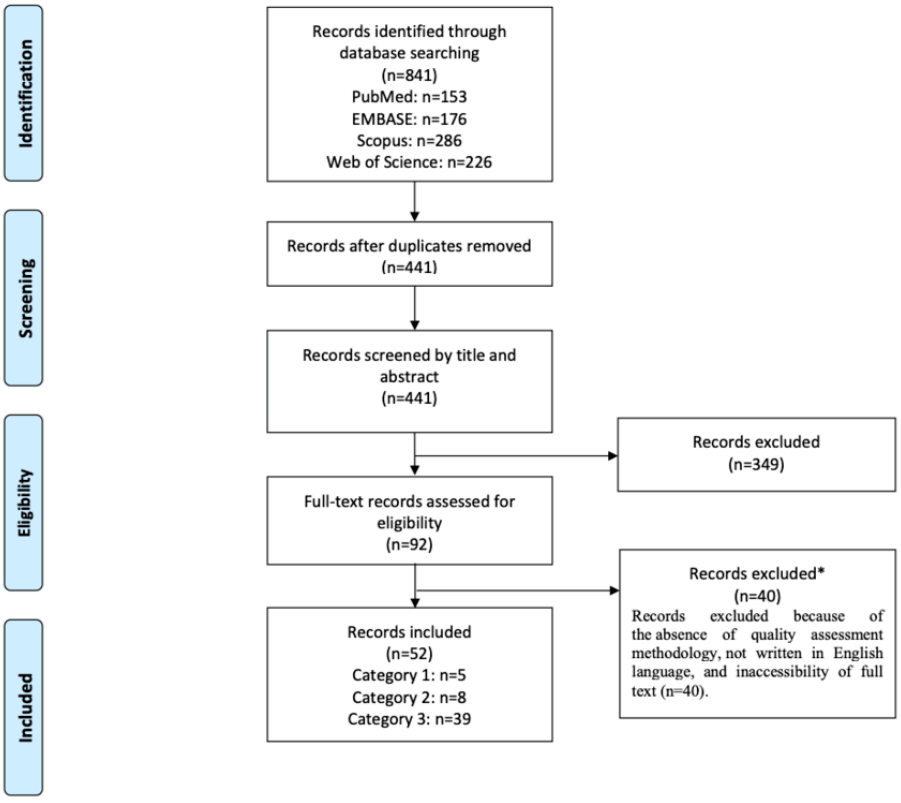
Under the guidance of an academic librarian and reflecting upon the advice of Arksey and O'Malley [23] for conducting literature reviews, the research question was split into the following four specific concepts: *methodology*, *assess*, *quality*, and *mHealth app*. Through the iterative process of keyword searching and preliminary search testing using Medical Subject Headings terms, synonyms of each concept were incorporated into the search string. The final search string is provided in Multimedia Appendix 1. An intrinsic link exists between mHealth apps and health websites. Therefore, variations of *health website* were included in the search string to identify papers that potentially covered both mHealth app and health website domains.

The search was conducted in December 2020 and was limited to studies published in English between 2005 and 2020. The year 2005 was chosen as the starting point for this review as the first iPhone was released on the market in 2007, and the app store was created in 2008 [27]. Geographical restrictions were not imposed on this search.

Study Selection

The reference management software EndNote X9 (Clarivate Analytics) was used to collate the initial literature search citations. Duplicates were removed before exporting the remaining citations to the Covidence systematic review software (v2409). The author FW initially screened titles and abstracts to determine whether a paper met the general study selection criteria. The full texts of the remaining papers were formally screened against the inclusion and exclusion criteria by FW. Any papers that the author FW was unsure about were screened by and clarified through engagement with the author JOD. The search results were presented in a PRISMA flow diagram (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the inclusion process.



Selection Criteria

The inclusion criteria for studies were as follows:

1. Papers that proposed a methodology to evaluate the quality of mHealth apps (regardless of assessing one aspect or several aspects of quality)
2. Papers that used a methodology for the quality evaluation of specific groups of mHealth apps
3. Papers published in English from 2005 onward in research journals

The exclusion criteria for studies were as follows:

1. Papers that proposed a methodology for the evaluation of non-mHealth apps
2. Papers that proposed a methodology solely for the evaluation of health websites

3. Papers that were not available through the library of University College Cork, interlibrary loans, or via direct communication with authors
4. Papers that were not available in the English language

Categorization of Reviewed Papers

Papers that successfully passed full-text screening were subdivided into three categories based on their thematic synergies. These were as follows:

- Category 1 (generic methodologies): papers that proposed generic methodologies to evaluate the quality of mHealth apps or mHealth websites
- Category 2 (health condition–specific methodologies): papers that proposed methodologies designed specifically to evaluate the quality of mHealth apps that focus on one medical condition (ie, App Quality Evaluation Tool to evaluate the quality of nutrition apps)
- Category 3 (use of existing methodologies): papers that used a prepublished methodology to evaluate certain groups of mHealth apps

Data Extraction

Data items extracted from all studies included paper title, author, year, aim or objective, name of methodology used or developed, target platform of methodology, disease focus of study, strengths, validity and reliability of methodology, weaknesses, and future work. The location of authors institute affiliations was classified based on the World Bank Classification system [28] into high-, middle-, or low-income countries. Any uncertainty was clarified through discussion with a second reviewer (JOD). This extraction form was initially piloted and amended where necessary.

The methodologies proposed by the papers in category 1 (ie, those that proposed a generic methodology for mHealth app evaluation) were compared with a reference classification checklist of criteria for assessing mHealth app quality proposed by Nouri et al [18].

Quality Evaluation

The aim of this rapid review is to assess the extent of published literature on mHealth quality assessment methodologies and related studies rather than to evaluate specific causes and effects. Therefore, as supported by the World Health Organization, risk of bias assessment was not conducted, as this review served as an information gathering process [29].

Results

The search and paper retrieval processes are illustrated in Figure 1. A total of 841 potentially relevant papers were identified. Of the 841 papers, following the removal of duplicates, 441 (52.4%) papers remained, with 52 (6.2%) papers meeting the criteria for inclusion.

Characteristics of Retrieved Papers

The papers were subdivided into three categories. Category characteristics and respective citations are indicated in Table 1.

Table 1. Summary of paper categories and their respective citations (N=52).

Category	Number of papers, n (%)	General explanation of category
Category 1: generic methodologies	5 (10)	• Papers that propose generic methodologies to evaluate the quality of mHealth apps and health websites [30] • Papers that propose generic methodologies to evaluate the quality of mHealth apps [31-34]
Category 2: health condition–specific methodologies	8 (15)	• Papers that propose methodologies designed specifically to evaluate the quality of mHealth apps which focus on one medical condition (ie, AQEL[a] to evaluate the quality of nutrition apps) [35-42]
Category 3: use of existing methodologies	39 (75)	• Papers that use a prepublished methodology to evaluate certain groups of mHealth apps (ie, the use of MARS[b] to evaluate the quality of traditional medicine mHealth apps available in Iran) [43-81]

[a]AQEL: App Quality Evaluation Tool.

[b]MARS: Mobile App Rating Scale.

Category 1

Of the five papers identified in category 1, 1 (20%) proposed a methodology to evaluate the quality of both mHealth apps and health websites [30], and 4 (80%) proposed methodologies solely used to evaluate the quality of mHealth apps [31-34]. The coverage of the Nouri et al [18] mHealth app evaluation criteria found in the methodologies within this category can be viewed in Table 2.

Table 2. Coverage of category 1 methodologies of the criteria for assessing the quality of mHealth apps proposed by Nouri et al [18].

mHealth app quality assessment criteria (Nouri et al [18])	Baumel et al (Enlight) [30]	Yasini et al [31]	Anderson et al (ACDC[a]) [32]	Stoyanov et al (MARS[b]) [33]	Martínez-Pérez et al [34]
Design					
Suitability of design	✓				
Aesthetics	✓			✓	
Appearance	✓				✓
Design consistency	✓		✓	✓	
Information or content					
Credibility	✓	✓	✓	✓	✓
Accuracy	✓	✓		✓	✓
Quality of information	✓		✓	✓	
Quantity of information	✓		✓	✓	
Usability					
Ease of use	✓	✓	✓	✓	✓
Operability		✓			
Visibility	✓				
User control and freedom					
Consistency and standards					
Error prevention					✓
Completeness	✓	✓			
Information needs	✓	✓			
Flexibility and customizability		✓			
Competency					
Style	✓				
Behavior	✓				
Structure					
Functionality					
Performance			✓	✓	✓
Health warnings			✓		
Feedback	✓		✓		
Connectivity and interpretability		✓	✓		
Record	✓				
Display					
Guide	✓				
Remind or alert	✓				
Communicate	✓				✓
Ethical issues					
Beneficence	✓	✓			
Nonmaleficence		✓			
Autonomy		✓			
Justice		✓			
Legal obligations	✓	✓			
Security and privacy					

mHealth app quality assessment criteria (Nouri et al [18])	Baumel et al (Enlight) [30]	Yasini et al [31]	Anderson et al (ACDC[a]) [32]	Stoyanov et al (MARS[b]) [33]	Martínez-Pérez et al [34]
Security	✓	✓			✓
Privacy	✓	✓			
User perceived value					
User perceived value	✓	✓	✓	✓	✓

[a]ACDC: App Chronic Disease Checklist.

[b]MARS: Mobile App Rating Scale.

Category 2

The methodologies proposed in category 2 of the papers focused specifically on asthma [35], pain management [36], medication adherence [37], medication-related problems [38], hard of hearing [39], diabetes mellitus [40], infant feeding [41], and nutritional [42] mHealth apps. The methodologies proposed within this category of papers were highly specific to one topic of medicine. Therefore, their respective dimensions of quality were not subjected to further investigation.

Category 3

Papers in category 3 were concerned with a variety of medical conditions. Quality assessment of nutrition-related mHealth apps was the most prevalent area of research within this category [46,47,50,55,58,63,65,76]. Other types of mHealth apps studied were those related to obesity and weight management [53,62,64], pain management [43,70,75], mental health [51,67], and oncology [57,82].

Within this category, of the 39 papers, 23 (59%) papers applied one methodology, and 16 (41%) papers used a combination of methodologies to evaluate the quality of mHealth apps. The Mobile App Rating Scale (MARS) construct [33] was the most commonly used methodology (24/39, 62%). It was used in 38% (15/39) of papers as the sole means of quality evaluation [45-49,52,57,69-71,73-76,78,79]. MARS was also used in an additional 23% of papers along with other methodologies, such as clinical guidelines [46,55,56,58,65,67,68,77,81]. The breakdown of methodologies and the frequency of their use can be viewed in Table 3.

Table 3. Breakdown of methodologies used by papers in category 3 (N=39).

Methodology used	Articles, n (%)
MARS[a] [45-49,52,57,69-71,73-76,78,79]	15 (38)
uMARS[b] [48,63]	2 (5)
Silberg scale [51,53,64]	3 (8)
AQEL[c] [50]	1 (3)
mHON[d] code [61]	1 (3)
IOM[e] aims [72]	1 (3)
Combination of methodologies (System Usability Scale and clinical guidelines) [46,47,54-56,58-60,62,65-68,77,80,81]	16 (41)

[a]MARS: Mobile App Rating Scale.

[b]uMARS: user version of Mobile App Rating Scale.

[c]AQEL: App Quality Evaluation Tool.

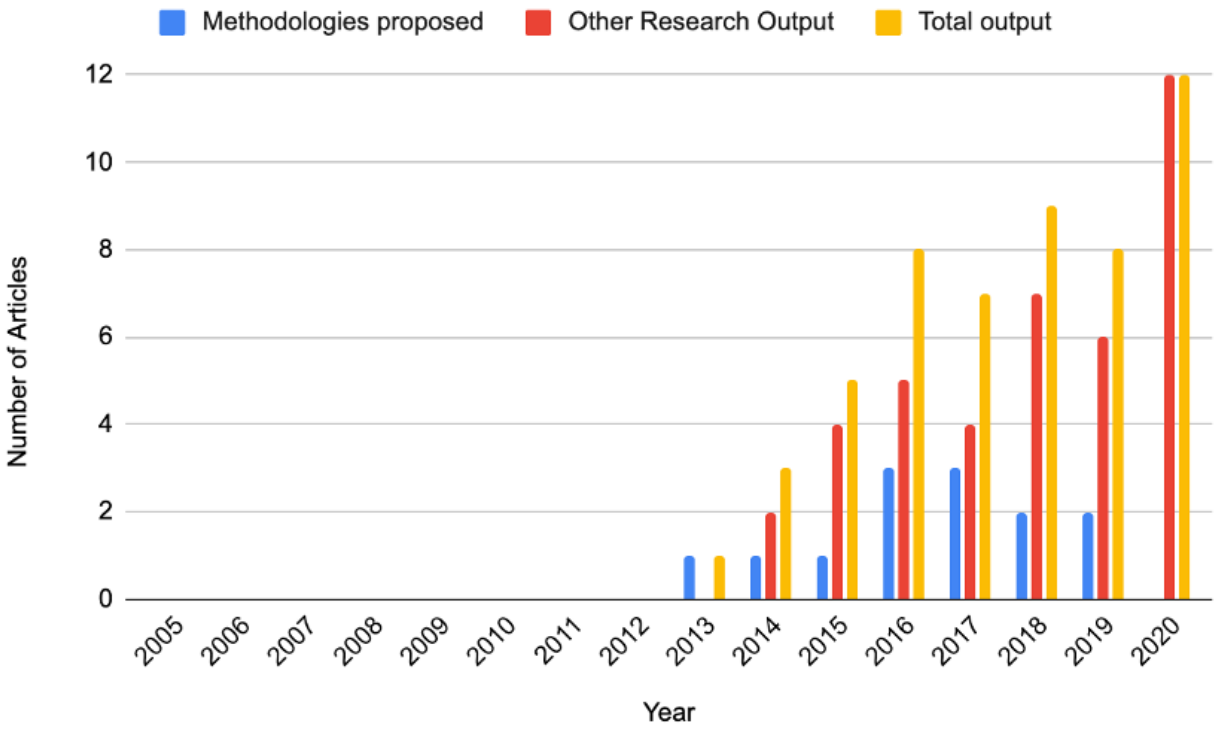
[d]mHON: Mobile applications–Health on the Net.

[e]IOM: Institute of Medicine.

Timeline of Published mHealth Assessment Methodologies and Studies

Research output on the topic of mHealth quality assessment has significantly increased in recent years. Since 2005, 52 papers have been published on this topic; of the 52 papers, 12 (23%) were published in 2020 alone. The research output of novel methodologies for evaluating mHealth apps (categories 1 and 2) and studies relating to the topic (category 3) since 2005 are illustrated in Figure 2.

Figure 2. Illustration of research output on mobile health app evaluation studies from 2005 to 2020.



International Input

The location of authors institute affiliations for all papers (categories 1, 2, and 3) was classified into high-, middle-, or low-income countries based on the World Bank Classification system [28]. All category 1 papers were published by authors affiliated with institutions in high-income countries. One paper in category 2 was published by authors affiliated with an institute in a low-or middle-income country [35]. Of the 39 papers in category 3, 2 (5%) were solely published by authors affiliated with institutes in LMICs [44,78]. A further 21% (8/39) of papers in this category had at least one author affiliated with institutes in LMICs [45,54,56,60,63,67,74,75]. The location of authors' affiliated institutes can be viewed in the Geo chart in Figure 3. A breakdown of countries and the number of authors affiliated with it can be viewed in Table 4.

Figure 3. Geo chart indicating research output affiliated to each country.

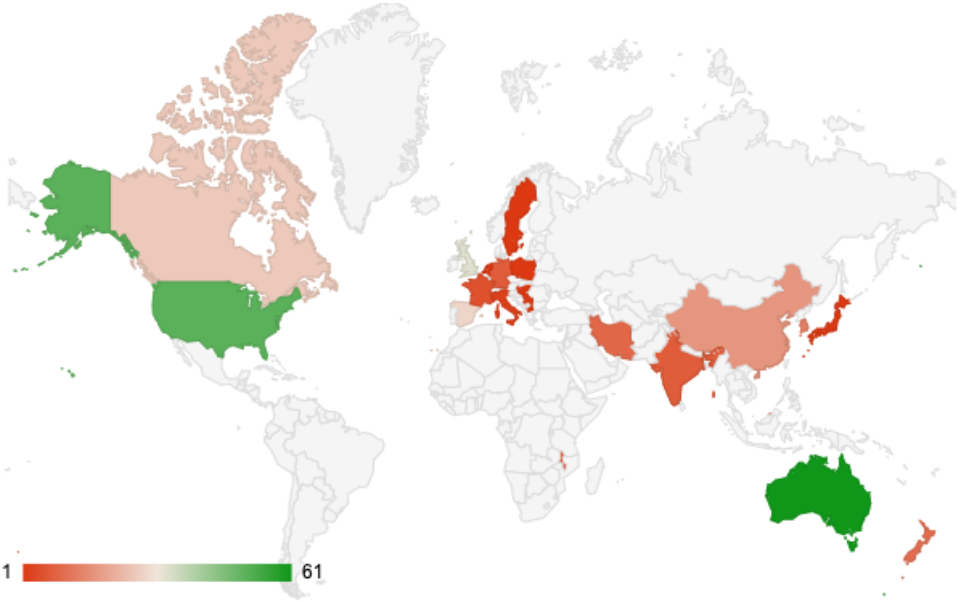


Table 4. A breakdown of research contribution based on the country of author affiliation institute.

Author affiliation and country (ordered alphabetically)	Number of authors affiliated to institutions in that country, n (%)
High-income countries	
Australia	61 (23.7)
Canada	26 (10.1)
France	5 (1.9)
Germany	7 (2.7)
Hungary	1 (0.4)
Italy	2 (0.8)
Japan	1 (0.4)
Korea, Republic	13 (5.1)
Netherlands	1 (0.4)
New Zealand	10 (3.9)
Poland	1 (0.4)
Spain	28 (10.9)
Singapore	16 (6.2)
Sweden	1 (0.4)
United Kingdom	33 (12.8)
United States	51 (19.8)
Total affiliations from high-income countries	257 (100)
Middle-income countries	
China	17 (48.6)
India	7 (20)
Iran, Islamic Republic	9 (25.7)
Macedonia	1 (2.9)
Serbia	1 (2.9)
Total affiliations from middle-income countries	35 (100)
Low-income countries	
Brunei	3 (42.9)
Malawi	4 (57)
Total affiliations from low-income countries	7 (100)

Discussion

Principal Findings

A variety of methodologies to assess the quality of mHealth apps have been identified in this review. Some adopted a generic approach and can be used to evaluate mHealth apps for various medical conditions. Other methodologies take a disease-centric approach and are only relevant when considering apps concerned with that particular disease. Despite a number of quality assessment methodologies being available, significant variations in the dimensions of quality that they address were identified. Given the subjective nature of quality and its subdimensions, it is not surprising to find this high degree of diversity.

As presented in category 3, the MARS construct proposed by Stoyanov et al [33] has been widely used by other authors to evaluate the quality of mHealth apps. MARS is a concise, easy-to-use tool that covers many of the Nouri et al [28] criteria for assessing mHealth app quality (Table 2). Despite its popularity, MARS fails to address some important key aspects of quality, most notably security and privacy. The use of mHealth apps may involve the processing of sensitive information by multiple parties. Therefore, a rising awareness and concern exist in relation to the safety of the information that they contain [13,83]. This underscores the importance of considering privacy and security when evaluating mHealth apps and highlights a significant limitation of the MARS construct. Only two of the five generic methodologies took both of these dimensions into consideration [30,31].

In contrast to MARS, the Enlight suite of assessments proposed by Baumel et al [30] provides a more thorough assessment of quality. It has been designed for both mHealth apps and health

website quality evaluation purposes. As presented in Table 2, Enlight has comprehensive coverage of the Nouri et al [28] criteria for mHealth app quality assessment. Rating measures within the Enlight suite are divided into two sections: quality assessments and checklists. The quality assessment section refers to aspects of quality that relate to the user's experience of an mHealth app. The checklists are not expected to directly impact the end user's experience of the product's efficacy; rather, these lists may expose the user (or provider) to acknowledged risks or benefits.

Respondent fatigue is a well-documented phenomenon in questionnaires [84]. Although the Enlight suite provides a far-reaching means to evaluate the quality of mHealth apps, its all-encompassing nature may, in reality, curtail its use. Along with the checklists section, 28 questions are contained within the Enlight Quality Assessment section. This is significantly greater than that of other generic methodologies. Hence, the use of the Enlight suite would take significantly longer than others to score the quality of mHealth apps. Undeniably, a greater balance is needed to maximize user uptake and engagement among the health care community. This is especially important, as in many cases health care professionals are not allotted additional time to assess new apps.

Although an abundance of mHealth apps is available, academic studies on their clinical impact are lacking. Concerningly, many mHealth apps are not based on any behavior change theory, and in many cases, their effectiveness has not been correctly evaluated [82,85]. With that said, the ability of apps to stimulate behavior change is becoming a growing area of interest [86]. The behavior change technique is not an explicit quality criterion proposed by Nouri et al [28]; however, the World Health Organization recognizes the importance of health outcome–based measures [87]. This review identified its considerations in three of the generic methodologies [30,32,33]. The App Chronic Disease Checklist (ACDC) construct includes behavior change as a singular point of consideration [32]. In contrast, the MARS construct assesses "the perceived impact of an app on the user's knowledge, attitudes, intentions to change as well as the likelihood of actual change in the targeted health behavior" in its *App-Specific* section [33]. Similarly, the *Enlight Suites Therapeutic Persuasiveness* section is specifically dedicated to addressing the topic of behavior change techniques [30]. Although behavior change technique in itself is a broad concept, it is reassuring to identify its consideration even to a certain extent within many methodologies.

Challenges in App Assessment

As mentioned in the introduction, a paucity of uniform definitions for quality and its respective subdimensions exists. A lack of clear-cut definitions not only poses a challenge to this research but also adds a level of ambiguity to mHealth app quality evaluations as a whole. Until precise definitions of each dimension of quality are provided, ongoing subjectivity regarding the interpretation of a dimension of quality with respect to an mHealth app may continue.

It is quite important to consider the validity and reliability of the assessment tools in health care [88]. Validity indicates how well a tool measures what it intends to measure, and reliability expresses the extent to which the obtained results are reproducible [88]. Most of the selected tools offered some form of face and content validity based on expert opinions [30-33]. Only 4% (2/52) of studies [30,33] provided reliability results. However, the selected studies did not conduct factor analysis, which can limit their construct validity. In addition, none of the tools provided any predictive validity, which is the extent to which the scores predict the ability of the mHealth app to improve the targeted health condition. Thus, the paucity of information on the validity and reliability of the available tools could limit their usefulness in practice.

Methodologies proposed within category 1 provide the user with a means to assess the quality of mHealth apps. However, no methodology within this category provides the user with a scoring mechanism or rubric to interpret the results. For example, when using the ACDC checklist, what does it mean if an app contains an overwhelming amount of information but scores perfect results in all other dimensions of quality? Does this render the app low quality? A lack of clear scoring mechanisms may hinder a user's interpretation of the evaluation process, making it an inconclusive exercise.

Applicability of mHealth App Evaluation Methodologies in High-, Middle-, and Low-Income Countries

Although 46% of new mHealth app publishers are from Europe [89], the apps they develop are often available in international markets. As the functionality of mHealth apps becomes more diverse and ownership of smartphones rises, it is likely that their adoption by those living in LMICs will continue to increase. The applicability of the aforementioned methodologies for assessing mHealth app quality in LMIC settings has not been widely considered. As a health care professional contemplates whether a specific mHealth app would be beneficial for their patient, the suitability of an app in the context of his or her patient must be considered. Various regulatory, technical, and user-prospective factors have been identified as obstacles to the integration of mHealth solutions in resource-poor settings [10].

Many regulatory factors that may affect mHealth use in LMICs also affect their use in high-income countries (HICs). Security and privacy of data are two examples. Table 2 highlights that these factors are currently considered in many quality assessment methodologies. Continued access to the internet represents a technological factor that may affect mHealth use in LMICs disproportionately to that in HICs [10]. Despite the penetration rate of mobile broadband signal doubling in LMICs over the past two decades [90], challenges such as use, cost, and speed continue to exist. As such, researchers may wish to consider the impact of inconsistent internet services on an app's functionality. The ACDC checklist [32] was the sole generic methodology to address the facilitation of an *offline mode*. The incorporation of questions such as this within methodologies helps to consider the reality faced by many within LMICs at present.

Socioeconomics can impact the use of mHealth solutions. With increased global demand, it represents an important parameter for consideration. Two factors within the domain of

socioeconomics, which may be important, are cultural appropriateness and literacy. Cultural appropriateness is essential for designing user interfaces or web interfaces for international and country-specific audiences that will be accepted and liked by users [91]. Cultural appropriateness applies to mHealth app evaluations not only in LMICs but also in HICs. If the content of an mHealth app is unsuitable for a particular audience, its download may become a contentious or fruitless exercise. For example, an app designed for prenatal care in Ireland may not be appropriate for use in sub-Saharan Africa. As far as the authors are aware, no generic methodology has explicitly examined the cultural appropriateness of an mHealth app. However, vague considerations were made in the MARS construct [33] and the Enlight suite [30]. In these cases, the suitability of certain aspects of an app, such as information and visual content with respect to the target audience, were mentioned. Given the broadening cultural diversity of app users, perhaps a more formal effort to consider cultural appropriateness exists for the benefit of those in LMICs and HICs.

Health literacy is a concern for many low- and middle-income populations. Within the domain of literacy, readability refers to the comprehension level required by an individual to correctly understand and engage with written material [92]. Past research indicates that many mHealth apps are written at excessively high reading grade levels [66,93]. Poor readability may increase the scope for misinterpretation and render an app inaccessible to many potential end users [66,93]. Nouri et al [28] considered readability as a subcriterion of ease of use [18]. Only two of the five generic methodologies explicitly consider these subcriteria [31,32]. Although the average reading level in LMICs is rising, in many cases, it is still behind that of HICs [94]. Given the proportion of mHealth app development from HICs, a salient need for health care professionals in LMICs to consider the readability of these apps in terms of their potential end users is important.

Future Work

The authors identify several directions for future work in this area of research. First, the review could be extended to papers published in languages other than English, providing a more accurate representation of quality assessment methodologies currently available at an international level.

The Enlight suite provides a thorough means for evaluating the quality of mHealth apps; however, its fundamental usability and ability to consider an app in the context of various populations could be enhanced. An area of active research by the authors is the revision and enhancement of this tool based on the knowledge of this rapid review. Through a Delphi study and supporting survey techniques, the suite is in the process of being modified to make it more user friendly and comprehensive in LMIC settings.

This study highlights several challenges associated with the use of quality assessment methodologies in practice. There is scope to formalize methodology reliability processes, yielding more transparency and comparability in assessments. A scoring mechanism or rubric may be considered in future methodologies that provides users with a means to summarize an app based on the aggregated dimensions of quality that it fulfills.

On a practical level, this research provides additional emphasis on the importance of mHealth app quality assessments. Methodologies such as the Enlight suite and MARS construct are suitable for the purposes outlined in this paper. However, going forward, these methodologies may also be used in consultation with health care professionals for reasons of app development, providing a template for quality assurance.

Strengths

This review has several strengths. To the best of the authors' knowledge, this is the first review to consider the applicability of generic methodologies to evaluate the quality of mHealth apps in LMIC settings. Furthermore, it highlights the affiliations of authors institutes, indicating where significant research input has come from in the past. This review begins to consider further parameters that one may wish to incorporate into methodologies in the future to improve their relevancy across resource-poor settings.

Limitations

This research is not without limitations. A decision was made by the research team to exclude methodologies for the evaluation of *health websites only* and non-mHealth apps. This decision was based on the fact that such methodologies often consider parameters that are not applicable to mHealth apps themselves. In an effort to retrieve all relevant papers, terms relating to these concepts were included within the search string. However, only those papers that formally met the inclusion or exclusion criteria were considered in this review.

Although reviewed by a second author where necessary, paper retrieval, selection, and data extraction were completed by one reviewer (FW). Nouri et al provided generalized definitions or examples of its respective quality assessment criteria [18]. A considered approach was taken by the author FW, whereby a methodology with reasonable coverage of the criteria was positively reflected in data extraction. A lack of universal definitions for quality and its respective subdimensions posed a challenging factor for data extraction, comparison, and synthesis on this topic.

The investigators acknowledge that the country affiliation of methodology authors may have limited relevance toward the application of those methodologies within their respective locations. Nevertheless, given the international market demand and varying socioeconomics, the investigators believe that this approach serves as one of many, which may help indicate the suitability of mHealth app quality assessment methodologies in LMICs.

Finally, only articles published in English were included in this review. This may have some impact on our results presented in Figure 3 and Table 4, as methodologies published in other languages were not identified.

Conclusions

Quality assessment of mHealth apps is a complex task. Significant heterogeneity exists between the aspects of quality that are considered by the methodologies identified by this rapid review. Some key aspects of quality remain unaddressed by certain methodologies despite their growing popularity.

Although engagement with authors affiliated to institutes in LMIC exists on this topic, limited consideration has been made for the use of current methodologies in LMIC settings.

Owing to the variety of stakeholders involved in mHealth (eg, software engineers, information technology departments or companies, health care professionals, and patients), the challenges of finding or developing an all-encompassing methodology to assist health care professionals in assessing the quality of a given app is easily appreciated. With the ever-increasing role of mHealth apps in health care, it is time to consider policy development at the international level. An inclusive and intuitive mHealth app assessment methodology is required to ensure the reliable use of mHealth apps worldwide.

Acknowledgments

All those who met the authorship criteria are listed above as the authors. Each certify that they have participated sufficiently in this body of work. The lead author, FW, received an Irish Health Research Board Scholarship, which helped support this project (Scholarship Number: SS-2020-089).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string applied to databases.
[DOCX File , 14 KB - mhealth_v9i10e28384_app1.docx]

References

1. Ming LC, Hameed MA, Lee DD, Apidi NA, Lai PS, Hadi MA, et al. Use of medical mobile applications among hospital pharmacists in Malaysia. Ther Innov Regul Sci 2016 Jul 30;50(4):419-426. [doi: 10.1177/2168479015624732] [Medline: 30227023]
2. Mobile health (mHealth) market to reach USD 311.98 billion by 2027. Reports and Data. 2020. URL: https://www.globenewswire.com/news-release/2020/04/28/2023512/0/en/Mobile-Health-mHealth-Market-To-Reach-USD-311-98-Billion-By-2027-Reports-and-Data.html [accessed 2021-02-07]
3. Duplaga M, Tubek A. mHealth - areas of application and the effectiveness of interventions. Zdrowie Publiczne i Zarządzanie 2018;16(3):155-166. [doi: 10.4467/20842627oz.18.018.10431]
4. Ghanchi J. mHealth apps: how are they revolutionizing the healthcare industry? Mddionline. 2019 Sep. URL: https://www.mddionline.com/digital-health/mhealth-apps-how-are-they-revolutionizing-healthcare-industry [accessed 2021-01-30]
5. Safavi K, Kalis B. How can leaders make recent digital health gains last? US findings. Accenture 2020 Digital Health Consumer Survey. 2020. URL: https://www.accenture.com/_acnmedia/PDF-130/Accenture-2020-Digital-Health-Consumer-Survey-US.pdf#zoom=40 [accessed 2021-01-30]
6. Feroz A, Kadir MM, Saleem S. Health systems readiness for adopting mhealth interventions for addressing non-communicable diseases in low- and middle-income countries: a current debate. Glob Health Action 2018;11(1):1496887 [FREE Full text] [doi: 10.1080/16549716.2018.1496887] [Medline: 30040605]
7. Measuring the information society report. International Telecommunication Union. 2015. URL: http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf, [accessed 2021-02-07]
8. Mobile-cellular telephone subscriptions. In: Key ICT Indicators for Developed and Developing Countries and the World (Totals and Penetration Rates). Geneva: International Telecommunication Union; 2017.
9. About International Telecommunication Union (ITU). International Telecommunication Union. URL: https://www.itu.int/en/about/Pages/default.aspx [accessed 2021-01-30]
10. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Glob Health Action 2017 Jun 25;10(sup3):1327686 [FREE Full text] [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
11. 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]
12. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
13. Sampat B, Prabhakar B. Privacy risks and security threats in mHealth apps. J Int Technol Inf Manag 2017;26(4):126-153 [FREE Full text]
14. Moshi MR, Tooher R, Merlin T. Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: a systematic review. Int J Technol Assess Health Care 2018 Jan;34(5):464-475. [doi: 10.1017/S026646231800051X] [Medline: 30201060]

15. Kuehnhausen M, Frost V. Trusting smartphone apps? To install or not to install, that is the question. In: Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support. 2013 Presented at: IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA); Feb 25-28, 2013; San Diego, CA, USA. [doi: 10.1109/cogsima.2013.6523820]
16. IOM definition of quality. Institute of Medicine. 2013. URL: http://iom.nationalacademies.org/Global/News%20Announcements/Crossing-the-Quality-Chasm-The-IOM-Health-Care-Quality-Initiative [accessed 2021-02-07]
17. ISO 8000-2:2020(en) Data quality — Part 2: Vocabulary. Online Browsing Platform. 2011. URL: https://www.iso.org/obp/ui/#iso:std:iso:8000:-2:ed-4:v1:en:term:3.1.3 [accessed 2021-02-07]
18. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
19. NHS apps library. National Health Service. URL: https://www.nhs.uk/apps-library/ [accessed 2021-01-15]
20. German doctors to prescribe health apps in 'World First'. HealthManagement.org. 2020. URL: https://healthmanagement.org/c/it/news/german-doctors-to-prescribe-health-apps-in-world-first [accessed 2021-01-15]
21. The MAPS Toolkit: mHealth Assessment and Planning for Scale. Geneva: World Health Organization; 2015:1-94.
22. Kelly SE, Moher D, Clifford TJ. Defining rapid reviews: a modified delphi consensus approach. Int J Technol Assess Health Care 2016 Jan;32(4):265-275. [doi: 10.1017/S0266462316000489] [Medline: 27724987]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Webster J, Watson R. Analyzing the past to prepare for the future: writing a literature review. MIS Quarterly 2002 Jun;26(2):xiii-xxiii [FREE Full text]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Bramer W, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev 2017 Dec 06;6(1):245 [FREE Full text] [doi: 10.1186/s13643-017-0644-y] [Medline: 29208034]
27. A brief history of the Apple App store in its first 10 years. Digital Trends. 2021. URL: https://www.digitaltrends.com/news/apple-app-store-turns-10/ [accessed 2021-06-21]
28. World Bank Country And Lending Groups – World Bank Data Help Desk. Datahelpdesk.worldbank.org. 2021. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> [accessed 2021-01-12]
29. Rapid Reviews To Strengthen Health Policy And Systems: A Practical Guide. Geneva: World Health Organization; 2017.
30. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
31. Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth quality: a process to seal the qualified mobile health apps. Stud Health Technol Inform 2016;228:205-209. [Medline: 27577372]
32. Anderson K, Burford O, Emmerton L. App chronic disease checklist: protocol to evaluate mobile apps for chronic disease self-management. JMIR Res Protoc 2016 Nov 04;5(4):e204 [FREE Full text] [doi: 10.2196/resprot.6194] [Medline: 27815233]
33. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
34. Martínez-Pérez B, de la Torre-Díez I, Candelas-Plasencia S, López-Coronado M. Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. J Med Syst 2013 Oct;37(5):9976. [doi: 10.1007/s10916-013-9976-x] [Medline: 24014255]
35. Guan Z, Sun L, Xiao Q, Wang Y. Constructing an assessment framework for the quality of asthma smartphone applications. BMC Med Inform Decis Mak 2019 Oct 15;19(1):192 [FREE Full text] [doi: 10.1186/s12911-019-0923-8] [Medline: 31615493]
36. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
37. Ali EE, Teo AK, Goh SX, Chew L, Yap KY. MedAd-AppQ: a quality assessment tool for medication adherence apps on iOS and android platforms. Res Social Adm Pharm 2018 Dec;14(12):1125-1133. [doi: 10.1016/j.sapharm.2018.01.006] [Medline: 29402755]
38. Loy JS, Ali EE, Yap KY. Quality assessment of medical apps that target medication-related problems. J Manag Care Spec Pharm 2016 Oct;22(10):1124-1140. [doi: 10.18553/jmcp.2016.22.10.1124] [Medline: 27668561]
39. Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of deaf and hard-of-hearing mobile apps: evaluation using the Mobile App Rating Scale (MARS) with additional criteria from a content expert. JMIR Mhealth Uhealth 2019 Oct 30;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]

40. Kaltheuner M, Drossel D, Heinemann L. DiaDigital apps: evaluation of smartphone apps using a quality rating methodology for use by patients and diabetologists in Germany. J Diabetes Sci Technol 2019 Jul 28;13(4):756-762 [FREE Full text] [doi: 10.1177/1932296818803098] [Medline: 30264585]
41. Zhao J, Freeman B, Li M. How do infant feeding apps in China measure up? A content quality assessment. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e186 [FREE Full text] [doi: 10.2196/mhealth.8764] [Medline: 29212627]
42. DiFilippo KN, Huang W, Chapman-Novakofski KM. A new tool for Nutrition App Quality Evaluation (AQEL): development, validation, and reliability testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163 [FREE Full text] [doi: 10.2196/mhealth.7441] [Medline: 29079554]
43. 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 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
44. Rangrazejeddi F, Anvari S, Farrahi SR, Sharif RS. The most popular Iranian smartphone applications for traditional medicine: A quality assessment. Shiraz E-Med J 2018;19(Suppl):e66306 [FREE Full text]
45. Jovičić S, Siodmiak J, Watson ID, European Federation of Clinical ChemistryLaboratory Medicine Working Group on Patient Focused Laboratory Medicine. Quality evaluation of smartphone applications for laboratory medicine. Clin Chem Lab Med 2019 Feb 25;57(3):388-397. [doi: 10.1515/cclm-2018-0710] [Medline: 30496130]
46. Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300 [FREE Full text] [doi: 10.2196/17300] [Medline: 32459187]
47. Taki S, Elliott R, Russell G, Laws R, Campbell K, Denney-Wilson E. Smartphone applications and websites on infant feeding: a systematic analysis of quality, suitability and comprehensibility. Obes Res Clin Pract 2014 Dec;8(1):101. [doi: 10.1016/j.orcp.2014.10.184]
48. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the mobile application rating scale. JMIR Mhealth Uhealth 2019 May 01;7(5):e13019 [FREE Full text] [doi: 10.2196/13019] [Medline: 31066712]
49. Chapman C, Champion KE, Birrell L, Deen H, Brierley M, Stapinski LA, et al. Smartphone apps about crystal methamphetamine ("Ice"): systematic search in app stores and assessment of composition and quality. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e10442 [FREE Full text] [doi: 10.2196/10442] [Medline: 30463834]
50. DiFilippo KN, Huang WD, Chapman-Novakofski KM. Mobile apps for the Dietary Approaches to Stop Hypertension (DASH): app quality evaluation. J Nutr Educ Behav 2018 Jun;50(6):620-625. [doi: 10.1016/j.jneb.2018.02.002] [Medline: 29526661]
51. Zhang MW, Ho RC, Loh A, Wing T, Wynne O, Chan SW, et al. Current status of postnatal depression smartphone applications available on application stores: an information quality analysis. BMJ Open 2017 Nov 14;7(11):e015655 [FREE Full text] [doi: 10.1136/bmjopen-2016-015655] [Medline: 29138195]
52. Jupp JC, Sultani H, Cooper CA, Peterson KA, Truong TH. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. Pediatr Blood Cancer 2018 Nov;65(11):e27278. [doi: 10.1002/pbc.27278] [Medline: 29943893]
53. Jeon E, Park H, Min YH, Kim H. Analysis of the information quality of Korean obesity-management smartphone applications. Healthc Inform Res 2014 Jan;20(1):23-29 [FREE Full text] [doi: 10.4258/hir.2014.20.1.23] [Medline: 24627815]
54. Xiao Q, Wang Y, Sun L, Lu S, Wu Y. Current status and quality assessment of cardiovascular diseases related smartphone apps in China. Stud Health Technol Inform 2016;225:1030-1031. [Medline: 27332467]
55. Lambert K, Mullan J, Mansfield K, Owen P. Should we recommend renal diet-related apps to our patients? An evaluation of the quality and health literacy demand of renal diet-related mobile applications. J Ren Nutr 2017 Nov;27(6):430-438. [doi: 10.1053/j.jrn.2017.06.007] [Medline: 29056159]
56. Regmi K, Kassim N, Ahmad N, Tuah N. Assessment of content, quality and compliance of the STaR mobile application for smoking cessation. Tob Prev Cessat 2017 Jul 3;3(July):120 [FREE Full text] [doi: 10.18332/tpc/75226] [Medline: 32432194]
57. Hartin P, Cleland I, Nugent C, Tschanz J, Clark C, Norton M. Assessing app quality through expert peer review: a case study from the gray matters study. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 16-20, 2016; Orlando, FL, USA p. 4379-4382. [doi: 10.1109/embc.2016.7591697]
58. McAleese D, Papadaki A. A content analysis of the quality and behaviour change techniques of smartphone apps promoting the Mediterranean diet. Proc Nutr Soc 2020 Jun 10;79(OCE2):E271. [doi: 10.1017/s0029665120002190]
59. Bergeron D, Iorio-Morin C, Bigder M, Dakson A, Eagles ME, Elliott CA, Canadian Neurosurgery Research Collaborative. Mobile applications in neurosurgery: a systematic review, quality audit, and survey of Canadian neurosurgery residents. World Neurosurg 2019 Jul;127:e1026-e1038. [doi: 10.1016/j.wneu.2019.04.035] [Medline: 30980978]
60. Crehan C, Kesler E, Dube Q, Lufesi N, Giaconne M, Normand C, et al. The acceptability, feasibility and usability of the neotree application in malawi: an integrated data collection, clinical management and education mhealth solution to improve

quality of newborn care and thus newborn survival in health facilities in resource-poor settings. Br Med J 2018;103(Suppl 1):117. [doi: 10.1136/archdischild-2018-rcpch.278]
61. Cabrera A, Ranasinghe M, Frossard C, Postel-Vinay N, Boyer C. An exploratory empirical analysis of the quality of mobile health apps. In: Proceedings of the 18th International Conference on WWW/Internet. 2019 Presented at: 18th International Conference on WWW/Internet; Nov 7-9, 2019; Cagliari, Italy. [doi: 10.33965/icwi2019_201913l013]
62. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
63. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China app store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261 [FREE Full text] [doi: 10.2196/13261] [Medline: 31364606]
64. Zhang MW, Ho RC, Hawa R, Sockalingam S. Analysis of the information quality of bariatric surgery smartphone applications using the Silberg scale. Obes Surg 2016 Jan 30;26(1):163-168. [doi: 10.1007/s11695-015-1890-5] [Medline: 26424704]
65. Lambert K, Owen P, Koukoumas A, Mesiti L, Mullan J, Mansfield K, et al. Content analysis of the quality of diet-related smartphone applications for people with kidney disease. Nephrol 2015 Sep:57 [FREE Full text]
66. Ayyaswami V, Padmanabhan DL, Crihalmeanu T, Thelmo F, Prabhu AV, Magnani JW. Mobile health applications for atrial fibrillation: a readability and quality assessment. Int J Cardiol 2019 Oct 15;293:288-293 [FREE Full text] [doi: 10.1016/j.ijcard.2019.07.026] [Medline: 31327518]
67. Li Y, Zhao Q, Cross WM, Chen J, Qin C, Sun M. Assessing the quality of mobile applications targeting postpartum depression in China. Int J Ment Health Nurs 2020 Oct 28;29(5):772-785. [doi: 10.1111/inm.12713] [Medline: 32223070]
68. Bearne LM, Sekhon M, Grainger R, La A, Shamali M, Amirova A, et al. Smartphone apps targeting physical activity in people with rheumatoid arthritis: systematic quality appraisal and content analysis. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e18495 [FREE Full text] [doi: 10.2196/18495] [Medline: 32706727]
69. Pourrahmat M, De Vera MA. FRI0629 Patient-targeted smartphone apps for systemic lupus erythematosus: a systematic review and assessment of features and quality. Br Med J 2018;77:7106. [doi: 10.1136/annrheumdis-2018-eular.7106]
70. Hoffmann A, Faust-Christmann CA, Zolynski G, Bleser G. Toward gamified pain management apps: mobile application rating scale-based quality assessment of pain-mentor's first prototype through an expert study. JMIR Form Res 2020 May 26;4(5):e13170 [FREE Full text] [doi: 10.2196/13170] [Medline: 32452803]
71. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
72. Lee Y, Shin S, Kim J, Kim JH, Seo D, Joo S, et al. Evaluation of mobile health applications developed by a tertiary hospital as a tool for quality improvement breakthrough. Healthc Inform Res 2015 Oct;21(4):299-306 [FREE Full text] [doi: 10.4258/hir.2015.21.4.299] [Medline: 26618037]
73. Sullivan RK, Marsh S, Halvarsson J, Holdsworth M, Waterlander W, Poelman MP, et al. Smartphone apps for measuring human health and climate change co-benefits: a comparison and quality rating of available apps. JMIR Mhealth Uhealth 2016 Dec 19;4(4):e135 [FREE Full text] [doi: 10.2196/mhealth.5931] [Medline: 27993762]
74. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for Corona Virus Disease 2019 (COVID19) and the Assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
75. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):9209 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
76. Moseley I, Roy A, Deluty A, Brewer JA. Evaluating the quality of smartphone apps for overeating, stress, and craving-related eating using the mobile application rating scale. Curr Addict Rep 2020 Sep 01;7(S2):260-267. [doi: 10.1007/s40429-020-00319-7]
77. Jones C, O'Toole K, Jones K, Brémault-Phillips S. Quality of psychoeducational apps for military members with mild traumatic brain injury: an evaluation utilizing the mobile application rating scale. JMIR Mhealth Uhealth 2020 Aug 18;8(8):e19807 [FREE Full text] [doi: 10.2196/19807] [Medline: 32808937]
78. Kalhori SR, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of English mobile applications for gestational diabetes: app review using Mobile Application Rating Scale (MARS). Curr Diabetes Rev 2021 Jul 03;17(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
79. Amor M, Collado-Borrell R, Escudero-Vilaplana V. 1ISG-013 Smartphone applications for patients diagnosed with genitourinary tumours: analysis of the quality using the mobile application rating scale. Eur J Hosp Pharm 2020;27(1):A6. [doi: 10.1136/ejhpharm-2020-eahpconf.13]
80. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Experiences and results of applying tools for assessing the quality of a mHealth app named Heartkeeper. J Med Syst 2015 Nov;39(11):142. [doi: 10.1007/s10916-015-0303-6] [Medline: 26345452]

81. Hay-Smith J, Peebles L, Farmery D, Dean S, Grainger R. Apps-olutely fabulous? - The quality of PFMT smartphone app content and design rated using the Mobile App Rating Scale, Behaviour Change Taxonomy, and guidance for exercise prescription. In: Proceedings of the ICS 2019 Gothenburg Conference. 2019 Presented at: ICS 2019 Gothenburg Conference; Sep 3-6, 2019; Gothemburg, Sweden URL: https://www.ics.org/2019/abstract/547
82. 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 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
83. Nurgalieva L, O'Callaghan D, Doherty G. Security and privacy of mHealth applications: a scoping review. IEEE Access 2020 Jun 4;8:104247-104268. [doi: 10.1109/access.2020.2999934]
84. Lavrakas PJ. Respondent fatigue. In: Encyclopedia of Survey Research Methods. Thousand Oaks, California, United States: SAGE Publications Ltd; 2008.
85. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
86. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
87. mHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011:1-102.
88. Bolarinwa O. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Niger Postgrad Med J 2015;22(4):195-201 [FREE Full text] [doi: 10.4103/1117-1936.173959] [Medline: 26776330]
89. The rise of mHealth apps: a market snapshot. Liquid State. 2018. URL: https://liquid-state.com/mhealth-apps-market-snapshot/ [accessed 2021-01-15]
90. Mobile network coverage and evolving technologies. ICT Facts and Figures. 2016. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf [accessed 2016-11-15]
91. What is cultural appropriateness. IGI Global. URL: https://www.igi-global.com/dictionary/culturally-appropriate-web-user-interface/6361 [accessed 2021-01-30]
92. Albright J, de Guzman C, Acebo P, Paiva D, Faulkner M, Swanson J. Readability of patient education materials: implications for clinical practice. Appl Nurs Res 1996 Aug;9(3):139-143. [doi: 10.1016/S0897-1897(96)80254-0]
93. Dunn Lopez K, Chae S, Michele G, Fraczkowski D, Habibi P, Chattopadhyay D, et al. Improved readability and functions needed for mHealth apps targeting patients with heart failure: an app store review. Res Nurs Health 2021 Feb;44(1):71-80. [doi: 10.1002/nur.22078] [Medline: 33107056]
94. International student assessment (PISA). OECD Library. 2021. URL: https://www.oecd-ilibrary.org/education/reading-performance-pisa/indicator/english_79913c69-en [accessed 2021-02-07]

Abbreviations

ACDC: App Chronic Disease Checklist
HIC: high-income country
LMIC: low- and middle-income country
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PROSPERO: International Prospective Register of Systematic Reviews

Edited by L Buis; submitted 03.03.21; peer-reviewed by Y Wu, M Kolotylo-Kulkarni, S Donevant; comments to author 27.04.21; revised version received 22.06.21; accepted 21.07.21; published 12.10.21.

Please cite as:
Woulfe F, Fadahunsi KP, Smith S, Chirambo GB, Larsson E, Henn P, Mawkin M, O'Donoghue J
Identification and Evaluation of Methodologies to Assess the Quality of Mobile Health Apps in High-, Low-, and Middle-Income Countries: Rapid Review
JMIR Mhealth Uhealth 2021;9(10):e28384
URL: https://mhealth.jmir.org/2021/10/e28384
doi:10.2196/28384
PMID:34636737

Original Paper

Open-source Longitudinal Sleep Analysis From Accelerometer Data (DPSleep): Algorithm Development and Validation

Habiballah Rahimi-Eichi[1,2,3], PhD; Garth Coombs III[1], PhD; Constanza M Vidal Bustamante[1], BSc; Jukka-Pekka Onnela[4], PhD; Justin T Baker[2,3], MD, PhD; Randy L Buckner[1,3,5], PhD

[1]Department of Psychology, Harvard University, Cambridge, MA, United States

[2]Institute for Technology in Psychiatry, McLean Hospital, Belmont, MA, United States

[3]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

[5]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, United States

Corresponding Author:
Habiballah Rahimi-Eichi, PhD
Department of Psychology
Harvard University
52 Oxford Street
Northwest Building, East Wing, Room 280
Cambridge, MA, 02138
United States
Phone: 1 3057337293
Email: hrahimi@fas.harvard.edu

Abstract

Background: Wearable devices are now widely available to collect continuous objective behavioral data from individuals and to measure sleep.

Objective: This study aims to introduce a pipeline to infer sleep onset, duration, and quality from raw accelerometer data and then quantify the relationships between derived sleep metrics and other variables of interest.

Methods: The pipeline released here for the deep phenotyping of sleep, as the *DPSleep* software package, uses a stepwise algorithm to detect missing data; within-individual, minute-based, spectral power percentiles of activity; and iterative, forward-and-backward–sliding windows to estimate the major Sleep Episode onset and offset. Software modules allow for manual quality control adjustment of the derived sleep features and correction for time zone changes. In this paper, we have illustrated the pipeline with data from participants studied for more than 200 days each.

Results: Actigraphy-based measures of sleep duration were associated with self-reported sleep quality ratings. Simultaneous measures of smartphone use and GPS location data support the validity of the sleep timing inferences and reveal how phone measures of sleep timing can differ from actigraphy data.

Conclusions: We discuss the use of DPSleep in relation to other available sleep estimation approaches and provide example use cases that include multi-dimensional, deep longitudinal phenotyping, extended measurement of dynamics associated with mental illness, and the possibility of combining wearable actigraphy and personal electronic device data (eg, smartphones and tablets) to measure individual differences across a wide range of behavioral variations in health and disease. A new open-source pipeline for deep phenotyping of sleep, DPSleep, analyzes raw accelerometer data from wearable devices and estimates sleep onset and offset while allowing for manual quality control adjustments.

(JMIR Mhealth Uhealth 2021;9(10):e29849) doi:10.2196/29849

KEYWORDS

actigraphy; accelerometer; sleep; deep-phenotyping; smartphone; mobile phone

Introduction

Background and Challenge

Prolonged daily episodes of sleep behavior are expressed nearly ubiquitously in all members of our species, as they are innate and undergird both physical and mental health across the lifespan. Multiple studies have suggested that sleep loss or poor sleep quality are predictors (and potentially moderators and mediators) of mental illness symptoms and poor cognitive performance [1-6]. While the use of modern digital devices influences sleep timing, these devices also afford new low-burden opportunities to measure sleep [7-10]. This paper introduces an open-source sleep-analysis pipeline called DPSleep, referring to the deep phenotyping of sleep that we offer to the community as a platform to facilitate longitudinal studies of sleep using data from widely available wearable devices [11].

The current gold standard for documenting sleep timing and content is polysomnography (PSG), during which multiple physiological measures are recorded, usually in a clinic or laboratory setting [12,13] or recently in ambulatory settings [14,15]. Although PSG is a comprehensive assessment of sleep stages, there are limitations associated with cost and subject burden, and it is difficult to obtain in-patient or at-home versions of PSG for extended periods [16,17]. Actigraphy, defined as recording activity-related data, mainly acceleration, using wearable devices, has been suggested as an efficient and reliable alternative to measure certain features of sleep patterns in natural, at-home settings [18,19]. Actigraphy data, estimated from accelerometers, are a common output of many wearable and held devices, including wrist watches, ankle bands and wristbands, smartglasses, sewn-in or attached devices, and smartphones [15,20-24]. At the same time, openly available feature-extraction algorithms with the capability to retain and present the features from raw to derived measures are essential for a reproducible large-scale understanding of human sleep, even in the presence of proprietary algorithms associated with many of the devices.

Objective

In actigraphy-based sleep assessment using wristbands, acceleration is typically measured in 3 dimensions, where each axis—x, y, and z—reflects linear acceleration along with one dimension of a triaxial accelerometer; some devices also measure environmental factors such as ambient light, temperature, or physiological measures such as heart rate variability or electrodermal activity [25]. A body of literature investigating the sensitivity and specificity of wristbands to detect sleep parameters, including total sleep time, sleep onset latency, wake after sleep onset, and sleep efficiency, has evolved [26-28]. Although several of these approaches have been validated using PSG and self-reported sleep quality ratings under certain controlled conditions, there is ongoing research to validate the software and develop in-house algorithms for different applications and under real-world circumstances [29-31]. Several of the devices, in order to save memory and battery, provide preprocessed, 1-minute averaged acceleration data [6,32-35], whereas others provide continuous high-frequency data [36-38].

This study contributes to this evolving field by providing a comprehensive pipeline to analyze raw accelerometer data to estimate minute-based activity and detect the major Sleep Episode, defined as the longest continuous sleep episode of at least 100 minutes. The overall goals are to (1) develop an open-source processing pipeline to detect major Sleep Episodes from commonly available accelerometer data; (2) apply and validate the estimation procedure using real-world data, including individuals studied over extended periods who independently rated their sleep; and (3) apply the processing pipeline to exemplar data to illustrate its application. To this end, we aim to analyze data from undergraduate students who are studied over 6-9 months during college to illustrate sleep patterns that fluctuate with environmental demands and in relation to other self-report measures of sleep and mental state. We also aim to analyze data from 2 individuals who are outpatients with severe mental illness to illustrate the boundaries of the methods and their ability to measure dramatically altered sleep patterns.

Methods

Participants

Participants were enrolled into 2 distinct cohorts to obtain actigraphy data across a range of individuals. The samples are described separately.

Study 1: Undergraduate Study

In total, 6 undergraduate participants (all aged 19 years; 3 females; 4 White participants, 1 unspecified, and 1 Asian; all non-Hispanic) were recruited from a local private institution and participated for one academic year (165-268 days), including a buffer extending into the summer break. These individuals had successfully participated in a shorter, earlier pilot study that did not use the present actigraphy device or processing pipeline. Participants were compensated per hour for the lab visits and for completing daily app-based questionnaires and given milestone bonuses to encourage continued participation. Participants were required to be enrolled full-time in classes and own an iPhone or Android smartphone compatible with the study smartphone app, Beiwe, which is part of the open-source Beiwe platform for digital phenotyping [39]. The Beiwe app was configured to collect passive phone use, phone acceleration, and GPS data at an almost continuous rate, as well as active self-report data on a regular, daily basis. As each participant served as their own baseline, participants were not excluded for current or past psychiatric disorders or medication use. Mental health history was measured by self-report of current or past diagnoses of mental disorders, current or past use of prescribed psychotropic medications, and current or past concerns about mental health symptoms (undiagnosed). Only 1 participant reported any current or past mental health history (current psychotropic medication for anxiety). All study procedures were approved by the Institutional Review Board of Harvard University.

Study 2: Clinical Study

Two individuals (aged 62 and 24 years; 1 female) were recruited from an ongoing cohort following the clinical progression of severe mental illness at a local hospital. Individuals were diagnosed with psychotic disorders (bipolar, n=1; schizophrenia, n=1) using a structured clinical interview for Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV [40]. The diagnoses were initially acquired before DSM-V gaining traction. However, the primary diagnosis was reviewed periodically, and DSM-V-(revised version) [41] criteria were referenced when adjusting the diagnoses. Participant enrollment for this study targeted obtaining >1 year of data for each participant (duration: 543 and 309 days), which included a nearly continuous collection of smartphone and actigraphy data via the Beiwe platform [39] and wearable watch, respectively, from each participant. Participants were compensated for these data and for monthly in-person study visits during which clinical assessments were recorded to quantify disease progression using clinical gold standard measures. Milestone bonuses were provided to encourage continued participation. All study procedures were approved by the Institutional Review Board of Partners Healthcare.

Wrist Actigraphy and Ancillary Data Acquisition

The present pipeline was developed using triaxial acceleration data from a commercially available waterproof watch worn on the wrist (GENEActiv, Activinsights Ltd) and is intended as a general open resource for processing accelerometer data from any device that saves raw triaxial, high-frequency, continuous accelerometer data, sampled at a fixed and known rate. Missingness of data was assumed to occur completely at random. Data saved as minute-based or shorter activity estimates can also be accommodated. The frequency of data sampling was set to 30 Hz for study 1 and 20 Hz to preserve the memory in case the patients missed their study visits for study 2. Following the initial consent and receipt of the watch, individuals in studies 1 and 2 visited the lab every 4-5 weeks to return the watch and receive a new, fully charged watch with formatted memory; participants in study 2 were given the watches during their in-patient study visits. Participants were instructed to wear the watch continuously, including during sleep and while bathing. Using the same sampling rates across modalities of 30 Hz in study 1 and 20 Hz in study 2, the watch collected acceleration (g), light (lux), and ambient temperature (C). In addition, the wristband recorded key presses. Participants were instructed to press the key when they started to go to sleep and when they woke in the morning. The acceleration data are the primary data used for the automatic detection of episodes that would be scored as sleep with additional corrections from light and key press data when available and when necessary.

Data obtained from a smartphone (iPhone or Android) were used as an ancillary data source [39,42]. None of the automated processing or manual Sleep Episode adjustments used data from the smartphone. The smartphone data provided valuable independent information for validation. Individuals installed the research smartphone app, Beiwe, to collect active (questionnaire), passive phone use (via timestamping of lock-unlock events), accelerometer, and GPS data [39,43,44]. The GPS location of the phone was sampled every 10 minutes for a 2-minute duration, and phone acceleration was sampled 10 seconds on and 10 seconds off. At 5 PM each day, an in-app questionnaire appeared that asked about the quality of the previous night's sleep on a Likert scale (5-point Likert scale from 0 [exceptional] to 4 [terribly]), the amount of caffeine consumed in the previous 24 hours (5-point Likert scale from 0 [none] to 4 [five +]; Multimedia Appendix 1), and a set of questions about their mood and social and academic activities. Answers to these questions and passive estimates of phone use were used to validate the identification of the major Sleep Episodes from the independent watch actigraphy data.

DPSleep: A Processing Pipeline for Deep Phenotyping of Sleep

Raw Actigraphy Data and Removal of Missing Data (Wrist-Off)

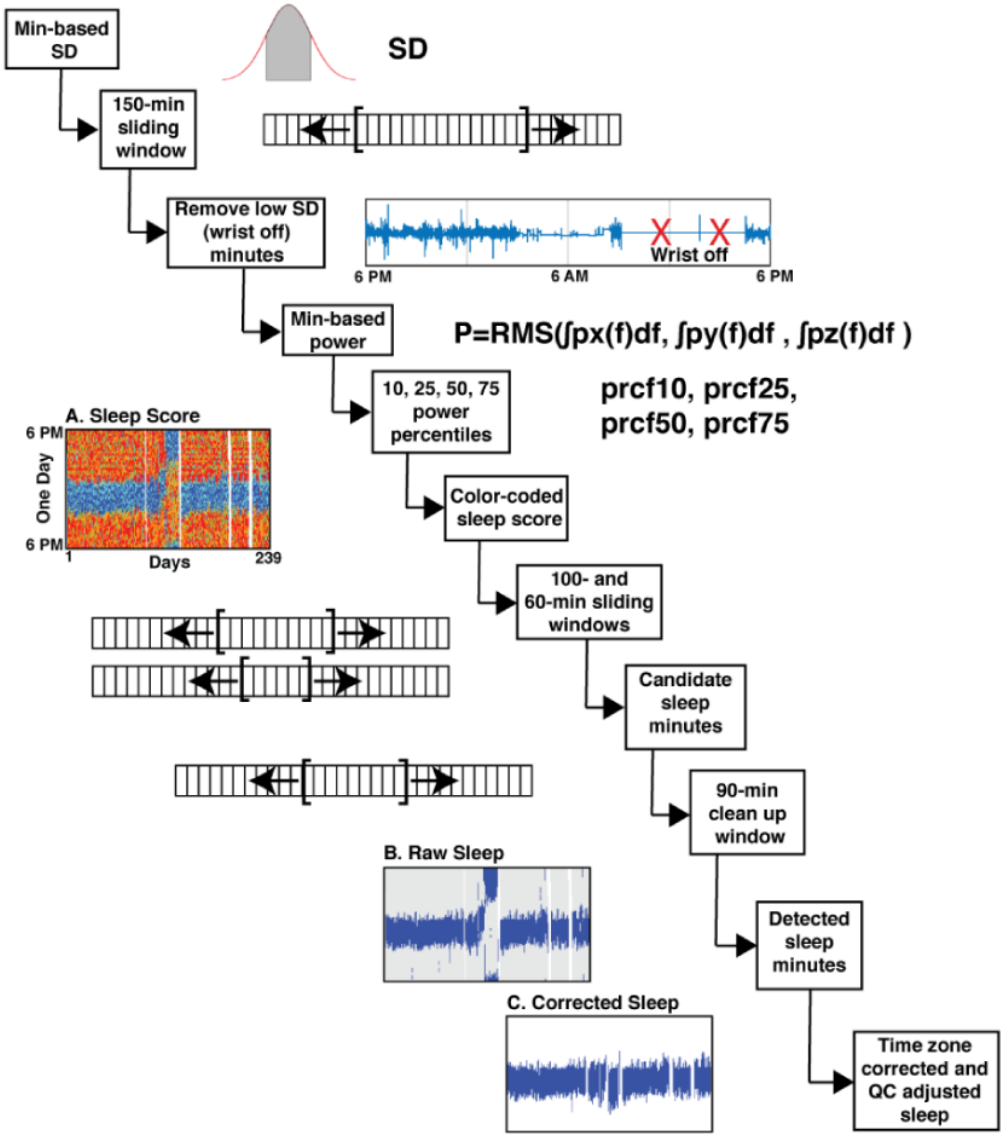
Raw wrist actigraphy data were originally saved as large, compressed files, each containing multiple weeks of data. Each file comprises a table with columns of acceleration, light, and temperature (depending on the device). To overcome the challenge of time-consuming access to specific rows during analysis, the first step in our processing pipeline is to parse and save the data into separate daily files from midnight to 11:59 PM. For days on which a watch change occurred (to allow continuous data sampling), a new, already charged watch was placed on the wrist. As the old watch keeps collecting nonwrist data until it is connected back to the data extraction station, the data from the new watch were formatted to overwrite any data from the previous watch at the same clock times.

Raw actigraphy data (30 Hz) are displayed for one daily data file illustrating each separate accelerometer trace (Figure S1 of Multimedia Appendix 2). This type of raw actigraphy data measures linear acceleration and is thus most sensitive to dynamic movements with varying amplitudes and frequencies during different activities such as walking, phone typing, or even tossing and turning in bed [45]. The amplitude of the fluctuations, measured in units of gravity (g), reflects the relative acceleration of the actigraphy device related to the gravity of the Earth. Each axis—x(g), y(g), and z(g)—reflects linear acceleration along with one dimension of a triaxial accelerometer. The raw data always include the Earth's gravity component, which can be reflected on different axes depending on the orientation of the device. The separate channels are highly correlated and, for our purposes here, provide redundant information that can be interrogated separately or combined to make estimates more robust. The amplitude of the accelerometer fluctuations varies across the awake hours but shows a stark reduction in fluctuation in all three axes during the sleep period. As explained below, the DPSleep processing pipeline is optimized to detect the reduction in fluctuations and estimate a single extended episode each day.

An immediate challenge is that, even for compliant participants with charged waterproof actigraphy watches, they occasionally removed their devices. Figure 1 shows an epoch in which the watch was off the wrist. Unlike the Sleep Episode, which contains residual periodic low levels of fluctuations, the

accelerometer traces are nearly flat during the wrist-off periods. The first step is to detect and remove the wrist-off minutes. As the focus of these analyses is to identify the major Sleep Episode, a window size of 150 minutes is used to detect the wrist-off minutes. The SD of the acceleration time series for each of the three axes was calculated for each minute. A forward-and-backward–moving window of SD for each axis was calculated for a window size of 150 minutes. Then, the root mean square (RMS) of the moving average SD values for the three axes is compared with a small experimental threshold of 0.0185 to detect the wrist-off minutes. This threshold was determined based on more than 600 hours (5 volunteers from our research group for 5-7 days each) of annotated data collected in-house. The minute of data is considered wrist-off if the RMS of the variance of the three channels is below this threshold. The remaining data were considered for further analyses.

Figure 1. The DPSleep processing pipeline. The sequential steps in the processing pipeline are illustrated with example data for key steps. The pipeline begins with raw accelerometer data and derives estimates of individualized sleep scores (A) and raw (B) and corrected Sleep Episodes (C). Sleep variables are calculated based on the corrected Sleep Episodes as estimated in C. Specifically, as the first step, the SD of the acceleration along three axes is averaged over 150-minute windows forward and backward to find the wrist-off minutes. The power density spectrum of the acceleration signal is calculated at different frequencies, and the area under the curve estimates power as the root mean squared integrated over the three axes. Minutes are classified based on 10, 25, 50, and 75 percentile thresholds and can be visualized (A; blue, cyan, green, orange, and red display the minutes based on increasing power scores). A series of forward and backward moving average windows is used to flag the candidate Sleep Episodes that are then further filtered to derive the raw sleep estimate (B). Quality control and adjustment for the local time zone then yield to the final corrected sleep estimate (C). QC: quality control; RMS: root mean square.



Scoring Activity Level for Each Minute

After removing the wrist-off minutes from the analysis, the power density spectrum of the acceleration signal in each minute was calculated using Welch formula [46]. The power of the signal in each minute is the area under the curve of the power density spectrum. Within-individual power thresholds were then determined for 10, 25, 50, and 75 percentiles. A 15-minute-wide

window of the average RMS power that combines across the three axes is used to find the percentile thresholds specific to each individual. Data from each minute of the study were classified based on their spectral power in comparison with these percentile cut-offs, and color-coded daily maps of the activity scores for each individual were generated (Figure 1, center-left). Highly active minutes are colored in red and orange, minutes with medium activity are colored in green, and low active minutes are colored in blue and cyan. Daily maps provide intuitive information about the sleep patterns of the individual. As the watch is recording continuous actigraphy without reference to the external world, when the time zone is changed due to traveling, the sleep pattern is shifted and will need to be accommodated at a later stage of processing, as can be seen in Figure 1 at approximately day 150. Standard or daylight saving time transitions require correction.

Estimating Major Sleep Episodes

To automatically estimate the timing of the Sleep Episode, we used multiple moving windows that slide over a weighted transformation of the minute-based activity levels. The minutes with less than 25% activity (25th percentile of the empirical distribution of activity for the person) are assigned a sleep score of 1, and those with higher than 50% activity are penalized by negative experimental scores (-0.75 and -1). Then, two moving windows of narrow (60 minutes) and wide (100 minutes) sizes were used to sweep the scores and find a provisional nocturnal episode of sleep, while the narrower window adjusts the beginning and end estimates of the major Sleep Episode. The separation of tasks between two window sizes was found, in pilot analyses, to better capture the distinct targeted events, where detecting the nocturnal Sleep Episode benefitted from the larger window, but the precision of the sleep onset and offset time estimates benefitted from the smaller window. A *clean-up* 90-minute wide moving window was then used to connect adjacent short candidate Sleep Episodes with more than three-quarter sleep-scored minutes in each 90-minute window. However, this process, on some occasions, left two separate candidate Sleep Episodes that were separated by a period of activity during the middle of the night. As a convention, the automated algorithm joined discontinuous Sleep Episodes into one longer episode if they fell within 22.5 minutes (a quarter of the 90-minute window) of one another. This is a decision of convention and occurred in 2.34% (34/1448) of the cases in which sleep was measured in healthy young adults.

The outcome of these steps is an estimate of a single provisional Sleep Episode for each day. The estimate, via the filtering approaches used, usually underestimates the full Sleep Episode duration by not including minutes on either temporal side of the sliding windows that have low activity. To mitigate this bias, as a final step, the initial estimate was expanded or shrunk to include all adjacent minutes that show less than 25% activity so long as they were after (when available) the evening button press (indicating the participant's intended start of attempting to sleep) and before their waking button press. A button press was considered available if the device successfully recorded a button press within 60 minutes of the estimate. The compliance of the individuals to provide informative (within 60 minutes of the estimate) button presses for sleep or bed ranged from 76.9% to 90.4% across the 6 participants in study 1 (mean 84.0%, SD 5.5%) and was 0.3% and 22.6% for the 2 participants in study 2. The duration of the major Sleep Episode after these corrections is recorded in the data output files as the automatically generated *SleepDuration* with its beginning (*SleepOnset*) and end (*SleepOffset*) times.

Figure S2 of Multimedia Appendix 2 displays examples of the initial automatic estimate of the Sleep Episode (yellow bottom bar) and the final corrected Sleep Episode (middle green bar) in the third row of panels A, B, and C. A third estimate, shown as a blue bar, expands from the final Sleep Episode to include adjacent minutes when the activity is below the 50% threshold, such as when individuals are resting in bed but not yet asleep. We store the beginning of the expanded epoch as the *BedrestOnset*, the end as the *BedrestOffset*, and the duration as the *BedrestDuration.* For some purposes, the time from the beginning of the *BedrestOnset* to *SleepOnset* can be used as a distinct measure (eg, *SleepOnsetLatency*). The variables are listed in Textbox 1. The estimated light exposure level from the watch is also shown in Figure S2 of Multimedia Appendix 2, as well as the timestamps of the recorded button presses. The light level is not used by the algorithm to estimate the Sleep Episode but is visualized because it can aid manual adjustments that are applied during quality control. An assumption of the present approach is that a single Sleep Episode will occur that is usually 100 minutes or more; the limitations of this simplifying assumption will be discussed.

Textbox 1. Sleep variables generated by the DPSleep pipeline.

Parameter name and description

- OffWrist: percentage of wrist-off time per 24 hours (%)
- SleepOnset: beginning of the Sleep Episode (hh:mm)
- SleepOffset: end of Sleep Episode (hh:mm)
- SleepDuration: difference between SleepOnset and SleepOffset (min)
- BedrestOnset: beginning of Bedrest Episode (hh:mm)
- BedrestOffset: end of Bedrest Episode (hh:mm)
- BedrestDuration: difference between BedrestOnset and BedrestOffset (min)
- SleepOnsetLatency: difference between BedrestOnset and SleepOnset (min)
- SleepEfficiency: percentage of SleepDuration or BedrestDuration (%)
- ActiveMinutes: number of minutes during Sleep Episode with activity higher than the 40th percentile
- ImmobileMinutes: number of minutes during Sleep Episode with activity lower than the 40th percentile
- ActiveBouts: number of bouts (sequences) during Sleep Episode with continuous ActiveMinutes, with 1-Min immobility tolerance
- SleepImmobility: percentage of immobile minutes or sleep duration (%)
- LightMinutes: number of minutes during Sleep Episode with light greater than 1 lux
- LightBouts: number of bouts during Sleep Episode with continuous LightMinutes, with 1-minute darkness tolerance
- PhoneMinutes: number of minutes during sleep episodes with any phone event, including locked, unlocked, or in use
- PhoneBouts: number of bouts during Sleep Episode with continuous PhoneMinutes, letting 1-minute no-event tolerance

Smartphone Data

In addition to wearing the actigraphy watch, most of the individuals in studies 1 and 2 installed the Beiwe app on their smartphones. The Beiwe app was configured to passively collect phone on or off times and the GPS location of the phone. An independent pipeline was used to securely analyze the GPS data and extract the places most visited by the individual during the study and estimate their major locations every 12 minutes. Then, a daily map was color-coded based on the presence of the individual at those points of interest. The GPS map provides the time zone of the locations where the subject has visited and evidence of location stability or movement around the time of the major Sleep Episode.

When participants traveled across time zones, a challenge arose as a matter of practice: when the new time zone was behind the old time zone, the data shifted back and overwrote the previously recorded data, and when the new time zone was ahead of the old one, the data shifted forward, and there was a missing data gap. The time shifts occurring during the actual days of travel, especially when travel occurs by plane, are challenging to incorporate, and we considered these days as missing data with the days before and after being retained with data shifted to reflect the time zone experienced by the participant. The participants in study 1 traveled two to four times during the course of the study, whereas the participants in study 2 did not travel. A similar time-shifting issue occurs for standard or daylight saving time transitions. In these cases, the time is shifted forward or backward accordingly, and the two transition nights are considered as missing data. Alternate goals, such as estimating circadian rhythms, may be better served by analyzing the data in a continuous fashion and will be different from the present focus, where the discrete daily patterns require these practical adjustments.

The smartphone data from Beiwe also provided relevant data for validation, including accelerometer and phone use data (via the recorded lock-unlock events). The DPSleep pipeline allows the integration and visualization of smartphone data when available. To integrate accelerometer data for the validation purposes of this study, a simple SD analysis was applied to the phone acceleration along the x-axis as a representative of the acceleration score to recognize minutes of high movement. The distribution of the acceleration score in all minutes during the study for each individual was used to find minutes with greater than 75 and 90 percentile movement (normalized to the individual). These minutes were color-coded in yellow and red, respectively, contrasting the lower acceleration minutes in gray and plotted in relation to the daily Sleep Episode estimates (Figure S3 of Multimedia Appendix 2). In addition, the locked-unlocked events of the phone document when the phone is in use. Phone in use time was defined as the time between every consecutive unlocked-locked event lasting no more than 15 minutes. This is to soften the strong assumption of the phone being in use all the time after the unlocked event and before it is locked again. The daily map is then color-coded to show the locked-unlocked minutes in red and blue, respectively, in addition to in-use minutes in green (Figure S3 of Multimedia Appendix 2). These data are used in this study to build confidence in the DPSleep estimates of the major Sleep Episodes. They may also be useful for understanding the relationship between digital technology use and sleep patterns, for example, as might occur if individuals use their phone sporadically at night.

Sleep Estimation Quality Control

DPSleep should not be expected to deliver a perfectly accurate output when operating in automatic mode. Several assumptions are made, and the structure of an individual sleep night can be complex. Instead, DPSleep provides an elaborative day-by-day report, examples of which appear in Figure S2 of Multimedia Appendix 2. The user can decide about the confidence of the estimated Sleep Episode and revise the results manually, if necessary. DPSleep includes editing tools. Every page of the report presents the data about one day from 6 PM on the previous day to 6 PM on the original day. The report also includes, when available, smartphone data, as illustrated in Figure S3 of Multimedia Appendix 2. The daily report is a significant help to the investigators to decide about and increase the precision of the sleep estimation results based on the availability and richness of the data.

All results in this paper have been quality-controlled by 2 individuals independently relying on only the watch actigraphy data and not any ancillary data from the smartphone to make modifications. Thus, the data are analyzed here, as would be from any typical study that only obtained watch actigraphy data. The guidelines used for quality control are described in Multimedia Appendix 3. Figure S4 of Multimedia Appendix 2 illustrates the plots of sleep duration across nights before and after manual adjustments for each of the 6 individuals in study 1. As shown, most nights show identical values before and after quality control, meaning no adjustment was required, whereas many others showed slight adjustments. For several nights, a large adjustment was required (eg, in Figure S4 P6 of Multimedia Appendix 2, there is an outlier value; in Figure S4 P5 of Multimedia Appendix 2 there are several values that were substantially corrected). The automated values showed a correlation with the final corrected values that ranged between r=0.92 and r=0.98. Figure S5 of Multimedia Appendix 2 illustrates examples of errors that require manual adjustments. Manual adjustments were made for 18.02% (261/1448) of the nights. As illustrated in Figure S4 of Multimedia Appendix 2, although approximately one in five nights were adjusted, most were small adjustments that would not impact most analyses. Approximately 9.05% (131/1448) of the nights were adjusted to change the major Sleep Episode estimate by greater than 1 hour.

Data Security

Throughout the analysis, the pipeline is designed to handle the data securely without exposing any identifiable information. Days are reported as the days of the study relative to the individual's consent days. We consider the GPS data identifiable not only when the actual coordinates are presented but also the patterns of the daily maps as presented with significant (for the individual) locations. Therefore, the pipeline was designed to work with the encrypted data and avoid saving any of the coordinates or maps as unencrypted scratch files. The final results are saved directly in an encrypted format.

Within-Individual Statistical Modeling

To analyze the longitudinal association between actigraphy-based sleep estimation and self-reported data, a simple individual-level linear model was used with self-reported sleep quality as a predictor and actigraphy-based sleep duration as the outcome. This simple linear model was selected after using a mixed linear model analysis to account for the time and autocorrelation of the observations. The categorical day of the week was included as a covariate to account for weekly structure related to course schedules and weekday-weekend differences. Only data collected during the academic year were included (ie, fall and spring semesters, including exam periods and Thanksgiving and spring breaks, but excluding winter and summer breaks) to account for differences between the school year and extended school breaks. All analyses were conducted in R (R Foundation for Statistical Computing) using the *stats* package *lm()* function [47].

Code Availability

The DPSleep pipeline software package is available [11] as an open-source package to be downloaded and used by the research community.

Results

Longitudinal Activity-Based Sleep Estimates Within Individuals

Longitudinal sleep patterns and daily sleep maps are the key outcomes of the DPSleep pipeline (Figure 1; see the *Methods* section). To show the performance of the pipeline, four examples of processed data are shown for individuals from study 1 (see the *Methods* section; Figures 2 and 3; Figure S6 and S7 of Multimedia Appendix 2). The data are from the full academic year with travel to and from campus and across daylight saving time. Each figure contains five plots, and three plots on the left are described here. Plot A is the color-coded daily map of activity scores that show the minutes with high (>50% orange; >75% red), low (<25% cyan; <10% blue), and medium (25%-50% green) activity. Plot B represents the raw estimated Sleep Episodes for each day. Plot C is the time zone–adjusted and quality control–adjusted longitudinal plot of sleep behavior. The third plot represents the data used for all the quantitative analyses and validations.

Figure 2. Longitudinal Sleep Episode estimates in P1 related to phone use. Three left panels display longitudinal activity score and Sleep Episode estimates for 243 days. Day 1 is near to the ninth day after the beginning of the semester. Winter break falls near days 102 to 133. Each vertical line displays data for a 24-hour day. (A) The top panel displays the continuous activity scores colored by the threshold for each day. For each day, the plot begins at 6 PM at the bottom and ends at 6 PM at the top, allowing the nighttime sleep period to plot in the center of the graph. Over days, the low activity band (in blue) is relatively consistent except for a dramatic shift at day 118, which is attributed to travel outside of the time zone. (B) The middle panel displays the same data revealing the automated detection of Sleep Episodes and nap periods. (C) The bottom panel C displays the time zone–adjusted and fully quality-controlled estimates of the final Sleep Episodes. On the right panels, the Sleep Episodes of panel C are plotted (light blue) in relation to independently estimated phone events. (D) The phone status is shown with colored hashing for when the phone is locked (red), unlocked (blue), and in use (green). Note that there is no phone use during the estimated Sleep Episodes. The phone use events, on many nights, occur up until and just before the beginning of the Sleep Episode. However, on most mornings, there is a gap between when the Sleep Episode ends and phone use begins, which often begins abruptly around 9 AM, possibly reflecting that phone use begins with an alarm-triggered event. (E) Phone acceleration data are plotted and also reveal no phone movement during the Sleep Episodes.

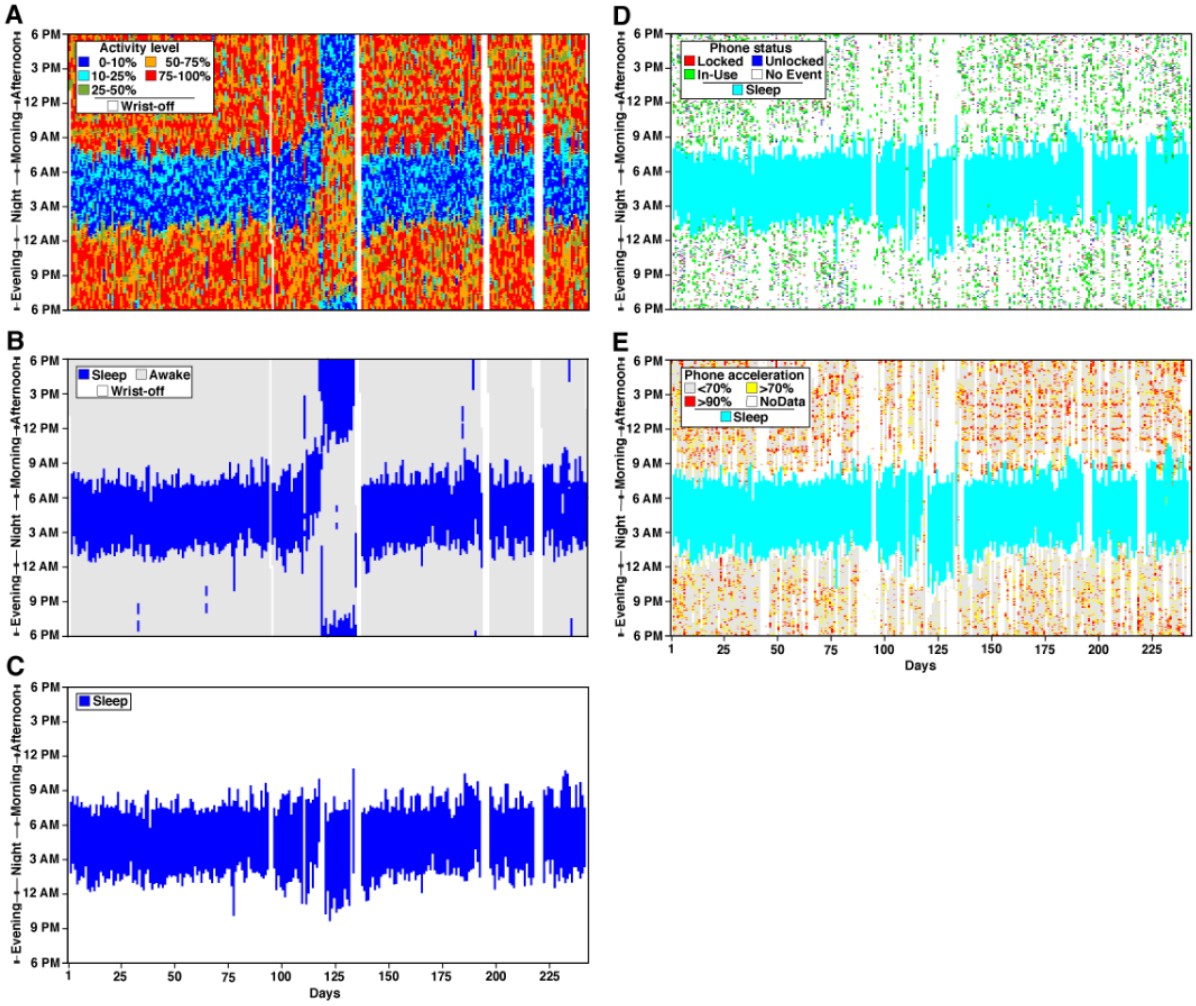


Figure 3. Longitudinal Sleep Episode estimates in P4 in relation to phone use. Longitudinal activity score, Sleep Episode estimates, and phone use data for 257 days are plotted. Day 2 is near to the beginning of the semester. Winter break falls near days 112 to 143. (A) The top-left panel shows the continuous activity scores colored by the threshold for each day. (B) The middle-left panel displays the same data revealing the automated detection of Sleep Episodes and nap periods. (C) The bottom-left panel demonstrates the time zone–adjusted and fully quality-controlled estimates of the final Sleep Episodes. On the right panels, the Sleep Episodes of panel C are plotted (light blue) in relation to independently estimated phone events. (D) The phone status is shown with colored hashing for when the phone is locked (red), unlocked (blue), and in use (green). (E) Phone acceleration data are plotted and also reveal no phone movement during the Sleep Episodes. This individual shows highly irregular Sleep Episodes. Note that the phone status and phone acceleration events track the irregular Sleep Episodes.

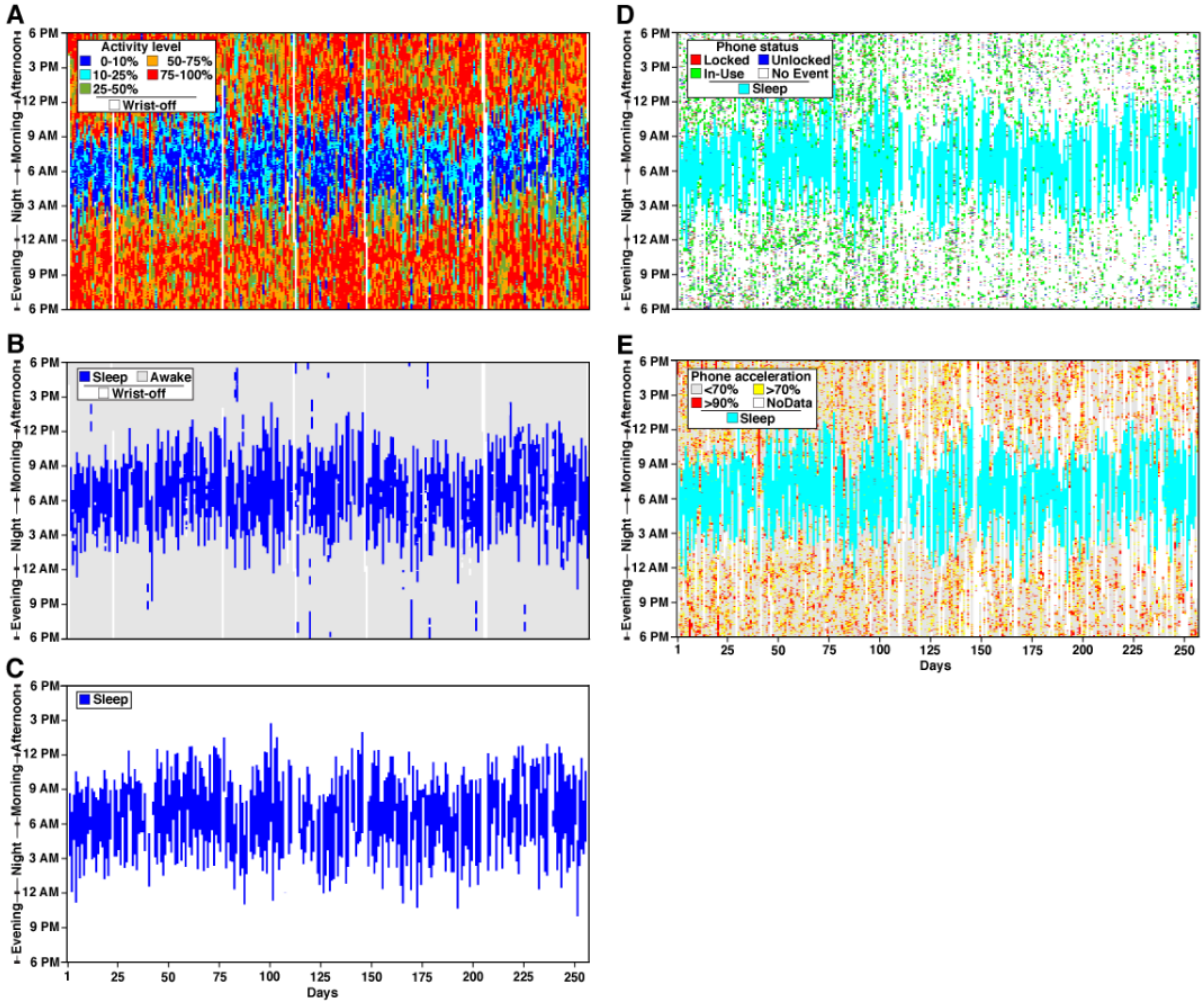


Figure 2 shows the sleep data for P1 with a highly regular sleep pattern where the subject goes to bed sometime between 1 AM and 2 AM and wakes up consistently near 8 AM. On weekends, the participant wakes up about an hour later than usual. Figure 3 shows the sleep data for P4, which displays the most irregular sleep patterns. There are some nights when this participant goes to bed at around midnight and does not wake up before 9 AM, and on other nights, the individual does not go to sleep before 6 AM and sleeps for only a few hours. The irregular sleep pattern does not appear to be related to the academic calendar because the participant displays a similar sleep pattern during breaks. In addition, the sleep data for P2 and P3 with a less regular sleep pattern compared with P1 are shown in Figure S6 and S7 of Multimedia Appendix 2. The effect of beginning the academic year is notable, with a gradual transition to a later sleep time (first 25 days). Break and travel also significantly affect the sleep schedule during days 125-150.

Smartphone Use Tracks Sleep Episodes

One way of validating the estimated Sleep Episode arises from the independent phone data collected simultaneously through the Beiwe app on each individual's cell phone. Even though these data do not continuously measure activity, as the phone can be put down, they do indicate the minutes during which the individual is clearly not sleeping. The right parts of Figures 2 and 3 and Figure S6 and S7 of Multimedia Appendix 2 plot the smartphone data for the 4 participants analyzed above. In each plot, panel D shows the phone status via locked-unlocked events. When the phone is not used, there should be no change in the phone status. Panel E displays phone acceleration that, similar to the wristwatch, provides a measure of dynamic movement when the phone is picked up, used, or is moving with the participant's body.

In all participants, the phone activity measures were generally outside the time of the estimated Sleep Episodes and tracked

the variations in sleep period from night to night. P2 and P3 (Figure S6 and S7 of Multimedia Appendix 2) show this quite clearly because their sleep patterns change gradually during the first 25 nights of measurement. Phone activity measures tracked these transitions. P4, who showed the most erratic sleep patterns, also demonstrated clear evidence that phone use was frequent and intensive only outside the time of the major Sleep Episodes (Figure 3). Beyond the general correspondence between phone and sleep, there were also interesting features in the details of the phone use that are relevant for comparing the activity- and phone-based data types.

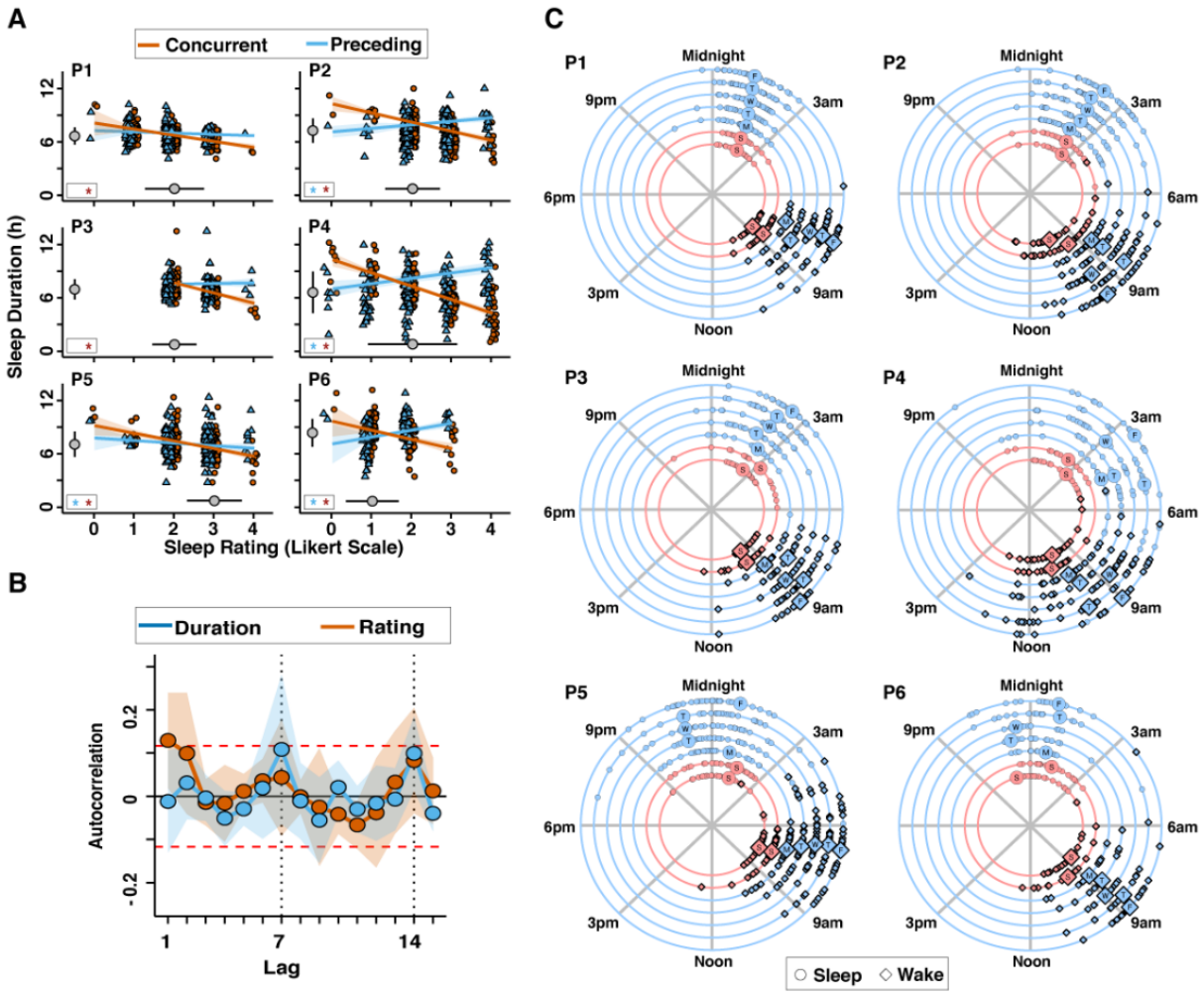
As an additional visualization, Figure S8 of Multimedia Appendix 2 shows data from P1 with the horizontal axis representing clock time (48 consecutive hours, with the second 24 hours on each horizontal line repeated in the first 24 hours of the next horizontal line) and day number down the y-axis. This is the plotting convention often used by investigators interested in circadian rhythms [48]. DPSleep allows plotting using either lateral (as in most figures in this paper) or horizontal (eg, Figure S8 of Multimedia Appendix 2) conventions.

Actigraphy Measures of Sleep Duration Track Self-report Sleep Quality

In each individual, the DPSleep pipeline yielded an estimate of sleep duration (*SleepDuration* in Textbox 1), the time difference between *SleepOnset* and *SleepOffset*. Each day, the individuals also reported the quality of the previous night's sleep on a Likert scale (Multimedia Appendix 3). Figure S9A of Multimedia Appendix 2 shows the variation in sleep duration across days in each of the 6 individuals from study 1, and their self-reported sleep rating. Missing data excluded from the analysis included data from winter break (the gaps near day 120) and noncompliant data (eg, the wristband was removed or the survey was submitted the following day).

An individual-level linear model analysis was used to evaluate the longitudinal association between self-reported sleep quality (for both concurrent night and nights preceding) as the predictor and actigraphy-based sleep duration as the outcome. There was a significant correlation between the concurrent night's self-reported sleep rating and the actigraphy-measured sleep duration in each of the 6 participants (Figure 4A): short sleep duration predicts self-reported ratings of poor sleep. The observation that sleep rating and sleep duration measures track one another in healthy young individuals provides evidence of validity. No correlation or a weakly positive correlation was observed for the preceding night.

Figure 4. Correlation, autocorrelation, and weekly patterns of sleep rating and sleep duration. (A) Individual-level linear model applied to each subject shows the association between self-report sleep rating and actigraphy-based sleep duration. In each case, the self-report sleep rating negatively predicted sleep duration when the sleep rating targeted the night of sleep (red line and circles). When the sleep rating was shifted to the next day so that the rating no longer matched the night of sleep duration, the relationship between sleep rating and sleep duration in most participants showed either no relation (P1 and P3) or a small positive relation, perhaps a form of sleep rebound (P2, P4, and P6; blue triangles and lines). P5 showed negative association for both nights. The larger gray circles to the left of each plot show the mean (SD) sleep duration across the study for each participant. The larger gray circles below each plot show the most frequent mode (SD) sleep rating across the study for each participant. The bottom-left box in each panel indicates which associations are significant (P<.05). (B) The autocorrelation, or the correlation of each variable's time course with itself at varying lags, is plotted for actigraphy-based sleep duration (blue) and self-report sleep rating (red) averaged across the 6 participants of study 1. Shading illustrates the SE of the mean. Increased autocorrelation values at lags of 7 and 14 days indicate that sleep duration and sleep rating for a given weekly night (eg, Friday or Saturday) are more similar to the same weekly nights on different weeks than to the immediately adjacent nights falling on different weekdays. Although autocorrelation is generally weak, there is a time-dependent structure within the data that should be considered when the data are modeled. (C) Sleep onset and offset (wake) times show differences depending on the day of the week. Each participant's sleep onset (circles) and offset (diamonds) times are displayed (large circles or diamonds =medians) on a 24-hour clock format. Between-subject sleep onset and offset times are variable, as well are weekday patterns within the same individuals. These structured patterns, which can vary from participant to participant, should be considered when the data are modeled.



Sleep Duration and Sleep Ratings Show Structured Weekly Variation

One of the outcomes of accurate estimation of major Sleep Episodes in individuals is the ability to investigate nonstationary effects that fluctuate with the weekly schedule, holidays, and academic calendar. An autocorrelation analysis was performed to evaluate the weekly variation. Actigraphy-based sleep duration showed almost no lag effect, such that one night's duration showed little association with the next (Figure 4B). Interestingly, there was a notable lag effect on days 7 and 14, suggesting a strong autocorrelation between the same days of the week. Self-report sleep rating shows similar lag day 7 and lag day 14 effects and a temporal autocorrelation at lag days 1 and 2, suggesting a poor or good sleep rating predicted a similar rating on subsequent nights despite no evidence of autocorrelation in the actigraphy-based sleep duration.

When data were analyzed by day of the week (Figure S9B of Multimedia Appendix 2), most participants showed relative stability in their sleep duration. We used a circle plot to show the sleep onset and offset distributions for each participant separately on weekdays (Figure 4C). Each point indicates the

beginning or end of a major Sleep Episode and their medians. Although there were participants with stable average sleep onset (P1) and offset (P6), the plots confirm different effects of weekdays for different participants, such as the effect of the weekend from Friday to Monday on late sleep in P5; the effect of Monday, Tuesday, and Thursday on very late sleep in P4; and the effect of Monday, Tuesday, and Thursday on earlier wake up in P2, that is, specific structured sleep patterns are highly idiosyncratic to individuals. The daily caffeine consumption was also assessed. Figure S10A of Multimedia Appendix 2 shows the variation in caffeine consumption ranging from 0 (none) to 4 (5+ drinks) for each of the 6 participants in study 1; Figure S10B of Multimedia Appendix 2 shows the weekly variation; Figure S11 of Multimedia Appendix 2 shows the individual-level linear model for each participant modeling the relationship between caffeine consumption and sleep duration. Although the effect is quantitatively small, 4 of the 6 participants show a statistically significant relationship such that more caffeine consumption predicts shorter sleep duration for the upcoming night.

Example Use Case: Sleep Patterns Show State Variation in Severe Mental Illness

To explore the feasibility and utility of sleep measures from extensive longitudinal assessments in patients, data from 2 individuals managing severe mental illness from study 2 were analyzed using the DPSleep pipeline. Individuals participated for 543 and 309 days, with 89.1% (484/543) and 59.2% (183/309) of completed days of data obtained after data loss due to missingness and quality control, respectively.

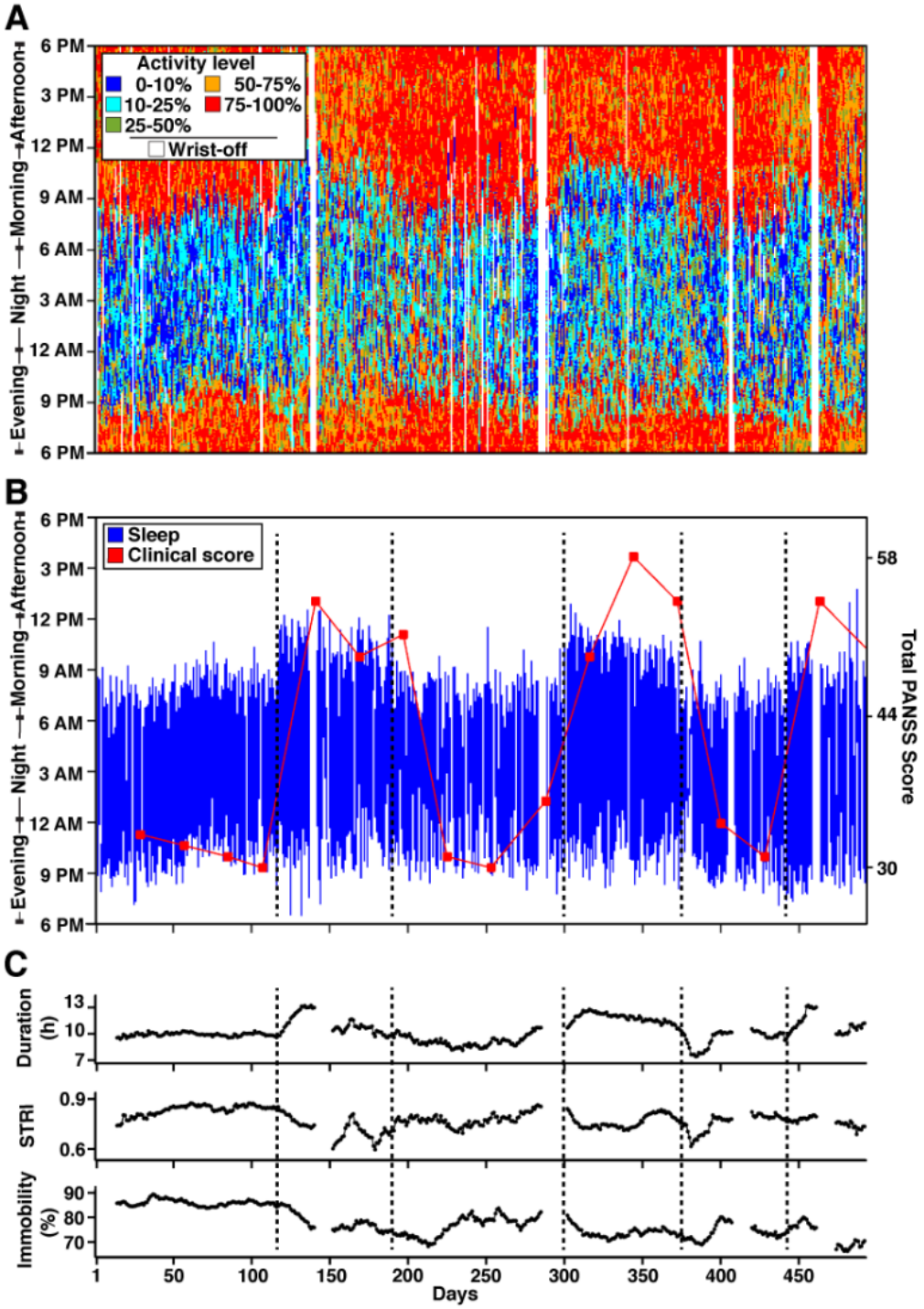
Figure 5 illustrates the data from P11, who is managing mood fluctuations originally diagnosed with bipolar disorder. Of interest are the slowly changing sleep patterns that can be immediately visualized in the activity score plots (Figure 5A). Two separate features of the data are interesting and require distinct measures for quantification. The first is that the shifts to low-level activity, indicative of sleep, begin earlier and end considerably later across two long episodes that begin near day 110 and day 300. The reduced activity scores extended until noon on many days. The change in sleep onset and sleep offset and increase in sleep duration was quantified with a 14-previous-day sliding window with less than three missing value tolerance in the DPSleep output shown in Figure 5C. The Sleep Timing Regularity Index (STRI) was also calculated as a 0-1 similarity index for 24-hour sleep and wake minutes of every day compared with an assumed day with the averaged sleep onset and offset. The STRI is a modified version of the Sleep Regularity Index (SRI) introduced by Phillips et al [2]. The SRI "calculates the percentage probability of an individual being in the same state (asleep vs awake) at any 2 minutes 24 hours apart", thus focusing on day-to-day, circadian fluctuations in sleep. The STRI, our modified version of the SRI, compares each study day with each participant's *average sleep day*.

Computing a participant’s average sleep day involves several steps to select the minutes from each day during which a participant is most frequently asleep such that the total duration is equal to the participant's average daily sleep duration. The daily STRI is then computed by comparing each daily Sleep Episode with the average Sleep Episode at the minute level and computing the proportion of minutes for which these two Sleep Episodes matched in sleep. This value, demonstrated in Figure 5C with a backward 14-day sliding window, shows noticeable drops that roughly occur at the beginning of the period of longer sleep intervals. Data with more than three missing values during the sliding window were considered missing.

What is further notable is that the major Sleep Episodes include higher activity score periods than during typical sleep, suggesting interrupted and inconsistent sleep. This feature is picked up in the derived measures of *SleepImmobility* percentage, defined in Textbox 1 as the percentage of *ImmobileMinutes* in the major Sleep Episode illustrated as similar sliding window values in Figure 5C. Adopting a previously described quantitative framing of Immobility [34], the *ImmobileMinutes* are defined in Textbox 1 as minutes with lower than a cut-off threshold (40th percentile activity here), which is the threshold to show the activity level in Figure S3 of Multimedia Appendix 2 (see the *Methods* section). To further illustrate the use of these data, the clinical severity of illness, measured using the Positive and Negative Syndrome Scale Total Score, is shown overlaid on top of the major Sleep Episode in Figure 5B. The periods of sleep disruption and extended sleep offsets correspond to the periods of high illness severity.

In addition, Figure S12 of Multimedia Appendix 2 illustrates data from P12, who lives with psychosis associated with schizophrenia. The activity scores displayed irregular and slowly drifting patterns, including a shift to low activity episodes that occurred later in the day around day 55 and then an earlier shift beginning near day 70 (Figure S12A of Multimedia Appendix 2). The period of most severe illness symptoms occurred near the most irregular sleep periods after day 75. It is to note that this individual has generally poorer clinical scores (Figure S12B of Multimedia Appendix 2) as compared with P11 and a generally lower STRI (Figure S12C of Multimedia Appendix 2). These data illustrate the complexity and richness of information that can be obtained through extensive longitudinal analysis of the actigraphy data and how different individuals can be from one another.

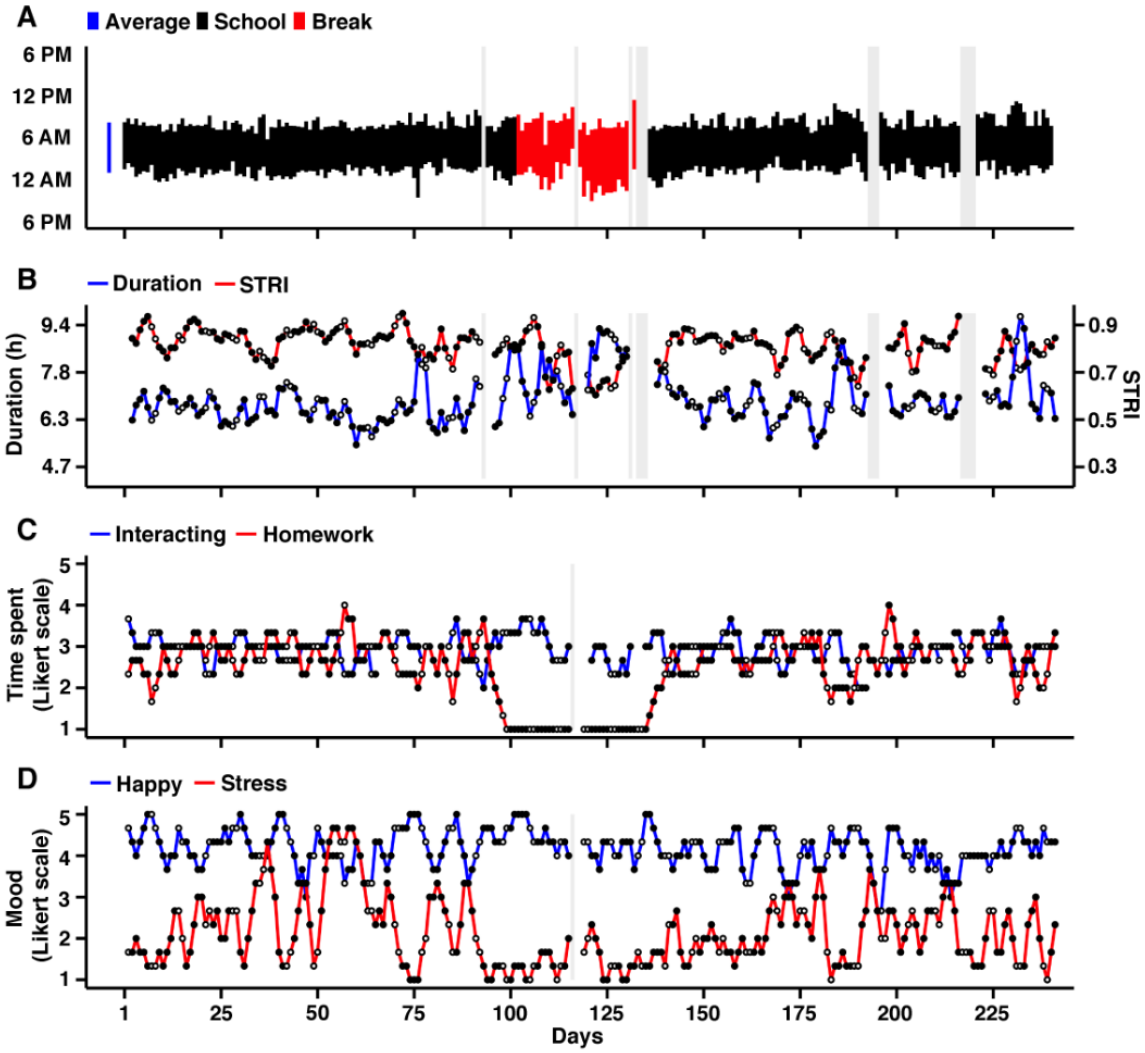
Figure 5. Example longitudinal sleep pattern over 500 days in a patient with severe mental illness. Three panels display longitudinal activity score, Sleep Episode estimates, and clinical severity score and quantitative metrics derived from the actigraphy for P11 of study 2. (A) The top panel displays the continuous activity scores colored by the threshold for each day. (B) The middle panel displays the time zone–adjusted and fully quality-controlled estimates of the Sleep Episodes with the severity of the clinical score overlaid by a red line. The clinical score reflects the total score on the Positive and Negative Syndrome Scale. (C) Temporally smoothed (14-day backward moving average) estimates of three activity-based measures are plotted: the estimated sleep duration (labeled Duration), the Sleep Timing Regularity Index, and the SleepImmobility percentage (labeled Immobility). Gaps in the plots reflect missing days; if more than 2 days were missing, the temporal average that would include those days is absent. Clear state changes in the sleep patterns can be observed (demarcated by dashed black lines in B) that are temporally coincident with negative changes in the clinical score. PANSS: Positive and Negative Syndrome Scale; STRI: Sleep Timing Regularity Index.



Example Use Case: Sleep Patterns Related to Deep Dynamic Phenotyping

To demonstrate how sleep measures can be combined with additional forms of digital phenotyping information, 2 individuals from study 1 are displayed in Figure 6 and Figure S13 of Multimedia Appendix 2. In each plot, the sleep patterns for each individual are illustrated through the course of an academic year, along with the quantified measures of sleep duration and STRI. In addition, self-report measures of social and academic behaviors (time on homework and time interaction) and mood (happy and stress) were displayed aligned to the estimates of the longitudinal Sleep Episodes. The examples illustrate clear dynamics of sleep, including weekend (empty circles) or weekday (filled circles) effects, changes between the active semester and breaks, and intermittent deviations for regular patterns. These data illustrate the potential of using low-burden wearables in combination with smartphone-based digital phenotyping to capture a great deal of information about life rhythms and changes due to environmental demands.

Figure 6. Example year in the life of college student P1. Four panels display multimodality longitudinal data from the full academic year in a college undergraduate (P1 of study 1). (A) Actigraphy-based estimates of the Sleep Episode are displayed, colored by academic period (black=academic school year; red=winter break). Missing data with gray backgrounds reflect missing (wrist-off) or quality removed data (eg, on a travel day across time zones). The mean Sleep Episode is shown to the left in blue. (B) Temporally smoothed (3-day backward moving average with no missing tolerance) estimated of the sleep duration (Duration) and Sleep Timing Regularity Index. (C) Temporally smoothed estimates of time spent interacting and time spent doing homework from the self-report questionnaire. (D) Self-report estimates of mood, including happiness (Happy) and stress (Stress). This individual displays stable and regular sleep patterns with periodic deviations. Note the deviations in time spent interacting and time spent doing homework during the break period. STRI: Sleep Timing Regularity Index.



An Example of Limitations

In showing examples of the utility of the DPSleep processing pipeline and example apps, we wanted to also show, as the last point, a clear example of a limitation. The present pipeline and quality control adjustments make assumptions that have been selected because they work most of the time in the range of contexts for which they were tested. However, real-life situations are complex. Figure S14 of Multimedia Appendix 2 shows an interesting example. In this example, the daily data from Figure

S2 of Multimedia Appendix 2 are replotted along with phone use and GPS location data from the same participant of the third example. What is notable is that our original estimate, based only on watch actigraphy data and the button press, combined with the rule used to join episodes of low activity as a continuous major Sleep Episode, likely mischaracterizes this night's true continuous Sleep Episode. The phone data reveal that the nighttime period of activity is followed by missing GPS data, intermittent phone use, and eventual arrival at a new location. This individual most likely woke up early and got on a bus or train. Our automated estimation procedures, as well as our quality control, using only watch actigraphy, mischaracterizes this Sleep Episode.

Discussion

Principal Findings

This work describes and demonstrates the utility of an open-source, longitudinal sleep-analysis platform called DPSleep. The platform was applied to two cohorts of participants that possessed extended data over months to years and included clinically healthy undergraduates and outpatients living with severe mental illness. The goal of these diverse explorations was to validate the approach and demonstrate its utility across multiple real-world participant cohorts. The results revealed that the approach captured the major Sleep Episode and detected dynamic patterns of sleep behavior, including in individuals presenting with episodic clinical illness (eg, Figure 5 and Figure S12 of Multimedia Appendix 2).

Several previous studies have investigated the validity and accuracy of wrist-worn actigraphy devices to estimate sleep parameters and compare them against gold standard PSG-derived measures and self-reported sleep quality [49]. DPSleep uses frequency-based analysis and begins with high-frequency raw data from the wrist-worn accelerometer. The high precision of the sampling frequency and the continuity of the collected data in our samples provided ideal data sets for frequency-based analysis, a robust and efficient tool to analyze the activity of the individual that complements alternative procedures such as zero crossing mode, time above threshold, or digital integration mode [19]. Frequency-based analysis inherently and automatically discounts the alternating gravity effect on different axes, without the need to eliminate the amplitude of the signal [50]; at the same time, it disentangles the natural high-frequency shakes of the body from the actual repositioning dynamic, which is the focus of sleep research. As explained earlier, the raw accelerometer data contain the gravity component offset that can be systematically removed in the frequency analysis.

A challenge with our approach was related to participant burden, as the participants had to visit the lab every 4-5 weeks to get a refreshed battery and download data. An alternative solution to attenuate the burden on subjects is to use wireless data transfer and cloud storage; however, this approach has its own limitations, based on the wireless storage size and battery usage in addition to data security concerns. We expect advances in widely available commercial technologies to gradually alleviate existing challenges. Another challenge was the time changes caused by travel or standard or daylight saving switches during the study. To reduce the uncertainty around participants' true sleep behavior when the time was shifted, days with time transitions were removed and considered missing data, although future steps could leverage the phone-based GPS data to exclude the relocating epochs from sleep and adjust the data on those days with more confidence.

In our study, self-reported sleep quality showed a strong correlation with the activity-based sleep duration measure; therefore, the focus of our sleep algorithm development was mainly to optimize this parameter, while building a platform to explore other relevant sleep-related parameters in future work. Despite existing sleep detection algorithms that look for 5-15 low activity minutes [50,51], DPSleep starts with a structural analysis of the daily activity to first detect the large episode of the day with the lowest average activity, and then the edges of the episode are adjusted using smaller moving windows and heuristic rules. This approach is an efficient solution to eliminate the need for any kind of sleep diary or ambient light assumptions because the former is not usually available and accurate, and the latter could be misleading because the wristband can be blocked by a long sleeve at any time of the day. Moreover, this approach distinguishes short naps or inactive periods during the day from the long, continuous, and disrupted Sleep Episodes without any assumptions about sleep time. Structural analysis of the activity scores during the whole study, using the individual's statistics and moving average windows with different sizes, suggests the most likely episode for the individual's sleep during the day. In addition to moving average windows to automatically detect the major Sleep Episode, DPSleep provides a helpful tool for investigators to decide, with reasonable confidence, about idiosyncratic sleep behaviors such as no sleep or very short Sleep Episodes. Additional button presses, if available, and adjacent activities are used to tune this episode and connect smaller pieces to shape the entire Sleep Episode.

Caveats and Limitations

As illustrated in Figure S14 of Multimedia Appendix 2, our approach can make mistakes. Although we illustrate how inexpensive and easy-to-obtain actigraphy data can be analyzed to estimate the major Sleep Episodes and dynamic patterns over many days, the real world is messy. Atypical behavioral patterns (eg, many short naps without a clear extended primary Sleep Episode) and behaviors that yield extreme accelerometer readings are challenging for our approach. Specifically, using our methods, actigraphy-based sleep detection is not appropriate for measuring sleep when an individual is on a shaking platform such as a plane, train, or bus. This is an unavoidable situation in longitudinal studies; however, as the on-plane sleep effect was not the focus of our study, and those days were very few compared with the days of the entire study, we were able to detect those days using GPS data and eliminate them from the results. To solve the continuous shaking challenge in similar situations or in studies on individuals with Parkinson disorder, sleep detection devices based on ambulatory circadian monitoring are recommended [28]. More broadly, as actigraphy research grows and repositories of annotated data from common and less common activities become available, machine learning

techniques can be leveraged to further refine the present approach (and others) handle a wider variety of situations. The validation and exploration of DPSleep, as illustrated in these initial explorations, provides a tool that can be used today and continues to be refined and expanded through its open-source release. Future directions include examining whether the association between mood and actigraphy-based sleep and activities can serve as a biomarker of psychiatric illness, and eventually use these objective measures to predict clinical patients' mood fluctuations.

Acknowledgments

The authors thank Laura Farfel, Marisa Marotta, Erin Phlegar, Lauren DiNicola, Arpi Youssoufian, Elyssa Barrick, Nora Mueller, Crystal Blankenbaker, and Joanna Tao for their help in collecting data and Elizabeth Klerman for valuable comments and editing of the manuscript. They thank Timothy O'Keefe, Harris Hoke, Lily Jeong, and Sarah Guthrie who provided valuable assistance in neuroinformatics support. The authors acknowledge Kenzie W Carlson who helped with the Beiwe component of the studies and Katherine Miclau, Amira Song, Emily Iannazzi, and Hannah Becker who helped with the quality control procedure. They thank Patrick Mair for providing valuable consultation on the statistical approaches. This study was funded by a generous gift from Kent and Liz Dauten, National Institute of Mental Health grants P50MH106435 and U01MH116925, and National Institutes of Health award DP2MH103909.

Authors' Contributions

HRE, RLB, and JTB created the pipeline. GC, JTB, and RLB designed the study. HRE, GC, CMVB, JTB, and RLB analyzed the data and interpreted the experiments. GC and JTB collected the data. HRE, GC, CMVB, and RLB wrote the manuscript. JPO and JTB critically commented on the manuscript.

Conflicts of Interest

JPO is a cofounder and board member of a commercial entity, established in 2020, which operates in digital phenotyping. JTB has received consulting fees and equity from Mindstrong Health, which are unrelated to the current paper.

Multimedia Appendix 1
Self-reported questionnaire.
[DOCX File , 23 KB - mhealth_v9i10e29849_app1.docx]

Multimedia Appendix 2
Extended figures.
[DOCX File , 26376 KB - mhealth_v9i10e29849_app2.docx]

Multimedia Appendix 3
Quality control instruction.
[DOCX File , 25 KB - mhealth_v9i10e29849_app3.docx]

References

1. Maglione J, Ancoli-Israel S, Peters K, Paudel ML, Yaffe K, Ensrud KE, Study of Osteoporotic Fractures Research Group. Subjective and objective sleep disturbance and longitudinal risk of depression in a cohort of older women. Sleep 2014 Jul 01;37(7):1179-1187 [FREE Full text] [doi: 10.5665/sleep.3834] [Medline: 25061246]
2. Phillips AJ, Clerx WM, O'Brien CS, Sano A, Barger LK, Picard RW, et al. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. Sci Rep 2017 Jun 12;7(1):3216 [FREE Full text] [doi: 10.1038/s41598-017-03171-4] [Medline: 28607474]
3. Blake MJ, Trinder JA, Allen NB. Mechanisms underlying the association between insomnia, anxiety, and depression in adolescence: implications for behavioral sleep interventions. Clin Psychol Rev 2018 Jul;63:25-40. [doi: 10.1016/j.cpr.2018.05.006] [Medline: 29879564]
4. Snyder E, Cai B, DeMuro C, Morrison MF, Ball W. A new single-item sleep quality scale: results of psychometric evaluation in patients with chronic primary insomnia and depression. J Clin Sleep Med 2018 Nov 15;14(11):1849-1857 [FREE Full text] [doi: 10.5664/jcsm.7478] [Medline: 30373688]
5. André C, Tomadesso C, de Flores R, Branger P, Rehel S, Mézenge F, et al. Brain and cognitive correlates of sleep fragmentation in elderly subjects with and without cognitive deficits. Alzheimers Dement (Amst) 2019 Feb 8;11:142-150 [FREE Full text] [doi: 10.1016/j.dadm.2018.12.009] [Medline: 30788411]
6. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. JAMA Psychiatry 2019 Feb 01;76(2):190-198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]

7. Adams S, Daly J, Williford D. Adolescent sleep and cellular phone use: recent trends and implications for research. Health Serv Insights 2013 Oct 3;6:99-103 [FREE Full text] [doi: 10.4137/HSI.S11083] [Medline: 25114565]
8. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92 [FREE Full text] [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
9. Cabré-Riera A, Torrent M, Donaire-Gonzalez D, Vrijheid M, Cardis E, Guxens M. Telecommunication devices use, screen time and sleep in adolescents. Environ Res 2019 Apr;171:341-347. [doi: 10.1016/j.envres.2018.10.036] [Medline: 30716511]
10. Tashjian SM, Mullins JL, Galván A. Bedtime autonomy and cellphone use influence sleep duration in adolescent. J Adolesc Health 2019 Jan;64(1):124-130. [doi: 10.1016/j.jadohealth.2018.07.018] [Medline: 30366713]
11. Rahimi-Eichi HR, Buckner BL, Baker JT. Deep Phenotyping of Sleep (DPSleep) processing pipeline. GitHub. 2021. URL: https://github.com/harvard-nrg/dpsleep.git [accessed 2021-09-23]
12. Sadeh A. Sleep assessment methods. In: Sleep and Development: Familial and Socio-Cultural Considerations. Oxford, United Kingdom: Oxford Scholarship Online; 2011.
13. Hirshkowitz M. Polysomnography and beyond. In: Principles and Practice of Sleep Medicine. Amsterdam, Netherlands: Elsevier; Jan 2017.
14. Fonareva I, Amen AM, Zajdel DP, Ellingson RM, Oken BS. Assessing sleep architecture in dementia caregivers at home using an ambulatory polysomnographic system. J Geriatr Psychiatry Neurol 2011 Mar;24(1):50-59 [FREE Full text] [doi: 10.1177/0891988710397548] [Medline: 21320949]
15. Smith C, Galland B, Taylor R, Meredith-Jones K. ActiGraph GT3X+ and Actical Wrist and Hip Worn Accelerometers for sleep and wake indices in young children using an automated algorithm: validation with polysomnography. Front Psychiatry 2019 Jan 14;10:958 [FREE Full text] [doi: 10.3389/fpsyt.2019.00958] [Medline: 31992999]
16. Newell J, Mairesse O, Verbanck P, Neu D. Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. Psychiatry Res 2012 Dec 30;200(2-3):795-801. [doi: 10.1016/j.psychres.2012.07.045] [Medline: 22901399]
17. Kaplan K, Hirshman J, Hernandez B, Stefanick ML, Hoffman AR, Redline S, Osteoporotic Fractures in Men (MrOS), Study of Osteoporotic Fractures SOF Research Groups. When a gold standard isn't so golden: lack of prediction of subjective sleep quality from sleep polysomnography. Biol Psychol 2017 Feb;123:37-46 [FREE Full text] [doi: 10.1016/j.biopsycho.2016.11.010] [Medline: 27889439]
18. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011 Aug;15(4):259-267. [doi: 10.1016/j.smrv.2010.10.001] [Medline: 21237680]
19. Stone K, Ancoli-Israel S. Actigraphy. In: Principles and Practice of Sleep Medicine. Amsterdam, Netherlands: Elsevier; 2017.
20. Schenk A, Witbrodt B, Hoarty C, Carlson RH, Goulding EH, Potter JF, et al. Cellular telephones measure activity and lifespace in community-dwelling adults: proof of principle. J Am Geriatr Soc 2011 Feb;59(2):345-352 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03267.x] [Medline: 21288235]
21. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
22. Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 2017 Oct 16;3(1):37 [FREE Full text] [doi: 10.1038/s41537-017-0038-0] [Medline: 29038553]
23. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 01;312(3):358-367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
24. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. Sleep Med Rev 2020 Feb;49:101227. [doi: 10.1016/j.smrv.2019.101227] [Medline: 31901524]
25. Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. Metabolism 2018 Jul;84:99-108. [doi: 10.1016/j.metabol.2017.10.008] [Medline: 29080814]
26. Soric M, Turkalj M, Kucic D, Marusic I, Plavec D, Misigoj-Durakovic M. Validation of a multi-sensor activity monitor for assessing sleep in children and adolescents. Sleep Med 2013 Feb;14(2):201-205. [doi: 10.1016/j.sleep.2012.11.003] [Medline: 23238268]
27. Shin M, Swan P, Chow CM. The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions. Sleep Sci 2015;8(1):9-15 [FREE Full text] [doi: 10.1016/j.slsci.2015.02.003] [Medline: 26483937]
28. Madrid-Navarro C, Cuesta FJ, Escamilla-Sevilla F, Campos M, Ruiz Abellán F, Rol MA, et al. Validation of a device for the ambulatory monitoring of sleep patterns: a pilot study on Parkinson's disease. Front Neurol 2019 Apr 11;10:356 [FREE Full text] [doi: 10.3389/fneur.2019.00356] [Medline: 31031690]

29. Cellini N, McDevitt EA, Mednick SC, Buman MP. Free-living cross-comparison of two wearable monitors for sleep and physical activity in healthy young adults. Physiol Behav 2016 Apr 01;157:79-86 [FREE Full text] [doi: 10.1016/j.physbeh.2016.01.034] [Medline: 26821185]
30. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
31. John D, Tang Q, Albinali F, Intille S. An open-source monitor-independent movement summary for accelerometer data processing. J Meas Phys Behav 2019 Dec;2(4):268-281 [FREE Full text] [doi: 10.1123/jmpb.2018-0068] [Medline: 34308270]
32. Yin J, Julius A, Wen JT, Oishi MM, Brown LK. Actigraphy-based parameter tuning process for adaptive notch filter and circadian phase shift estimation. Chronobiol Int 2020 Nov;37(11):1552-1564. [doi: 10.1080/07420528.2020.1805460] [Medline: 32867527]
33. Bakken L, Kim H, Finset A, Lerdal A. Subjective sleep quality in relation to objective sleep estimates: comparison, gender differences and changes between the acute phase and the six-month follow-up after stroke. J Adv Nurs 2014 Mar;70(3):639-650. [doi: 10.1111/jan.12228] [Medline: 23967876]
34. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. Sleep 2015 Aug 01;38(8):1323-1330 [FREE Full text] [doi: 10.5665/sleep.4918] [Medline: 26118555]
35. Bellone GJ, Plano SA, Cardinali DP, Chada DP, Vigo DE, Golombek DA. Comparative analysis of actigraphy performance in healthy young subjects. Sleep Sci 2016;9(4):272-279 [FREE Full text] [doi: 10.1016/j.slsci.2016.05.004] [Medline: 28154740]
36. Dillon C, Fitzgerald A, Kearney P, Perry IJ, Rennie KL, Kozarski R, et al. Number of days required to estimate habitual activity using wrist-worn GENEActiv accelerometer: a cross-sectional study. PLoS One 2016 May 5;11(5):e0109913 [FREE Full text] [doi: 10.1371/journal.pone.0109913] [Medline: 27149674]
37. Crowley P, Skotte J, Stamatakis E, Hamer M, Aadahl M, Stevens ML, et al. Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: a proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). Int J Behav Nutr Phys Act 2019 Aug 16;16(1):65 [FREE Full text] [doi: 10.1186/s12966-019-0835-0] [Medline: 31419998]
38. Jones SE, van Hees VT, Mazzotti DR, Marques-Vidal P, Sabia S, van der Spek A, et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. Nat Commun 2019 Apr 05;10(1):1585 [FREE Full text] [doi: 10.1038/s41467-019-09576-1] [Medline: 30952852]
39. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
40. First M. User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders SCID-I: Clinician Version. Washington, DC: American Psychiatric Press; 1997.
41. First M, Williams J, Karg R, Spitzer R. The Structured Clinical Interview for DSM-5. American Psychiatric Association. URL: https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5 [accessed 2021-08-25]
42. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
43. Torous J, Firth J, Mueller N, Onnela JP, Baker JT. Methodology and reporting of mobile heath and smartphone application studies for schizophrenia. Harv Rev Psychiatry 2017;25(3):146-154 [FREE Full text] [doi: 10.1097/HRP.0000000000000133] [Medline: 28234658]
44. Barnett I, Onnela J. Inferring mobility measures from GPS traces with missing data. Biostatistics 2020 Apr 01;21(2):98-112 [FREE Full text] [doi: 10.1093/biostatistics/kxy059] [Medline: 30371736]
45. Farrahi V, Niemelä M, Kangas M, Korpelainen R, Jämsä T. Calibration and validation of accelerometer-based activity monitors: a systematic review of machine-learning approaches. Gait Posture 2019 Feb;68:285-299 [FREE Full text] [doi: 10.1016/j.gaitpost.2018.12.003] [Medline: 30579037]
46. Iwata H, Umebayashi K, Lehtomaki J, Narieda S. Welch FFT segment size selection method for FFT based wide band spectrum measurement. IEICE Trans Commun 2018 Jan;E101.B(7):1733-1743. [doi: 10.1587/transcom.2017ebp3069]
47. R: A language environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL: https://www.R-project.org/ [accessed 2021-09-20]
48. Formentin C, Garrido M, Montagnese S. Assessment and management of sleep disturbance in cirrhosis. Curr Hepatol Rep 2018;17(1):52-69 [FREE Full text] [doi: 10.1007/s11901-018-0390-1] [Medline: 29876197]
49. Khademi A, El-Manzalawy Y, Master L, Buxton OM, Honavar VG. Personalized sleep parameters estimation from actigraphy: a machine learning approach. Nat Sci Sleep 2019 Dec 11;11:387-399 [FREE Full text] [doi: 10.2147/NSS.S220716] [Medline: 31849551]

50. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One 2015 Nov 16;10(11):e0142533 [FREE Full text] [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]
51. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 2018 Aug 28;8(1):12975 [FREE Full text] [doi: 10.1038/s41598-018-31266-z] [Medline: 30154500]

Abbreviations

DSM: Diagnostic and Statistical Manual of Mental Disorders
PSG: polysomnography
RMS: root mean square
SRI: Sleep Regularity Index
STRI: Sleep Timing Regularity Index

Edited by G Eysenbach; submitted 28.04.21; peer-reviewed by C Ralph-Nearman; comments to author 19.05.21; revised version received 17.06.21; accepted 02.08.21; published 06.10.21.

Please cite as:
Rahimi-Eichi H, Coombs III G, Vidal Bustamante CM, Onnela JP, Baker JT, Buckner RL
Open-source Longitudinal Sleep Analysis From Accelerometer Data (DPSleep): Algorithm Development and Validation
JMIR Mhealth Uhealth 2021;9(10):e29849
URL: https://mhealth.jmir.org/2021/10/e29849
doi:10.2196/29849
PMID:34612831

Original Paper

Automatic Mobile Health Arrhythmia Monitoring for the Detection of Atrial Fibrillation: Prospective Feasibility, Accuracy, and User Experience Study

Onni E Santala[1,2], MD; Jari Halonen[1,3], MD, PhD; Susanna Martikainen[4], PhD; Helena Jäntti[5], MD, PhD; Tuomas T Rissanen[6], MD, PhD; Mika P Tarvainen[7,8], PhD; Tomi P Laitinen[1,8], MD, Prof Dr; Tiina M Laitinen[8], PhD; Eemu-Samuli Väliaho[1,2], BM, BHC; Juha E K Hartikainen[1,3], MD, Prof Dr; Tero J Martikainen[9*], MD, PhD; Jukka A Lipponen[7*], PhD

[1]School of Medicine, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland
[2]Doctoral School, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland
[3]Heart Center, Kuopio University Hospital, Kuopio, Finland
[4]Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland
[5]Center for Prehospital Emergency Care, Kuopio University Hospital, Kuopio, Finland
[6]Heart Center, North Karelia Central Hospital, Joensuu, Finland
[7]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland
[8]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland
[9]Department of Emergency Care, Kuopio University Hospital, Kuopio, Finland
[*]these authors contributed equally

Corresponding Author:
Onni E Santala, MD
School of Medicine
Faculty of Health Sciences
University of Eastern Finland
Yliopistonranta 1
P O BOX 1627
Kuopio, FI-70211
Finland
Phone: 358 503010879
Email: elmeris@uef.fi

Abstract

Background: Atrial fibrillation (AF) is the most common tachyarrhythmia and associated with a risk of stroke. The detection and diagnosis of AF represent a major clinical challenge due to AF's asymptomatic and intermittent nature. Novel consumer-grade mobile health (mHealth) products with automatic arrhythmia detection could be an option for long-term electrocardiogram (ECG)-based rhythm monitoring and AF detection.

Objective: We evaluated the feasibility and accuracy of a wearable automated mHealth arrhythmia monitoring system, including a consumer-grade, single-lead heart rate belt ECG device (heart belt), a mobile phone application, and a cloud service with an artificial intelligence (AI) arrhythmia detection algorithm for AF detection. The specific aim of this proof-of-concept study was to test the feasibility of the entire sequence of operations from ECG recording to AI arrhythmia analysis and ultimately to final AF detection.

Methods: Patients (n=159) with an AF (n=73) or sinus rhythm (n=86) were recruited from the emergency department. A single-lead heart belt ECG was recorded for 24 hours. Simultaneously registered 3-lead ECGs (Holter) served as the gold standard for the final rhythm diagnostics and as a reference device in a user experience survey with patients over 65 years of age (high-risk group).

Results: The heart belt provided a high-quality ECG recording for visual interpretation resulting in 100% accuracy, sensitivity, and specificity of AF detection. The accuracy of AF detection with the automatic AI arrhythmia detection from the heart belt ECG recording was also high (97.5%), and the sensitivity and specificity were 100% and 95.4%, respectively. The correlation

between the automatic estimated AF burden and the true AF burden from Holter recording was >0.99 with a mean burden error of 0.05 (SD 0.26) hours. The heart belt demonstrated good user experience and did not significantly interfere with the patient's daily activities. The patients preferred the heart belt over Holter ECG for rhythm monitoring (85/110, 77% heart belt vs 77/109, 71% Holter, *P*=.049).

Conclusions: A consumer-grade, single-lead ECG heart belt provided good-quality ECG for rhythm diagnosis. The mHealth arrhythmia monitoring system, consisting of heart-belt ECG, a mobile phone application, and an automated AF detection achieved AF detection with high accuracy, sensitivity, and specificity. In addition, the mHealth arrhythmia monitoring system showed good user experience.

Trial Registration: ClinicalTrials.gov NCT03507335; https://clinicaltrials.gov/ct2/show/NCT03507335

(JMIR Mhealth Uhealth 2021;9(10):e29933) doi:10.2196/29933

KEYWORDS

atrial fibrillation; ECG; algorithm; stroke; mHealth; user experience; Awario analysis Service; Suunto Movesense; cardiology; digital health; mobile health; wearable device; heart belt; arrhythmia monitor; heart monitor

Introduction

Atrial fibrillation (AF) is the most common tachyarrhythmia diagnosed in clinical practice. The aging of the global population is expected to increase the prevalence of AF by 2.3-fold by 2030 as compared with 2010 [1]. The most serious complication of AF is embolic stroke. The stroke risk can be reduced by as much as 60% with oral anticoagulation therapy in high-risk AF patients [2-6]. Evidently, there is a need to develop new approaches to diagnose AF in patients who would benefit from anticoagulation.

According to the current recommendations from the European Society of Cardiology (ESC), an electrocardiography (ECG) documentation interpreted by a physician is required to establish the diagnosis of AF [7]. AF screening recommendations include opportunistic or systematic screening in patients ≥65 years of age or with other characteristics suggesting an increased risk of stroke [7]. The clinical challenge is that AF is often paroxysmal or asymptomatic. Thus, it remains often undiagnosed when using traditional 12-lead ECG or Holter recordings [8,9]. Even the widely available patient-triggered mobile health (mHealth) products have not been able to resolve this challenge in asymptomatic patients [10]. User experience also is particularly important in the development of mHealth products, as these technologies can be too technical and complex for elderly people (ie, those who would benefit most from AF screening and anticoagulation therapy) [11]. In the development of medical technology, user involvement has positive effects, such as increased awareness of users' needs and experiences, better design, and clearer interfaces, as well as improved functionality, usability, and quality [12].

In this proof-of-concept study, we evaluated the feasibility and accuracy of the entire sequence of operations from ECG recording to artificial intelligence (AI) arrhythmia detection and the diagnosis of AF. This novel mHealth arrhythmia self-monitoring system includes a commonly available consumer-grade, single-lead heart rate belt ECG device (heart belt), a mobile phone application, and a cloud service with an AI arrhythmia detection algorithm.

The specific aims of the study were to (1) evaluate the feasibility and quality of a single lead heart belt for ECG recording, (2) determine the accuracy of the heart belt ECG in AF diagnosis, (3) assess the accuracy of the AI arrhythmia detection algorithm for AF screening, and (4) evaluate the user experience with the heart belt in a subgroup of patients >65 years of age (high-risk group).

Methods

Study Design

This study was part of a larger study entity, Atrial Fibrillation Detection: 24 Hour Study (AFIB24h), in which several different measurement techniques for detecting AF were studied. The study was performed as a single-center study at Kuopio University Hospital. The local ethics committee approved the study protocol (July 23, 2017), and the study was registered in the ClinicalTrials.gov database (NCT03507335).

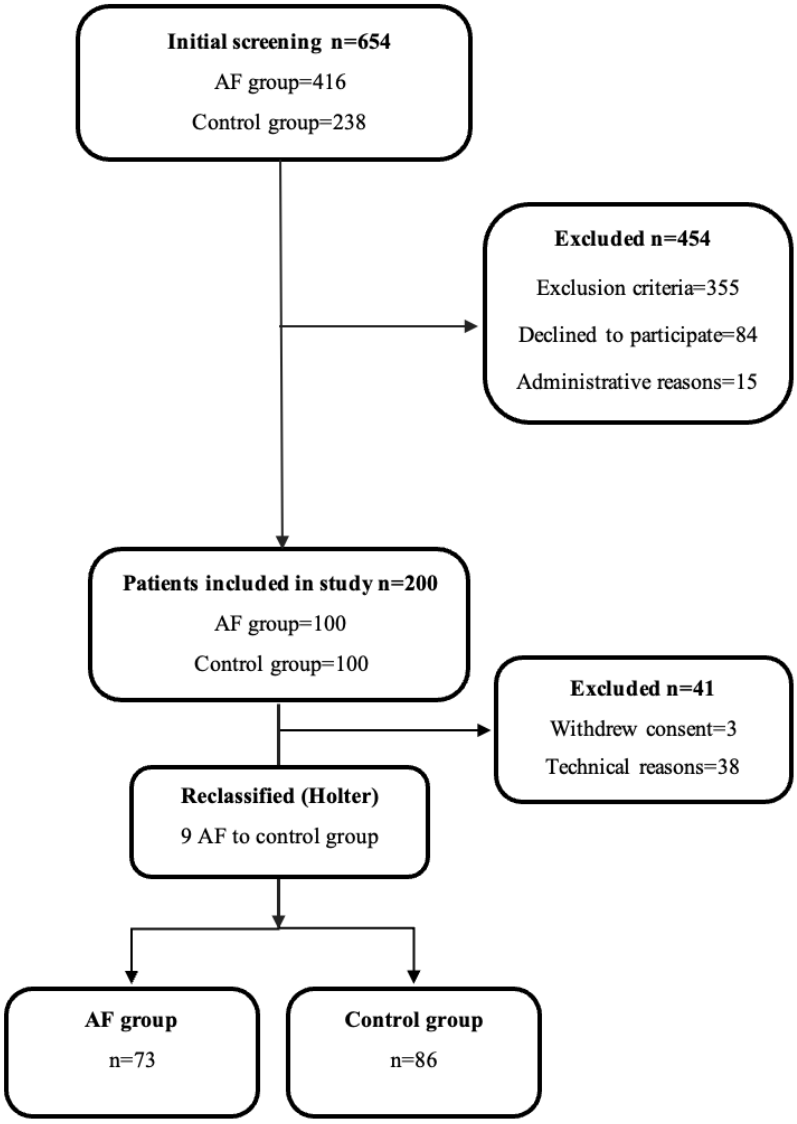
Recruitment

The inclusion criteria were AF or sinus rhythm (SR) based on a 12-lead resting ECG recorded during admission to the hospital. The exclusion criteria were (1) estimated stay in the hospital <24 hours, (2) BMI ≥35 kg/m^2, (3) left bundle branch block (LBBB) or right bundle branch block (RBBB), (4) implanted cardiac pacemaker, and (5) a medical condition requiring immediate treatment.

A total of 654 patients were screened in the emergency department between April 2018 and December 2019. In the initial screening, 454 patients were excluded for the reasons summarized in Figure 1. Of the remaining 200 patients, 100 patients with AF were assigned to the AF group, and 100 patients with SR were assigned to the control group. However, a further 41 patients were excluded: 38 patients due to technical reasons and 3 patients who withdrew their consent. In addition, the rhythm of some patients had converted from the time of 12-lead ECG recording prior to study measurements. Consequently, the final rhythm classification made from Holter ECG recording reclassified 9 patients from the AF group to the control group. Thus, the final study population consisted of 159 patients, of whom 73 were in the AF group and 86 were in the control group. The clinical characteristics of the patients were collected using a standardized data collection protocol and confirmed or complemented from the medical records. All

participants provided written informed consent to participate in the study. In addition, the performance of the AI arrhythmia detection algorithm for identifying short AF episodes was tested using 173 ECG recordings from 4 public AF-detection datasets (Multimedia Appendix 1).

Figure 1. Study flow chart. AF: atrial fibrillation.



ECG Recordings

A consumer-grade, single-lead heart belt (Suunto Movesense, Suunto, Vantaa, Finland) was attached to the patient's chest approximately 2 cm below the lower end of the sternum (Figure 2). The heart belt ECG data were transferred via Bluetooth connection to a mobile phone from where the data were transmitted to a cloud service for both visual and automatic analyses (Figure 3). A heart belt ECG was recorded for 24 hours. A simultaneously registered 3-lead Holter ECG recording (Faros 360, Bittium, Oulu, Finland) was used as the gold standard for rhythm classification (Figure 2).

Figure 2. Electrocardiogram (ECG) recordings using a (1) single-lead heart belt ECG recording and (2) 3-lead Holter ECG recording. LA: left arm; LL: left limb; RA: right arm; V3: V3 lead of the 12-lead ECG.

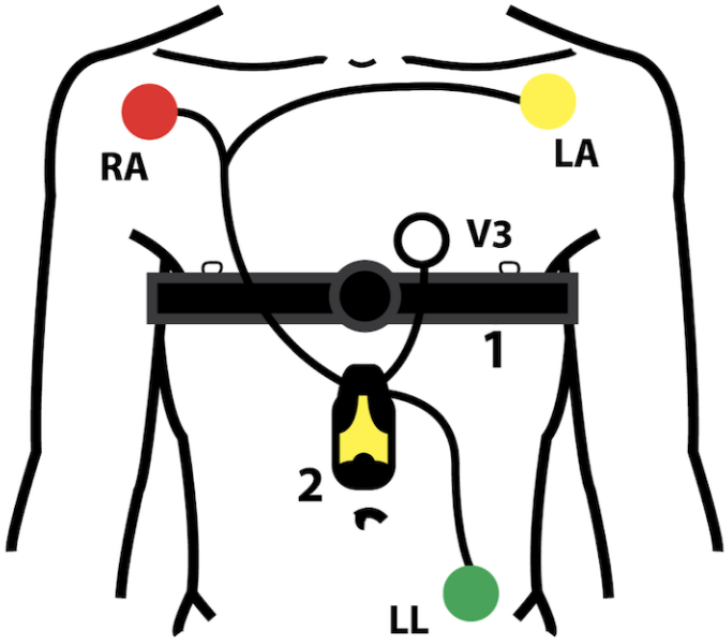


Figure 3. Schematic presentation of the heart belt electrocardiogram (ECG)-based automatic arrhythmia detection.

ECG Analysis

The heart belt ECG and Holter recordings were both analyzed using Medilog Darwin Professional V2.8.1 software (Schiller Global, Baar, Switzerland). The ECG recordings were reviewed in a random order independently by 4 investigators blinded to the 12-lead ECG and classified as either an AF or non-AF rhythm. Subsequently, commercial AI arrhythmia analysis software (Awario, Heart2Save, Kuopio, Finland) was used for automatic AF screening from the heart belt ECG recordings. The AI arrhythmia detection algorithm classified the heart belt ECG data into SR, AF, or noninterpretable. The accuracy of the visual and automatic rhythm classifications from the heart belt ECG recording was further assessed by comparing it with the gold standard Holter ECG recording. The rhythm analyses were performed from recordings in which both the heart belt ECG and Holter ECG recordings were interpretable.

Heart Belt User Experience

The patients were asked to complete a user experience questionnaire developed for this study at the end of the ECG registration. Patients evaluated their heart belt and Holter experience separately, answering the following questions for each device: how you rank the device (comfortable=1, reasonably comfortable=2, neutral=3, slightly uncomfortable=4, uncomfortable=5); did the device interfere with sleep, eating, toileting, or movement (Yes or No); and would you be willing to use the device at home for rhythm monitoring (Yes or No). The user experience was analyzed in a subgroup of patients >65 years of age (ie, in those patients in whom the AF screening is recommended by the ESC guidelines).

Statistical Analysis

The AF and control groups were compared using *t* tests for continuous variables and χ^2 tests or Fisher exact tests for dichotomous variables. The following parameters were used to quantify the performance of (1) detecting AF per patient (subject-based) and (2) total accumulated AF duration across all patients (time-based) from heart belt ECG recordings visually and with an AI arrhythmia detection algorithm: accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). The absolute difference between the AF burden derived from the heart belt and the Holter was described using the mean AF burden error (time). The AF burden determined by the heart belt was compared with the reference AF burden from the Holter using a Bland-Altman plot [13]. In the survey assessing user experience, the users' opinions of the heart belt and Holter were compared using Wilcoxon signed rank tests and McNemar tests. All significance tests were two-tailed, and $P \leq .05$ was considered statistically significant. The data were analyzed using SPSS version 25.

Results

Clinical Characteristics

In comparison with the control group, AF patients were older (mean 77, SD 10 years vs mean 68, SD 16 years; $P<.001$), presented more often with a history of paroxysmal AF ($P<.001$) or congestive heart failure ($P<.001$), and were more often on anticoagulation ($P<.001$), digoxin ($P=.045$), and beta-blocker ($P=.02$) therapy (Table 1). Furthermore, the AF patients also reported more often the presence of palpitations ($P=.02$) and respiratory distress ($P=.001$).

Table 1. Patient demographics.

Characteristics	Control group (n=86)	AF[a] group (n=73)	Significance (2-sided)
Age (years), mean (SD)	68 (16)	77 (10)	<.001
BMI (kg/m^2), mean (SD)	26 (4)	27 (4)	.31
Male gender, n (%)	34 (40)	38 (52)	.11
Medical history, n (%)			
Earlier AF episode	17 (20)	58 (80)	<.001
Coronary heart disease	22 (26)	22 (30)	.52
Diabetes mellitus	19 (22)	17 (23)	.86
Hypertension	51 (59)	53 (73)	.08
Congestive heart failure	11 (13)	34 (47)	<.001
Previous heart surgery	5 (6)	10 (14)	.09
Medication, n (%)			
Anticoagulation therapy	22 (26)	61 (84)	<.001
Beta-blocker	40 (47)	48 (66)	.02
Digoxin	4 (5)	10 (14)	.045
Anti-arrhythmic medication	0 (0)	2 (3)	.21
Symptoms prior to hospital admission, n (%)			
Decrease in general condition	48 (56)	43 (59)	.70
Fatigue	47 (55)	44 (60)	.48
Palpitations	22 (26)	32 (44)	.02
Respiratory distress	24 (28)	39 (53)	.001
Chest pain	16 (19)	13 (18)	.90

[a]AF: atrial fibrillation.

Quality of the Heart Belt ECG Data

In the analysis of heart belt ECG data (all subjects), 2707 hours (2707.44/3416.81, 79.24%) of visual analysis and 2748 hours (2747.73/3416.81, 80.42%) of automatic analysis were deemed interpretable. Based on the visual assessment, 1226 hours (1225.59/1566.45, 78.24%) of the AF group recordings and 1482 hours (1481.85/1850.40, 80.08%) of the control group recordings were judged as being interpretable. Correspondingly, 1224 hours (1223.91/1566.39, 78.14%) of the AF recordings and 1524 hours (1523.82/1850.40, 82.35%) of the control group recordings were interpretable when analyzed with the AI arrhythmia detection algorithm. The subject-based median for visually interpretable data was 87% (25th percentile=76%; 75th percentile=95%), very similar to the automatic analysis, (median 89%; 25th percentile=76%; 75th percentile=97%; Figure 4).

The accuracy of visual and automated rhythm analyses was evaluated from the ECG data deemed interpretable in both heart belt and Holter ECG recordings (2655.72 hours). Representative examples of heart belt ECG measurements are presented in Figure 5.

Figure 4. Percentage of interpretable electrocardiograms (ECGs) in individual subject recordings, which are sorted using an automatic quality value.

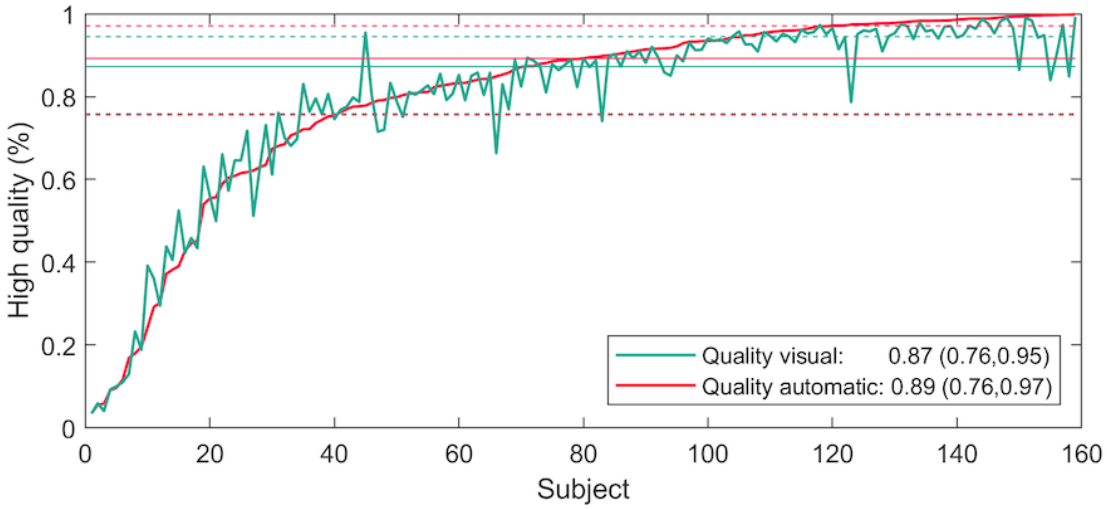
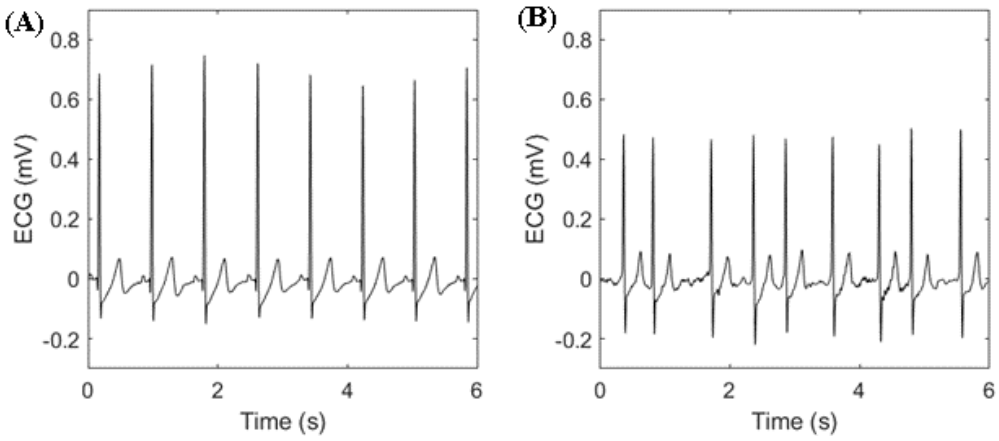


Figure 5. Examples of heart belt electrocardiogram (ECG) recordings for (A) sinus rhythm and (B) atrial fibrillation.



Accuracy of Visual Assessment From Heart Belt ECG

The quality of the ECG signal from the mHealth system was tested by classifying the heart belt ECG recordings visually into AF and non-AF rhythms. In the subject-based analysis, all the patients with AF were correctly identified, and correspondingly, none of the subjects with SR were given a false AF diagnosis (Table 2). In the time-based analysis, the accuracy, sensitivity, and specificity of diagnosing AF from the heart belt ECG recordings were all >99.9%. Correspondingly, the PPV and NPV of detecting the presence or absence of AF were both >99.9%.

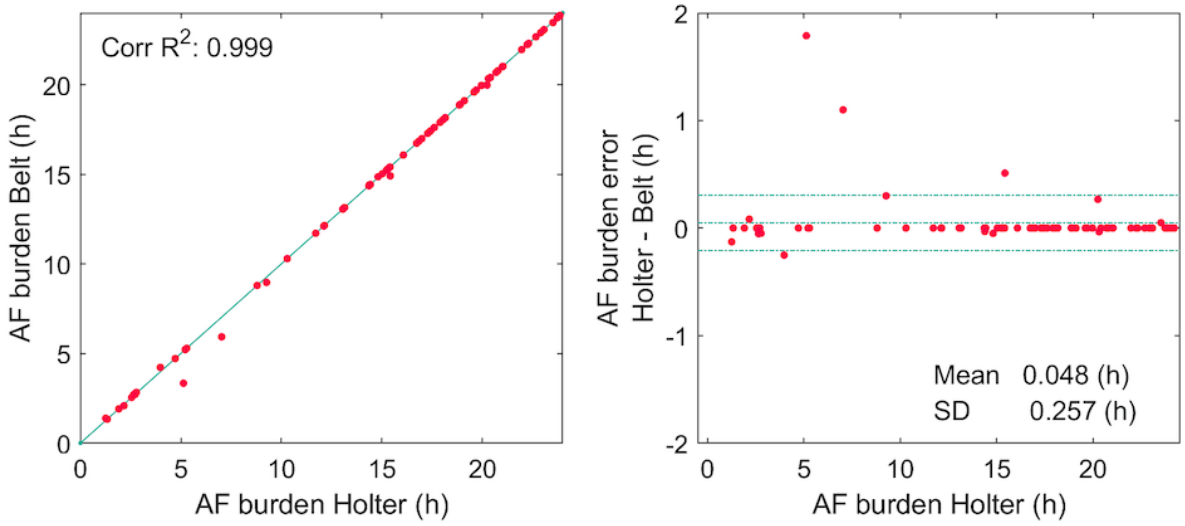
Table 2. Subject- and time-based atrial fibrillation (AF) detection accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) based on visual and automatic artificial intelligence (AI) arrhythmia algorithms.

Type of algorithm	Accuracy, %	Sensitivity, %	Specificity, %	PPV, %	NPV, %
Visual subject-based	100	100	100	100	100
Visual time-based	>99.9	>99.9	>99.9	>99.9	>99.9
Algorithm subject-based	97.5	100	95.4	94.8	100
Algorithm time-based	99.2	98.5	99.7	99.6	>99.9

Accuracy of the AI Arrhythmia Detection Algorithm From the Heart Belt ECG

The AI arrhythmia detection algorithm detected AF correctly in all the patients in the AF group and suggested the presence of AF in 4 patients in the SR group (false-positive AF detection), resulting in a subject-based accuracy of 97.5%, sensitivity of 100%, specificity of 95.4%, PPV of 94.8%, and NPV of 100% (Table 2). The time-based accuracy, sensitivity, and specificity values for the AI arrhythmia detection algorithm for identifying AF were 99.2%, 98.5%, and 99.7%, respectively, and the PPV and NPV for detecting the presence or absence of AF were 99.6% and >99.9%, respectively (Table 2). The correlation between the true and automatically estimated AF burden from AF patients was >0.99, and the mean burden error was 0.05 (SD 0.26) hours (Figure 6).

Figure 6. Correlation between atrial fibrillation (AF) burden estimated by the artificial intelligence (AI) arrhythmia detection algorithm and the reference AF burden from (A) Holter recording and (B) Bland Altman plot of the AF burden estimate.



The supplementary analyses performed for 4 freely available ECG datasets including different arrhythmias found that the sensitivity of the AI arrhythmia detection algorithm to detect short AF episodes was moderate. The AI arrhythmia detection algorithm detected 83.43% (2421/2902) of all AF episodes with a duration of >30 seconds, but the sensitivity increased significantly with the duration of an AF episode. The AI arrhythmia detection algorithm detected 95.10% (641/674) of AF episodes lasting >5 minutes and 98.49% (327/332) of episodes with a duration >15 minutes (Figure S1 in Multimedia Appendix 1). The overall time-based accuracy, sensitivity, and specificity for AF detection were high, at 97.8%, 97.1%, and 98.4%, respectively (Table S1 in Multimedia Appendix 1). The correlation between the reference AF burden and the estimated AF burden was >0.99, and the mean burden error was 0.05 (SD 0.82) hours (Figure S2 in Multimedia Appendix 1).

User Experience With the Heart Belt

A total of 112 patients over 65 years of age filled in the user experience survey. Using a scale of 1 to 5 (1=comfortable and 5=uncomfortable), the patients rated the heart belt as a median of 3 (25th percentile=2; 75th percentile=4; Table 3). In terms of discomfort, 9 (9/112, 8.0%) patients reported that the heart belt device interfered with their sleep, 3 (3/112, 2.7%) with eating, 4 (4/112, 3.6%) with toileting, and 4 (4/112, 3.6%) with normal movement. Finally, 85 (85/110, 77.3%) of the patients reported that they would be willing to use the heart belt device, slightly more than the 77 (77/109, 70.6%) who stated that they would wear the Holter device at home for rhythm monitoring (P=.049; Table 3).

Table 3. User experience with the heart belt and Holter devices.

Question	Holter	Heart belt	Significance (2-sided)
How would you rank the device[a], median (25th percentile, 75th percentile)	3 (2,3)[b]	3 (2,4)[c]	.06[b]
Device interference, n (%)			
Device interfered with sleep	9 (8.0)[d]	9 (8.0)[d]	1.0[e]
Device interfered with eating	1 (0.9)[d]	3 (2.7)[d]	.63[e]
Device interfered with toileting	4 (3.6)[d]	4 (3.6)[d]	1.0[e]
Device interfered with movement	1 (0.9)[d]	4 (3.6)[d]	.40[e]
Experience with device usability, n (%)			
I would use the device at home for rhythm monitoring	77 (70.6)[f]	85 (77.3)[c]	.049[f]

[a]Comfortable=1, reasonably comfortable=2, neutral=3, slightly uncomfortable=4, uncomfortable=5.

[b]n=108.

[c]n=110.

[d]n=112.

[e]n=111.

[f]n=109.

[g]n=107.

Discussion

Principal Findings

We demonstrated that a novel mHealth arrhythmia monitoring system using a consumer-grade heart belt ECG device, a mobile phone application, and an automated AI arrhythmia analysis was both feasible and accurate for 24-hour ECG monitoring and rhythm diagnostics. The heart belt provided a high-quality ECG signal for visual evaluation, achieving an AF diagnostic accuracy of 100%. In addition, the AI arrhythmia detection algorithm identified AF patients with a sensitivity of 100% and a specificity of 95.4% (4 false positives).

In our study, 80% of the heart belt ECG recordings were of sufficient quality to permit visual and automatic rhythm diagnostics. The proportion of analyzable data using the heart belt operating with dry electrodes was comparable with results obtained in earlier studies using adhesive-coated wet electrodes or shirt-type ECG recording devices [14-16]. In these publications, the amount of analyzable data varied from 92% to 99% of the total wear time, which ranged from 48 hours to several days.

AF diagnosis requires confirmation by a physician [7]. Thus, the ECG signal provided by the mHealth system needs to be of high quality. In previous studies, the sensitivity of AF detection by health care professionals using mHealth ECG recordings has ranged from 73% to 100%, with specificity from 84% to 100% [17-20]. In our study, the accuracy of AF diagnosis was superior to previous studies. In the visual assessment, all AF patients were identified, and no patient with SR was misdiagnosed as having AF. Thus, the mHealth monitoring system described here represents a useful tool for identification of AF.

Visual analysis of long-term ECG recording for AF screening is very time and manpower consuming. Thus, there is an unmet need for automatic AF detection. In previous studies, the sensitivity of identifying AF using arrhythmia detection algorithms from mHealth ECG recording has ranged from 87% to 99%, with a specificity from 80% to 97% [21-26]. In our study, when using the AI arrhythmia detection algorithm, the AF detection sensitivity (100%) was superior, and the specificity (95.4%) was comparable or even superior to those in previous studies. Only 4 of 86 patients had a false AF alarm, and, in most cases, this was due to frequent (>10,000 per 24 hours) supraventricular or ventricular extrasystoles. Although the accuracy of current “state-of-the-art” AF detection algorithms is high, the presence of false positives (approximately 2-5/100) indicates that AF diagnosis needs to be confirmed by a physician. Nonetheless, the benefits of automated AI arrhythmia detection in ambulatory screening are evident; they can (1) exclude poor quality data, (2) detect AF patients with high sensitivity and specificity, and (3) exclude >90% of patients who do not require medical attention. In addition to AF diagnosis, the frequency and duration of an AF episode as well as the AF burden are associated with the risk of stroke [27-31]. We found that the AF burden estimated by the AI arrhythmia detection algorithm correlated almost perfectly with the true AF burden (r>0.99).

Previous studies have reported poor compliance with lead-based, long-term ECG monitoring devices, such as Holter devices, mobile telemetry devices, and event monitors, because of their difficulty in use; interference with the patient’s work, travel, or lifestyle; and skin irritation [32-35]. In contrast, novel mHealth methods have been more comfortable, causing less interference with daily living; therefore, participants have preferred the new mHealth ECG devices instead of the traditional Holter for

rhythm monitoring [36,37]. In our study, the heart belt user experience was found to be significantly better than the Holter ECG in a subgroup of elderly patients (over 65 years of age). Indeed, a higher proportion (77.3% vs 70.6%) of elderly patients preferred the heart belt to the Holter device, the gold standard of rhythm monitoring. Patient-reported discomfort caused by the heart belt was very low. The advantage of heart belts over traditional measurement methods is that they have been designed to allow freedom of movement and users find them easy to use. It should be mentioned here that the Holter device used in this study was very small, weighing only 18 grams.

Several new photoplethysmography and ECG-based wearable mHealth technologies, such as smartphones and watches, have been studied for AF detection [18-20,38-41]. These wearable technologies could provide a practical and cost-effective solution for AF detection and AF burden assessment [7]. According to a recent survey [42], there is consensus on recommending mHealth wearable devices or apps as an alternative to traditional methods such as Holter monitoring for detecting AF in symptomatic patients or post stroke or TIA. However, health care professionals do not feel that health care systems are ready for mass consumer-initiated AF screening with these techniques, as there still appears to be a need to better define suitable screening population and an appropriate management pathway for consumers with positive results [42]. From the perspective of health care professionals, the presented approach with the heart belt monitoring has the potential to be a small change from traditional Holter and therefore to be widely accepted. Our study suggests that the presented mHealth monitoring technology enables long-term AF screening and can be considered as being user-friendly. An AI arrhythmia monitoring system can warn the patient of a possible AF, store the ECG from the AF episode, and send it to the physician for a final rhythm confirmation. In addition to the AF diagnosis, the system could provide an estimate of the AF burden and help in the initiation of appropriate treatment for stroke prevention and rhythm control.

Study Limitations

We acknowledge some limitations in our study. First, morbid obesity could degrade the signal quality and thus, produce more failed measurements. In addition, in cases with RBBB and LBBB, the presence of 2 broad R-peak QRS complexes could increase the variation of the R-R interval, which could result in false AF detection by the automatic algorithm. For these reasons, these patients were excluded, and further studies in these subgroups will be needed. Second, the study was conducted in a proof-of-concept style to test the feasibility of the entire sequence of operations, from ECG recording to automatic AF analysis. Examining a greater number of patients was not possible due to technical aspects related to telecommunication links to the server as well as some application problems with this proof-of-concept system. In addition, poor electrode contacts and incorrect placement of the heart belt caused some noninterpretable periods of recording. In addition, our mHealth system for AF monitoring should be studied in an out-of-hospital setting to assess the signal quality, the accuracy of AF detection, and AF burden estimate as well as the overall usability.

Conclusions

A consumer-grade, single-lead ECG heart belt provided good-quality ECG for rhythm diagnosis. The mHealth arrhythmia monitoring system, consisting of heart-belt ECG, a mobile phone application, and automated AF detection achieved AF detection with high accuracy, sensitivity, and specificity. In addition, the mHealth arrhythmia monitoring system showed good user experience.

Acknowledgments

This work was supported by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (project 5101137, Kuopio, Finland).

We thank the nursing and medical staff of Kuopio University Hospital and especially Menna Kuosmanen BM, Matias Rapo BM, Antti Redsven BM, Roope Sepponen BM, Salla Autio BM, and Vilppu Kekkonen BM for their work with data acquisition; Oona Silvennoinen for illustration assistance; and biostatistician Tuomas Selander MSc for statistical advice.

Authors' Contributions

All authors contributed to the study and preparation of this manuscript. TJM, JAL, HJ, MPT, JH, JEKH, TTR, OES, and ESV designed the study. OES and ESV were responsible for collecting the data. JAL, OES, ESV, TPL, TML, and SM contributed to data analysis and figure and table preparation. OES, JH, SM, HJ, TTR, MPT, TPL, TML, ESV, JEKH, TJM, JAL were responsible for editing and further improvements to the manuscript. OES was responsible for literature search, drafting the first version of the manuscript, final editing, and preparation of the manuscript for submission. All authors read and approved the final manuscript.

Conflicts of Interest

OS received research support from the Finland's state research fund (VTR). JAL, TTR, TJM, SM, HJ, JH, and MPT are shareholders of a company (Heart2Save) that designs ECG-based software for medical equipment. SM, JAL, MPT, and HJ report personal fees from Heart2Save.

Multimedia Appendix 1
Supplementary material A.

[PDF File (Adobe PDF File), 348 KB - mhealth_v9i10e29933_app1.pdf]

References

1. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of atrial fibrillation in the U.S. adult population. Am J Cardiol 2013 Oct 15;112(8):1142-1147. [doi: 10.1016/j.amjcard.2013.05.063] [Medline: 23831166]
2. Kishore A, Vail A, Majid A, Dawson J, Lees KR, Tyrrell PJ, et al. Detection of atrial fibrillation after ischemic stroke or transient ischemic attack: a systematic review and meta-analysis. Stroke 2014 Feb;45(2):520-526. [doi: 10.1161/STROKEAHA.113.003433] [Medline: 24385275]
3. Henriksson KM, Farahmand B, Åsberg S, Edvardsson N, Terént A. Comparison of cardiovascular risk factors and survival in patients with ischemic or hemorrhagic stroke. Int J Stroke 2012 Jun;7(4):276-281. [doi: 10.1111/j.1747-4949.2011.00706.x] [Medline: 22151650]
4. Grond M, Jauss M, Hamann G, Stark E, Veltkamp R, Nabavi D, et al. Improved detection of silent atrial fibrillation using 72-hour Holter ECG in patients with ischemic stroke: a prospective multicenter cohort study. Stroke 2013 Dec;44(12):3357-3364. [doi: 10.1161/STROKEAHA.113.001884] [Medline: 24130137]
5. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 2007 Jun 19;146(12):857-867. [doi: 10.7326/0003-4819-146-12-200706190-00007] [Medline: 17577005]
6. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. Lancet 2014 Mar 15;383(9921):955-962. [doi: 10.1016/S0140-6736(13)62343-0] [Medline: 24315724]
7. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 01;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
8. Kirchhof P, Bax J, Blomstrom-Lundquist C, Calkins H, Camm AJ, Cappato R, et al. Early and comprehensive management of atrial fibrillation: executive summary of the proceedings from the 2nd AFNET-EHRA consensus conference 'research perspectives in AF'. Eur Heart J 2009 Dec;30(24):2969-277c. [doi: 10.1093/eurheartj/ehp235] [Medline: 19535417]
9. Healey JS, Alings M, Ha A, Leong-Sit P, Birnie DH, de Graaf JJ, ASSERT-II Investigators. Subclinical Atrial Fibrillation in Older Patients. Circulation 2017 Oct 03;136(14):1276-1283. [doi: 10.1161/CIRCULATIONAHA.117.028845] [Medline: 28778946]
10. Diederichsen SZ, Haugan KJ, Kronborg C, Graff C, Højberg S, Køber L, et al. Comprehensive Evaluation of Rhythm Monitoring Strategies in Screening for Atrial Fibrillation: Insights From Patients at Risk Monitored Long Term With an Implantable Loop Recorder. Circulation 2020 May 12;141(19):1510-1522. [doi: 10.1161/CIRCULATIONAHA.119.044407] [Medline: 32114796]
11. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
12. Shah SGS, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
14. Turakhia MP, Hoang DD, Zimetbaum P, Miller JD, Froelicher VF, Kumar UN, et al. Diagnostic utility of a novel leadless arrhythmia monitoring device. Am J Cardiol 2013 Aug 15;112(4):520-524 [FREE Full text] [doi: 10.1016/j.amjcard.2013.04.017] [Medline: 23672988]
15. Fukuma N, Hasumi E, Fujiu K, Waki K, Toyooka T, Komuro I, et al. Feasibility of a T-Shirt-Type Wearable Electrocardiography Monitor for Detection of Covert Atrial Fibrillation in Young Healthy Adults. Sci Rep 2019 Aug 13;9(1):11768 [FREE Full text] [doi: 10.1038/s41598-019-48267-1] [Medline: 31409855]
16. Solomon MD, Yang J, Sung SH, Livingston ML, Sarlas G, Lenane JC, et al. Incidence and timing of potentially high-risk arrhythmias detected through long term continuous ambulatory electrocardiographic monitoring. BMC Cardiovasc Disord 2016 Feb 17;16:35 [FREE Full text] [doi: 10.1186/s12872-016-0210-x] [Medline: 26883019]
17. Desteghe L, Raymaekers Z, Lutin M, Vijgen J, Dilling-Boer D, Koopman P, et al. Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. Europace 2017 Jan;19(1):29-39. [doi: 10.1093/europace/euw025] [Medline: 26893496]
18. Tarakji KG, Wazni OM, Callahan T, Kanj M, Hakim AH, Wolski K, et al. Using a novel wireless system for monitoring patients after the atrial fibrillation ablation procedure: the iTransmit study. Heart Rhythm 2015 Mar;12(3):554-559. [doi: 10.1016/j.hrthm.2014.11.015] [Medline: 25460854]

19. Lau JK, Lowres N, Neubeck L, Brieger DB, Sy RW, Galloway CD, et al. iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke. Int J Cardiol 2013 Apr 30;165(1):193-194. [doi: 10.1016/j.ijcard.2013.01.220] [Medline: 23465249]
20. Himmelreich JC, Karregat EP, Lucassen WA, van Weert HC, de Groot JR, Handoko ML, et al. Diagnostic Accuracy of a Smartphone-Operated, Single-Lead Electrocardiography Device for Detection of Rhythm and Conduction Abnormalities in Primary Care. Ann Fam Med 2019 Sep 09;17(5):403-411 [FREE Full text] [doi: 10.1370/afm.2438] [Medline: 31501201]
21. Haeberlin A, Roten L, Schilling M, Scarcia F, Niederhauser T, Vogel R, et al. Software-based detection of atrial fibrillation in long-term ECGs. Heart Rhythm 2014 Jun;11(6):933-938. [doi: 10.1016/j.hrthm.2014.03.014] [Medline: 24632179]
22. Lian J, Wang L, Muessig D. A simple method to detect atrial fibrillation using RR intervals. Am J Cardiol 2011 May 15;107(10):1494-1497. [doi: 10.1016/j.amjcard.2011.01.028] [Medline: 21420064]
23. Duverney D, Gaspoz J, Pichot V, Roche F, Brion R, Antoniadis A, et al. High accuracy of automatic detection of atrial fibrillation using wavelet transform of heart rate intervals. Pacing Clin Electrophysiol 2002 Apr;25(4 Pt 1):457-462. [doi: 10.1046/j.1460-9592.2002.00457.x] [Medline: 11991371]
24. Tateno K, Glass L. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals. Med. Biol. Eng. Comput 2001 Nov;39(6):664-671. [doi: 10.1007/bf02345439]
25. Hindricks G, Pokushalov E, Urban L, Taborsky M, Kuck K, Lebedev D, XPECT Trial Investigators. Performance of a new leadless implantable cardiac monitor in detecting and quantifying atrial fibrillation: Results of the XPECT trial. Circ Arrhythm Electrophysiol 2010 Apr;3(2):141-147. [doi: 10.1161/CIRCEP.109.877852] [Medline: 20160169]
26. Giebel GD, Gissel C. Accuracy of mHealth Devices for Atrial Fibrillation Screening: Systematic Review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
27. Chen LY, Chung MK, Allen LA, Ezekowitz M, Furie KL, McCabe P, American Heart Association Council on Clinical Cardiology, Council on CardiovascularStroke Nursing; Council on Quality of Care and Outcomes Research, Stroke Council. Atrial Fibrillation Burden: Moving Beyond Atrial Fibrillation as a Binary Entity: A Scientific Statement From the American Heart Association. Circulation 2018 May 15;137(20):e623-e644. [doi: 10.1161/CIR.0000000000000568] [Medline: 29661944]
28. Boriani G, Glotzer TV, Santini M, West TM, De Melis M, Sepsi M, et al. Device-detected atrial fibrillation and risk for stroke: an analysis of >10,000 patients from the SOS AF project (Stroke preventiOn Strategies based on Atrial Fibrillation information from implanted devices). Eur Heart J 2014 Feb;35(8):508-516 [FREE Full text] [doi: 10.1093/eurheartj/eht491] [Medline: 24334432]
29. Senoo K, Lip GYH, Lane DA, Büller HR, Kotecha D. Residual Risk of Stroke and Death in Anticoagulated Patients According to the Type of Atrial Fibrillation: AMADEUS Trial. Stroke 2015 Sep;46(9):2523-2528. [doi: 10.1161/STROKEAHA.115.009487] [Medline: 26205373]
30. Vanassche T, Lauw MN, Eikelboom JW, Healey JS, Hart RG, Alings M, et al. Risk of ischaemic stroke according to pattern of atrial fibrillation: analysis of 6563 aspirin-treated patients in ACTIVE-A and AVERROES. Eur Heart J 2015 Feb 01;36(5):281-27a. [doi: 10.1093/eurheartj/ehu307] [Medline: 25187524]
31. Banerjee A, Taillandier S, Olesen JB, Lane DA, Lallemand B, Lip GYH, et al. Pattern of atrial fibrillation and risk of outcomes: the Loire Valley Atrial Fibrillation Project. Int J Cardiol 2013 Sep 10;167(6):2682-2687. [doi: 10.1016/j.ijcard.2012.06.118] [Medline: 22795403]
32. Roten L, Schilling M, Häberlin A, Seiler J, Schwick NG, Fuhrer J, et al. Is 7-day event triggered ECG recording equivalent to 7-day Holter ECG recording for atrial fibrillation screening? Heart 2012 Apr;98(8):645-649. [doi: 10.1136/heartjnl-2011-301455] [Medline: 22397942]
33. Ackermans PAJ, Solosko TA, Spencer EC, Gehman SE, Nammi K, Engel J, et al. A user-friendly integrated monitor-adhesive patch for long-term ambulatory electrocardiogram monitoring. J Electrocardiol 2012 Mar;45(2):148-153. [doi: 10.1016/j.jelectrocard.2011.10.007] [Medline: 22153334]
34. Rothman SA, Laughlin JC, Seltzer J, Walia JS, Baman RI, Siouffi SY, et al. The diagnosis of cardiac arrhythmias: a prospective multi-center randomized study comparing mobile cardiac outpatient telemetry versus standard loop event monitoring. J Cardiovasc Electrophysiol 2007 Mar;18(3):241-247. [doi: 10.1111/j.1540-8167.2006.00729.x] [Medline: 17318994]
35. Vasamreddy CR, Dalal D, Dong J, Cheng A, Spragg D, Lamiy SZ, et al. Symptomatic and asymptomatic atrial fibrillation in patients undergoing radiofrequency catheter ablation. J Cardiovasc Electrophysiol 2006 Feb;17(2):134-139. [doi: 10.1111/j.1540-8167.2006.00359.x] [Medline: 16533249]
36. Smith WM, Riddell F, Madon M, Gleva MJ. Comparison of diagnostic value using a small, single channel, P-wave centric sternal ECG monitoring patch with a standard 3-lead Holter system over 24 hours. Am Heart J 2017 Mar;185:67-73. [doi: 10.1016/j.ahj.2016.11.006] [Medline: 28267477]
37. Barrett PM, Komatireddy R, Haaser S, Topol S, Sheard J, Encinas J, et al. Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring. Am J Med 2014 Jan;127(1):95.e11-95.e17 [FREE Full text] [doi: 10.1016/j.amjmed.2013.10.003] [Medline: 24384108]

38. Chan P, Wong C, Poh YC, Pun L, Leung WW, Wong Y, et al. Diagnostic Performance of a Smartphone-Based Photoplethysmographic Application for Atrial Fibrillation Screening in a Primary Care Setting. J Am Heart Assoc 2016 Jul 21;5(7):1 [FREE Full text] [doi: 10.1161/JAHA.116.003428] [Medline: 27444506]
39. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch Algorithm for Automated Detection of Atrial Fibrillation. J Am Coll Cardiol 2018 May 29;71(21):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.003] [Medline: 29535065]
40. Rozen G, Vaid J, Hosseini SM, Kaadan MI, Rafael A, Roka A, et al. Diagnostic Accuracy of a Novel Mobile Phone Application for the Detection and Monitoring of Atrial Fibrillation. Am J Cardiol 2018 May 15;121(10):1187-1191. [doi: 10.1016/j.amjcard.2018.01.035] [Medline: 29525063]
41. Orchard J, Lowres N, Freedman SB, Ladak L, Lee W, Zwar N, et al. Screening for atrial fibrillation during influenza vaccinations by primary care nurses using a smartphone electrocardiograph (iECG): A feasibility study. Eur J Prev Cardiol 2016 Oct;23(2 suppl):13-20. [doi: 10.1177/2047487316670255] [Medline: 27892421]
42. Boriani G, Schnabel RB, Healey JS, Lopes RD, Verbiest-van Gurp N, Lobban T, et al. Consumer-led screening for atrial fibrillation using consumer-facing wearables, devices and apps: A survey of health care professionals by AF-SCREEN international collaboration. Eur J Intern Med 2020 Dec;82:97-104 [FREE Full text] [doi: 10.1016/j.ejim.2020.09.005] [Medline: 32933842]

Abbreviations

AF: atrial fibrillation
AI: artificial intelligence
ECG: electrocardiogram
ESC: European Society of Cardiology
LBBB: left bundle branch block
mHealth: mobile health
NPV: negative predictive value
PPV: positive predictive value
RBBB: right bundle branch block
SR: sinus rhythm

Edited by L Buis; submitted 26.04.21; peer-reviewed by G Boriani, A Brandes; comments to author 17.06.21; revised version received 30.06.21; accepted 27.08.21; published 22.10.21.

Please cite as:
Santala OE, Halonen J, Martikainen S, Jäntti H, Rissanen TT, Tarvainen MP, Laitinen TP, Laitinen TM, Väliaho ES, Hartikainen JEK, Martikainen TJ, Lipponen JA
Automatic Mobile Health Arrhythmia Monitoring for the Detection of Atrial Fibrillation: Prospective Feasibility, Accuracy, and User Experience Study
JMIR Mhealth Uhealth 2021;9(10):e29933
URL: https://mhealth.jmir.org/2021/10/e29933
doi:10.2196/29933
PMID:34677135

Original Paper

Augmented Reality for Guideline Presentation in Medicine: Randomized Crossover Simulation Trial for Technically Assisted Decision-making

Andreas Follmann[1], MD; Alexander Ruhl[1]; Michael Gösch[2], PhD; Marc Felzen[1,3], MD; Rolf Rossaint[1], MD; Michael Czaplik[1], MD, PhD

[1]Department of Anesthesiology, Faculty of Medicine, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany
[2]Tech2go Mobile Systems GmbH, Hamburg, Germany
[3]Medical Direction, Emergency Medical Service, City of Aachen, Aachen, Germany

Corresponding Author:
Andreas Follmann, MD
Department of Anesthesiology
Faculty of Medicine
Rheinisch-Westfälische Technische Hochschule Aachen University
Pauwelsstraße 30
Aachen, 52074
Germany
Phone: 49 241 80 36219
Email: afollmann@ukaachen.de

Abstract

Background: Guidelines provide instructions for diagnostics and therapy in modern medicine. Various mobile devices are used to represent the potential complex decision trees. An example of time-critical decisions is triage in case of a mass casualty incident.

Objective: In this randomized controlled crossover study, the potential of augmented reality for guideline presentation was evaluated and compared with the guideline presentation provided in a tablet PC as a conventional device.

Methods: A specific Android app was designed for use with smart glasses and a tablet PC for the presentation of a triage algorithm as an example for a complex guideline. Forty volunteers simulated a triage based on 30 fictional patient descriptions, each with technical support from smart glasses and a tablet PC in a crossover trial design. The time to come to a decision and the accuracy were recorded and compared between both devices.

Results: A total of 2400 assessments were performed by the 40 volunteers. A significantly faster time to triage was achieved in total with the tablet PC (median 12.8 seconds, IQR 9.4-17.7; 95% CI 14.1-14.9) compared to that to triage with smart glasses (median 17.5 seconds, IQR 13.2-22.8, 95% CI 18.4-19.2; *P*=.001). Considering the difference in the triage time between both devices, the additional time needed with the smart glasses could be reduced significantly in the course of assessments (21.5 seconds, IQR 16.5-27.3, 95% CI 21.6-23.2) in the first run, 17.4 seconds (IQR 13-22.4, 95% CI 17.6-18.9) in the second run, and 14.9 seconds (IQR 11.7-18.6, 95% CI 15.2-16.3) in the third run *(P*=.001). With regard to the accuracy of the guideline decisions, there was no significant difference between both the devices.

Conclusions: The presentation of a guideline on a tablet PC as well as through augmented reality achieved good results. The implementation with smart glasses took more time owing to their more complex operating concept but could be accelerated in the course of the study after adaptation. Especially in a non–time-critical working area where hands-free interfaces are useful, a guideline presentation with augmented reality can be of great use during clinical management.

(JMIR Mhealth Uhealth 2021;9(10):e17472) doi:10.2196/17472

KEYWORDS

augmented reality; smart glasses; wearables; guideline presentation; decision support; triage

Introduction

Guidelines as well as standard operating procedures consist of recommendations for action in corresponding diseases or injuries and they contribute to a standardized, high-quality patient care, preferably independent from the user's experiences [1]. Many studies have certainly proven the benefit of guidelines but have also shown their deficiencies in appropriate implementation. Most of these decision trees are complex and are often misapplied or not applied at all [2,3]. Other studies have shown that the simultaneous visual presentation of the guidelines as well as electronic decision-support systems have improved their implementation [4,5]. Various guidelines are therefore available as printed versions or posters in clinics.

Electronic tools such as the display on tablet computers are beneficial [6]. A display using augmented reality in smart glasses has the decisive advantage of a hands-free concept, which enables operation of the smart glasses and the simultaneous use of hands as important tools for examination or treatment. Moreover, these smart glasses can collect and save all the guidelines to use case-related data iteratively [7]. Smart glasses have already been applied in experimental implementations, especially in the field of telesupervision for intraoperative teleconsultation from a surgeon's point of view [8] or to transfer medical knowledge from medically highly advanced countries to low-income countries [9]. Smart glasses can display patient-related information such as radiographic images [10] as well as the name of the currently consulted patient and his/her vital parameters [11].

In emergency and disaster medicine, decisions are taken in a short amount of time. However, the rapid expansion of medical literature has led to a high publication rate of various guidelines, further limiting the rapid implementation into practice [12,13]. The introduction of electronic-based decision-making systems could provide a decisive advantage in the use of clinical guidelines. This advantage is additionally strengthened by the fact that emergency medicine has been an early adaptor for a variety of technology-based tools [14,15]. Carenzo et al [16] tested the Google glasses for the first time in a large-scale emergency with 100 theoretical emergency patients and used a triage algorithm that was implemented into an app, and they reported that smart glasses have a high potential to be used as decision-support systems.

To provide iterative technical support for triage in case of mass casualty incidents and as an example for any other guideline-related decision, we have developed a guideline presentation as a decision support running on smart glasses (Recon Jet, Recon Instruments). Guidelines were displayed on a small monitor on the smart glasses. With simple operating gestures performed on an optical touchpad, menu items can be selected. Although the results showed a clearly improved implementation of the guideline, they also showed that more time was necessary for the processing of all decisions [17]. This finding suggests that a lot of time is needlessly wasted due to the unusual application of the smart glasses' operating concept. Only after a long-term application, the user achieves a learning effect, which accelerates the app performance. The use of a device with the usual display and choice of decisions such as a tablet computer could save a lot of time. However, these devices are not appropriate for use in emergency cases: hands cannot be used freely, body fluids can come into contact with the devices compared to smart glasses because of the short distance between the device and patient, and thereby cause further transmission of germs.

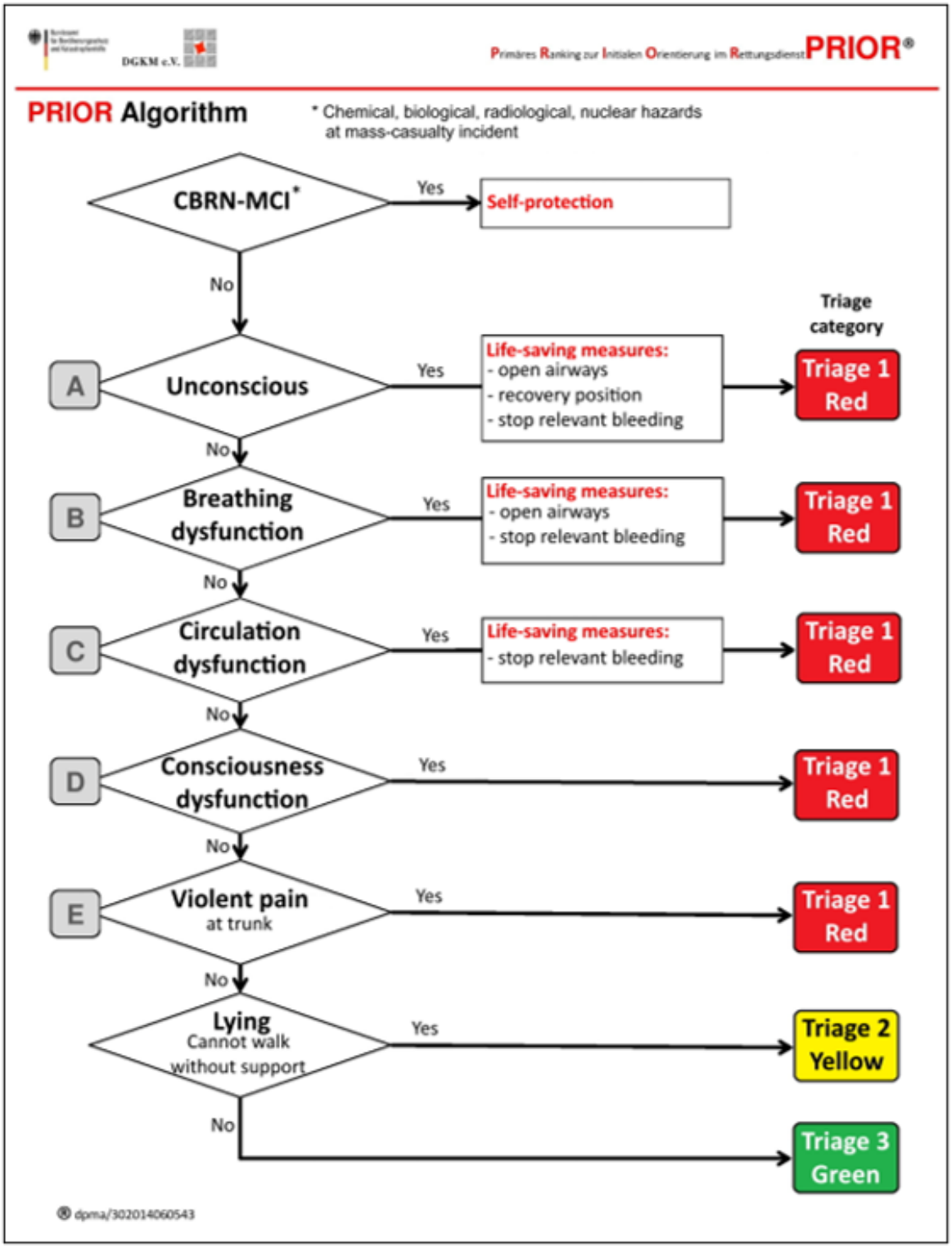
The objective of this study was to examine if smart glasses are in an inferior position compared to tablet computers in the case of guideline presentation. This study will evaluate if the application of a guideline displayed by smart glasses is performed much slower than that performed using a tablet computer and if the time required can be reduced by the learning effect after longer use. Another purpose of this study was to analyze if the test person's experience in emergency medicine has an influence on the time and accuracy of the triage. Especially in case of a mass casualty incident with shortage of staff, it is important to transfer medical staff from other departments with less experience in emergency medicine to the place of accident to support the experienced staff. Therefore, this study analyzes if staff with no or less experience in emergency or disaster medicine can also triage patients quickly and accurately.

Methods

Study Design and Triage

This study is a randomized controlled crossover simulation trial for guideline presentation. A so-called "PRIOR" (primary ranking for initial orientation in rescue service) triage algorithm [18] was applied and displayed on a tablet PC or smart glasses (Figure 1). By using the triage algorithm and following its flowchart, the test person's task was to determine 1 of the 3 triage categories for each of the 30 theoretical patient cases.

Figure 1. English translation of the algorithm for PRIOR (primary ranking for initial orientation in rescue service) triage [18].



First, the user of the PRIOR algorithm has to decide if there are any chemical, biological, radiological, or nuclear hazards concerning the injured person that should be triaged. If there are any of these hazards, the triaging person is instructed to set the focus on self-protection and not to triage the injured patient. This concept should prevent the triaging person from spreading the contamination. If these hazards do not exist, the test subject can triage the injured person by following the further paths of the algorithm. The main structure of the algorithm is based on the ABCDE-scheme (airway, breathing, circulation, disability, environment) that is applied in emergency medicine.

The user has to decide if the current criterion (eg, unconscious) can be applied concerning the injury of the triaged patient and choose between “Yes” or “No.” At the end of each path, the algorithm announces 1 of the 3 possible triage categories (red, yellow, or green), and the triage of 1 injured patient is finished. The 3 triage categories represent the urgency of a fast treatment of the injured person. An injured person that is triaged red

(emergency) should be treated immediately; yellow (urgent) and green (routine) triaged persons should be treated within 1 hour and 4 hours, respectively. [16].

Electronic Devices

The following 2 devices for technical support were used by the test subjects to triage 30 theoretical emergency patients after receiving approval from the local ethics committee (Aachen, Germany; EK320/16):

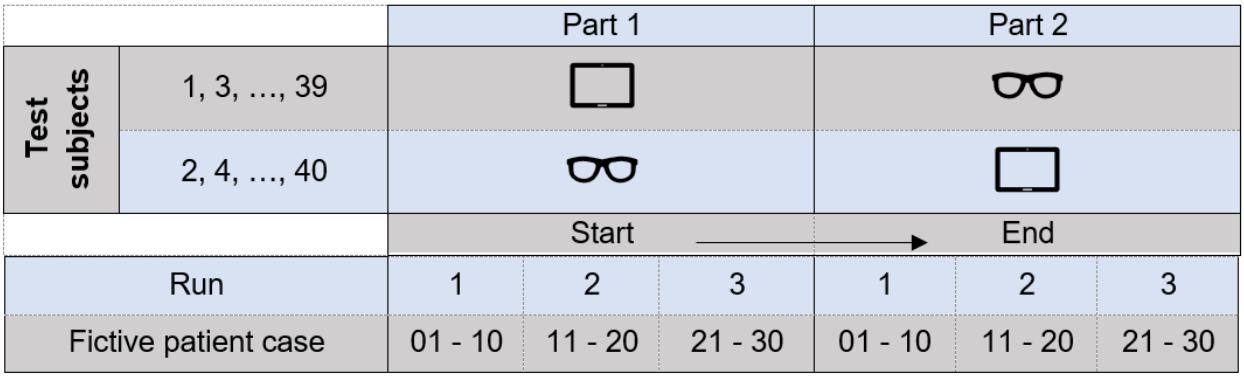
1. Tablet PC as an established control device (Samsung Galaxy Tab A 2016, Samsung Electronics AG)
2. Smart glasses as a new technical method of support (Recon Jet, Recon Instruments Corp).

The tablet PC was operated using a touchscreen, whereas the smart glasses were operated using several buttons (touchpad, switch with 2 buttons). The front switch is used to confirm a choice, while the back switch is used as the return key. Moreover, with the visual touchpad that is located at the right-hand outer side of the smart glasses, the marked choice can be changed by swiping with the index finger to the left or right.

Test Procedure

The triage time and the correct application of the guideline ("PRIOR") via smart glasses were analyzed and compared with the already established operating concepts of the tablet. Forty volunteer test subjects were included in this study by using the triage as an exemplary application of a guideline with the help of theoretical case descriptions. During the introduction, the test subjects were allowed to try out the relevant operating buttons on the smart glasses. Depending on the test subject number, which was randomly assigned, participants started either with the tablet (uneven subject number) or with the smart glasses (even subject number). First, the triage of 30 fictive patients was performed with either only the smart glasses (n=20) or only the tablet (n=20). After that, the test subjects performed the triage on the same patients in the identical order with the other electronic device (Figure 2). Moreover, after each run of 10 patient triages, a break of 1 minute was taken.

Figure 2. The test procedure in the crossover design with different devices: tablet PC (as a usual device for control group) and smart glasses.



During the trial, the test subjects only used the smart glasses or the tablet with the installed app, whereas the presentation of the cases, the monitoring of time, and the documentation of the correct use concerning the applied guideline was performed by the investigator. The fictive cases were presented on the screen of a desktop PC. The attended time of the guideline's application was defined as the primary outcome parameter, and the correct application was defined as the secondary outcome. The start of the time measurement was defined as the beginning of the presentation of the fictive patient and the end was defined by determining the triage category.

Fictive Patients in This Study

In this study, the cases designed for triage with technical support consisted of 30 different theoretical patients separated into 3 runs with 10 fictive patients each. The cases were text-based scenarios, and the text of each case was presented on a display of a desktop computer and the test persons had the task of reading the text of the case and to triage the patient that was described in the text. The description of the patients involved typical injuries that can be expected at a mass casualty incident, for example, the first run consisted of 10 patients with injuries because of a multi-vehicle accident. Each run contained the same triage results following the same paths in the decision tree for this guideline.

A comparison of all 3 runs (complexity of cases, triage categories, quantity of signs) for the subsequent evaluation was included to compare the triage results within the study process and between the devices. Different linguistic designs of comparable cases were used to mask the similarities preventing the recognition of comparable patients. The description of the cases consisted of short sentences, and no medical terms were used so that medical novices as well as medical experts could understand the description of the cases. In the following sentences, 3 exemplary descriptions of the theoretical patients that were applied in the study are listed with the triage category in brackets:

1. At a traffic collision, an involved person runs in your direction. He speaks normally and seems to be fully orientated (triage category: green).
2. The codriver of a passenger car is trapped in his car too. He cannot move his legs but he is awake and approachable. He breathes normally and has only little pain (triage category: yellow).

3. There is a trapped person under a seat of a bus. The person's breaths are very flat and very fast, but she is orientated (triage category: red).

Volunteers in This Study

The volunteers took part in this study in November and December 2017. They were recruited by an information letter with information about the workflow and intention of the study. The information letter was presented on a board where all public studies of the University Hospital of the RWTH (Rheinisch-Westfälische Technische Hochschule) Aachen are advertised. The subjects who were interested in participating in this study had to email or call the director of the study if they had any questions regarding the study or had the intention to take part in the study.

The participation in this study was voluntary and the participating students had no advantages or disadvantages by participating in this study along with their university studies. The only exclusion criterion of this study was the inability to use smart glasses because of visual impairment (inability to see the information on the display of the smart glasses or tablet screen clearly despite wearing glasses or contact lenses). Only 1 of the 41 test subjects was excluded because of visual impairment. Prior medical knowledge was not required because the used "PRIOR" triage algorithm can be applied by nonmedical subjects. Moreover, special technical knowledge or the application of technical devices in the past was also not an inclusion criterion. The test persons received a financial incentive of €10 (US $11.73).

Android App

In the Android app "AUDIME" (Augmented Disaster Medicine, Tech2go Mobile Systems GmbH), which was designed for application in smart glasses as well as on the tablet PC, individual decision criteria of the guidelines are displayed to support test subjects with the triage of the respective case. The test subject's task was to decide whether the respective criterion of the algorithm can be applied on the patient and to choose accordingly between "Yes" or "No." The app was programmed in a way that the next criterion is displayed only after the previous criterion has been answered. This stepwise processing of the exact decision tree was necessary to avoid skipping of questions. Regarding the operating of the "AUDIME" app, there were device-specific differences: with the tablet PC, questions of the algorithm were answered by one-time typing on the Yes button or No button on the tablet display (Figure 3). Because of the smart glasses' different operating concept, answers with reference to the questions of the algorithm were marked as "Yes" or "No" on the response fields. The marked answer could be changed by a one-time swipe on the touchpad of the smart glasses and chosen by a press on the confirmation key (Figure 4).

Figure 3. First question of the PRIOR (primary ranking for initial orientation in rescue service) triage algorithm on the user interface of the app on a tablet PC. The selection is done via a touch on the screen; answers are Yes or No. CBRN: chemical, biological, radiological, nuclear; MANV: mass casualty incident.

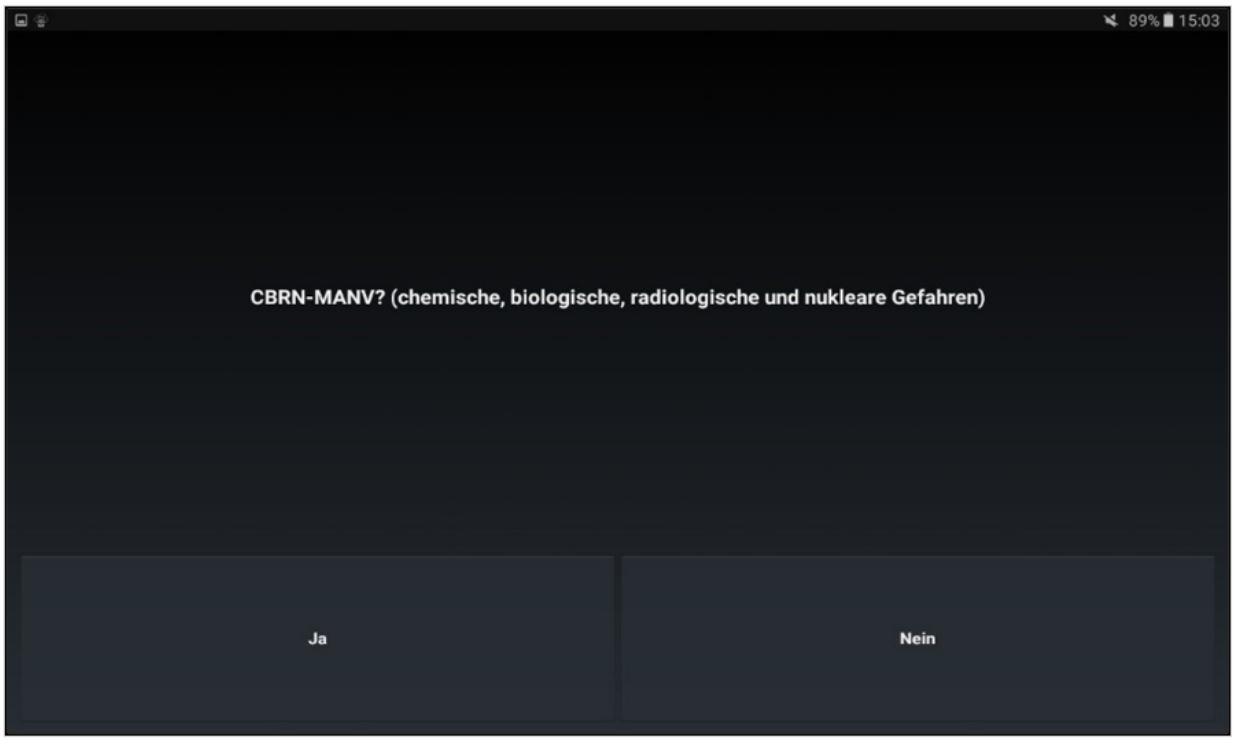
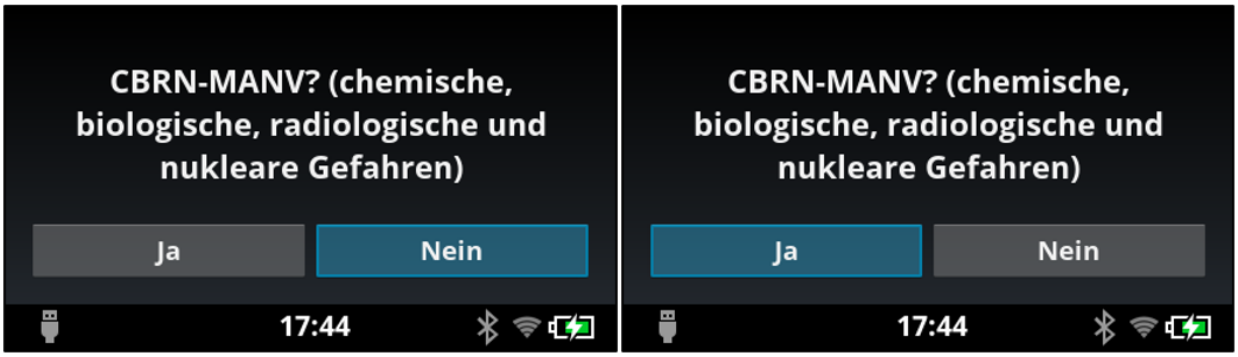


Figure 4. User interface of the app on the smart glasses. The answer (Yes or No) is chosen on an optical touchpad and confirmed with the confirmation key. CBRN: chemical, biological, radiological, nuclear; MANV: mass casualty incident.



Statistical Analysis

Data analysis was performed with a nonparametric distribution of the primary outcome parameters using the Mann-Whitney *U* test and the Wilcoxon test for independent samples (P=.05). SPSS Statistics 23.0.0.2 (IBM Corp) was used for statistical evaluation. All data included median or arithmetic mean, IQR, and 95% CI.

Results

Sample Characteristics

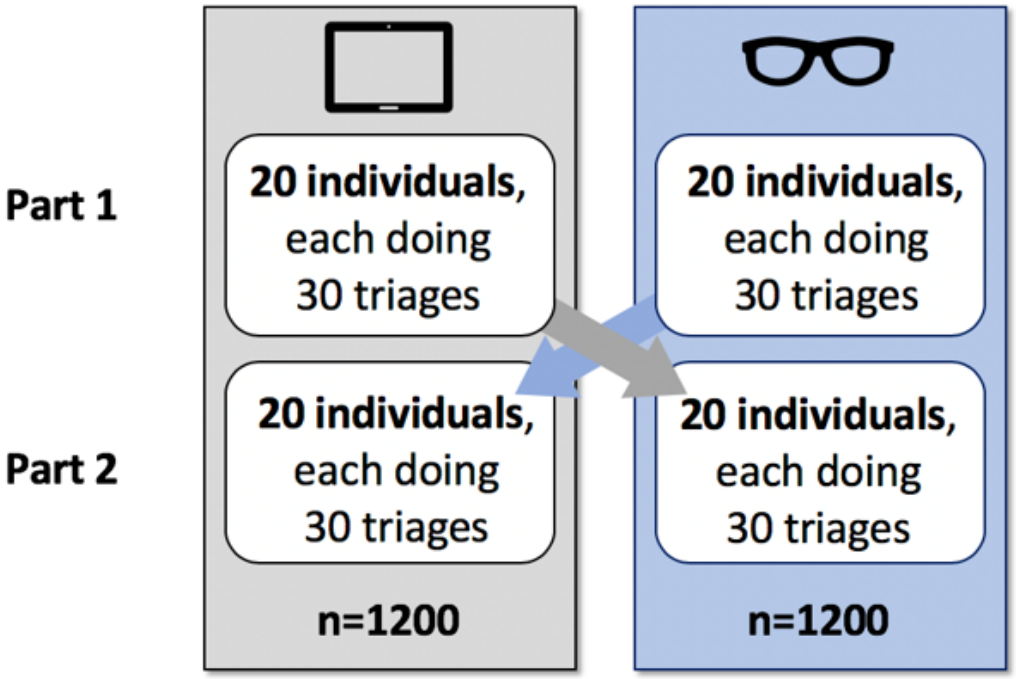
A total of 40 test subjects took part in this trial: 53% (21/40) were males and 48% (19/40) were females. The age of the test subjects ranged between 18 and 37 years, with a median age of 24 years (Table 1). Approximately 85% (34/40) of all the test subjects had medical knowledge because of their medical studies or a medical job. The proportion of students among all the participants was 93% (37/40); 29 volunteers were students of human medicine and 4 were students of dentistry. With regard to experience in emergency medicine, the subjects had the following experience: none (n=25), 1-100 hours (n=9), 101-1000 hours (n=5), and >1000 hours (n=1). None of the volunteers had experience with the use of smart glasses or the “AUDIME” Android app.

Table 1. Demographic data of the volunteers, depending on the device used in part 1.

Demographics	Tablet PC (n=20)	Smart glasses (n=20)
Females (n)	9	10
Males (n)	11	10
Age (years), median (IQR)	23.5 (19-37)	24.0 (18-35)

All 40 test subjects examined 30 fictive patients and applied the integrated guideline in each case by using both technical devices in a crossover, which resulted in 2400 triages: 1200 with the support of the smart glasses and 1200 with the support of the tablet PC (Figure 5).

Figure 5. Number of triages per device group: tablet PC (as a usual device for control group) and smart glasses.



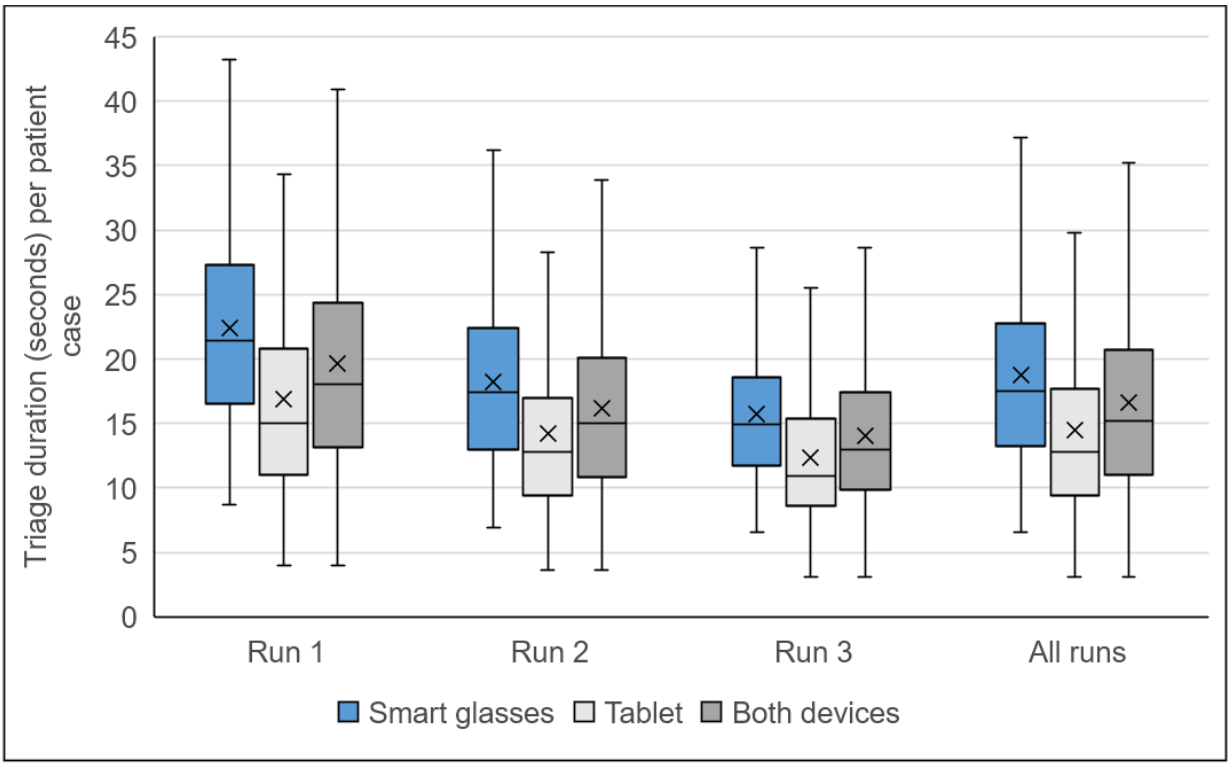
Duration of Triage

The median time to triage with smart glasses was 17.5 seconds (IQR 13.2-22.8, 95% CI 18.4-19.2) and that to triage with the tablet was 12.8 seconds (IQR 9.4-17.7, 95% CI 14.1-14.9). Thirty triages were subdivided into 3 sequential runs with 10 triages each. All 30 triages were finished with 1 device and subsequently with the other electronic device. The triage time within the same run with the tablet was significantly faster than that with the smart glasses for all 3 runs, which means that in total, triage using the tablet was significantly faster (Wilcoxon test, run 1: z=–10.178, P<.001; run 2: z=–10.373, P<.001; run 3: z=–10.697, P<.001; in total: z=–17.898, P<.001).

With the smart glasses, the median triage time was 21.5 seconds (IQR 16.5-27.3, 95% CI 21.6-23.2) in the first run, 17.4 seconds (IQR 13-22.4, 95% CI 17.6-18.9) in the second run, and 14.9 seconds (IQR 11.7-18.6, 95% CI 15.2-16.3) in the third run, whereas with the tablet, triage in the first, second, and third run took 15.0 seconds (IQR 11.0-20.8, 95% CI 16.1-17.7), 12.8 seconds (IQR 9.4-17.0, 95% CI 13.5-14.9), and 11.0 seconds (IQR 8.6-15.4, 95% CI 11.8-12.9), respectively.

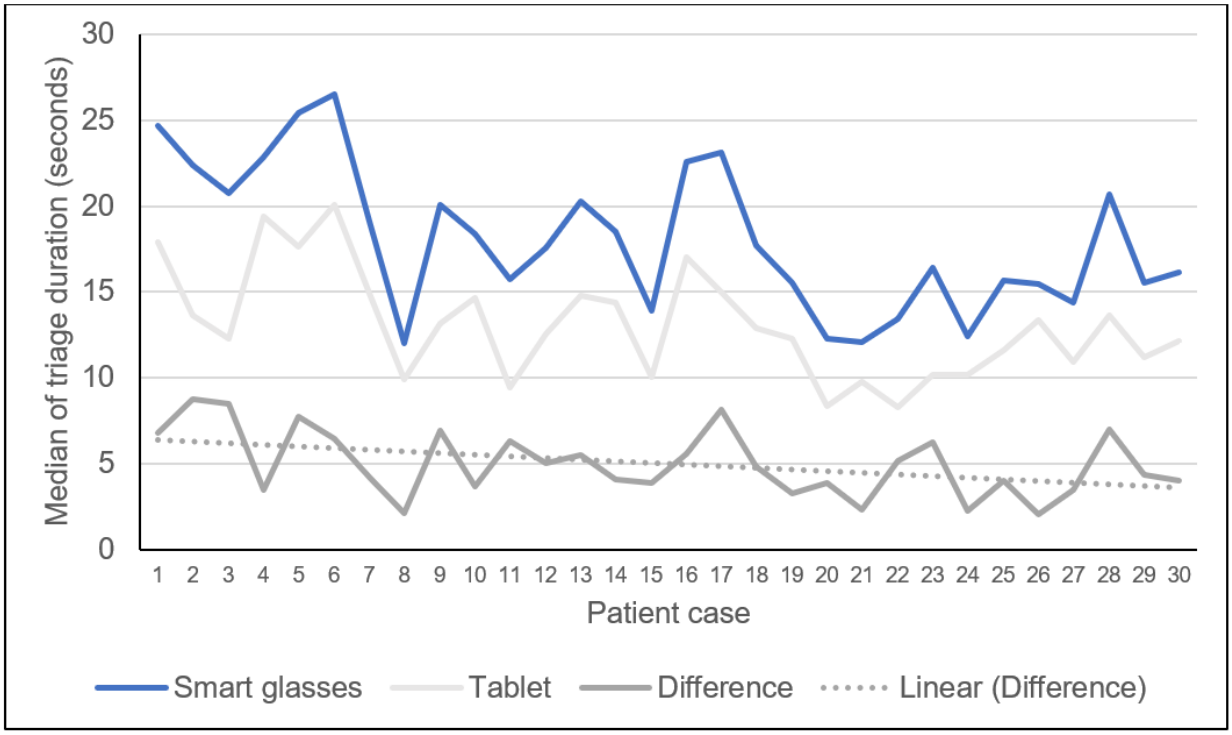
Both with the smart glasses and the tablet, significantly faster triage time was achieved in the subsequent runs compared to the triage time in the previous run (Wilcoxon test, run 1 vs 2 with smart glasses: z=–9.332, P<.001; run 2 vs 3 with smart glasses: z=–7.169, P<.001; run 1 vs 2 with tablet: z=–7.582, P<.001; run 2 vs 3 with tablet: z=–5.191, P<.001). Moreover, the scattering of the triage times in consideration of the interquartile range was less for both devices in the subsequent runs compared with that in the previous runs (Figure 6).

Figure 6. Box plot of triage duration (in seconds) per case over all trial parts depending on device and run.



With regard to the triage time for all the test subjects, the median triage time concerning all 30 patients with the tablet was lesser than that with the smart glasses (Figure 7). However, the time difference between both electronic devices decreased as the study progressed (with the increasing number of cases). Considering the respective medians, triage with the smart glasses took 5 seconds (IQR –0.9 to 11.8, 95% CI 4.5-6.5) in the first run, 4.1 seconds (IQR –0.3 to 8.5, 95% CI 3.3-4.8) in the second run, and only 3.5 seconds (IQR –0.2 to 7.0, 95% CI 2.8-4.0) longer than that with the tablet in the last run. Here, considering the difference in the triage time between both devices, the additional time needed with the smart glasses could be reduced significantly comparing the subsequent runs to its previous run (Wilcoxon test: run 1 vs 2, z=–3.142, P=.002; run 1 vs 3: z=–4.805, P<.001; run 2 vs 3: z=–2.181, P=.03).

Figure 7. Line diagram showing the medians of triage duration (in seconds) depending on device and case as well as the calculated time difference between both devices.



The first triage of all 30 fictive patients (before changing the device) with the used device showed a median time of 17.7 seconds (IQR 13.5-23.1, 95% CI 18.6-19.5), whereas the retriage (after device change) with the other device required 12.6 seconds (IQR 9.5-17.4, 95% CI 13.8-14.6) and was therefore significantly faster (Wilcoxon test: z=–20.090, P<.001). Similar results appeared comparing the triage time of the same device depending on the first or second application for triage by the test subject. That way, triage with smart glasses took about 3.5 seconds and triage with the tablet took almost 6 seconds longer compared to those test subjects who started their triage with this device with those test subjects who used this device secondly (Mann-Whitney U test, smart glasses: z=–8.696, P<.001; tablet: z=–16.372, P<.001). Thus, triage with smart glasses took the longest time in the case of the triage started with the smart glasses (median 19.1 seconds, IQR 15.3-24.7, 95% CI 19.8-21.0). Triage using the tablet was faster after having used the smart glasses (median 10.2 seconds, IQR 8.0-13.6, 95% CI 10.9-11.6). Moreover, test persons with experience in emergency medicine (n=15) needed a median time of 14.5 seconds (IQR 10.3-19.5, 95% CI 15.1-16.0) to triage, whereas inexperienced volunteers (n=25) were significantly slower (Mann-Whitney U test: z=–5.018, P<.001) with a median of 15.7 seconds (IQR 11.3-21.5, 95% CI 16.9-17.7).

Accuracy of the Devices

Approximately 86.3% (95% CI 0.84-0.88) of all triages were finished correctly by using smart glasses and 85.6% (95% CI 0.84-0.88) of all triages were finished correctly with the application of the tablet (Table 2). Within the same run as well as in total, there was no significant difference between both devices and their triage reliability (Wilcoxon test: run 1, z=–0.452, P=.65; run 2, z=–0.000, P>.99; run 3, z=–0.696, P=.46; in total, z=–0.727, P=.47). The triage accuracy with both devices during the trial showed the following values: 85% accuracy in the first run (95% CI 0.83-0.88), 82% accuracy in the second run (95% CI 0.80-0.85), and 90% accuracy in the third run (95% CI 0.88-0.92).

Table 2. Triage accuracy with the used device and runs.

Runs	Accuracy with smart glasses (%), mean (95% CI)	Accuracy with tablet PC (%), mean (95% CI)
Run 1	85.0 (81.5-88.5)	85.8 (82.3-89.2)
Run 2	82.3 (78.5-86.0)	82.3 (78.5-86.0)
Run 3	89.5 (86.5-92.5)	90.8 (87.9-93.6)
Total	86.3 (84.3-88.2)	85.6 (83.6-87.6)

Consequently, triage reliability was the highest in the last run. Furthermore, test subjects using smart glasses had a significantly more reliable triage in the third run than that in the first 2 runs (Wilcoxon test: run 1 vs 3, z=2.025, P=.04; run 2 vs 3, z=3.558, P<.001) and were significantly more precise with the tablet in the third run than that in the second run (Wilcoxon test: run 2

vs 3, z=–3.057, P=.002). Test persons with experience in emergency medicine (n=15) performed with an average accuracy of 91.6% (95% CI 0.90-0.93), whereas the inexperienced volunteers (n=25) triaged significantly less accurately (Mann-Whitney U test: z=–6.200, P<.001) with an accuracy of 82.5% (95% CI 0.81-0.84).

Discussion

Principal Results

In this study, a significantly faster triage time was achieved with the tablet compared to that with the smart glasses in all 3 runs as well as in total. While triage with the tablet took a median of 12.8 seconds, that with the smart glasses took a median of 17.5 seconds. Consequently, triage with the tablet was approximately 27% faster. Repetitive usage led to a learning curve with steadily decreasing time consumption. The accuracy of the triages was not different between using smart glasses or the tablet.

The differences between using both the devices can be attributed to the fact that the different operating concepts of both the devices have an enormous influence on the triage time and thus show great potential in the use of clinical guidelines. Therefore, future developments of these electronic devices should place a special focus on an easier operating concept to enable faster decision-making. A comparison of the triage time of smart glasses with triage times in other studies is not possible, although there have been studies with smart glasses in use without recordings of the triage times [16,19].

During this study, test subjects using both devices achieved an increasingly faster triage time as they became familiar with the guidelines and thus, the triage of patients. In addition, the use of smart glasses needed the subjects to adapt to an unfamiliar operating system, which later led to a constant reduction in the time difference between using smart glasses and tablet in numerous triages. In contrast, an adaptation to the tablet's operation could be neglected owing to an intuitive operation. Taking the respective medians into consideration, triage with smart glasses took 6.5 seconds longer in the first run, 4.7 seconds in the second run, and 4.0 seconds longer in the third run compared to triage with the tablet. However, the time difference between both the devices could be reduced in the following runs as the test subjects became familiar with the operation of the smart glasses and finished triages faster.

The triage result using the guideline with smart glasses was correct in 86.3% of the fictive cases and that with the tablet was correct in 85.6% of the cases. Thus, the operation concepts of the devices within the same test conditions did not have a significant impact on the triage accuracy. In other triage studies with electronic triage support (ie, personal digital assistant, tablets, or computer), a comparable triage accuracy of 79%-83% has been achieved. However, only 1 electronic device was used in these studies, and therefore, the effect of the operation concept on the reliability of triage could not be differentiated [20,21].

Volunteers with experience in emergency medicine achieved significantly (P<.001) faster and more reliable triage results (median 14.5 seconds, 91.6% accuracy) compared to the inexperienced test subjects (median 15.7 seconds, 82.5% accuracy). Therefore, this study showed that the time to triage and the accuracy depends on the experience in emergency medicine of the triaging person. Nevertheless, our study shows that subjects inexperienced in emergency medicine knowledge are also able to triage at a mass casualty incident by using electronic devices and the PRIOR triage algorithm because they also achieved fast and accurate triage results. Owing to the good triage results of the test subjects who did not have experience in disaster or emergency medicine, it was not a methodical flaw that these volunteers took part in the study, and experience in emergency medicine was not defined as an inclusion criterion in this study.

This study shows the exemplary display of standardized treatment cords in augmented reality and the related information transfer. Thus, this study proves that standardized algorithms can be displayed with the help of electronic devices and therefore enable users to have direct access to compressed and situation-related information. This is especially relevant for medical care as well as other departments in which standardized guidelines exist and simultaneously targeted effective action is of crucial importance. The use of electronic-based decision-making systems such as smart glasses can help in this regard to quickly overcome the initial alienation and complexity of many clinical guidelines and enable an effective implementation in clinical patient care. In addition, it has been shown that by using electronic support, even inexperienced persons become capable of gradually making decisions within these guidelines, which makes the use of these devices also suitable for education purposes.

The operating concept of smart glasses is still unfamiliar compared to the simple touchscreen on tablet computers and smartphones. Therefore, it is recommended that paramedics use smart glasses and the triage algorithm regularly to triage emergency patients to receive frequent practice with the use of the triage algorithm and electronic device. For this reason, the implementation of further guidelines and algorithms that are useful for paramedics in all daily emergency medicine is preferable. In addition, paramedics who use these electronic devices and guidelines daily are better prepared for the use of these technologies at a mass casualty incident. If smart glasses are developed further and become widely spread, augmented reality will become a superb opportunity for guideline-conforming operation with free usable hands. In addition to the presentation of guidelines, they can record relevant information by using an integrated camera or can be used as a tool for telemedical support.

Limitations

A limiting factor for the presentation of guidelines with smart glasses is the low battery life of many models. Moreover, numerous models, including the model used in this study, are not suited for spectacle wearers. The operating concept of the used smart glasses with the optical touchpad and the switch is surely complex in the beginning, but easier user interfaces can be expected in the future. Another limitation of our study design is the adaptation to the guideline apart from the adaptation to the electronic devices. These parameters cannot be evaluated

separately; therefore, the influence of the adaptation to the guideline on the speed of operation cannot be calculated. Through the preparation of comparable theoretical patient cases in 3 runs, it was possible to simulate and analyze the time and correctness of the triage. It certainly would have been more realistic to use, for example, a real disaster operation, though this would hardly be viable compared to a simulation. In this case, the examination of an alternative guideline that can be implemented into a real work environment can be planned.

Future studies should also focus on the age and experience of the test subjects, as this study is limited by test subjects' average age of 24 years, which is certainly a young age and includes an already existing experience and familiarity with the developed technological devices. As medical staff consists of people of various ages and age generations, future studies should also evaluate the influence of the age generation on the usage of electronic devices for medical algorithm and decision-making. Further, experienced paramedics or other persons that become more familiar with the algorithm because of many triages might not strictly use a triage algorithm but use it only in a case of need. Therefore, future studies should analyze how often paramedics use a triage algorithm in reality and if the use of the triage algorithm in the course of many triages is reduced because they become more familiar with the triage (algorithm).

Conclusion

In summary, the presentation of a guideline on a tablet as well as in the form of augmented reality achieved good results even with inexperienced users. The exemplary guideline was securely implemented. The implementation using smart glasses took more time owing to its more complex operating concept but could be accelerated in the course of the study after adaptation. The tablet, however, with its touchscreen as a familiar operating concept, achieved faster results. This operating concept has nevertheless the disadvantage of the use of both hands, which can be especially relevant in departments with special hygienic requirements or a high proportion of manual occupation. Especially in a non–time-critical working area, a guideline presentation with smart glasses can be implemented and a free-hand occupation can be executed. Further development of the operating concepts in the future or a highly intense training can speed up the processes, thereby making smart glasses a usable tool for guideline-conforming technical decision support.

Acknowledgments

Funding for the AUDIME project was provided by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF).

Authors' Contributions

AF, AR, MG, and MC conceived the study, designed the trial, and obtained research funding. AF, AR, RR, and MC supervised the conduct of the trial and data collection. AF and MC managed the data, including quality control. MF and RR provided statistical advice on the study design. AF, AR, and MC analyzed the data. AF drafted the manuscript and all authors contributed substantially to its revision. AF takes responsibility for the paper as a whole.

Conflicts of Interest

MG was the CEO of Tech2go Mobile Systems GmbH, Hamburg, Germany at the time of the study and has since left the company.

References

1. Brouwers MC, Florez ID, McNair SA, Vella ET, Yao X. Clinical Practice Guidelines: Tools to Support High Quality Patient Care. Semin Nucl Med 2019 Mar;49(2):145-152. [doi: 10.1053/j.semnuclmed.2018.11.001] [Medline: 30819394]
2. McKee G, Kerins M, Hamilton G, Hansen T, Hendriks J, Kletsiou E, et al. Barriers to ESC guideline implementation: results of a survey from the European Council on Cardiovascular Nursing and Allied Professions (CCNAP). Eur J Cardiovasc Nurs 2017 Dec;16(8):678-686. [doi: 10.1177/1474515117710097] [Medline: 28498092]
3. Francke AL, Smit MC, de Veer AJ, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. BMC Med Inform Decis Mak 2008 Sep 12;8:38 [FREE Full text] [doi: 10.1186/1472-6947-8-38] [Medline: 18789150]
4. Aigner W, Kaiser K, Miksch S. Visualization methods to support guideline-based care management. Stud Health Technol Inform 2008;139:140-159 [FREE Full text] [Medline: 18806325]
5. Heselmans A, Van de Velde S, Donceel P, Aertgeerts B, Ramaekers D. Effectiveness of electronic guideline-based implementation systems in ambulatory care settings - a systematic review. Implement Sci 2009 Dec 30;4:82 [FREE Full text] [doi: 10.1186/1748-5908-4-82] [Medline: 20042070]
6. 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]
7. Dougherty B, Badawy SM. Using Google Glass in Nonsurgical Medical Settings: Systematic Review. JMIR Mhealth Uhealth 2017 Oct 19;5(10):e159 [FREE Full text] [doi: 10.2196/mhealth.8671] [Medline: 29051136]
8. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google Glass in Surgical Settings: Systematic Review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]

9. Datta N, MacQueen IT, Schroeder AD, Wilson JJ, Espinoza JC, Wagner JP, et al. Wearable Technology for Global Surgical Teleproctoring. J Surg Educ 2015;72(6):1290-1295. [doi: 10.1016/j.jsurg.2015.07.004] [Medline: 26276303]
10. Chimenti PC, Mitten DJ. Google Glass as an Alternative to Standard Fluoroscopic Visualization for Percutaneous Fixation of Hand Fractures. Plastic and Reconstructive Surgery 2015;136(2):328-330. [doi: 10.1097/prs.0000000000001453]
11. Vorraber W, Voessner S, Stark G, Neubacher D, DeMello S, Bair A. Medical applications of near-eye display devices: an exploratory study. Int J Surg 2014 Dec;12(12):1266-1272 [FREE Full text] [doi: 10.1016/j.ijsu.2014.09.014] [Medline: 25462704]
12. Zupon A, Rothenberg C, Couturier K, Tan T, Siddiqui G, James M, et al. An appraisal of emergency medicine clinical practice guidelines: Do we agree? Int J Clin Pract 2019 Feb;73(2):e13289 [FREE Full text] [doi: 10.1111/ijcp.13289] [Medline: 30372798]
13. Venkatesh AK, Savage D, Sandefur B, Bernard KR, Rothenberg C, Schuur JD. Systematic review of emergency medicine clinical practice guidelines: Implications for research and policy. PLoS One 2017;12(6):e0178456 [FREE Full text] [doi: 10.1371/journal.pone.0178456] [Medline: 28628660]
14. Munzer BW, Khan MM, Shipman B, Mahajan P. Augmented Reality in Emergency Medicine: A Scoping Review. J Med Internet Res 2019 Apr 17;21(4):e12368 [FREE Full text] [doi: 10.2196/12368] [Medline: 30994463]
15. Broach, John, Alexander H, Matthew G, Jeffrey L, Edward WB, et al. Usability and reliability of smart glasses for secondary triage during mass casualty incidents. 2018 Jan 03 Presented at: Annual Hawaii International Conference on System Sciences; 2018; Hawaii p. 1416. [doi: 10.24251/hicss.2018.175]
16. Carenzo L, Barra FL, Ingrassia PL, Colombo D, Costa A, Della Corte F. Disaster medicine through Google Glass. Eur J Emerg Med 2015 Jun;22(3):222-225. [doi: 10.1097/MEJ.0000000000000229] [Medline: 25460812]
17. Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M. Technical Support by Smart Glasses During a Mass Casualty Incident: A Randomized Controlled Simulation Trial on Technically Assisted Triage and Telemedical App Use in Disaster Medicine. J Med Internet Res 2019 Jan 03;21(1):e11939. [doi: 10.2196/11939] [Medline: 30609988]
18. Bubser F, Callies A, Schreiber J. PRIOR: Vorsichtungssystem für Rettungsassistenten und Notfallsanitäter. Rettungsdienst. 2014. URL: https://docplayer.org/9972864-Prior-vorsichtungssystem-fuer-rettungsassistenten.html [accessed 2021-09-20]
19. Cicero MX, Walsh B, Solad Y, Whitfill T, Paesano G, Kim K, et al. Do You See What I See? Insights from Using Google Glass for Disaster Telemedicine Triage. Prehosp. Disaster med 2015 Jan 09;30(1):4-8. [doi: 10.1017/s1049023x1400140x]
20. Bolduc C, Maghraby N, Fok P, Luong TM, Homier V. Comparison of Electronic Versus Manual Mass-Casualty Incident Triage. Prehosp. Disaster med 2018 Apr 17;33(3):273-278. [doi: 10.1017/s1049023x1800033x]
21. Twomey M, Wallis L, Thompson M, Myers J. The South African triage scale (adult version) provides valid acuity ratings when used by doctors and enrolled nursing assistants. African Journal of Emergency Medicine 2012 Mar;2(1):3-12 [FREE Full text] [doi: 10.1016/j.afjem.2011.08.014]

Abbreviations

AUDIME: Augmented Disaster Medicine
PRIOR: primary ranking for initial orientation in rescue service

Edited by L Buis; submitted 14.12.19; peer-reviewed by P Bradl, B Metelmann, E Melnick; comments to author 26.08.20; revised version received 21.10.20; accepted 15.07.21; published 18.10.21.

Please cite as:
Follmann A, Ruhl A, Gösch M, Felzen M, Rossaint R, Czaplik M
Augmented Reality for Guideline Presentation in Medicine: Randomized Crossover Simulation Trial for Technically Assisted Decision-making
JMIR Mhealth Uhealth 2021;9(10):e17472
URL: https://mhealth.jmir.org/2021/10/e17472
doi:10.2196/17472
PMID:34661548

Corrigenda and Addenda

Correction: A Mobile App (FallSA) to Identify Fall Risk Among Malaysian Community-Dwelling Older Persons: Development and Validation Study

Devinder Kaur Ajit Singh[1*], PhD; Jing Wen Goh[1*], BSc; Muhammad Iqbal Shaharudin[1,2*], MSc; Suzana Shahar[1*], PhD

[1]Center for Healthy Ageing and Wellness, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[2]Faculty of Health Sciences, Cawangan Pulau Pinang, Kampus Bertam, Universiti Teknologi Majlis Amanah Rakyat, Penang, Malaysia

[*]all authors contributed equally

Corresponding Author:
Devinder Kaur Ajit Singh, PhD
Center for Healthy Ageing and Wellness
Faculty of Health Sciences
Universiti Kebangsaan Malaysia
Jalan Raja Muda Abdul Aziz
Kuala Lumpur, 50300
Malaysia
Phone: 60 392897000
Email: devinder@ukm.edu.my

Related Article:

Correction of: https://mhealth.jmir.org/2021/10/e23663

(JMIR Mhealth Uhealth 2021;9(10):e34368) doi:10.2196/34368

In "A Mobile App (FallSA) to Identify Fall Risk Among Malaysian Community-Dwelling Older Persons: Development and Validation Study" (JMIR Mhealth Uhealth 2021;9(10):e23663), the authors noted one error.

In the originally published manuscript, the name of author *Devinder Kaur Ajit Singh* was formatted incorrectly as the given name "Devinder" and surname "Kaur Ajit Singh." This has now been corrected to the given names "Devinder Kaur Ajit" and surname "Singh." The citation information for this author has also been corrected from "Kaur Ajit Singh D" to "Singh DKA."

The correction will appear in the online version of the paper on the JMIR Publications website on October 28, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 19.10.21; this is a non–peer-reviewed article;accepted 20.10.21; published 28.10.21.

Please cite as:
Singh DKA, Goh JW, Shaharudin MI, Shahar S
Correction: A Mobile App (FallSA) to Identify Fall Risk Among Malaysian Community-Dwelling Older Persons: Development and Validation Study
JMIR Mhealth Uhealth 2021;9(10):e34368
URL: https://mhealth.jmir.org/2021/10/e34368
doi:10.2196/34368
PMID:

Original Paper

Digital Biomarkers for Depression Screening With Wearable Devices: Cross-sectional Study With Machine Learning Modeling

Yuri Rykov[1], PhD; Thuan-Quoc Thach[2], PhD; Iva Bojic[3], PhD; George Christopoulos[4], PhD; Josip Car[3,5], MD, PhD

[1]Neuroglee Therapeutics, Singapore, Singapore

[2]Department of Psychiatry, The University of Hong Kong, Hong Kong SAR, China (Hong Kong)

[3]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Division of Leadership, Management and Organisation, Nanyang Business School, College of Business, Nanyang Technological University, Singapore, Singapore

[5]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Josip Car, MD, PhD
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Rd, Clinical Sciences Building
Level 18
Singapore, 308232
Singapore
Phone: 65 +85291725838
Email: josip.car@gmail.com

Abstract

Background: Depression is a prevalent mental disorder that is undiagnosed and untreated in half of all cases. Wearable activity trackers collect fine-grained sensor data characterizing the behavior and physiology of users (ie, digital biomarkers), which could be used for timely, unobtrusive, and scalable depression screening.

Objective: The aim of this study was to examine the predictive ability of digital biomarkers, based on sensor data from consumer-grade wearables, to detect risk of depression in a working population.

Methods: This was a cross-sectional study of 290 healthy working adults. Participants wore Fitbit Charge 2 devices for 14 consecutive days and completed a health survey, including screening for depressive symptoms using the 9-item Patient Health Questionnaire (PHQ-9), at baseline and 2 weeks later. We extracted a range of known and novel digital biomarkers characterizing physical activity, sleep patterns, and circadian rhythms from wearables using steps, heart rate, energy expenditure, and sleep data. Associations between severity of depressive symptoms and digital biomarkers were examined with Spearman correlation and multiple regression analyses adjusted for potential confounders, including sociodemographic characteristics, alcohol consumption, smoking, self-rated health, subjective sleep characteristics, and loneliness. Supervised machine learning with statistically selected digital biomarkers was used to predict risk of depression (ie, symptom severity and screening status). We used varying cutoff scores from an acceptable PHQ-9 score range to define the depression group and different subsamples for classification, while the set of statistically selected digital biomarkers remained the same. For the performance evaluation, we used k-fold cross-validation and obtained accuracy measures from the holdout folds.

Results: A total of 267 participants were included in the analysis. The mean age of the participants was 33 (SD 8.6, range 21-64) years. Out of 267 participants, there was a mild female bias displayed (n=170, 63.7%). The majority of the participants were Chinese (n=211, 79.0%), single (n=163, 61.0%), and had a university degree (n=238, 89.1%). We found that a greater severity of depressive symptoms was robustly associated with greater variation of nighttime heart rate between 2 AM and 4 AM and between 4 AM and 6 AM; it was also associated with lower regularity of weekday circadian rhythms based on steps and estimated with nonparametric measures of interdaily stability and autocorrelation as well as fewer steps-based daily peaks. Despite several reliable associations, our evidence showed limited ability of digital biomarkers to detect depression in the whole sample of working adults. However, in balanced and contrasted subsamples comprised of depressed and healthy participants with no risk of depression (ie, no or minimal depressive symptoms), the model achieved an accuracy of 80%, a sensitivity of 82%, and a specificity of 78% in detecting subjects at high risk of depression.

Conclusions: Digital biomarkers that have been discovered and are based on behavioral and physiological data from consumer wearables could detect increased risk of depression and have the potential to assist in depression screening, yet current evidence shows limited predictive ability. Machine learning models combining these digital biomarkers could discriminate between individuals with a high risk of depression and individuals with no risk.

(JMIR Mhealth Uhealth 2021;9(10):e24872) doi:10.2196/24872

KEYWORDS

depression; digital biomarkers; screening; wearable electronic device; fitness tracker; circadian rhythm; rest-activity rhythm; heart rate; machine learning

Introduction

Background

Depression is the third-largest contributor to years lost to disability and affects 264 million people globally [1]. Despite its high prevalence, depression remains undiagnosed and untreated in half of all cases [2,3]. At the same time, the evolving COVID-19 pandemic and related economic crisis are worsening the population's mental health [4-6].

Wearable activity trackers are increasingly widely used and provide an opportunity to harness sensor data for detection of different health conditions [7]. Wearable trackers can monitor physiological and behavioral data, including steps, heart rate, energy expenditure, sleep patterns, respiration rate, blood oxygen saturation, skin temperature, and skin conductance, among others. Close interrelationships between everyday behavior, physiology, and mental well-being makes digital phenotyping with wearables especially attractive for detection of mental disorders and discovering respective digital biomarkers [8-11]. These digital biomarkers could be used for risk prediction of depression and to scale up population mental health screening. Moreover, due to the unprecedented granularity of available data, digital phenotyping can advance our understanding of etiology and subtypes of mental disorders and complement established diagnostic criteria. Complementary smartphone apps could be further used for prevention and treatment of mental disorders, to deliver digital health interventions and personalized cognitive behavioral therapy [12-14]. In this work, we focus on digital biomarkers based on wearable sensor data for depression screening.

Related Work

Current evidence is mainly comprised of studies investigating separate associations between depressive symptomatology and various actigraphy and sensor-based metrics, which capture meaningful aspects of behavior and physiology, to reveal objective risk or diagnostic markers. These aspects include physical activity, sleep characteristics, circadian rhythms, and physiological parameters. Systematic reviews and meta-analyses of actigraphy studies demonstrated that patients with depression were significantly less physically active than healthy controls [15,16]. Meta-analyses of prospective cohort studies showed that physical activity had a protective effect against the future onset of depression [17], while sedentary time increased the risk of depression [18]. For example, odds of depression development were 1.8 to 2.7 times lower in participants with more time spent in moderate to vigorous physical activity compared to the least active participants [19].

Sleep disturbance is another common symptom and risk factor for depression that can be measured objectively together with other sleep characteristics. Meta-analyses of polysomnography (PSG) studies showed that depression is associated with less total sleep time (TST), increased sleep onset latency (SOL), lower sleep efficiency (SE), lower duration and fraction of slow-wave sleep, lower fraction of light sleep, reduced latency between sleep onset and onset of rapid eye movement (REM) sleep, greater duration and fraction of REM sleep, greater number of awakenings, and increased wake after sleep onset (WASO) time compared to healthy controls [20,21]. Although PSG is a gold standard for sleep measurement, contemporary wearables can measure sleep and identify sleep stages. A systematic review of actigraphy studies suggested that the most persistent sleep abnormality in depressive patients is longer WASO, whereas other parameters had mixed effects [16]. Separate actigraphy studies showed lower SE, longer SOL, later sleep offset [22,23], and a greater sleep fragmentation in adults with depression [24].

Contrary to measuring physical activity and sleep as separate factors, circadian rhythms characterize the activity pattern of a full 24-hour cycle. Circadian rhythm metrics quantify regularity, shape, and timing of repeated daily processes and can be derived from fine-grained wearable data. Actigraphy studies suggested that people with clinically significant depressive symptoms had lower mesor, a rhythm-adjusted mean activity level [24-26]; reduced total locomotor activity [27,28]; lower amplitude [25,26]; shorter active periods [29,30]; delayed acrophase [22,30]; and less robust circadian activity rhythms [22,25,31]. Longitudinal studies showed that lower rhythm robustness predicted worsening of depressive symptoms in the future [26]. Depressive symptoms were also associated with heart rate–based mesor [32] and amplitude [32] as well as increased heart rate in the night and morning hours [33]. Few studies using nonparametric rhythm measures demonstrated that participants with more severe depressive symptoms had lower interdaily stability (IS) and higher intradaily variability (IV) of circadian activity rhythm [31,34], reduced relative amplitude (RA) between the highest and the lowest activity levels [28,35], and reduced RA of skin temperature [36]. Although findings on separate circadian rhythm indicators are mixed, the evidence suggests that depression is often associated with disturbed, irregular, and delayed circadian behavioral and physiological rhythms.

Another stream of evidence is comprised of a few studies that developed predictive models for depression detection from multimodal wearable data using data-driven approaches and machine learning. For example, Jacobson et al extracted 9929 digital markers using signal processing methods from a publicly available actigraphy data set of 55 individuals to detect people with depressive disorder [37]. A machine learning model correctly predicted the diagnostic status in 89% of the cases with a sensitivity of 94% and a specificity of 83% using leave-one-cohort-out cross-validation. Only spectral density features were important predictors [38]. In another study, authors extracted distribution characteristics of wearable data, including acceleration, skin conductance, and temperature, collected for 1 month and sampled from different time windows within 24-hour cycles, which resulted in 204 features in total [39]. A machine learning model with this feature set reached 87% accuracy in classifying 47 participants with high or low mental health scores using leave-one-out cross-validation. Nighttime skin conductance features were most important in prediction models. Tazawa et al extracted distribution characteristics of per-hour data from multidimensional wearable data, including steps, energy expenditure, body motion, heart rate, skin temperature, sleep time, and ultraviolet light, which resulted in 63 features [33]. A machine learning model with features based on a 7-day period achieved an accuracy of 76%, a sensitivity of 73%, and a specificity of 79% in prediction of depression screening status in 86 participants using 10-fold cross-validation. Authors found that features of skin temperature and sleep were most important for the model's predictive ability.

Objectives

The first stream of evidence from actigraphy and PSG studies lacks research on the predictive ability of various metrics to assess the risk of depression, whereas the other stream of data-driven studies often ignores established risk markers and leaves associations between depressive symptomatology and wearable data uninterpreted. Absence of robust and interpretable digital biomarkers complicates further development of a comprehensive and explainable algorithm for depression screening in the general population. To address these gaps, we examined the associations between depressive symptomatology and digital biomarkers based on sensor data from consumer-grade fitness trackers, including established and novel markers, and explored the predictive ability of these digital biomarkers in depression screening.

Methods

Study Design and Participants

In a cross-sectional study, we collected data from a multiethnic working population in Singapore. A total of 290 adult volunteers (aged ≥21 years) among full-time employees of Nanyang Technological University (NTU) were recruited to participate in the study from August to October 2019. Participants responded to online questionnaires and wore activity trackers for at least 14 days. Subjects received financial compensation for participating in the study. The study protocol and informed consent form were approved by the NTU Institutional Review Board (application reference: RB-2016-03-033).

Fitbit Charge 2 devices—consumer-grade fitness trackers—were used for data collection. The accuracy of Fitbit data has been investigated in several studies [40-44]. According to a systematic review [42], studies have consistently indicated that Fitbit devices were likely to provide accurate measures of daily step counts. However, energy expenditure is less accurately estimated by Fitbit devices. In general, Fitbit wearables have similar accuracy for activity assessment as research-grade devices, but they overestimate moderate to vigorous physical activity in free-living conditions. According to a recent meta-analysis, sleep-staging Fitbit devices, such as the Charge 2, showed no significant difference in measured values of WASO, TST, and SE, but they slightly underestimated SOL in comparison to the gold standard PSG [45]. Also sleep-stage transition dynamics measured by Fitbit devices were found to be inaccurate compared to PSG [46]. Participants were instructed to wear trackers all the time and to remove them only when taking a shower or charging. Data from trackers were synchronized with the Fitbit mobile app and further transferred to the Fitabase server.

Depression Screening and Self-reported Covariates

We used validated self-report questionnaires for depression screening and to collect sociodemographic, lifestyle, and health characteristics. REDCap (Research Electronic Data Capture), the research-grade online survey platform, was used for survey administration. The 9-item Patient Health Questionnaire (PHQ-9) was used for depression screening. Participants completed the PHQ-9 at the beginning and at the end of the observation period. We used different cutoff points and approaches to define provisionally depressed participants in the analysis based on common standards and the actual distribution of scores across two assessments. The average PHQ-9 score of two assessments was used in statistical analysis.

We collected a range of covariates to control for possible confounders. Demographics, including age, gender, ethnic group (ie, Chinese, Malay, Indian, or other), marital status, education (ie, university degree or below), and income level (ie, above or below SGD 4000 [US $3000]), were collected. Additionally, we collected data on alcohol consumption, smoking status, and self-rated health. We assessed sleep characteristics, including subjective sleep quality, using the Pittsburgh Sleep Quality Index [47]; sleep hygiene using the Sleep Hygiene Index [48]; daytime sleepiness using the Epworth Sleepiness Scale [49]; and perceived loneliness using the revised UCLA (University of California, Los Angeles) Loneliness Scale [50]. All covariates were collected at the beginning of activity tracking.

Wearable Data Preprocessing

Fitbit devices measure steps; energy expenditure, measured in metabolic equivalents (METs); and heart rate, measured in beats per minute (bpm), and they identify sleep stages (ie, wake, light, deep, or REM sleep). Steps and energy expenditure are available at by-the-minute intervals, sleep stages are available at 30-second intervals, and heart rate data are available at 5- or 10-second intervals. We examined the completeness of the data set using heart rate data to verify the actual device use time as the most sensitive to inappropriate use or nonuse. Missing and complete time points of heart rate data were determined from

the full period of participant tracking within the study. Time points of activity and sleep data corresponding to complete heart rate data points were sampled and considered as clean data. We used data from days with at least 20 hours of complete recording for further analysis. Participants with a minimum of 10 complete days were included for further analysis (see Multimedia Appendix 1 for the distribution of complete days among participants). Outliers were identified using the Tukey rule for outliers—outliers are values more than 1.5 times the IQR from the quartiles—either below the first quartile or above the third quartile.

Extraction of Digital Biomarkers

We extracted a range of digital biomarkers characterizing physical activity, sleep, circadian rhythms, and physiological parameters from the raw data. This set of digital biomarkers relied on previous findings and was substantially extended with novel metrics hypothetically indicative of depressive symptoms [15-18,20,21,24-26,30,32,34,36].

First, we extracted physical activity metrics based on steps and energy expenditure data, including daily steps, time spent at different intensity levels of physical activity, and sedentary time. The daily durations of light, moderate, and vigorous physical activity were determined according to the physical activity guidelines of the US Centers for Disease Control and Prevention [51], where moderate activity corresponds to energy expenditures from 3.0 to 6.0 METs, vigorous activity is above 6.0 METs, and light physical activity is below 3.0 METs. We sampled minutes within these intervals separately and calculated the mean daily sum of these minutes. Sedentary time was defined as any waking behavior with energy expenditure less than 1.5 METs [52]. Hence, to determine sedentary time, we excluded all sleep intervals and calculated a daily mean of total minutes with ≤1.5 METs. Daily steps and sedentary time were calculated for all days, and for weekdays and for weekends separately.

Second, we extracted sleep metrics, including average values and coefficients of variation (CVs) of time in bed, TST, SE, SOL, and WASO. TST is the difference between the length of time in bed and the length of wake time, and SE is the ratio of TST to time in bed. SOL is the length of time in minutes until the first minute of sleep onset from bedtime (ie, from the beginning of a sleep episode). WASO was calculated as the number of wake minutes in the middle half of a sleep episode (ie, in the second and third quartiles of a sleep episode). Additionally, we computed the mean and SD of sleep offset and sleep midpoint time as measured in hours since midnight. All sleep metrics were calculated for all days, and for weekdays separately.

Third, we extracted cosinor-based and nonparametric measures of circadian rhythms using steps and heart rate data. These metrics were extracted based on data from all days and based on weekdays only separately. Cosinor-based metrics were estimated using the extended cosinor model with antilogistic transformation and included mesor, acrophase, amplitude, pseudo-*F* statistic, and α and β parameters. Nonparametric measures included IS, IV, interdaily coefficient of variation (ICV), diurnal activity level (mean activity of the 10 consecutive most active hours of the day [M10]), nocturnal activity level (mean value of the 5 consecutive least active hours of the day [L5]), and RA. In addition, we calculated lagged autocorrelation and peak detection–based metrics. Mesor is a rhythm-adjusted mean value that estimates central tendency of the distribution of an oscillating variable, with lower values indicating reduced activity levels [53]. Acrophase is the time of day of the cosine curve peak. Amplitude is the difference between the peak value of the curve and mesor, where lower amplitude indicates a more dampened rhythm. The pseudo-*F* statistic is the goodness-of-fit estimate of the fitted model, which indicates so-called robustness of the rhythm. α is the relative width of the curve at the middle of the peak, and β is an indicator of the steepness of the rise and fall of the curve.

IS is a measure of stability and regularity of activity patterns across a series of 24-hour cycles and is calculated as the ratio of variance of the average 24-hour activity profile to the total variance of data aggregated by the hour from all days [54]. Higher IS indicates more stable, more regular circadian rhythm:

where N is the total number of data points, p is the number of data points per day (24 in this case), x_h represents values of each hour from the mean 24-hour profile, x_i represents each given hour of raw data, and  is the mean of all data.

IV quantifies the fragmentation of rest and activity periods within a 24-hour cycle and is calculated as the mean square of differences between successive hourly aggregated data normalized by the total variance of all days [54]. Higher IV indicates a more fragmented rhythm and reflects shorter alternating periods of rest and activity rather than one extended active period and one extended rest period:

where N is the total number of data points, x_i represents the value of a given hour, and  is the mean of all data.

RA reflects the difference between the most and the least active periods and is calculated as the difference between M10 and L5 divided by sum of M10 and L5, where higher values show greater amplitude:

ICV is a novel rhythm stability measure and is calculated as the 24-hour mean of CVs, which is the ratio of the SD to the mean in each hour between days. A higher coefficient indicates higher variation and less stable rhythm. We proposed this metric as an alternative to IS, which aims to assess the same phenomena with a different approach:

where p is the number of data points per day (24 in this case; data are aggregated by the hour), x_i represents values corresponding to each hour from all days, x_h represents values

of each hour from the mean 24-hour profile, and *N* is the number of days.

Autocorrelation is another measure of rhythm stability, calculated as the lagged autocorrelation of time series. Autocorrelation was calculated for time series aggregated into 15-minute, 30-minute, and 60-minute intervals with a day-length lag:

where AC is autocorrelation, k is the day-length lag (eg, 24 if data are aggregated by the hour), x_i represents values of each interval,  is the mean of all data, and N is total number of data points.

We applied the robust peak-detection algorithm based on z scores to times series to identify peaks in steps and heart rate data [55]. We calculated the daily mean number of peaks and SD.

Finally, extracted heart rate–based metrics included the overall average heart rate, resting heart rate (RHR), delta heart rate, daytime and nighttime heart rate, a variation of these measures using SD and CV, and the root mean square of successive differences (RMSSD) of heart rate. RHR was calculated as the average heart rate for 15-minute intervals with zero steps. Daytime heart rate was obtained by averaging heart rate values between 2 PM and 4 PM, whereas nighttime heart rate was obtained by averaging values sampled from three consecutive 2-hour intervals: 12 AM to 2 AM, 2 AM to 4 AM, and 4 AM to 6 AM. Unlike daytime heart rate, nighttime heart rates were sampled from more time intervals because nocturnal physiological processes seemed to be more sensitive to depression [56]. We calculated SDs and CVs for all heart rates. Delta heart rate is the difference between average heart rate and RHR. RMSSD values of heart rate were calculated from raw data and from data aggregated into hourly intervals.

In total, we extracted 126 digital biomarkers. The full list of digital biomarkers with detailed descriptions is presented in Multimedia Appendix 2.

Statistical Analysis and Predictive Modeling With Machine Learning

Spearman rank correlation was used to evaluate the strength of associations between digital biomarkers and severity of depressive symptoms. The false discovery rate (FDR) was used for multiple testing correction of P values [57]. A multiple hierarchical linear regression was used to determine the strength of association between digital biomarkers and severity of depressive symptoms adjusted for covariates, including sociodemographics, alcohol consumption, smoking, self-rated health, subjective sleep characteristics, and loneliness.

Next, we trained a series of supervised machine learning models predicting symptom severity and depression screening status of participants (ie, depressed or healthy) to evaluate the predictive ability of digital biomarkers identified at the previous step of statistical analysis. Digital biomarkers were selected based on Spearman correlation at different levels of significance, including P<.01, P<.05, and P values between .01 and .05. Extreme gradient boosting algorithm (DART [Dropouts meet Multiple Additive Regression Trees]) was used for model training, as it was the most efficient method in similar studies [33,37,58]. For performance evaluation of the symptom severity prediction model, we used repeated k-fold cross-validation with 4 folds and 25 repeats and obtained R^2, mean absolute error (MAE), and root mean square error (RMSE) for holdout folds. For evaluation of the model predicting depression screening status, we used k-fold cross-validation with 4 folds and 25 repeats and obtained accuracy, sensitivity, specificity, positive predictive value, negative predictive value, Cohen κ, and area under the curve (AUC). The k-fold cross-validation was used as an alternative to performance evaluation with the testing set, which is usually between 20% and 30% of a sample and, therefore, using 4-fold cross-validation corresponds to 25% of a sample in holdout folds for performance evaluation. Relative feature importance, which indicates the contribution of each predictor in the overall improvement of classification accuracy across all decision trees, was reported. We used RStudio from R (version 3.6.0; The R Foundation) for all analyses. R scripts for feature extraction and statistical analyses are available in Multimedia Appendix 3.

Results

Characteristics of the Data and Participants

We collected 5180 days of observational data in total, or 17 days and 20 hours per participant, on average. The mean compliance with wearing the activity tracker among participants within the observation period was 84.8% (SD 10.4), whereas the total fraction of complete data was 84.0% (6,265,469 minutes of complete data from 5180 days). A total of 77.7% (4025/5180) of days contained at least 20 hours of complete data; however, only 277 participants out of 290 (95.5%) had at least 10 days with enough data (see Table S1 in Multimedia Appendix 1). Thus, 5,464,248 minutes of complete data from 3948 days in total (83.7%) were included for further analysis, and 801,221 minutes of complete data (12.8%) were discarded. Further, we identified and excluded outliers: 1 participant was excluded due to outlying heart rate, 4 participants were excluded due to exceeding average steps, and another 5 participants were excluded due to insufficient sleep data with less than three nights on weekdays. The final sample comprised of 267 participants with 14 days of tracking data, on average; the participant flowchart is presented in Multimedia Appendix 4.

The mean age of the participants was 33 (SD 8.6, range 21-64) years. Out of 267 participants, a mild female bias was displayed (n=170, 63.7%). The majority of the participants were Chinese (n=211, 79.0%), single (n=163, 61.0%), with a university degree (n=238, 89.1%), and with a monthly income below SGD 4000 (US $3000; n=154, 57.7%). Only 15 participants (5.6%) were current smokers, while 158 participants (59.2%) reported drinking alcohol. Finally, most participants rated their health as “good” (n=124, 46.4%) or “very good” (n=105, 39.3%); fewer participants rated their health as “fair” (n=21, 7.9%) and the lowest number of participants felt that their health was “excellent” (n=17, 6.4%). The mean step count was 10,252 (SD

2822) steps per day and the mean heart rate was 74 (SD 7.1) bpm. Figure 1 shows the smoothed 24-hour daily profile of heart rate and steps averaged in 1-hour windows in the study sample. Table 1 summarizes the characteristics of the participants.

Figure 1. Average 24-hour profiles of heart rate and steps measured by the wearable activity tracker. The purple and orange lines represent the means and the gray shaded areas represent the SDs. bpm: beats per minute.

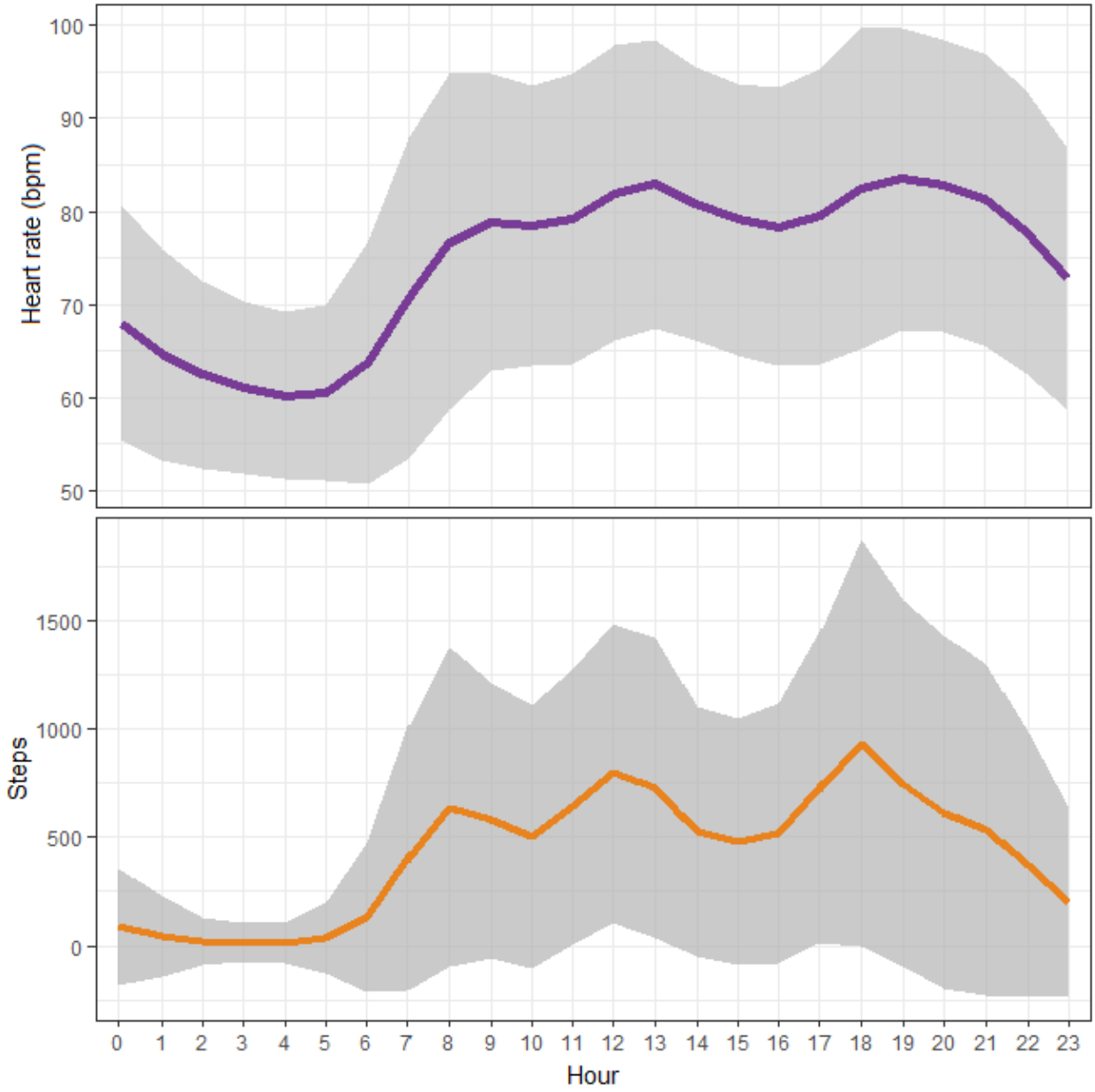


Table 1. Sociodemographic characteristics, self-reported health outcomes, and basic wearable metrics of participants.

Participant characteristics	Value (N=267)
Age (years), mean (SD)	32.8 (8.6)
Gender, n (%)	
Male	97 (36.3)
Female	170 (63.7)
Ethnicity, n (%)	
Chinese	211 (79.0)
Indian	22 (8.2)
Malay	10 (3.7)
Other	24 (9.0)
Marital status, n (%)	
Married	104 (39.0)
Single	163 (61.0)
Education, n (%)	
University degree	238 (89.1)
Below university degree	29 (10.9)
Monthly income (SGD[a]), n (%)	
<4000	154 (57.7)
≥4000	113 (42.3)
Alcohol consumption, n (%)	
No	109 (40.8)
Yes	158 (59.2)
Smoking, n (%)	
Current smoker	15 (5.6)
Nonsmoker	252 (94.4)
Overall self-rated health, n (%)	
Fair	21 (7.9)
Good	105 (46.4)
Very good	124 (39.3)
Excellent	17 (6.4)
Sleep Hygiene Index score, mean (SD)	16.9 (5.7)
Pittsburgh Sleep Quality Index score, mean (SD)	5.4 (2.7)
Daytime sleepiness score (ESS[b]), mean (SD)	7.2 (4.2)
Loneliness score (UCLA[c] Loneliness Scale), mean (SD)	37.8 (9.8)
Complete actigraphy days, mean (SD)	14.2 (2.1)
Steps per day, mean (SD)	10,252 (2822)
Heart rate (beats per minute), mean (SD)	74.4 (7.1)

[a]A currency exchange rate of SGD 1=US $0.75 is applicable.

[b]Epworth Sleepiness Scale.

[c]UCLA: University of California, Los Angeles.

Table 2 and Figure 2 display the distribution of PHQ-9 scores across two assessments and show the change in scores over a 2-week period. PHQ-9 scores are highly and linearly correlated (r=0.73, P<.001), showing that the change in depressive symptoms over time is quite smooth and gradual rather than quick and sharp. However, participants experienced an overall

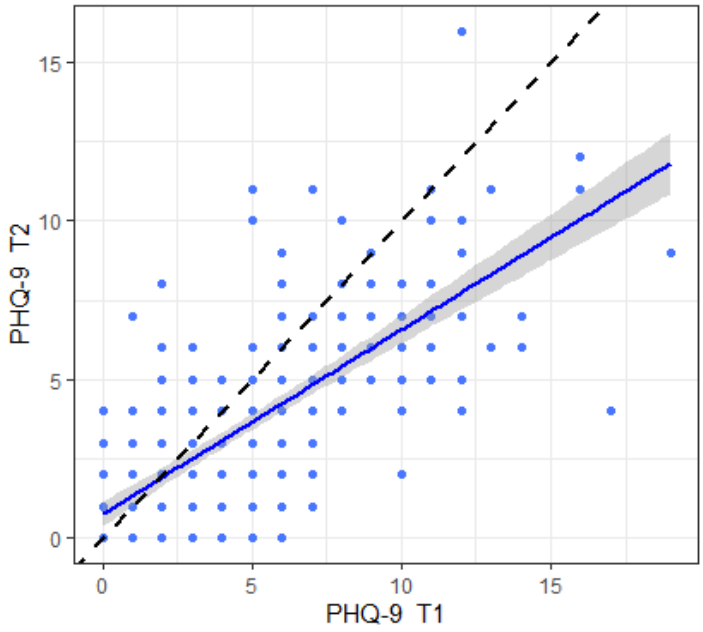
decrease in depressive symptoms over time; for example, 20 participants with moderate (score 10-14) depressive symptoms at the first assessment shifted to the mild range (score 5-9) at the second assessment, while just 5 participants with mild symptoms at the first assessment appeared to have moderate symptoms at the second assessment. Figure 2 shows the same trend. Hence, using the widely adopted cutoff score of ≥10 for both assessments significantly reduced the prevalence rate of depressed participants: only 8 participants had PHQ-9 scores of ≥10 at both assessments. Given that the overall change in symptom severity is gradual, we used less strict criteria and defined a participant as provisionally depressed if he or she scored 10 at one or both assessments and was otherwise healthy. Having a PHQ-9 score of ≥10 at any time point means that the score at an adjacent time point is likely to be comparable in magnitude. In total, 38 out of 267 (14.2%) participants had PHQ-9 scores of ≥10 at either of the two assessments and were identified as depressed. In addition, we used alternative cutoff points to define depression in participants. First, a meta-analysis by Manea et al [59] showed that a PHQ-9 cutoff score of ≥8 also has acceptable screening properties; in our sample, there were 22 (8.2%) participants who had scores of ≥8 at both assessments. Second, given the linearity of PHQ-9 change over time, we used an average PHQ-9 score of ≥8 as another criterion; 42 (15.7%) participants had an average PHQ-9 score ≥8.

Table 2. The distribution of PHQ-9 scores across two assessments.

PHQ-9[a] score at the first assessment	Participants within each category of PHQ-9 scores at the second assessment (N=267), n (%)				
	Normal: score of 0-4	Mild: score of 5-9	Moderate: score of 10-14	Moderately severe: score of ≥15	Total
Normal: score of 0-4	143 (53.6)	14 (5.2)	0 (0)	0 (0)	157 (58.8)
Mild: score of 5-9	32 (12.0)	40 (15.0)	5 (1.9)	0 (0)	77 (28.8)
Moderate: score of 10-14	3 (1.1)	20 (7.5)	5 (1.9)	1 (0.4)	29 (10.9)
Moderately severe: score of ≥15	1 (0.4)	1 (0.4)	2 (0.7)	0 (0)	4 (1.5)
Total	179 (67.0)	75 (28.0)	12 (4.5)	1 (0.4)	267 (100)

[a]PHQ-9: 9-item Patient Health Questionnaire.

Figure 2. Scatterplot of 9-item Patient Health Questionnaire (PHQ-9) scores at two assessments (T1 and T2). The blue line is the linear projection of the relationship between two scores with CIs and the dashed diagonal line represents no change in scores between two assessments.



Associations Between Digital Biomarkers and Depressive Symptoms

Correlation analysis revealed that 36 digital biomarkers were significantly associated with the average PHQ-9 score—absolute coefficients of Spearman rank correlation were weak and varied from 0.12 to 0.26—and only 17 of them remained significant predictors after the FDR correction (Table 3). Severity of depressive symptoms was correlated to variation in nighttime heart rate in several time intervals, regularity of circadian rhythms measured with nonparametric and cosinor measures, daily peaks, and timing and variation in sleep offset and midpoint. Physical activity metrics were not associated with PHQ-9 scores.

Table 3. Correlation between digital biomarkers and 9-item Patient Health Questionnaire scores.

Category and digital biomarkers	Spearman rank correlation coefficient	*P* value	Adjusted *P* value[a]
Heart rate metrics			
DHR[b].cv[c]	0.121	.047	.18
NHR[d].0204.sd[e]	0.262	<.001	.001
NHR.0204.cv	0.257	<.001	.001
NHR.0406.sd	0.182	.003	.04
NHR.0406.cv	0.185	.002	.04
NHR.0002.sd	0.149	.01	.08
NHR.0002.cv	0.138	.02	.11
Circadian rhythm metrics: nonparametric			
IS[f].st[g].wd[h]	–0.165	.007	.049
IS.hr[i].wd	–0.199	.001	.02
AC[j].st.60m[k]	–0.125	.04	.15
AC.st.15m	–0.175	.004	.04
AC.st.30m	–0.155	.01	.07
AC.st.60m.wd	–0.159	.009	.06
AC.st.15m.wd	–0.177	.004	.04
AC.st.30m.wd	–0.177	.004	.04
AC.hr.60m.wd	–0.150	.01	.08
AC.hr.30m.wd	–0.144	.02	.08
ICV[l].st.wd	0.176	.004	.04
ICV.hr	0.177	.004	.04
ICV.hr.wd	0.237	<.001	.004
peaks.st	–0.205	<.001	.02
peaks.st.wd	–0.202	<.001	.02
Circadian rhythm metrics: cosinor based			
acro[m].st	0.133	.03	.12
F.st.wd	–0.121	.046	.17
beta.hr	0.169	.006	.04
acro.hr	0.146	.02	.09
F.hr	–0.126	.04	.15
beta.hr.wd	0.126	.04	.15
acro.hr.wd	0.154	.01	.07
F.hr.wd	–0.137	.03	.11
Sleep metrics			
sleep.offset	0.144	.02	.09
sleep.midpoint	0.172	.005	.04
sleep.offset.wd	0.134	.03	.12
sleep.offset.wd.sd	0.199	.001	.02
sleep.midpoint.wd	0.146	.02	.09
sleep.midpoint.wd.sd	0.146	.02	.09

[a]*P* values were adjusted for multiple testing correction. Only significant associations were reported.

[b]DHR: daytime heart rate between 2 PM and 4 PM.

[c]cv: coefficient of variation.

[d]NHR: nighttime heart rate in a specified 2-hour time interval; 2 AM-4 AM (0204), 4 AM-6 AM (0406), and 12 AM-2 AM (0002).

[e]sd: standard deviation.

[f]IS: interdaily stability.

[g]st: steps based.

[h]wd: weekdays based.

[i]hr: heart rate based.

[j]AC: autocorrelation.

[k]m: minutes; 15-minute, 30-minute, or 60-minute time interval in which raw data were aggregated.

[l]ICV: interdaily coefficient of variation.

[m]acro: acrophase.

Further linear regression analysis showed that 11 digital biomarkers—seven nonparametric measures of weekday circadian rhythm regularity, based on steps and heart rate, and four metrics of heart rate variation at nighttime intervals—were significantly associated with severity of depressive symptoms independent of sociodemographic confounders, including age, gender, ethnicity, marital status, education, and income levels (see Table S2 in Multimedia Appendix 1). Only three digital biomarkers—weekday steps–based IS, autocorrelation (based on 15-minute intervals), and CV of heart rate between 4 AM and 6 AM—were consistently associated with symptom severity independent of all confounders, which additionally included alcohol consumption, smoking, self-rated health, loneliness, and subjective sleep characteristics (Table 4). Weekday steps–based IS and autocorrelation were negatively associated with PHQ-9 scores (for IS in the fully adjusted model: β=−0.30 per 10% change, 95% CI −0.51 to −0.08, *P*=.01; for autocorrelation: β=−0.38 per 10% change, 95% CI −0.75 to −0.01, *P*=.04), so participants with lower stability of circadian activity rhythms had more severe symptoms. Heart rate variation between 4 AM and 6 AM was positively associated with PHQ-9 scores (β=0.90 per 10% change, 95% CI 0.04-1.78, *P*=.04), where a greater magnitude of heart rate variation indicated higher PHQ-9 scores. This digital biomarker correlated to depressive symptoms independent of sleep offset time, which might cause greater variation (Table S3 in Multimedia Appendix 1).

Table 4. Coefficients of multiple regression analysis for digital biomarkers and 9-item Patient Health Questionnaire (PHQ-9) scores. Fully adjusted models.

Predictor	β[a] (95% CI)	*P* value
Model 1		
Digital biomarker: IS.st.wd[b]	–2.974 (–5.202 to –0.747)	.009
Age	–0.114 (–0.163 to –0.064)	<.001
Gender (male)	–0.165 (–0.811 to 0.481)	.62
Ethnic group (Indian)	–0.297 (–1.393 to 0.798)	.59
Ethnic group (Malay)	–0.036 (–1.629 to 1.556)	.96
Ethnic group (others)	1.012 (–0.021 to 2.045)	.06
Marital status (single)	–0.633 (–1.341 to 0.076)	.08
Education level (university degree)	–1.023 (–2.079 to 0.034)	.06
Monthly income level (SGD 4000[c] and above)	0.283 (–0.422 to 0.988)	.43
Alcohol consumption (yes)	–0.201 (–0.834 to 0.433)	.53
Smoking status (nonsmoker)	–0.768 (–2.075 to 0.54)	.25
Self-rated health (fair)	1.528 (–0.103 to 3.159)	.07
Self-rated health (good)	0.597 (–0.631 to 1.824)	.34
Self-rated health (very good)	–0.026 (–1.246 to 1.194)	.97
UCLA (University of California, Los Angeles) Loneliness Scale score	0.094 (0.062 to 0.127)	<.001
Sleep Hygiene Index (SHI) score	0.082 (0.022 to 0.143)	.008
Pittsburgh Sleep Quality Index (PSQI) score	0.362 (0.227 to 0.496)	<.001
Epworth Sleepiness Scale (ESS) score	0.06 (–0.014 to 0.134)	.11
Intercept	3.407 (–0.167 to 6.98)	.06
Model 2		
Digital biomarker: AC.st.15m.wd[d]	–3.843 (–7.567 to –0.119)	.04
Age	–0.113 (–0.163 to –0.063)	<.001
Gender (male)	–0.063 (–0.712 to 0.586)	.85
Ethnic group (Indian)	–0.19 (–1.285 to 0.904)	.73
Ethnic group (Malay)	–0.038 (–1.64 to 1.563)	.96
Ethnic group (others)	1.126 (0.09 to 2.162)	.03
Marital status (single)	–0.635 (–1.348 to 0.077)	.08
Education level (university degree)	–1.056 (–2.12 to 0.007)	.051
Monthly income level (SGD 4000 and above)	0.309 (–0.402 to 1.02)	.39
Alcohol consumption (yes)	–0.229 (–0.87 to 0.413)	.48
Smoking status (nonsmoker)	–0.741 (–2.056 to 0.573)	.27
Self-rated health (fair)	1.574 (–0.066 to 3.213)	.06
Self-rated health (good)	0.604 (–0.631 to 1.838)	.34
Self-rated health (very good)	–0.031 (–1.259 to 1.196)	.96
UCLA Loneliness Scale score	0.092 (0.059 to 0.124)	<.001
SHI score	0.091 (0.031 to 0.151)	.003
PSQI score	0.351 (0.216 to 0.486)	<.001
ESS score	0.062 (–0.013 to 0.136)	.11
Intercept	2.526 (–0.921 to 5.973)	.15

Predictor	β[a] (95% CI)	*P* value
Model 3		
Digital biomarker: NHR.0406.cv[e]	9.096 (0.333 to 17.859)	.04
Age	–0.114 (–0.164 to –0.064)	<.001
Gender (male)	–0.246 (–0.909 to 0.416)	.47
Ethnic group (Indian)	–0.188 (–1.282 to 0.907)	.74
Ethnic group (Malay)	–0.193 (–1.802 to 1.415)	.81
Ethnic group (others)	0.985 (–0.058 to 2.028)	.06
Marital status (single)	–0.661 (–1.373 to 0.051)	.07
Education level (university degree)	–0.837 (–1.913 to 0.238)	.13
Monthly income level (SGD 4000 and above)	0.301 (–0.409 to 1.011)	.40
Alcohol consumption (yes)	–0.1 (–0.736 to 0.536)	.76
Smoking status (nonsmoker)	–0.72 (–2.035 to 0.595)	.28
Self-rated health (fair)	1.811 (0.166 to 3.456)	.03
Self-rated health (good)	0.763 (–0.479 to 2.005)	.23
Self-rated health (very good)	0.215 (–1.026 to 1.457)	.73
UCLA score	0.091 (0.059 to 0.124)	<.001
SHI score	0.103 (0.043 to 0.163)	.001
PSQI score	0.332 (0.195 to 0.47)	<.001
ESS score	0.061 (–0.013 to 0.135)	.11
Intercept	0.598 (–2.891 to 4.088)	.74

[a]Unstandardized coefficients (β) with their 95% CIs and exact *P* values of digital biomarkers are reported as predictors of PHQ-9 scores in multiple regression models.

[b]IS.st.wd: steps-based interdaily stability on weekdays.

[c]A currency exchange rate of SGD 1=US $0.75 is applicable.

[d]AC.st.15m.wd: steps-based autocorrelation with weekday data aggregated into 15-minute intervals.

[e]NHR.0406.cv: coefficient of variation of nighttime heart rate between 4 AM and 6 AM.

Depression Screening Using Digital Biomarkers and Machine Learning

The performance of the symptom severity prediction models was evaluated using the whole data set of 267 participants. The range of mean correlations between actual and predicted PHQ-9 scores across trained models was 0.14 to 0.27 (R^2=0.03-0.08), the range of mean RMSE was 3.10 to 3.20, and the range of average MAE in holdout samples was 2.54 to 2.63 (Table 5, models A1-C1). Adding sociodemographic characteristics did not substantially improve performance (Table 5, models A2-C2). However, applying more conservative feature selection criteria, including digital biomarkers most correlated to the outcome, showed relatively better results compared to the less conservative criteria and including digital biomarkers less correlated to the outcome. The selected digital biomarkers are presented in Table 3 and listed in Table S4 in Multimedia Appendix 1.

Table 5. Performance of symptom-severity prediction model.

Model (feature set)	R^2, mean (SD)	Pearson correlation, mean (SD)	Root mean square error, mean (SD)	Mean absolute error, mean (SD)
A1[a]	0.06 (0.04)	0.23 (0.09)	3.12 (0.22)	2.57 (0.15)
B1[b]	0.06 (0.04)	0.22 (0.09)	3.14 (0.22)	2.58 (0.15)
C1[c]	0.03 (0.03)	0.14 (0.08)	3.20 (0.21)	2.63 (0.14)
A2[d]	0.08 (0.05)	0.27 (0.09)	3.10 (0.22)	2.54 (0.16)
B2[d]	0.06 (0.04)	0.24 (0.09)	3.12 (0.22)	2.57 (0.15)
C2[d]	0.04 (0.03)	0.18 (0.08)	3.16 (0.20)	2.60 (0.13)

[a]Model A1 includes digital biomarkers selected at a significance level of <.01.

[b]Model B1 includes digital biomarkers selected at a significance level of <.05.

[c]Model C1 includes digital biomarkers selected at a significance level of <.05 and ≥.01.

[d]Models A2, B2, and C2 additionally include sociodemographic characteristics: age, gender, ethnic group, and marital status.

Next, we trained models for classification of the outcome of depression screening using different PHQ-9 score cutoff points to determining which participants were depressed or healthy. For the default cutoff point (ie, either baseline or follow-up PHQ-9 score of ≥10), the accuracy of the models in holdout folds was 86% (equal to the no information rate 86%), the sensitivity range was 3% to 13%, the specificity range was 98% to 100%, and the AUC range was 0.51 to 0.66 (Table 6, models A1-C1). Adding sociodemographic characteristics did not improve classification accuracy (Table 6, models A2-C2). However, applying more conservative feature selection criteria, including digital biomarkers most correlated to the outcome, showed relatively better results compared to the less conservative criteria and including digital biomarkers less correlated to the outcome.

For the second cutoff point option (ie, both baseline and follow-up PHQ-9 scores of ≥8), the accuracy of the models in holdout folds was 92% (no information ratio 92%), the sensitivity range was 0% to 5%, the specificity was 100%, and the AUC range was 0.54 to 0.67 (Table 6). Finally, for the third cutoff point option (ie, average PHQ-9 score of ≥8), the accuracy range of the models in holdout folds was 85% to 87% (no information ratio 84%), the sensitivity range was 2% to 24%, the specificity range was 97% to 100%, and the AUC range was 0.62 to 0.74 (Table 6). Thus, the performance of models classifying the outcome of depression screening was relatively better when the third cutoff point was applied.

In general, the accuracy of models predicting severity of depressive symptoms and depression screening status based on digital biomarkers in the whole sample was poor, probably indicating a significant heterogeneity of data within groups of depressed and healthy participants.

Table 6. Performance of the prediction models of depression screening status for different cutoff points and varying sets of selected digital biomarkers.

Outcome cutoff point and model (feature set)	No information rate	Accuracy	Sensitivity	Specificity	PPV[a]	NPV[b]	κ	AUC[c]
PHQ-9[d] score of ≥10 at assessment 1 or 2 (n=38 in depressed group)								
A1[e]	0.86	0.86	0.05	0.99	0.50	0.86	0.07	0.66
B1[f]	0.86	0.86	0.03	1.00	0.50	0.86	0.04	0.63
C1[g]	0.86	0.86	0.03	1.00	1.00	0.86	0.04	0.51
A2[h]	0.86	0.86	0.03	1.00	0.50	0.86	0.04	0.64
B2[h]	0.86	0.86	0.13	0.98	0.56	0.87	0.17	0.64
C2[h]	0.86	0.86	0.05	1.00	0.67	0.86	0.08	0.61
PHQ-9 score of ≥8 at assessment 1 or 2 (n=22 in depressed group)								
A1	0.92	0.92	0.05	1.00	1.00	0.92	0.08	0.64
B1	0.92	0.92	0.00	1.00	N/A[i]	0.92	0.00	0.54
C1	0.92	0.92	0.05	1.00	1.00	0.92	0.08	0.56
A2	0.92	0.92	0.05	1.00	1.00	0.92	0.08	0.67
B2	0.92	0.92	0.00	1.00	N/A	0.92	0.00	0.58
C2	0.92	0.92	0.05	1.00	1.00	0.92	0.08	0.62
PHQ-9 average score of ≥8 (n=42 in depressed group)								
A1	0.84	0.85	0.12	0.99	0.63	0.86	0.16	0.70
B1	0.84	0.85	0.19	0.97	0.57	0.87	0.23	0.67
C1	0.84	0.85	0.02	1.00	1.00	0.85	0.04	0.62
A2	0.84	0.87	0.24	0.99	0.77	0.87	0.31	0.74
B2	0.84	0.85	0.17	0.98	0.64	0.86	0.21	0.70
C2	0.84	0.85	0.07	1.00	1.00	0.85	0.12	0.62

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the curve.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]Model A1 includes digital biomarkers selected at a significance level of <.01.

[f]Model B1 includes digital biomarkers selected at a significance level of <.05.

[g]Model C1 includes digital biomarkers selected at a significance level of <.05 and ≥.01.

[h]Models A2, B2, and C2 additionally include sociodemographic characteristics: age, gender, ethnic group, and marital status.

[i]N/A: not applicable, due to division by zero.

Detecting Individuals at High Risk of Depression Against Those With No Risk

We retrained models using random downsampling of healthy participants from the lowest range of PHQ-9 scores (0-4) to address the class imbalance in our data and to increase the contrast between compared groups, similar to Sano et al [39]. We believe that excluding participants with midrange (ie, borderline) PHQ-9 scores results in the larger difference between groups and, therefore, increases the discriminatory power of digital biomarkers. We excluded participants with zero scores (n=24) due to concerns in honesty of their responses to minimize the bias in self-reported outcome. We used three contrasted subsamples varying by the PHQ-9 cutoff points determining the depressed group similar to the classification in the whole sample; additionally, we used the contrasted subsample comprised of the top 20% and the bottom 20% of participants by average PHQ-9 scores. For the default cutoff point (ie, either baseline or follow-up PHQ-9 score of ≥10), the contrasted sample included 78 participants (38 depressed and 40 healthy); for the second cutoff point (ie, both baseline and follow-up PHQ-9 scores of ≥8), the contrasted sample included 44 participants (22 depressed and 22 healthy); and for the third cutoff point (ie, average PHQ-9 scores of ≥8), the contrasted sample included 84 participants (42 depressed and 42 healthy). Finally, the fourth subsample included 96 participants (48 depressed and 48 healthy), where the range of PHQ-9 scores for the healthy group was 0.5 to 1.5 and for the depressed group was 7.5 to 14. Subsamples mainly did not differ from the whole sample in terms of sociodemographic characteristics, with two

exceptions: the second subsample had a slightly younger age, and the fourth subsample did not have a gender bias (see Tables S5 to S8 in Multimedia Appendix 1).

Importantly, feature selection for these models was done using the entire sample (ie, selected digital biomarkers were based on statistical associations found in the entire sample, not in subsamples, and remained the same for all models). Otherwise, if feature selection is done each time for a new subsample with a different cutoff point, sets of digital biomarkers would be arbitrary and would vary depending on a particular sample composition. This allows us to mitigate overgeneralization that is possible in skewed samples, because models were trained with digital biomarkers inferred from the entire sample. For performance evaluation of these models, we used stratified repeated cross-validation with 4 folds and 25 repeats.

Similar to the previous step of modeling, using more conservative feature selection criteria consistently resulted in better performance compared to the less conservative criteria. The best model was based on the contrasted subsample with default cutoff point (ie, either baseline or follow-up PHQ-9 score of ≥10) and has correctly predicted the depression screening status in 80% of participants from holdout folds, with a sensitivity of 82% and a specificity of 78% (Table 7 and Figure 3, A and B). Alternatively, the model based on the subsample comprised of the top 20% and bottom 20% of participants by average PHQ-9 score achieved the highest AUC of 0.8, and had accuracy, sensitivity, and specificity values of 77% each (Table 7 and Figure 3, C and D). Additionally, using the contrasted subsample with the top 20% and bottom 20% of participants, we trained models without statistical feature selection but with different feature subsets varying by categories (ie, activity metrics, nonparametric circadian rhythm metrics, cosinor-based metrics, heart rate metrics, and sleep metrics). Among these models, the model with heart rate metrics had the best accuracy, sensitivity, and AUC—70%, 71%, and 72%, respectively—and the model with nonparametric circadian rhythm metrics had the best specificity of 76% (see Table S9 in Multimedia Appendix 1); however, models with correlation-based feature selection outperformed all of these models.

Relative importance of digital biomarkers was extracted and averaged from the best models (model A from Table 7) for each of four contrasted subsamples (Figure 4). The scatterplots in Figure 5 illustrate how different combinations of digital biomarkers can discriminate between provisionally depressed and healthy participants in the contrasted sample.

Table 7. Performance of the prediction of depression screening status in contrasted subsamples.

Contrasted subsample and model (feature set)	No information rate	Accuracy	Sensitivity	Specificity	PPV[a]	NPV[b]	κ	AUC[c]
PHQ-9[d] score of ≥10 at assessment 1 or 2 (n=78)								
A[e]	0.51	0.80	0.82	0.78	0.78	0.82	0.59	0.75
B[f]	0.51	0.77	0.79	0.75	0.75	0.79	0.54	0.74
C[g]	0.51	0.71	0.74	0.68	0.68	0.73	0.41	0.70
PHQ-9 score of ≥8 at assessment 1 or 2 (n=44)								
A	0.50	0.71	0.73	0.68	0.70	0.71	0.41	0.71
B	0.50	0.71	0.77	0.64	0.68	0.74	0.41	0.68
C	0.50	0.71	0.73	0.68	0.70	0.71	0.41	0.64
PHQ-9 average score of ≥8 (n=84)								
A	0.50	0.77	0.79	0.76	0.77	0.78	0.55	0.76
B	0.50	0.71	0.71	0.71	0.71	0.71	0.43	0.71
C	0.50	0.61	0.64	0.57	0.60	0.62	0.21	0.57
Top 20% and bottom 20% by average PHQ-9 score (n=96)								
A	0.50	0.77	0.77	0.77	0.77	0.77	0.54	0.80
B	0.50	0.74	0.73	0.75	0.75	0.74	0.48	0.76
C	0.50	0.65	0.63	0.67	0.65	0.64	0.29	0.62

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the curve.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]Model A includes digital biomarkers selected at a significance level of <.01.

[f]Model B includes digital biomarkers selected at a significance level of <.05.

[g]Model C includes digital biomarkers selected at a significance level of <.05 and ≥.01.

Figure 3. A and B. Performance evaluation of model A based on the contrasted subsample with default cutoff point (ie, either baseline or follow-up PHQ-9 score of ≥10). A. Confusion matrix of predicted and observed outcomes. B. Area under the curve (AUC) with 95% CI. C and D. Performance evaluation of model A based on the contrasted subsample comprised of the top 20% and bottom 20% of participants by average PHQ-9 score. C. Confusion matrix of predicted and observed outcomes. D. AUC with 95% CI. PHQ-9: 9-item Patient Health Questionnaire.

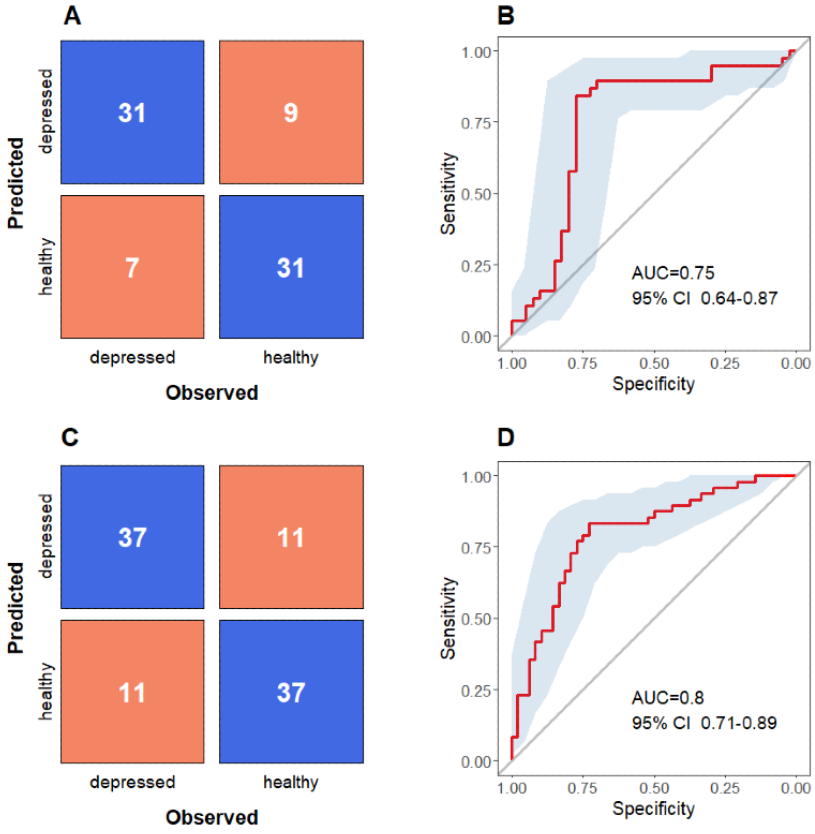


Figure 4. Relative importance of digital biomarkers averaged from four models. 15m, 30m, and 60m: 15- , 30-, and 60-minute time interval in which raw data were aggregated; AC: autocorrelation; cv: coefficient of variation; hr: heart rate based; ICV: interdaily coefficient of variation; IS: interdaily stability; NHR: nighttime heart rate in a specified 2-hour time interval (0204: 2 AM-4 AM; 0406: 4 AM-6 AM); sd: standard deviation; st: steps based; wd: weekdays based.

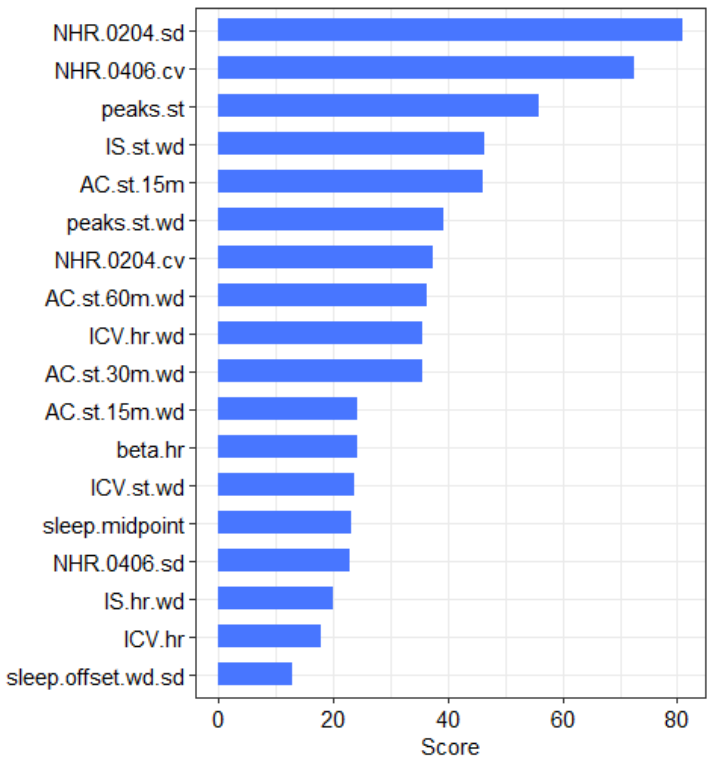
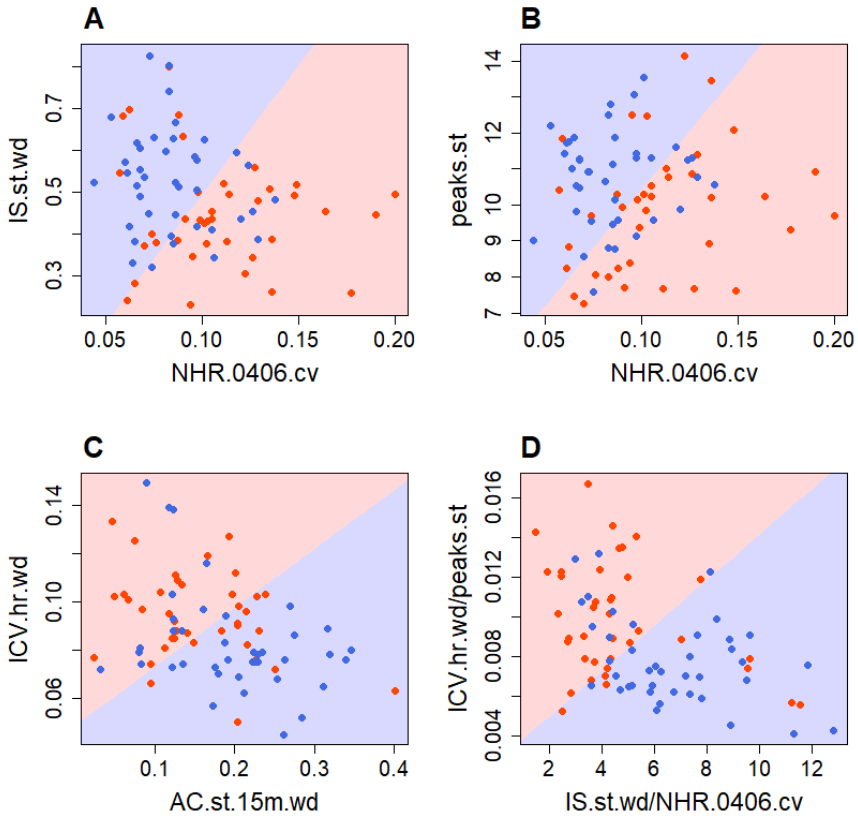


Figure 5. Digital biomarkers of depressed and healthy participants. All scatterplots show the contrasted subsample with the default cutoff point. Red dots represent depressed participants and blue dots represent healthy participants; the background coloring represents a decision boundary of linear discriminant analysis. A. IS.st.wd: weekday steps–based interdaily stability; NHR.0406.cv: variation of heart rate between 4 AM and 6 AM. B. peaks.st: daily steps–based peaks; NHR.0406.cv: variation of heart rate between 4 AM and 6 AM. C. ICV.hr.wd: interdaily coefficient of variation of heart rate on weekdays; AC.st.15m.wd: autocorrelation of weekday steps–based rhythm (steps aggregated in 15-minute intervals). D. ICV.hr.wd/peaks.st: interdaily coefficient of variation of heart rate on weekdays divided by daily steps–based peaks; IS.st.wd/NHR.0406.cv: weekday steps–based interdaily stability divided by variation of heart rate between 4 AM and 6AM.



Discussion

Principal Findings

In this study, using statistical analysis and machine learning, we demonstrated that some known and novel digital biomarkers based on behavioral and physiological data from consumer wearables could indicate increased risk of depression in a multiethnic working population. We found that greater severity of depressive symptoms was robustly associated with greater variation of nighttime heart rate between 4 AM and 6 AM; it was also associated with lower regularity of weekday circadian activity rhythms based on steps and measured with nonparametric measures of IS and autocorrelation. Effects of these digital biomarkers on symptom severity were stable and independent of all confounders including strong predictors of depression, such as sleep quality and loneliness. Additionally, we found that lower regularity of heart rate circadian rhythm, fewer steps-based daily peaks, greater steepness of the heart rate rhythm curve, later sleep midpoint, and greater variation of sleep offset time were associated with a greater severity of depressive symptoms, yet these associations were less reliable and became nonsignificant in regression models with covariates. Despite several reliable associations, our evidence showed limited ability of digital biomarkers to detect depression in the whole sample of working adults. However, in balanced and contrasted subsamples comprised of provisionally depressed participants and healthy participants with no risk of depression, the model achieved an accuracy of 80%, a sensitivity of 82%, and a specificity of 78% in detecting subjects at high risk of depression. Similar performance has been achieved across all models trained using alternative contrasted subsamples. Thus, predictive models based on a combination of these digital biomarkers could quite accurately discriminate individuals with a high risk of depression from individuals with no risk.

Comparison With Previous Research

We compared our study to research investigating relationships between specific actigraphy metrics and depressive disorder. Firstly, regarding circadian rhythm metrics, we found that a lower steps-based weekday IS—a nonparametric measure of rhythm regularity—was robustly associated with a greater severity of depressive symptoms independent of confounders, supporting and extending existing evidence [31,60], including one large-sample study [34]. However, there are several studies that did not find IS to be related to depressive symptoms or other mental disorders [27,35,36]. In support of this association, we also found steps-based autocorrelation during weekdays—a complementary and alternative nonparametric measure of rhythm regularity—to be robustly correlated to symptom severity independent of all confounders. Rhythm stability metrics based on heart rate were also correlated to severity of

depressive symptoms, yet these associations became nonsignificant in fully adjusted regression models. Likewise, among cosinor-based metrics, our data indicated lower pseudo-*F* statistic values and later acrophase, both based on steps and heart rate, in participants with more severe symptoms, which is consistent with existing evidence [22,23,25,26,29,30], yet these correlations became nonsignificant after the FDR correction and in multiple regression analysis. Perhaps this supports the advantage of nonparametric indicators over cosinor-based metrics, where the former appear more robust and sensitive to indicate depressive symptomatology in nonclinical samples. In general, results of our study together with previous evidence demonstrated that people with more severe depressive symptoms tended to have less stable circadian activity rhythms.

There are several key novel approaches regarding circadian rhythm analysis in our study. First, we used step counts and heart rate data from consumer wearables instead of "activity counts"—a measure of total linear acceleration—from research-grade devices commonly used in other studies. There are few studies that used alternative source data for circadian rhythm analysis; for example, heart rate [32] or skin temperature [36]. Second, we analyzed weekday circadian rhythms separately that were found to be stronger predictors rather than rhythms based on all days. Although weekday rhythm is mainly determined by work routine, the ability to adherently follow this routine better discriminates between depressed and healthy individuals, where healthy people demonstrated a greater regularity. Third, we showed the value of novel rhythm stability metrics—autocorrelation and ICV—as risk markers of depression. Finally, we first showed that a greater severity of depressive symptoms was associated with a fewer steps-based daily peaks, which perhaps reflects fewer distinct activities happening over a day, indicating a diagnostic symptom of anhedonia (ie, loss of interest in activities).

Secondly, our findings suggest that a greater variation of nighttime heart rate between 2 AM and 4 AM and between 4 AM and 6 AM indicates greater severity of depressive symptoms, which is aligned with previous electrocardiogram research that showed that changes in heart rate during sleep may be a valid physiological marker of depression [56,61]. However, study participants were recruited from a sleep disorder clinic and had sleep complaints apart from diagnosed depressive disorder [56].

Thirdly, in contrast to some previous findings [15-17,19,28,33,62], our data did not show reduced levels of locomotor activity in depressed participants in terms of less time spent in moderate to vigorous physical activity, fewer daily steps, or more sedentary time. This might be due to the overestimation of the time spent in high-intensity activities by consumer wearables [42] and due to the overall high level of physical activity in the Singapore population [63], where low physical activity may be a rare depression risk marker. In addition, our data did not show a relationship with cosinor-based metrics estimating the level of activity, including mesor and rhythm amplitude [24,25], or with nonparametric measures, including M10, L5, and RA.

Finally, the analysis of sleep data showed that later sleep midpoint and offset time were associated with more severe depressive symptoms, which is consistent with the existing evidence [16,20,22,23]; however, we did not find that shortened sleep duration, increased SOL, lower SE, and longer WASO were related to more severe symptoms, contributing to the mixed results from previous actigraphy studies [16,22-24,34]. This discrepancy may be explained by the lack of participants with clinical depression in our sample or by the limited accuracy of Fitbit wearables in measuring sleep compared to PSG [45,46].

The results of our study are also comparable to a few previous studies that used machine learning with wearable sensor data for depression detection. Jacobson et al achieved a high accuracy in detecting depressed individuals, but their approach has some limitations [37]. First, although their model classified clinically diagnosed patients and healthy controls, 5 out of 23 depressed patients in their sample were hospitalized, which significantly limits generalizability of their actigraphy data–based model. Second, they extracted and explored thousands of features which, without correction for multiple comparisons, by chance might correlate to the outcome variable in the given sample but may not in other samples; therefore, it is highly likely that significant digital biomarkers will be inconsistent across different samples. Third, the number of spectral analysis–based features depended on the minimum duration of actigraphy data available across participants; therefore, some features are not universal and would be unavailable for a shorter observation period. Finally, most features were extracted mechanically without relying on domain knowledge or previous findings and remained uninterpreted. For example, interpretation of spectral density features, which were the only important predictors in their model, remained unclear. Contrary to this study, our approach relies on interpretable digital biomarkers and meaningful behavioral and physiological phenomena underlying these markers.

In another study, Tazawa at al achieved an accuracy of 76% in detection of depressed individuals based on 236 assessments of 85 participants, which is very similar to the performance of our models trained with contrasted subsamples [33]. Although our best model had 9% higher sensitivity (82% vs 73%), which is more important than higher specificity if using these models for passive screening to address underdiagnosis of depression, the direct comparison between studies is problematic due to the specific downsampling used in our models. In addition, there are important differences between studies in both available sensor data and outcome measurement. First, Tazawa and colleagues used the Hamilton Depression Rating Scale for symptom assessment, and their sample included patients with clinical depression. Second, they had more types of sensor data, including skin temperature, which were the most indicative of depression. Finally, they mainly used distribution characteristics of the per-hour data and correlations between different data types as digital biomarkers but did not use circadian rhythm metrics.

It is worth mentioning the study by Sano et al, whose model based on wearable sensor data achieved a comparable classification accuracy of 87% [39]. The important methodological similarity between the studies is that Sano et al

similarly used a contrasted subsample for training of the models, yet theirs was comprised of an even smaller fraction and number of participants (ie, top 12% and bottom 12% of participants; n=47). Despite methodological similarities, the studies are different in terms of population, outcome measurement, and extraction of digital biomarkers. They studied college students and used the mental component summary score from the 12-item Short Form Health Survey for mental well-being assessment, which was not intended to screen for depression, unlike the PHQ-9 used in our study. Furthermore, they collected skin conductivity and skin temperature data in addition to accelerometer data and used data distribution characteristics (eg, mean and median) at different times of day as digital biomarkers, but they did not analyze and harness circadian rhythm metrics. Overall, the key novel approach in our study, as compared to existing efforts, was the use of data from widespread consumer wearables and the use of circadian rhythm metrics as digital biomarkers in predictive modeling with machine learning.

Possible Mechanisms

Regularity of circadian rhythm and variation of nighttime heart rate were the most robust digital biomarkers; below, we outline possible psychosocial and neurobiological mechanisms linking them to depressive disorder. The relationships between depression and circadian rhythms in behavior and physiology are probably bidirectional, but underlying neurobiological mechanisms remain unknown [64,65]. Existing evidence shows that disturbed rest-activity rhythms, as, for example, in shift workers, lead to desynchronization of internal molecular clocks, thereby disturbing circadian biochemical processes, secretion of hormones, metabolic functions, and physiological parameters [66-68]. In turn, disturbed master clocks at the molecular level could lead to neurobiological dysfunction that may generate depressive mood [69]. It has been documented that patients with major depression have elevated nocturnal body temperature, increased cortisol, lower melatonin, and lower norepinephrine levels [70,71]. On the other hand, mood disorders affect circadian activity rhythms through psycho-cognitive pathways: a depressed individual can experience increased apathy, impaired deliberative cognitive control, greater impulsivity, and other affects, which may result in inconsistent behavior, disturbed routine, and disturbed circadian rhythms. We may speculate that observed associations either support the social rhythm hypothesis or probably capture nuanced behavioral manifestations of depressive symptomatology [72-74]. Regarding the variation of heart rate in nighttime intervals, this digital biomarker could indicate depressive disorder because an increased arousal of autonomic nervous system that is possible with depression is likely to affect heart rate dynamics during sleep [56].

Strengths and Limitations

Strengths of our study include a relatively long period of continuous sensing and activity tracking in free-living settings, a relatively large workplace-based sample, use of correction for multiple testing in statistical analysis, use of a wide range of covariates for model adjustments in regression analysis, and use of cross-validation with multiple resampling in machine learning modeling. Moreover, as digital biomarkers, we used only metrics that meaningfully characterize everyday behavior and human physiology relevant to depressive disorder, avoiding extraction of a multitude uninterpretable features and black box approaches.

This study also has several limitations. First, we studied working adults who represent the generally healthy population and had not been diagnosed with depression. The absence of participants with clinical depression might cause a lack of contrast in behavioral and physiological data between depressed and healthy participants, which impedes discovery of reliable digital biomarkers. In addition, we used the PHQ-9 for depression screening, which is a self-reported scale with limited accuracy. Second, our participants were mostly highly educated university employees with sedentary jobs who might have specific psycho-behavioral characteristics. Third, the cross-sectional design of the study does not allow causal inferences. Fourth, the poor predictability of depressive symptomatology in the whole sample highlights a possible limitation, in principle, of using this set of digital biomarkers alone for depression screening universally due to prominent interindividual differences and heterogeneity that appear in naturalistic settings [36]. Finally, predictive models were retrained using balanced and contrasted subsamples equally comprised of depressed participants and healthy participants with no risk of depression; therefore, they probably suffer from overfitting and would perform worse on new samples. However, the set of selected digital biomarkers used in these models remained the same as in models with the full data set.

Conclusions and Future Research

Further discovery of digital biomarkers from wearable sensors has the potential to facilitate early, unobtrusive, continuous, and cost-effective detection of depression in the general population. This study showed that some known and novel digital biomarkers based on data from consumer wearables could indicate increased risk of depression in the working population. The predictive model based on a combination of these digital biomarkers could discriminate individuals with high risk of depression from individuals with no risk. Further research should examine and validate digital biomarkers with longitudinal design, because dynamic changes and deviations from a baseline are more likely to indicate risk of depression rather than one-time snapshots. Second, although the idea of inferring universal digital biomarkers is very tempting, the development of semipersonalized models adjusted for differential baseline characteristics can bring more accurate and clinically relevant predictions. Third, wearable sensor data can be enriched with smartphone data, which will enable more comprehensive digital phenotyping for depression detection [39,75].

Acknowledgments

This research was supported, in part, by the Singapore Ministry of National Development and the National Research Foundation, Prime Minister's Office, under the Land and Liveability National Innovation Challenge (L2NIC) Research Programme (L2NIC Award No. L2NIC FP1-2013-2). Analysis was conducted under the auspices of the Healthy Campus Initiative by the Ministry of Health Office of Transformation. Staff time was supported by Centre for Population Health Sciences funding from the Nanyang Technological University (NTU). Any opinions, findings, and conclusions expressed in this material are those of the authors and do not reflect the views of the Singapore Ministry of National Development and National Research Foundation, Prime Minister's Office, Singapore, or the Ministry of Health, Singapore. The data that support the findings of this study are available upon reasonable request to the corresponding author and pending eligibility to access data governed by the NTU and the Singapore Ministry of National Development.

Authors' Contributions

JC, GC, and YR formulated the study idea and design. YR led the data processing and analysis and drafted the manuscript. TQT and IB helped with data processing and commented on the methods and versions of the manuscript. JC and GC provided overall supervision, secured the funding, and commented on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables with statistics.
[PDF File (Adobe PDF File), 145 KB - mhealth_v9i10e24872_app1.pdf]

Multimedia Appendix 2
Digital biomarkers.
[PDF File (Adobe PDF File), 150 KB - mhealth_v9i10e24872_app2.pdf]

Multimedia Appendix 3
R-code scripts for data processing and statistical analysis.
[7Z File , 10 KB - mhealth_v9i10e24872_app3.7z]

Multimedia Appendix 4
Participant flowchart. HR: heart rate.
[PNG File , 72 KB - mhealth_v9i10e24872_app4.png]

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: A meta-analysis. Lancet 2009 Aug 22;374(9690):609-619. [doi: 10.1016/S0140-6736(09)60879-5] [Medline: 19640579]
3. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
4. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
5. Xiong J, Lipsitz O, Nasri F, Lui LMW, 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]
6. Wu T, Jia X, Shi H, Niu J, Yin X, Xie J, et al. Prevalence of mental health problems during the COVID-19 pandemic: A systematic review and meta-analysis. J Affect Disord 2021 Feb 15;281:91-98 [FREE Full text] [doi: 10.1016/j.jad.2020.11.117] [Medline: 33310451]
7. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018 Dec 12;1(1):67 [FREE Full text] [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]

8. Insel TR. Digital phenotyping: Technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
9. 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]
10. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. Physiol Meas 2018 May 15;39(5):05TR01 [FREE Full text] [doi: 10.1088/1361-6579/aabf64] [Medline: 29671754]
11. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: A survey. Pervasive Mob Comput 2018 Dec;51:1-26. [doi: 10.1016/j.pmcj.2018.09.003]
12. Skinner AL, Attwood AS, Baddeley R, Evans-Reeves K, Bauld L, Munafò MR. Digital phenotyping and the development and delivery of health guidelines and behaviour change interventions. Addiction 2017 Jul;112(7):1281-1285. [doi: 10.1111/add.13746] [Medline: 28472848]
13. Torous J, Hsin H. Empowering the digital therapeutic relationship: Virtual clinics for digital health interventions. NPJ Digit Med 2018 May 16;1(1):16 [FREE Full text] [doi: 10.1038/s41746-018-0028-2] [Medline: 31304301]
14. Kollins S, DeLoss D, Cañadas E, Lutz J, Findling R, Keefe R, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): A randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178 [FREE Full text] [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
15. Burton C, McKinstry B, Szentagotai Tătar A, Serrano-Blanco A, Pagliari C, Wolters M. Activity monitoring in patients with depression: A systematic review. J Affect Disord 2013 Mar 15;145(1):21-28. [doi: 10.1016/j.jad.2012.07.001] [Medline: 22868056]
16. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: A systematic review and meta-analysis. J Affect Disord 2019 Jun 15;253:257-269. [doi: 10.1016/j.jad.2019.04.087] [Medline: 31060012]
17. Schuch FB, Vancampfort D, Firth J, Rosenbaum S, Ward PB, Silva ES, et al. Physical activity and incident depression: A meta-analysis of prospective cohort studies. Am J Psychiatry 2018 Jul 01;175(7):631-648. [doi: 10.1176/appi.ajp.2018.17111194] [Medline: 29690792]
18. Zhai L, Zhang Y, Zhang D. Sedentary behaviour and the risk of depression: A meta-analysis. Br J Sports Med 2015 Jun;49(11):705-709. [doi: 10.1136/bjsports-2014-093613] [Medline: 25183627]
19. Vallance JK, Winkler EAH, Gardiner PA, Healy GN, Lynch BM, Owen N. Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). Prev Med 2011 Oct;53(4-5):284-288. [doi: 10.1016/j.ypmed.2011.07.013] [Medline: 21820466]
20. Benca RM, Obermeyer WH, Thisted RA, Gillin JC. Sleep and psychiatric disorders. A meta-analysis. Arch Gen Psychiatry 1992 Aug 01;49(8):651-668; discussion 669. [doi: 10.1001/archpsyc.1992.01820080059010] [Medline: 1386215]
21. Baglioni C, Nanovska S, Regen W, Spiegelhalder K, Feige B, Nissen C, et al. Sleep and mental disorders: A meta-analysis of polysomnographic research. Psychol Bull 2016 Sep;142(9):969-990 [FREE Full text] [doi: 10.1037/bul0000053] [Medline: 27416139]
22. Robillard R, Naismith SL, Smith KL, Rogers NL, White D, Terpening Z, et al. Sleep-wake cycle in young and older persons with a lifetime history of mood disorders. PLoS One 2014 Feb 25;9(2):e87763 [FREE Full text] [doi: 10.1371/journal.pone.0087763] [Medline: 24586290]
23. Robillard R, Carpenter JS, Rogers NL, Fares S, Grierson AB, Hermens DF, et al. Circadian rhythms and psychiatric profiles in young adults with unipolar depressive disorders. Transl Psychiatry 2018 Oct 09;8(1):213 [FREE Full text] [doi: 10.1038/s41398-018-0255-y] [Medline: 30301878]
24. Hori H, Koga N, Hidese S, Nagashima A, Kim Y, Higuchi T, et al. 24-h activity rhythm and sleep in depressed outpatients. J Psychiatr Res 2016 Jun;77:27-34. [doi: 10.1016/j.jpsychires.2016.02.022] [Medline: 26978182]
25. Maglione JE, Ancoli-Israel S, Peters KW, Paudel ML, Yaffe K, Ensrud KE, Study of Osteoporotic Fractures Research Group. Depressive symptoms and circadian activity rhythm disturbances in community-dwelling older women. Am J Geriatr Psychiatry 2014 Apr;22(4):349-361 [FREE Full text] [doi: 10.1016/j.jagp.2012.09.003] [Medline: 23567424]
26. Smagula SF, Ancoli-Israel S, Blackwell T, Boudreau R, Stefanick ML, Paudel ML, Osteoporotic Fractures in Men (MrOS) Research Group. Circadian rest-activity rhythms predict future increases in depressive symptoms among community-dwelling older men. Am J Geriatr Psychiatry 2015 May;23(5):495-505 [FREE Full text] [doi: 10.1016/j.jagp.2014.06.007] [Medline: 25066948]
27. Berle JO, Hauge ER, Oedegaard KJ, Holsten F, Fasmer OB. Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. BMC Res Notes 2010 May 27;3:149 [FREE Full text] [doi: 10.1186/1756-0500-3-149] [Medline: 20507606]
28. Difrancesco S, Lamers F, Riese H, Merikangas KR, Beekman ATF, van Hemert AM, et al. Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. Depress Anxiety 2019 Oct;36(10):975-986 [FREE Full text] [doi: 10.1002/da.22949] [Medline: 31348850]
29. Smagula SF, Boudreau RM, Stone K, Reynolds CF, Bromberger JT, Ancoli-Israel S, Osteoporotic Fractures in Men (MrOS) Research Group. Latent activity rhythm disturbance sub-groups and longitudinal change in depression symptoms among

older men. Chronobiol Int 2015;32(10):1427-1437 [FREE Full text] [doi: 10.3109/07420528.2015.1102925] [Medline: 26594893]
30. Smagula SF, Krafty RT, Thayer JF, Buysse DJ, Hall MH. Rest-activity rhythm profiles associated with manic-hypomanic and depressive symptoms. J Psychiatr Res 2018 Jul;102:238-244 [FREE Full text] [doi: 10.1016/j.jpsychires.2018.04.015] [Medline: 29705489]
31. Krafty RT, Fu H, Graves JL, Bruce SA, Hall MH, Smagula SF. Measuring variability in rest-activity rhythms from actigraphy with application to characterizing symptoms of depression. Stat Biosci 2019 Jul;11(2):314-333 [FREE Full text] [doi: 10.1007/s12561-018-09230-2] [Medline: 31788133]
32. Jarczok MN, Aguilar-Raab C, Koenig J, Kaess M, Borniger JC, Nelson RJ, et al. The heart's rhythm 'n' blues: Sex differences in circadian variation patterns of vagal activity vary by depressive symptoms in predominantly healthy employees. Chronobiol Int 2018 Jul;35(7):896-909. [doi: 10.1080/07420528.2018.1439499] [Medline: 29543518]
33. Tazawa Y, Liang K, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: Screening and assessing patient severity utilizing machine-learning. Heliyon 2020 Mar;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
34. Luik AI, Zuurbier LA, Direk N, Hofman A, Van Someren EJ, Tiemeier H. 24-hour activity rhythm and sleep disturbances in depression and anxiety: A population-based study of middle-aged and older persons. Depress Anxiety 2015 Sep;32(9):684-692. [doi: 10.1002/da.22355] [Medline: 25693731]
35. Rock P, Goodwin G, Harmer C, Wulff K. Daily rest-activity patterns in the bipolar phenotype: A controlled actigraphy study. Chronobiol Int 2014 Mar;31(2):290-296 [FREE Full text] [doi: 10.3109/07420528.2013.843542] [Medline: 24517177]
36. Lorenz N, Spada J, Sander C, Riedel-Heller SG, Hegerl U. Circadian skin temperature rhythms, circadian activity rhythms and sleep in individuals with self-reported depressive symptoms. J Psychiatr Res 2019 Oct;117:38-44. [doi: 10.1016/j.jpsychires.2019.06.022] [Medline: 31279242]
37. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019 Feb 1;2(1):3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
38. Jacobson NC. Digital phenotyping mood disorders reproducible R code. Nicholas C Jacobson. 2018 Sep 12. URL: http://www.nicholasjacobson.com/post/digital_biomarkers_mood_disorders/ [accessed 2021-09-30]
39. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
40. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014 Dec 23;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
41. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2018 Jul;52(13):844-850. [doi: 10.1136/bjsports-2016-096990] [Medline: 28483930]
42. 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 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
43. Rowe VT, Neville M. Measuring reliability of movement with accelerometry: Fitbit versus ActiGraph. Am J Occup Ther 2019 Mar 01;73(2):1-6. [doi: 10.5014/ajot.2019.030692] [Medline: 30915976]
44. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
45. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
46. Liang Z, Chapa-Martell MA. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e13384 [FREE Full text] [doi: 10.2196/13384] [Medline: 31172956]
47. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
48. Mastin DF, Bryson J, Corwyn R. Assessment of sleep hygiene using the Sleep Hygiene Index. J Behav Med 2006 Jun;29(3):223-227. [doi: 10.1007/s10865-006-9047-6] [Medline: 16557353]
49. Johns M, Hocking B. Daytime sleepiness and sleep habits of Australian workers. Sleep 1997 Oct;20(10):844-849. [doi: 10.1093/sleep/20.10.844] [Medline: 9415943]
50. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [doi: 10.1037//0022-3514.39.3.472] [Medline: 7431205]
51. US Department of Health and Human Services. Promoting Physical Activity: A Guide for Community Action. Champaign, IL: Human Kinetics; 1999.

52. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
53. Cornelissen G. Cosinor-based rhythmometry. Theor Biol Med Model 2014 Apr 11;11:16 [FREE Full text] [doi: 10.1186/1742-4682-11-16] [Medline: 24725531]
54. Gonçalves BSB, Adamowicz T, Louzada FM, Moreno CR, Araujo JF. A fresh look at the use of nonparametric analysis in actimetry. Sleep Med Rev 2015 Apr;20:84-91. [doi: 10.1016/j.smrv.2014.06.002] [Medline: 25065908]
55. van Brakel JPG. Peak signal detection in realtime timeseries data: Robust peak detection algorithm using z-scores. Stack Overflow. 2014. URL: https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data [accessed 2021-09-30]
56. Saad M, Ray LB, Bujaki B, Parvaresh A, Palamarchuk I, De Koninck J, et al. Using heart rate profiles during sleep as a biomarker of depression. BMC Psychiatry 2019 Jun 07;19(1):168 [FREE Full text] [doi: 10.1186/s12888-019-2152-1] [Medline: 31174510]
57. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
58. Olson RS, Cava WL, Mustahsan Z, Varik A, Moore JH. Data-driven advice for applying machine learning to bioinformatics problems. In: Proceedings of the Pacific Symposium on Biocomputing 2018. 2018 Presented at: Pacific Symposium on Biocomputing 2018; January 3-7, 2018; Big Island, Hawaii p. 192-203 URL: https://psb.stanford.edu/psb-online/proceedings/psb18/olson.pdf
59. 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 Mar 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
60. Smagula SF. Opportunities for clinical applications of rest-activity rhythms in detecting and preventing mood disorders. Curr Opin Psychiatry 2016 Nov;29(6):389-396 [FREE Full text] [doi: 10.1097/YCO.0000000000000283] [Medline: 27636598]
61. Stampfer HG. The relationship between psychiatric illness and the circadian pattern of heart rate. Aust N Z J Psychiatry 1998 Apr;32(2):187-198. [doi: 10.3109/00048679809062728] [Medline: 9588297]
62. Minaeva O, Booij SH, Lamers F, Antypa N, Schoevers RA, Wichers M, et al. Level and timing of physical activity during normal daily life in depressed and non-depressed individuals. Transl Psychiatry 2020 Jul 30;10(1):259 [FREE Full text] [doi: 10.1038/s41398-020-00952-w] [Medline: 32732880]
63. Executive Summary on National Population Health Survey 2016/17. The Singapore Ministry of Health. URL: https://www.moh.gov.sg/resources-statistics/reports/national-population-health-survey-2016-17 [accessed 2021-09-30]
64. Germain A, Kupfer DJ. Circadian rhythm disturbances in depression. Hum Psychopharmacol 2008 Oct;23(7):571-585 [FREE Full text] [doi: 10.1002/hup.964] [Medline: 18680211]
65. Kronfeld-Schor N, Einat H. Circadian rhythms and depression: Human psychopathology and animal models. Neuropharmacology 2012 Jan;62(1):101-114. [doi: 10.1016/j.neuropharm.2011.08.020] [Medline: 21871466]
66. De Bacquer D, Van Risseghem M, Clays E, Kittel F, De Backer G, Braeckman L. Rotating shift work and the metabolic syndrome: A prospective study. Int J Epidemiol 2009 Jun 07;38(3):848-854. [doi: 10.1093/ije/dyn360] [Medline: 19129266]
67. Abbott SM, Malkani RG, Zee PC. Circadian disruption and human health: A bidirectional relationship. Eur J Neurosci 2020 Jan 03;51(1):567-583 [FREE Full text] [doi: 10.1111/ejn.14298] [Medline: 30549337]
68. Reinke H, Asher G. Crosstalk between metabolism and circadian clocks. Nat Rev Mol Cell Biol 2019 Apr 11;20(4):227-241. [doi: 10.1038/s41580-018-0096-9] [Medline: 30635659]
69. Walker WH, Walton JC, DeVries AC, Nelson RJ. Circadian rhythm disruption and mental health. Transl Psychiatry 2020 Jan 23;10(1):28 [FREE Full text] [doi: 10.1038/s41398-020-0694-0] [Medline: 32066704]
70. Koenigsberg HW, Teicher MH, Mitropoulou V, Navalta C, New AS, Trestman R, et al. 24-h monitoring of plasma norepinephrine, MHPG, cortisol, growth hormone and prolactin in depression. J Psychiatr Res 2004;38(5):503-511. [doi: 10.1016/j.jpsychires.2004.03.006] [Medline: 15380401]
71. Monteleone P, Maj M. The circadian basis of mood disorders: Recent developments and treatment implications. Eur Neuropsychopharmacol 2008 Oct;18(10):701-711. [doi: 10.1016/j.euroneuro.2008.06.007] [Medline: 18662865]
72. Ehlers CL, Frank E, Kupfer DJ. Social zeitgebers and biological rhythms. A unified approach to understanding the etiology of depression. Arch Gen Psychiatry 1988 Oct 01;45(10):948-952. [doi: 10.1001/archpsyc.1988.01800340076012] [Medline: 3048226]
73. Grandin LD, Alloy LB, Abramson LY. The social zeitgeber theory, circadian rhythms, and mood disorders: Review and evaluation. Clin Psychol Rev 2006 Oct;26(6):679-694. [doi: 10.1016/j.cpr.2006.07.001] [Medline: 16904251]
74. Fuchikawa T, Eban-Rothschild A, Nagari M, Shemesh Y, Bloch G. Potent social synchronization can override photic entrainment of circadian rhythms. Nat Commun 2016 May 23;7(1):11662 [FREE Full text] [doi: 10.1038/ncomms11662] [Medline: 27210069]

75. Torous J, Staples P, Barnett I, Sandoval LR, Keshavan M, Onnela JP. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. NPJ Digit Med 2018;1:15 [FREE Full text] [doi: 10.1038/s41746-018-0022-8] [Medline: 31304300]

Abbreviations

AUC: area under the curve
bpm: beats per minute
CV: coefficient of variation
DART: Dropouts meet Multiple Additive Regression Trees
FDR: false discovery rate
ICV: interdaily coefficient of variation
IS: interdaily stability
IV: intradaily variability
L2NIC: Land and Liveability National Innovation Challenge
L5: mean value of the 5 consecutive least active hours of the day
M10: mean activity of the 10 consecutive most active hours of the day
MAE: mean absolute error
MET: metabolic equivalent
NTU: Nanyang Technological University
PHQ-9: 9-item Patient Health Questionnaire
PSG: polysomnography
RA: relative amplitude
REDCap: Research Electronic Data Capture
REM: rapid eye movement
RHR: resting heart rate
RMSE: root mean square error
RMSSD: root mean square of successive differences
SE: sleep efficiency
SOL: sleep onset latency
TST: total sleep time
UCLA: University of California, Los Angeles
WASO: wake after sleep onset

Edited by L Buis; submitted 09.10.20; peer-reviewed by S Smagula, E Lee, R Picard, E Moon; comments to author 21.12.20; revised version received 05.04.21; accepted 15.07.21; published 25.10.21.

Please cite as:
Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J
Digital Biomarkers for Depression Screening With Wearable Devices: Cross-sectional Study With Machine Learning Modeling
JMIR Mhealth Uhealth 2021;9(10):e24872
URL: https://mhealth.jmir.org/2021/10/e24872
doi:10.2196/24872
PMID:34694233

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org