JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 6 (2018), Issue 11 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Original Papers

Reviews

Corrigenda and Addenda

Original Paper

Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis

Hussein Jamaladin[1], MSc; Tom H van de Belt[2], PhD; Lianda CH Luijpers[1], BSc; Falco R de Graaff[1], BSc; Sebastian JH Bredie[2,3], MD, PhD; Nel Roeleveld[1], PhD; Marleen MHJ van Gelder[1,2], PhD

[1]Department for Health Evidence, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[2]Radboud REshape Innovation Center, Radboud University Medical Center, Nijmegen, Netherlands

[3]Department of Internal Medicine, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Hussein Jamaladin, MSc
Department for Health Evidence
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein-Noord 21
Nijmegen,
Netherlands
Phone: 31 024 366684
Email: hussein.jamaladin@radboudumc.nl

Abstract

Background: Using a mobile app for self-management could make it easier for patients to get insight into their blood pressure patterns. However, little is known about the availability, quality, and features of mobile apps targeting blood pressure.

Objective: The goal of the research was to determine the availability, functionality, and quality of mobile apps that could be used for blood pressure monitoring purposes.

Methods: A systematic app search was performed based on the standards for systematic reviews. We searched the Dutch official app stores for Android and iOS platforms using predefined keywords and included all English and Dutch mobile apps targeting blood pressure. Two independent assessors determined eligibility and quality of the apps using the 5-point Mobile App Rating Scale (MARS). Quality scores of the apps with and without 17 a priori selected characteristics were compared using independent sample t tests.

Results: A total of 184 apps (104 Android, 80 iOS) met the inclusion criteria. The mean overall MARS score was 2.63 (95% CI 2.55-2.71) for Android and 2.64 (95% CI 2.56-2.71) for iOS. The apps Bloeddruk (4.1) and AMICOMED BP (3.6) had the highest quality scores on the Android and iOS platforms, respectively. Of the app characteristics recorded, only pricing, in-app advertisements, and local data storage were not associated with the quality scores. In only 3.8% (7/184) of the apps, involvement of medical experts in its development was mentioned, whereas none of the apps was formally evaluated with results published in a peer-reviewed journal.

Conclusions: This study provides an overview of the best apps currently available in the app stores and important key features for self-management that can be used by health care providers and patients with hypertension to identify a suitable app targeting blood pressure monitoring. However, the majority of the apps targeting blood pressure monitoring were of poor quality. Therefore, it is important to involve medical experts in the developmental stage of health-related mobile apps to improve the quality of these apps.

(JMIR Mhealth Uhealth 2018;6(11):e187) doi:10.2196/mhealth.9888

KEYWORDS

mobile phone; mHealth; app review; high blood pressure; self-management; mobile app

Introduction

Hypertension contributes to the burden of various diseases including heart disease [1] and leads to premature mortality and morbidity [2]. Globally, 1 in 5 adults has hypertension [3], and 12.8% of total deaths worldwide are caused by high blood pressure [4]. Hypertension is a chronic disease that may result in severe complications such as stroke [5,6] and chronic kidney disease [7,8]. High blood pressure is treated in the long term pharmacologically and/or through lifestyle adjustments [9,10]. Therefore, continuous preventive care and daily management of patients is important in controlling blood pressure. Self-measured blood pressure monitoring leads to better blood pressure control [11] and enhances medication adherence [12,13]. In addition, self-measured blood pressure eliminates white coat hypertension and can be useful in the detection of masked hypertension [14].

The ownership of mobile phones is increasing rapidly and by the end of 2017, over one-third of consumers worldwide owned a mobile phone [15,16]. The health app market is growing and thousands of new health apps are published every year. In 2016, approximately 100,000 new health-related apps were published, resulting in a total of 259,000 health apps currently available in the major app stores [17]. The health app download rates were estimated to reach 3.2 billion in 2016, an increase of 7% compared to 2015 [17]. The main target areas of app developers are chronic diseases such as hypertension and diabetes [17]. A survey performed by Accenture showed that the use of health apps and wearables increased by almost 50% among consumers in 2016 compared to 2014 [18]. Furthermore, patients and physicians agree about the potential benefits of health apps and wearables [18].

Using a mobile app for self-management purposes could make it easier for patients with hypertension to have insight into and control their blood pressure. These apps may have several useful features: logbook or diary features facilitate logging of blood pressure measurements in an organized way, while reminder functions could facilitate monitoring and medication adherence. In addition, health apps may provide valuable background information for patients about the disease, its treatment, how to measure blood pressure adequately, and lifestyle management. Analysis tools (eg, graphs and trend analysis) may provide an overview of the course of blood pressure over time. Furthermore, some apps can export blood pressure readings and other user data to be sent by email. This enables patients to share their measurements with their health care providers and relatives.

Although mobile apps have the potential to be beneficial for patients with hypertension, little is known about the availability, quality, and features of mobile apps targeting blood pressure. Therefore, the aim of this study was to perform a systematic review of apps to determine the availability, functionality, and quality of mobile apps that could be used to collect readings of blood pressure for monitoring purposes.

Methods

App Search and Selection

We performed a systematic app search based on the standards for systematic reviews. Although we followed the standards for systematic reviews of scientific literature, these guidelines are not completely applicable to app reviews. All apps available in the Google Play store for Android and iOS App Store targeting blood pressure monitoring in which blood pressure measurements could be entered manually were potentially eligible for inclusion. On March 1, 2016, the Dutch app stores were searched using the following search terms: blood pressure, diastolic, health, heart rate, hypertension, hypotension, pressure, systolic, and their Dutch equivalents (bloeddruk, diastole, gezondheid, hartslag, hypertensie, hypotensie, and systole). Apps in languages other than English or Dutch and duplicates and irrelevant apps, such as games, were excluded. Some apps had a free version and a pro version. In cases where there was no difference in the functionality and relevant features between the two versions, only the free version was included in this app review. Two independent assessors (HJ and FRdG) selected the eligible apps based on app titles, description of the app in the app store, and screenshots provided. Discrepancies were discussed until a final decision was reached.

Data Extraction

The selected apps were downloaded on either a Samsung Galaxy S6 (Android version 6.0.1) or an iPhone 5c (iOS version 9.3.5) for complete assessment of eligibility and characteristics. Two independent assessors (HJ and LCHL) tested each app on each platform in duplicate for a minimum of 10 minutes before performing the final assessment. Using a standardized form, the assessors recorded technical app information and app features. Recorded technical information included the name of the app, app developer, version number, platform, affiliations of the app developers, price, number of ratings of all versions in the app store, star ratings in app store, whether Web access was required, data storage location (local and/or cloud), and whether the app was free of advertisements. App features included the ability of the app to register age, gender, height, weight, time, and date of blood pressure reading, measurement site (eg, left or right arm), and measurement position (eg, sitting or standing). We also registered the presence of a reminder function, analysis functions, data export, wireless transfer of measurement data from a blood pressure monitor, and whether user data were password-protected. Based on national and international guidelines for the management of hypertension [9,19-24] and recommendations from the literature addressing the management of hypertension [25-28], we selected 6 key app features that are essential for self-management. The key features included the ability to export data, send reminders, analyze data, record time and date of blood pressure reading, record weight, and provide information/education. In addition, we searched the app descriptions in the app stores for the involvement of medical experts in the development of the app. Furthermore, we searched PubMed and Google Scholar in March 2017 to determine whether the apps were trialed or evaluated with results published in peer-reviewed journals.

App Quality Rating

The quality of the apps was evaluated using a standard assessment protocol based on the Mobile App Rating Scale (MARS) [29], a questionnaire that measures app quality using 23 questions divided into 4 objective categories (engagement, functionality, aesthetics, and information quality) and 1 subjective category. Each question was rated on a 5-point scale (1-inadequate, 2-poor, 3-acceptable, 4-good, 5-excellent). The MARS overall score was calculated by averaging the means of the 4 objective categories. The developers of MARS recommend a training to standardize the assessors' ratings [29], so the assessors watched the MARS training video available on YouTube [29]. Afterwards 10 randomly selected apps were used for training purposes. The assessors discussed each item of the MARS scale and reached consensus on the scores during the training. After these 10 apps, the assessors did not discuss any apps and rated them independently.

Statistical Analysis

We calculated the scores of the MARS separately per assessor and averaged the scores at total level. The distributions of the scores were checked for normality. We measured the interrater reliability of the MARS scores using the intraclass correlation coefficient (ICC). Based on the ICC guidelines developed by Shrout and Fleiss [30], we used a 2-way mixed effects, average measures model with a consistency of agreement definition [31]. Cronbach alpha was used to assess the internal consistency reliability (ie, the extent to which all items in a scale measure the same concept) of the MARS [32].

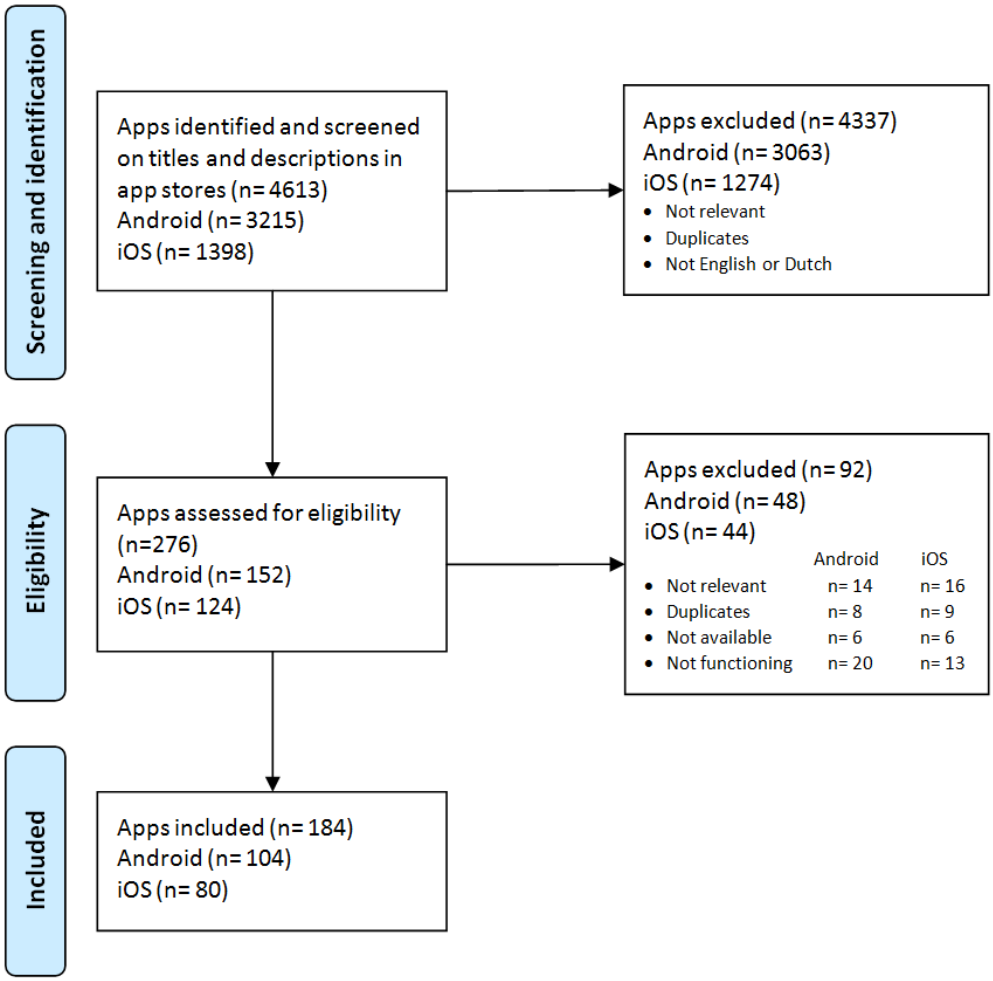
Descriptive statistics were used to summarize and evaluate the app features. Differences in proportions were tested using chi-square tests. To determine whether specific characteristics were associated with quality scores, MARS scores of apps with and without the a priori selected characteristics were compared using independent sample *t* tests. Pearson correlation coefficients were calculated to compare the MARS overall scores with star ratings obtained from the app stores of Android and iOS. Only apps with 10 or more user ratings were included in this analysis. Statistical significance was set at $P<.05$. The data were analyzed using SPSS Statistics version 22 (IBM Corp).

Results

App Selection

A total of 4613 apps were identified using the search terms. Screening on app titles and descriptions in the app stores resulted in 276 potentially eligible apps. Further assessment, performed after downloading and testing of the selected apps, resulted in the inclusion of 184 apps, of which 104 were Android apps and 80 iOS apps. Some apps were available on both platforms so were included twice in this study. Figure 1 illustrates the selection procedure.

Figure 1. Flow diagram of the review and selection process.



App Characteristics

The app characteristics are summarized in Table 1. The majority of the apps included (77.7%, 143/184) were free of charge. Android apps were more often free compared to iOS apps (89.4% [93/104] vs 62.5% [50/80], *P*<.001). However, iOS apps were more often free of advertisements compared to Android apps (80.0% [40/50] vs 59.1% [55/93], *P*=.01). Less than half of the apps tracked personal background data, with the exception of weight tracking in Android apps. Only a small proportion of apps recorded how the blood pressure measurements were performed (measurement side, 15.2%, 28/184; position, 14.1%, 26/184). Compared to iOS apps, Android apps more often recorded measurement side (20.2% [21/104] vs 8.8% [7/80], *P*=.03) and position (19.2% [20/104] vs 7.5% [6/80], *P*=.02). Only 2 apps did not register time and date of the measurement. A reminder function was present in 28.8% (53/184) of apps, and data export was possible in 65.2% (120/184) of apps. The latter was more often available in iOS apps than in Android apps (75.0% [60/80] vs 57.7% [60/104], *P*=.02). A total of 26 apps, of which 15 were Android and 11 iOS, contained 5 predefined key features. None of the apps contained the key feature information/education.

A total of 182 apps were developed by commercial (31) or unknown (151) developers. Only 2 apps (Heartkeeper and Blood pressure monitoring for pregnancy) were developed by universities or nongovernmental organizations. Of all apps included, only 3.8% (7/184) stated the involvement of medical experts in the development of the app in their own app store description. None of the apps was trialed or evaluated with results published in a peer-reviewed journal.

App Quality

The MARS overall and subscale scores were normally distributed. For iOS apps, the overall interrater reliability of the MARS was fair (ICC=0.42, 95% CI 0.21-0.59) and the Cronbach alpha was 0.59. For Android apps, the interrater reliability was slightly higher (ICC= 0.53, 95% CI 0.38-0.66) and the Cronbach alpha was 0.70. Table 2 shows the ICCs and MARS scores for the 5 categories and the MARS overall score per platform.

Table 1. Summary of characteristics of included apps, stratified by platform.

Characteristics	All apps (n=184), n (%)	Android (n=104), n (%)	iOS (n=80), n (%)	*P* values
Pricing				
Free	143 (77.7)	92 (89.4)	50 (62.5)	<.001
No advertisements[a]	95 (66.4)	55 (59.1)	40 (80.0)	.01
Personal data				
Age	75 (40.8)	42 (40.4)	33 (41.3)	.91
Gender	75 (40.8)	41 (39.4)	34 (42.5)	.67
Height	71 (38.6)	44 (42.3)	27 (33.7)	.24
Weight[b]	99 (53.8)	62 (59.6)	37 (46.2)	.07
Blood pressure measurements				
Side (left or right arm)	28 (15.2)	21 (20.2)	7 (8.8)	.03
Position (eg, sitting, lying)	26 (14.1)	20 (19.2)	6 (7.5)	.02
Date and time[b]	182 (98.9)	103 (99.0)	79 (98.8)	.85
Other features				
Reminder function[b]	53 (28.8)	30 (28.8)	23 (28.7)	.99
Analysis tool[b]	158 (85.9)	92 (88.5)	66 (82.5)	.25
Data export[b]	120 (65.2)	60 (57.7)	60 (75.0)	.02
Data upload from blood pressure meter	27 (14.7)	12 (11.5)	15 (18.7)	.17
Needs Web access to function	10 (5.4)	8 (7.7)	2 (2.5)	.12
Password protection	43 (23.4)	20 (19.2)	23 (28.7)	.13
Data storage (local)	181 (98.4)	102 (98.1)	79 (98.8)	.72
Data storage (cloud)	19 (10.3)	15 (14.4)	4 (5.0)	.04

[a]Only free apps included; paid apps were presumed to have no advertisements.

[b]Key app features for self-management based on guidelines and literature.

Table 2. Mobile App Rating Scale scores.

MARS[a] subscale	Android			iOS		
	Mean (95% CI)	ICC[b] (95% CI)	Alpha[c]	Mean (95% CI)	ICC (95% CI)	Alpha
Engagement	2.28 (2.17-2.39)	.62 (.48-.72)	.76	2.26 (2.16-2.36)	.47 (.28-.63)	.64
Functionality	3.54 (3.46-3.62)	.23 (.03-.41)	.37	3.55 (3.46-3.65)	.35 (.13-.53)	.51
Aesthetics	3.06 (2.96-3.17)	.51 (.35-.64)	.68	3.16 (3.07-3.25)	.20 (.03-.40)	.33
Information	1.63 (.52-1.73)	.57 (.42-.69)	.73	1.57 (1.46-1.68)	.54 (.35-.68)	.70
Subjective quality	2.54 (2.35-2.74)	.56 (.41-.68)	.72	2.63 (2.42-2.84)	.49 (.29-.64)	.65
MARS overall score[d]	2.63 (2.55-2.71)	.53 (.38-.66)	.70	2.64 (2.56-2.71)	.42 (.21-.59)	.59

[a]MARS: Mobile App Rating Scale.

[b]ICC: intraclass correlation coefficient.

[c]Alpha: Cronbach alpha.

[d]Average of 4 objective subscales.

Table 3. Quality scores comparison of apps with and without a specific characteristic.

Characteristics	Present		Not present		Difference
	n (%)	Mean (SD)	n (%)	Mean (SD)	Mean (95% CI)
Pricing					
Free	143 (77.7)	2.66 (.38)	41 (22.3)	2.55 (.33)	.11 (–.24 to .02)
No advertisements[a]	95 (66.4)	2.67 (.38)	48 (33.6)	2.64 (.37)	.03 (.10 to .16)
Personal data					
Age	75(40.8)	2.78 (.39)	109 (59.2)	2.53 (.32)	.25 (.15 to .35)
Gender	75 (40.8)	2.76 (.36)	109 (59.2)	2.54 (.35)	.22 (.11 to .32)
Height	71 (38.6)	2.74 (.35)	113 (61.4)	2.56 (.37)	.18 (.07 to .28)
Weight[b]	99 (53.8)	2.74 (.37)	85 (46.2)	2.50 (.33)	.24 (.14 to .34)
Blood pressure measurements					
Side (left or right arm)	28 (15.2)	2.83 (.40)	156 (84.8)	2.60 (.35)	.23 (.09 to .38)
Position (eg, sitting, lying)	26 (14.1)	2.87 (.42)	158 (85.9)	2.60 (.35)	.27 (.12 to .42)
Date and time[b]	182 (98.9)	2.64 (.36)	2 (1.1)	2.04 (.61)	.60 (.08 to 1.01)
Other features					
Reminder function[b]	53 (28.8)	2.82 (.36)	131 (71.2)	2.56 (.35)	.26 (.15 to .38)
Analysis tool[b]	158 (85.9)	2.67 (.37)	26 (14.1)	2.40 (.26)	.27 (.12 to .42)
Data export[b]	120 (65.2)	2.71 (.39)	64 (34.8)	2.48 (.28)	.23 (.13 to .35)
Data upload from blood pressure meter	27 (14.7)	2.94 (.35)	157 (85.3)	2.58 (.35)	.36 (.22 to. 51)
Needs Web access to function	10 (5.4)	2.88 (.40)	174 (94.6)	2.62 (.36)	.26 (.03 to .50)
Password protection	43 (23.4)	2.78 (.38)	141 (76.6)	2.59 (.38)	.19 (.07 to .32)
Data storage (local)	181 (98.4)	2.62 (.36)	3 (1.6)	3.10 (.52)	–.48 (–.89 to –.05)
Data storage (cloud)	19 (10.3)	3.04 (.46)	165 (89.7)	2.58 (.33)	.46 (.29 to .62)

[a]Only free apps included; paid apps were presumed to have no advertisements.

[b]Key app features for self-management based on guidelines and literature.

On a scale from 1 to 5, the mean MARS score for the 4 objective categories was 2.6 for both platforms. The MARS scores for the separate categories were also very similar for Android and iOS. Subjective quality scored 2.5 and 2.6 for Android and iOS, respectively. Of the 5 categories, functionality received the highest score (Android 3.5, iOS 3.6) and information the lowest (Android 1.6, iOS 1.6). The complete list is available in Multimedia Appendix 1.

Among the 80 iOS apps, 7 had 10 or more user ratings in the app store, but we did not observe a correlation between the star ratings and the MARS overall score (r=0.29; P=.53). For Android, 78 apps received ≥10 user ratings; these were not correlated with the MARS scores either (r=0.17; P=.15). Table 3 shows the quality scores comparison for apps with and without each characteristic. Incorporation of all characteristics resulted in higher quality scores except for pricing, in-app advertisements, and local data storage.

App Top 5

The 5 apps with the highest MARS overall scores per platform are listed in Multimedia Appendix 2 together with their characteristics. On Android, Bloeddruk (developer: Klimaszewski Szymon) and Beurer HealthManager (developer: Beurer) were the best-scoring apps, with MARS overall scores of 4.1 and 3.7, respectively. AMICOMED BP (developer: AMICOMED) was the best-scoring app on the iOS platform with an overall score of 3.6. All apps in the top 5 for each platform were free.

Discussion

Principal Findings

In this study, we observed a lower MARS overall score compared to other reviews focusing on apps for other self-management aspects [33-35]. However, comparable to our study, functionality was previously reported as the objective category with the highest MARS score [33,35]. Our results also showed that some of the app features have a large influence on the overall quality score. The app features with the most positive influence on the app quality score are the ability of using the cloud for data storage, wireless data upload from blood pressure meters, ability to export data, ability to analyze data, ability to send reminders, and ability to record personal data, such as age and weight. More than half of the apps can export data and approximately 15% of the apps were able to upload data from blood pressure meters. If present, the latter feature makes it easier and more convenient to measure and record blood pressure.

Only approximately a quarter of the apps in our study had a reminder function, but reminder features can be very important in facilitating adherence [36,37]. The authors of a previous study on hypertension apps reported similar results [38]. All of the selected key features except information/education resulted in a large positive influence on the app overall quality score. The information/education feature was often absent or of poor quality in the apps included in this study. Only 2 apps were developed by a university or nongovernmental organizations and none of the apps was evaluated with results published in the literature. This, combined with the low scores for information on the MARS scale, suggests the lack of involvement from medical experts in the process of app development, which was also reported in previous studies [39-42]. The high MARS scores for functionality combined with the low MARS scores for information suggests that most apps function well but lack important information. This lack of information may result into incorrect use of the app (eg, incorrect interpretation of blood pressure readings, resulting in potential nonadherence to therapy) by users who are not sufficiently literate in digital health. Therefore, health apps should be validated before use. We found suboptimal ICCs and Cronbach alphas for the functionality subscale in Android and the aesthetics subscale in iOS. An explanation for the suboptimal ICCs and Cronbach alphas may be a lower agreement between assessors for the functionality and aesthetics subscales due to discrepancies on subjective criteria between the assessors, which arose after the training used to standardize the assessors' ratings. Another explanation may be that the MARS is not a perfect instrument to assess these features. The developers of MARS also reported a lower ICC for the functionality subscale [29].

There were no differences in quality scores between paid and free apps, which has also been reported previously [35]. Notably, the top 5 apps, identified by the highest MARS overall scores, were all free of charge. However, developers may earn money with their apps by selling the data shared by users and/or by promoting other products that can be combined with their apps, such as blood pressure meters. PatientsLikeMe is a well-known example. This platform is free of charge and very useful, but user data is sold [43]. That may explain why these apps are offered free of charge in the app stores by the app developers. We did not observe a correlation between the star ratings and the MARS overall scores. However, it is difficult to assess reliability of star ratings in app stores, since the criteria and qualifications of assessors are not always clear. For example, reviewers may leave ratings that do not reflect their true opinions or only selected users leave a rating (selection bias). The authors of a recent study on app store user ratings and reviews of a blood pressure app (Instant Blood Pressure) reported that these types of ratings were unreliable [44].

In this study, we identified a large number of apps ineligible for self-management and many apps of poor quality. These apps may potentially be harmful to users. Apps providing patients with erroneous information or apps that do not do what they are supposed to do are examples of such harmful apps. At the American Medical Association interim meeting in 2016, Executive Vice President James Madara mentioned a blood pressure app that failed at high rates in detecting elevated blood pressure and yet was one of the most frequently downloaded health apps for 2 years [45]. It is important to separate good apps from the harmful ones and to stimulate the development of high quality apps. Performing systematic app reviews and/or developing guidelines for health app developers could reduce the development of poor-quality health apps. It is important to regulate the development of health apps internationally, because apps are available in multiple national app stores. Therefore, setting up an international institute to regulate the development of health apps or certifying health apps may be necessary. The Health On the Net Foundation (HON) is a good example of such an institute. HON assesses the quality of health information online and provides certification to websites with reliable health information [46].

Strengths and Limitations

A major strength of this study is that we searched the 2 main app stores systematically using 15 search terms in English and Dutch and included both paid and free apps. This resulted in a

large number of apps that were first screened on titles and descriptions in the app stores. All apps identified through this process were assessed by 2 independent reviewers. In addition, we assessed the quality of the apps objectively using MARS [29], which has previously been used to evaluate app quality in several app reviews [35,47,48].

This review was limited to Dutch app stores, and we included apps in English or Dutch only. It is possible that other national app stores may contain a larger, smaller, or different assortment of apps. Although it is not feasible to search all national app stores from a single country, most apps are released worldwide and are not country-specific. Furthermore, we limited our search to the major app platforms Android and iOS. These platforms, however, accounted for approximately 98% of the mobile phone market share in 2015 [49]. In addition, we excluded apps that need a prescription by a health care provider or permission for use from the developer. Therefore, we may have missed potentially eligible apps, but these are not generally available to the target population. Another limitation was the compatibility of apps. As a large variety of mobile phones with several software versions are available in the markets, some apps may not have been compatible with the devices used in this study. However, it is not feasible to assess all apps using a large spectrum of mobile phones. We used the most recent software versions to ensure the maximum stability and safety.

Perspectives

Mobile apps may be a useful tool for self-management for patients with hypertension. In addition, mobile apps could be used to provide information to patients and increase awareness about blood pressure–related health issues among patients. Also, many mobile apps can export blood pressure data, which could be used by health care providers to make more informed decisions regarding treatment [50,51]. Furthermore, patients will be more involved in their own treatment through the use of high-quality, dedicated mobile apps. Therefore, health care providers should stimulate the use of mobile apps by patients with hypertension. In that case, however, they have to be sure that the apps used by their patients do not contain any misleading or harmful information. App reviews could be a suitable instrument to separate the useful apps from the harmful ones. Nevertheless, a practical guideline for app reviews is not available. Therefore, it is crucial to develop an international guideline for performing app reviews. This study provides a list of the top 5 useable apps targeting blood pressure monitoring available on the 2 major mobile phone platforms. Health care providers and patients with hypertension can use the results presented in this study to identify a suitable high-quality app targeting blood pressure monitoring, provided that blood pressure measurements are valid.

Conclusion

In this review, we identified only a few apps with sufficient quality for blood pressure self-management purposes. The use of these sufficient quality apps should be stimulated to improve patient care. This study provides an overview of the best apps currently available in the app stores and important key features for self-management that can be used by health care providers and patients with hypertension to identify a suitable app targeting blood pressure monitoring. However, the majority of the apps targeting blood pressure monitoring were of poor quality, and the accuracy of the blood pressure measurements registered in the apps was not assessed. It is important to involve medical experts in the developmental stage of health-related mobile apps to improve the quality of these apps.

Acknowledgments

This study was funded by the Radboud Institute for Health Sciences, Radboud University Medical Center.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The complete list of the reviewed apps and Mobile App Rating Scale overall scores.

[PDF File (Adobe PDF File), 372KB - mhealth_v6i11e187_app1.pdf]

Multimedia Appendix 2

Top five hypertension apps on the iOS and Android platforms.

[PDF File (Adobe PDF File), 277KB - mhealth_v6i11e187_app2.pdf]

References

1. Johansson BB. Hypertension mechanisms causing stroke. Clin Exp Pharmacol Physiol 1999 Jul;26(7):563-565. [Medline: 10405790]
2. World Health Organization. A global brief on hypertension. Silent killer, global public health crisis. 2013. URL: http://apps.who.int/iris/bitstream/handle/10665/79059/WHO_DCO_WHD_2013.2_eng.pdf?sequence=1&isAllowed=y [accessed 2018-09-25]

3. World Health Organization. Raised blood pressure (SBP ≥140 OR DBP ≥90), crude (%) data by WHO region. 2017. URL: http://apps.who.int/gho/data/view.main.NCDBPCREGv?lang=en [accessed 2018-01-12] [WebCite Cache ID 6wPpWYMJ2]
4. World Health Organization. Raised blood pressure: situation and trends. 2017. URL: http://www.who.int/gho/ncd/risk_factors/blood_pressure_prevalence_text/en/ [accessed 2018-09-25] [WebCite Cache ID 6wPpvm9WV]
5. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, et al. Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet 1990 Apr 07;335(8693):827-838. [Medline: 1969567]
6. Seshadri S, Wolf PA, Beiser A, Vasan RS, Wilson PW, Kase CS, et al. Elevated midlife blood pressure increases stroke risk in elderly persons: the Framingham Study. Arch Intern Med 2001 Oct 22;161(19):2343-2350. [Medline: 11606150]
7. Coresh J, Wei GL, McQuillan G, Brancati FL, Levey AS, Jones C, et al. Prevalence of high blood pressure and elevated serum creatinine level in the United States: findings from the third National Health and Nutrition Examination Survey (1988-1994). Arch Intern Med 2001 May 14;161(9):1207-1216. [Medline: 11343443]
8. Hsu C, McCulloch CE, Darbinian J, Go AS, Iribarren C. Elevated blood pressure and risk of end-stage renal disease in subjects without baseline kidney disease. Arch Intern Med 2005 Apr 25;165(8):923-928. [doi: 10.1001/archinte.165.8.923] [Medline: 15851645]
9. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). JAMA 2014 Feb 5;311(5):507-520. [doi: 10.1001/jama.2013.284427] [Medline: 24352797]
10. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston MN, Hubbard VS, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Jun 24;129(25 Suppl 2):S76-S99 [FREE Full text] [doi: 10.1161/01.cir.0000437740.48606.d1] [Medline: 24222015]
11. Cappuccio FP, Kerry SM, Forbes L, Donald A. Blood pressure control by home monitoring: meta-analysis of randomised trials. BMJ 2004 Jul 17;329(7458):145 [FREE Full text] [doi: 10.1136/bmj.38121.684410.AE] [Medline: 15194600]
12. McGrath BP. Home monitoring of blood pressure. Aust Prescr 2015 Feb;38(1):16-19 [FREE Full text] [Medline: 26648605]
13. Fletcher BR, Hartmann-Boyce J, Hinton L, McManus RJ. The effect of self-monitoring of blood pressure on medication adherence and lifestyle factors: a systematic review and meta-analysis. Am J Hypertens 2015 Oct;28(10):1209-1221. [doi: 10.1093/ajh/hpv008] [Medline: 25725092]
14. Imai Y, Ohkubo T, Hozawa A, Tsuji I, Matsubara M, Araki T, et al. Usefulness of home blood pressure measurements in assessing the effect of treatment in a single-blind placebo-controlled open trial. J Hypertens 2001 Feb;19(2):179-185. [Medline: 11212959]
15. eMarketer. 2014. 2 billion consumers worldwide to get smart(phones) by 2016 URL: https://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694 [accessed 2018-01-12] [WebCite Cache ID 6wPrlPJOD]
16. European Commission. 2014 Apr 10. Healthcare in your pocket: unlocking the potential of mHealth URL: http://europa.eu/rapid/press-release_IP-14-394_en.pdf [accessed 2018-09-25] [WebCite Cache ID 72hQqe2dt]
17. Research2guidance. 2016. mHealth App Developer Economics: the current status and trends of the mHealth app market URL: https://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2016.pdf [accessed 2018-01-18] [WebCite Cache ID 6wYpdDMon]
18. Accenture. 2016. Patients want a heavy dose of digital URL: https://www.accenture.com/t20160629T045303__w__/us-en/_acnmedia/PDF-6/Accenture-Patients-Want-A-Heavy-Dose-of-Digital-Infographic.pdf [accessed 2018-01-18] [WebCite Cache ID 6wYq2a0wQ]
19. Weber MA, Schiffrin EL, White WB, Mann S, Lindholm LH, Kenerson JG, et al. Clinical practice guidelines for the management of hypertension in the community: a statement by the American Society of Hypertension and the International Society of Hypertension. J Clin Hypertens (Greenwich) 2014 Jan;16(1):14-26 [FREE Full text] [doi: 10.1111/jch.12237] [Medline: 24341872]
20. Mancia G, Fagard R, Narkiewicz K, Redón J, Zanchetti A, Böhm M, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). J Hypertens 2013 Jul;31(7):1281-1357. [doi: 10.1097/01.hjh.0000431740.32696.cc] [Medline: 23817082]
21. Gabb G, Mangoni A, Anderson C, Cowley D, Dowden J, Golledge J, et al. Guideline for the diagnosis and management of hypertension in adults-2016. Med J Aust 2016 Jul 18;205(2):85-89. [Medline: 27456450]
22. Centers for Disease Control and Prevention. 2013. Self-measured blood pressure monitoring: action steps for public health practitioners URL: https://millionhearts.hhs.gov/files/MH_SMBP.pdf [accessed 2018-01-12] [WebCite Cache ID 6wPtlm4Bs]
23. Centers for Disease Control and Prevention. 2014. Self-measured blood pressure monitoring: actions steps for clinicians URL: https://millionhearts.hhs.gov/files/MH_SMBP_Clinicians.pdf [accessed 2018-09-25] [WebCite Cache ID 6wPuQM9o0]
24. The National Institute for Health and Care Excellence. 2011. Hypertension in adults: diagnosis and management URL: https://www.nice.org.uk/guidance/cg127/resources/hypertension-in-adults-diagnosis-and-management-pdf-35109454941637 [accessed 2018-09-25] [WebCite Cache ID 6wPuoZRsZ]

25. Pickering TG, Miller NH, Ogedegbe G, Krakoff LR, Artinian NT, Goff D, American Heart Association, American Society of Hypertension, Preventive Cardiovascular Nurses Association. Call to action on use and reimbursement for home blood pressure monitoring: executive summary: a joint scientific statement from the American Heart Association, American Society Of Hypertension, and Preventive Cardiovascular Nurses Association. Hypertension 2008 Jul;52(1):1-9 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.107.189011] [Medline: 18497371]
26. Bajorek B, Lemay K, Magin P, Roberts C, Krass I, Armour C. Patients' attitudes and approaches to the self-management of hypertension: perspectives from an Australian qualitative study in community pharmacy. High Blood Press Cardiovasc Prev 2017 Jun;24(2):149-155. [doi: 10.1007/s40292-017-0181-8] [Medline: 28168358]
27. Bolbrinker J, Zaidi TL, Gohlke H, Weisser B, Kreutz R. European guidelines on lifestyle changes for management of hypertension: awareness and implementation of recommendations among German and European physicians. Herz 2017 May 22. [doi: 10.1007/s00059-017-4575-0] [Medline: 28534177]
28. Kang H, Park HA. Development of hypertension management mobile application based on Clinical Practice Guidelines. Stud Health Technol Inform 2015;210:602-606. [Medline: 25991219]
29. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
30. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
31. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]
32. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
33. Wilson H, Stoyanov SR, Gandabhai S, Baldwin A. The quality and accuracy of mobile apps to prevent driving after drinking alcohol. JMIR Mhealth Uhealth 2016 Aug 08;4(3):e98 [FREE Full text] [doi: 10.2196/mhealth.5961] [Medline: 27502956]
34. Hartin PJ, Cleland I, Nugent CD, McClean SI, Tschanz J, Clark C, et al. Assessing app quality through expert peer review: a case study from the gray matters study. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:4379-4382. [doi: 10.1109/EMBC.2016.7591697] [Medline: 28269248]
35. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
36. Huang H, Li YJ, Chou Y, Hsieh Y, Kuo F, Tsai W, et al. Effects of and satisfaction with short message service reminders for patient medication adherence: a randomized controlled study. BMC Med Inform Decis Mak 2013 Nov 16;13:127 [FREE Full text] [doi: 10.1186/1472-6947-13-127] [Medline: 24238397]
37. Fenerty SD, West C, Davis SA, Kaplan SG, Feldman SR. The effect of reminder systems on patients' adherence to treatment. Patient Prefer Adherence 2012;6:127-135 [FREE Full text] [doi: 10.2147/PPA.S26314] [Medline: 22379363]
38. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015 Feb;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
39. Subhi Y, Bube SH, Rolskov BS, Skou TAS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
40. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):210. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
41. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile apps for weight management: a scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
42. Pereira-Azevedo N, Carrasquinho E, Cardoso DOE, Cavadas V, Osório L, Fraga A, et al. mHealth in Urology: a review of experts' involvement in app development. PLoS One 2015;10(5):e0125547 [FREE Full text] [doi: 10.1371/journal.pone.0125547] [Medline: 25984916]
43. PatientsLikeMe privacy policy. URL: https://www.patientslikeme.com/about/privacy [accessed 2018-01-12] [WebCite Cache ID 6wPzBXZJQ]
44. Plante TB, O'Kelly AC, Macfarlane ZT, Urrea B, Appel LJ, Miller IER, et al. Trends in user ratings and reviews of a popular yet inaccurate blood pressure-measuring smartphone app. J Am Med Inform Assoc 2018 Aug 01;25(8):1074-1079. [doi: 10.1093/jamia/ocy060] [Medline: 29878236]
45. Madara J. AMA Interim Meeting. 2016. URL: https://www.youtube.com/watch?v=Jpd1FqmRhLk [accessed 2018-01-18] [WebCite Cache ID 6wYwlyEUU]
46. Health On the Net Foundation. 2017. URL: http://www.hon.ch/Global/index.html [accessed 2018-01-18] [WebCite Cache ID 6wYxYpDU1]
47. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]

48. Masterson CRM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
49. Gartner says worldwide smartphone sales grew 9.7 percent in fourth quarter of 2015. 2016. URL: https://www.gartner.com/newsroom/id/3215217 [accessed 2018-01-18] [WebCite Cache ID 6wZ4MDSXX]
50. Staessen JA, Den Hond E, Celis H, Fagard R, Keary L, Vandenhoven G, Treatment of Hypertension Based on Home or Office Blood Pressure (THOP) Trial Investigators. Antihypertensive treatment based on blood pressure measurement at home or in the physician's office: a randomized controlled trial. JAMA 2004 Feb 25;291(8):955-964. [doi: 10.1001/jama.291.8.955] [Medline: 14982911]
51. Pickering TG, Miller NH, Ogedegbe G, Krakoff LR, Artinian NT, Goff D, American Heart Association, American Society of Hypertension, Preventive Cardiovascular Nurses Association. Call to action on use and reimbursement for home blood pressure monitoring: a joint scientific statement from the American Heart Association, American Society of Hypertension, and Preventive Cardiovascular Nurses Association. J Cardiovasc Nurs 2008;23(4):299-323. [doi: 10.1097/01.JCN.0000317429.98844.04] [Medline: 18596492]

Abbreviations

HON: Health on the Net
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale

Edited by G Eysenbach; submitted 19.01.18; peer-reviewed by R Masterson Creber, M Bardus, J Goris; comments to author 01.03.18; revised version received 26.04.18; accepted 06.07.18; published 14.11.18.

Please cite as:
Jamaladin H, van de Belt TH, Luijpers LCH, de Graaff FR, Bredie SJH, Roeleveld N, van Gelder MMHJ
Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2018;6(11):e187
URL: https://mhealth.jmir.org/2018/11/e187/
doi:10.2196/mhealth.9888
PMID:30429116

Original Paper

Key Elements of mHealth Interventions to Successfully Increase Physical Activity: Meta-Regression

Lisa V Eckerstorfer[1], MSc; Norbert K Tanzer[1], PhD; Claudia Vogrincic-Haselbacher[1], PhD; Gayannee Kedia[1], PhD; Hilmar Brohmer[1], MSc; Isabelle Dinslaken[1], MSc; Katja Corcoran[1,2], PhD

[1]Institute of Psychology, University of Graz, Graz, Austria

[2]BioTechMed-Graz, Graz, Austria

Corresponding Author:
Lisa V Eckerstorfer, MSc
Institute of Psychology
University of Graz
Universitaetsplatz 1
Graz, 8010
Austria
Phone: 43 316 380 ext 8524
Email: lisa.eckerstorfer@uni-graz.at

Abstract

Background: Mobile technology gives researchers unimagined opportunities to design new interventions to increase physical activity. Unfortunately, it is still unclear which elements are useful to initiate and maintain behavior change.

Objective: In this meta-analysis, we investigated randomized controlled trials of physical activity interventions that were delivered via mobile phone. We analyzed which elements contributed to intervention success.

Methods: After searching four databases and science networks for eligible studies, we entered 50 studies with N=5997 participants into a random-effects meta-analysis, controlling for baseline group differences. We also calculated meta-regressions with the most frequently used behavior change techniques (behavioral goals, general information, self-monitoring, information on where and when, and instructions on how to) as moderators.

Results: We found a small overall effect of the Hedges g=0.29, (95% CI 0.20 to 0.37) which reduced to g=0.22 after correcting for publication bias. In the moderator analyses, behavioral goals and self-monitoring each led to more intervention success. Interventions that used neither behavioral goals nor self-monitoring had a negligible effect of g=0.01, whereas utilizing either technique increased effectiveness by Δg=0.31, but combining them did not provide additional benefits (Δg=0.36).

Conclusions: Overall, mHealth interventions to increase physical activity have a small to moderate effect. However, including behavioral goals or self-monitoring can lead to greater intervention success. More research is needed to look at more behavior change techniques and their interactions. Reporting interventions in trial registrations and articles need to be structured and thorough to gain accurate insights. This can be achieved by basing the design or reporting of interventions on taxonomies of behavior change.

(JMIR Mhealth Uhealth 2018;6(11):e10076) doi:10.2196/10076

KEYWORDS

exercise; physical activity; mHealth; behavior change; meta-analysis; meta-regression

Introduction

People spend much time with their mobile phones (on average 2 hours and 27 minutes per day using apps and the Web [1]) and too little time being physically active [2,3], even though physical activity is beneficial for both body [4-6] and mind [6,7]. Recently, researchers have been looking for ways to use mobile phones to increase physical activity. These mobile health (mHealth) interventions have only been possible in recent years with the upsurge of mobile communication devices [8-11].

For researchers who strive to change health-related behavior, mHealth shows excellent promise for building life-changing interventions, but their efficacy is currently uncertain [12]. In the domain of physical activity, mHealth interventions yield small to moderate effects [13]. To date, researchers have not been able to pinpoint contextual elements that lead to greater

intervention success in increasing physical activity [13-19]. Indeed, a wide range of interventions are implemented that often overlap with each other. However, it is still unclear which basic principles and strategies lead to efficient changes. Tomlinson and colleagues [20] pointed out that there should be a joint effort among researchers to find and refine strategies for successful mHealth interventions. Hence, we know that mHealth interventions have great potential, but we need to establish how to maximize their efficacy. This meta-analysis study aims to identify content-related elements that predict intervention efficacy. The resulting knowledge is crucial for creating new interventions and guidelines.

One way to describe the content of interventions is to identify the behavior change techniques (BCTs) on which they rely. These are theory-based methods to change psychological determinants of behavior, such as agreeing on a behavioral contract or facilitating social comparison. Until now, it is unclear which BCTs contribute more to mHealth intervention success than others. This is mostly because until recently, only a limited number of studies have been available. Previous reviews and meta-analyses identified n=11 [15], n=19 [13], and n=18 [19] studies. These pools of studies only provide enough power to test two or three moderators in a meta-regression [21], so there was no feasible way to test the influence of each BCT—first, because there is a considerable number of BCTs and second, because some BCTs are seldom used or not used at all [13,22].

In the present meta-analysis, we decided to use a taxonomy by Michie and colleagues [23], which—in contrast to more general taxonomies for behavior change [24]—focuses on BCTs for diet and physical activity interventions. This taxonomy contains 40 BCTs, all of which we coded, but we only tested the influence of the 5 most frequently used ones in order to retain sufficient statistical power. In our sample, these are (1) behavioral goals, (2) general information, (3) self-monitoring, (4) information on where and when, and (5) instructions on how to. One change we wanted to make in comparison to previous work is to control for group differences at baseline. Even though we only used randomized controlled trials (RCTs), we expected to find a high number of feasibility and pilot tests with small sample sizes, in which not all the advantages of randomization can unfold properly. Our main research question is which of the 5 most frequently used BCTs have the potential to increase the efficacy of mobile phone-delivered physical activity interventions?

Methods

Searching for Studies

To find suitable studies, we searched Google scholar and 4 databases: PubMed, PsycINFO, ScienceDirect, and the ISI Web of Knowledge using search terms related to physical activity, mHealth and study design. An example for a search syntax is *([randomized controlled trial OR RCT OR randomised controlled trial OR clinical trial] AND [mobile phone OR smartphone OR mobile app OR mHealth] AND [exercise OR physical fitness OR physical activity])*. We did not restrict year of publication, but we did restrict language to English or German. Furthermore, we searched reference lists of published reviews on the topic [14,15,19] and posted invitations to research communities in health psychology, sports psychology, social psychology, and sports science to contribute relevant studies.

Selecting Studies

We did not restrict inclusion to specific populations. Instead, we accepted all studies that targeted physical activity—be it for healthy populations, sick populations, during pregnancy, or for children (Textbox 1). The general flow of study selection is presented in Figure 1.

We identified a total of 2067 studies. After removing duplicates, we screened the remaining 1817 records for eligibility according to their title and abstract. We assessed the full text of 205 studies and from these, 50 met the inclusion criteria [25-74]. When we screened titles and abstracts, we excluded poorly fitting records hierarchically. First, we reviewed content fit, and then we checked whether the research was original. Following that, we assessed whether there was an intervention and whether the intervention was delivered via mobile phone—and so on. In the second screening, we checked those points again—but more thoroughly—and we also assessed more complex questions, such as the suitability of the control group or whether there was enough information to compute effect sizes of group differences at baseline and the end of the intervention.

Textbox 1. Study inclusion and exclusion criteria.

Inclusion criteria

- The intervention was automatically delivered via mobile phone by either an app or texting
- The intervention targeted an increase in physical activity
- There was a control group, which was more passive than the intervention group and did not have personal communications with medical staff or researchers instead of receiving the mHealth intervention
- Allocation to experimental and control groups was randomized, though we accepted stratification (eg, by gender)
- At least one of the outcomes measured actual physical activity (via electronic trackers or self-reported) or assessed objective indicators of physical fitness (eg, peak oxygen intake)

Exclusion criteria

- Study designs were nonexperimental (eg, observational studies, reports and comments, case studies)
- The data necessary to calculate an effect size was not available

Figure 1. Flow diagram of the study selection process. RCT: randomized controlled trial; PA: physical activity; mHealth (mobile health).

Outcome Measures and Effect Size

In the present meta-analysis, we wanted to investigate how selected interventions influence physical activity; hence we focused on physical activity as outcomes. There are 2 main methods to measure physical activity: having people self-report their physical activity and electronically tracking people's movement. Self-reported and tracked physical activity generally correlate poorly or moderately with each other, but with a range from r=–.71 to .96 [75]. To check whether the way in which physical activity was measured in our studies impacts our main results, we investigated the influence of physical activity measurement on intervention success by including the type of outcome (tracked versus self-reported versus both) as a moderator in our design.

Since our pool of studies includes many pilot and feasibility tests, we decided to account for their small sample size by using the Hedges g to measure effect size. This is principally the same measurement as the Cohen d, but it uses a correction factor for small sample sizes [76]. The interpretation of the Hedges g follows the same rule of thumb that applies to the Cohen d:$|g|$=0.20 for a small effect, $|g|$=0.50 for a medium effect, and $|g|$=0.80 for a large effect.

Review Procedure and Moderators

The extracted studies were initially screened for eligibility and then articles were classified according to title and abstract. The categories are depicted in the list of "Records excluded" (Figure 1). The procedure was conducted hierarchically, first making sure the topic fitted, then checking the originality of the data (ie, if there was any data collected), implementation of an intervention, mHealth focus of that intervention, and so on. A consensus was reached through discussion. Information was then extracted from the articles and protocols into a Microsoft Excel spreadsheet. Of the 205 studies of which the full text was assessed for eligibility, 96 were coded twice to uphold the same rating standards. Again, a consensus was reached through discussion in several evaluation sessions.

To estimate study quality, we used the Effective Public Health Practice Project (EPHPP) Quality Assessment Tool [77]. The EPHPP tool has 6 sections, which are aggregated to a final grade on a 3-point scale. Low study quality is a known issue in electronic interventions [78,79], but we did not exclude those studies. Instead, we checked whether the study quality moderates intervention efficacy, because a correspondence can be a sign of publication bias.

Further, some of the studies we included did not (only) target physical activity, but health or weight, with an increase of physical activity being one of multiple intervention goals. We expected that mHealth interventions directly targeting physical activity rather than health or weight would be more efficient in increasing physical activity. Therefore, we assessed the studies' main objective as a moderator of intervention success.

Behavior Change Techniques

To identify the factors of mHealth intervention success, we used a taxonomy of 40 BCTs to code intervention contents [23]. We coded only BCTs that were employed in the intervention group and not in the control group. If a BCT was used in both groups, we counted it as absent from the intervention. Since we only expected to collect a sample of studies large enough to test 5 moderators (ie, around 25 to 50 studies), we decided to test the five most frequently used BCTs as moderators. The BCTs we tested were (1) behavioral goals, (2) general information, (3) self-monitoring, (4) information on where and when, and (5) instructions on how to.

Meta-Analysis and Meta-Regression

To assure quality in our meta-analysis, we consulted with experts, and followed the recommendations in Borenstein and colleagues [76], and consulted a quality assessment tool for meta-analyses [80]. All statistical analyses in the present study were conducted in R, version 3.4.2 [81] with the additional packages robumeta [82], metafor [83], MAd [84], tidyverse [85], readxl [86], compute.es [87], stats [81], ggplot2 [88], yarrr [89], and wesanderson [90]. We calculated the effect size (Hedges g) for the difference between the intervention and control group before the intervention (baseline) and at the end of the scheduled intervention (postintervention) from means, standard deviations and group size, P-values or proportions. To assess whether there were meaningful physical activity group differences before the interventions were administered, we first performed a random-effects meta-analysis with baseline group differences. Then, we correlated baseline group differences with postintervention group differences. Since there was more heterogeneity at baseline than expected by chance alone and baseline group differences were not independent of intervention success, we used baseline group differences as a covariate in further analyses, essentially setting physical activity group differences before the intervention to zero. We also removed outliers at baseline and repeated the overall effect size analyses with this reduced pool of studies.

For the meta-analysis with postintervention group differences and meta-regression we used robumeta, as recommended in a review of meta-analysis packages [91]. This package allows for dependent effect estimation (ie, combining multiple outcomes per study in a correlated or hierarchical fashion) [92]. We coded up to four outcomes per study and aggregated those with a correction for small sample sizes. Since we were not able to predict the correlation between outcomes, we performed sensitivity analyses. Varying the assumed correlation only led to negligible outcome changes in the second decimal place regarding the estimated coefficients and tau-squared (τ^2). For our main analyses, we first ran a random effects null-model to determine overall intervention success. Then, we checked—separately for each moderator—whether study quality, the intervention's main objective, physical activity tracking or any of the most frequently used BCTs were associated with greater intervention success. For each BCT, we compared interventions that used it with interventions that did not use it. In an exploratory fashion, we also investigated a combination of the two most efficient BCTs. We did not run any more moderator analyses to avoid exceeding the power of our pool of 50 studies.

Publication Bias

To evaluate a possible publication bias in the field, we created a funnel plot and performed the Egger asymmetry test [93] as

recommended in [76]. We also performed a trim and fill analysis [94] and the Orwin Failsafe-N test to assess whether the effect might be an artifact of bias [95]. Due to a large proportion of feasibility and pilot trials with small sample sizes in our pool of studies, we decided against restricting the analysis to large samples.

Results

Selected Studies

For our meta-analysis, we had a final pool of 50 studies with N=5997 people. The mean age of included samples was 40.6 (SD 16.7), and on average there were 62.7% women in each study (SD 29.2). We found that 40/50 (80%) studies were of good quality, 5 of the 50 (10%) were of moderate quality, and the remaining 10% (5/50) were of poor quality. Twenty-nine of the 50 (58%) studies targeted physical activity, 13/50 (26%) targeted health and 16% (8/50) targeted weight. Furthermore, 20 of the 50 (40%) studies used self-reported measures of physical activity, 17/50 (34%) tracked physical activity, and 13/50 (26%) studies used both methods. We coded a maximum of 4 outcomes per study (M=1.76), which led to 87 physical activity-related outcomes across the pool of 50 studies. Sample sizes and group differences at baseline as well as after the intervention are provided in Table 1. A table with more details regarding population, intervention, comparison, and outcomes can be found in Multimedia Appendix 1.

Not all BCTs were used equally often. Therefore, some BCTs are featured in more than half of the studies whereas others are not featured at all. Information for each BCT of the taxonomy is available in Multimedia Appendix 1. We decided to test the five most frequently used BCTs as moderators. As shown in Table 2, there are 18 to 30 studies out of 48 using each of these 5 BCTs. Table 2 also indicates the number of studies in each element in which physical activity was measured as self-reports, electronically tracked, or if both methods were used simultaneously. We ran a chi-square test to determine whether using a certain BCT was confounded by how physical activity was measured. The test did not reveal a significant association between using BCTs and the way in which physical activity was measured (χ^2_8=13.6, P=.09). Overall—in 48/50 (96%) studies—235 BCTs were used, which led to a mean of 5 BCTs per study.

Baseline Group Differences

A random-effects meta-analysis with baseline group differences revealed that the estimated effect g=0.05 was not significantly different from zero (95% CI –0.03 to 0.13, P=.22), but there was considerable heterogeneity between the studies (Q(49)=93.79, P<.001) and 43.33% of this heterogeneity was due to true effect sizes rather than sampling variance (I^2=43.33%). Overall, the variance of true effects was τ^2=0.03 on a scale of the Hedges g.

Correlating group differences at baseline with group differences postintervention showed a positive association of r=.44 (P<.001, see Figure 2). This means that when participants who received the intervention were more physically active at baseline than people in the control group, this advantage was also conveyed to group differences after the intervention. To consider this phenomenon, we took baseline differences into account in 2 separate ways. Firstly, we used baseline group differences as a covariate—setting group differences to zero. Secondly, we removed outcomes with group differences of $|g|$>0.50 at baseline and reran overall effect size analyses with this reduced pool of studies. In the reduced pool, there were 44 studies with 76 outcomes. We present a scatterplot of outcomes at baseline by outcomes postintervention (see Figure 2).

Point Estimate

We assumed a correlation of r=.80 between outcomes of the same study. We did not have the means to check whether this assumption was valid, so we also performed sensitivity analyses—varying the assumed correlation between 0 and 1—and we did not find meaningful changes in the results. A random effects model for group differences postintervention revealed an estimated effect size of g=0.33 (95% CI 0.22 to 0.44, P<.001). We show a forest plot of group differences after the intervention is presented (see Figure 3). However, including baseline differences as a covariate in the same analysis reduced the estimated effect size to g=0.29 (95% CI 0.20 to 0.37, P<.001) with a covariate slope estimate of g=0.63 (95% CI 0.11 to 1.14, P=.02), meaning that increasing baseline group differences by g=1.00 would lead to a greater intervention success of g=0.63. True means had a variance of τ^2=0.06 on a scale of the Hedges g, and 61.15% of the observed heterogeneity was due to true differences between studies rather than chance, so we went ahead with the planned moderator analyses. Estimating the overall effect size with a reduced pool of outcomes instead of baseline group differences as covariate painted much the same picture (estimated effect size g=0.28, 95% CI 0.19 to 0.38, P<.001, I^2=56.38%, τ^2=0.05).

Moderator Analyses

All moderator analyses were conducted using baseline group differences as a covariate. They did not reveal a significant influence of study quality (Δg=0.03, 95% CI –0.15 to 0.22, P=.70). In terms of an intervention's main objective, targeting weight rather than physical activity reduced the success of increasing physical activity, but there were no differences between health and physical activity or health and weight as the main objective ($\Delta g_{PA\ versus\ weight}$=–0.21, 95% CI –0.42 to –0.003, P=.047; $\Delta g_{health\ versus\ PA}$=0.13, 95% CI –0.09 to 0.36, P=.23; $\Delta g_{health\ vs\ weight}$=–0.08, 95% CI –0.29 to 0.14, P=.45). There was no difference between interventions in which people self-reported their physical activity and interventions whereby physical activity was tracked, and those interventions in which both methods were used simultaneously ($\Delta g_{self\text{-}reported\ versus\ tracked}$=0.01, 95% CI –0.21 to 0.24, P=.90; $\Delta g_{both\ versus\ self\text{-}reported}$=–0.11, 95% CI –0.34 to 0.11, P=.31; $\Delta g_{both\ versus\ tracked}$=–0.12, 95% CI –0.36 to 0.11, P=.27).

Table 1. Overview of included studies showing the Hedges *g* effect size of group differences.

Study	N	Baseline, *g*[a]	Postintervention, *g* (95% CI)
Abraham et al 2015 [25]	32	0.41	0.00 (–0.68 to 0.68)
Adams et al 2013 [26]	20	–0.77	0.44 (–0.41 to 1.29)
Allen et al 2013 [27]	23	–0.02	–0.11 (–0.89 to 0.67)
Allman-Farinelli et al 2016 [28]	248	–0.12	0.28 (0.03 to 0.52)
Cadmus-Bertram et al 2015 [29]	51	–0.10	0.13 (–0.31 to 0.57)
Choi et al 2016 [30]	29	0.11	0.58 (–0.15 to 1.31)
Chow et al 2015 [31]	710	–0.13	0.24 (0.04 to 0.44)
Cotten and Prapavessis 2016 [32]	56	0.02	0.34 (–0.11 to 0.78)
Cowdery et al 2015 [33]	39	–0.34	0.17 (–0.45 to 0.79)
Direito et al 2015 [34]	34	–0.54	–0.14 (–0.72 to 0.43)
Eckerstorfer et al (unpublished data)	95	–0.16	0.02 (–0.37 to 0.41)
Fassnacht et al 2015 [35]	45	–0.31	0.00 (–0.59 to 0.59)
Fjeldsoe et al 2016 [36]	216	–0.09	0.09 (0.56 to 1.25)
Fjeldsoe et al 2015 [73]	266	0.01	0.38 (0.09 to 0.68)
Fjeldsoe et al 2010 [37]	88	0.28	0.91 (–0.13 to 0.30)
Frederix et al 2015 [38]	139	–0.17	0.44 (0.14 to 0.73)
Fukuoka et al 2015 [39]	61	0.06	0.57 (0.17 to 0.97)
Garde et al 2015 [40]	47	–0.21	–0.07 (–0.55 to 0.42)
Gell and Wadsworth 2015 [41]	87	0.01	0.30 (–0.14 to 0.74)
Glynn et al 2014 [42]	66	–0.23	0.25 (–0.23 to 0.73)
Hales et al 2016 [43]	43	–0.07	0.00 (–0.59 to 0.59)
Hartman et al 2016 [44]	50	0.39	0.43 (–0.08 to 0.93)
Hebden et al 2014 [45]	51	0.05	0.01 (–0.47 to 0.49)
Hurling et al 2007 [46]	77	0.17	0.36 (–0.08 to 0.80)
Johnston et al 2016 [47]	151	0.07	0.07 (–0.23 to 0.36)
Joseph et al 2015 [48]	28	–0.04	0.17 (–0.41 to 0.75)
Kim and Glanz 2013 [49]	41	0.49	1.14 (0.55 to 1.72)
Kim et al 2015 [50]	196	–0.10	0.14 (–0.14 to 0.42)
Kim et al 2016 [51]	95	0.08	–0.06 (–0.45 to 0.33)
Kinnafick et al 2016 [52]	65	–0.12	–0.13 (–0.55 to 0.29)
Laing et al 2014 [53]	211	–0.12	0.23 (–0.05 to 0.51)
Lubans et al 2016 [54]	157	0.30	0.14 (–0.16 to 0.43)
Maddison et al 2015 [55]	143	0.05	0.25 (–0.02 to 0.52)
Maher et al 2015 [56]	98	0.05	0.54 (0.20 to 0.87)
Martin et al 2015 [57]	32	0.00	1.35 (0.73 to 1.97)
Nguyen et al 2013 [58]	84	1.30	1.88 (1.36 to 2.40)
Pfaeffli et al 2015 [59]	123	0.63	0.55 (0.07 to 1.03)
Poirier et al 2016 [60]	217	–0.15	0.32 (0.04 to 0.60)
Prestwich et al 2010 [61]	94	–0.14	0.36 (0.02 to 0.69)
Rubinstein et al 2016 [62]	553	0.03	0.05 (–0.15 to 0.25)
Schwerdtfeger et al 2012 [63]	43	–0.04	0.56 (–0.03 to 1.15)
Silveira et al 2013 [64]	31	1.32	1.36 (0.70 to 2.02)

Study	N	Baseline, g[a]	Postintervention, g (95% CI)
Suggs et al 2013 [65]	158	0.17	1.18 (0.84 to 1.52)
Tabak et al 2014 [66]	29	0.54	1.05 (0.29 to 1.81)
van der Weegen et al 2015 [67]	117	–0.25	0.43 (0.09 to 0.77)
van Drongelen et al 2014 [68]	390	0.12	0.19 (0.02 to 0.35)
Vorrink et al 2016 [69]	157	–0.08	–0.08 (–0.42 to 0.26)
Walsh et al 2016 [70]	55	0.23	0.29 (–0.23 to 0.81)
Wang et al 2016 [71]	59	0.30	–0.22 (–0.64 to 0.20)
Zach et al 2016 [72]	100	0.49	0.30 (–0.09 to 0.69)

[a]Hedges g refers to the effect size of group differences, where larger values indicate more physical activity of the intervention group compared to the control group.

Table 2. The number of studies using each tested behavior change technique overall and split for the way in which physical activity was measured.

Behavior change technique	N	Self-reported, n, (%)	Tracked, n (%)	Both, n (%)
Behavioral goals	30	10 (33)	8 (27)	12 (40)
Self-monitoring	26	5 (19)	8 (31)	13 (50)
General information	24	9 (38)	11 (46)	4 (17)
Information on where and when	19	6 (32)	11 (58)	2 (11)
Instructions on how to	18	4 (22)	8 (44)	6 (33)

Figure 2. Scatterplot of group differences before and after the intervention for each outcome separately. Point size indicates the number of participants in each study. ES: effect size.

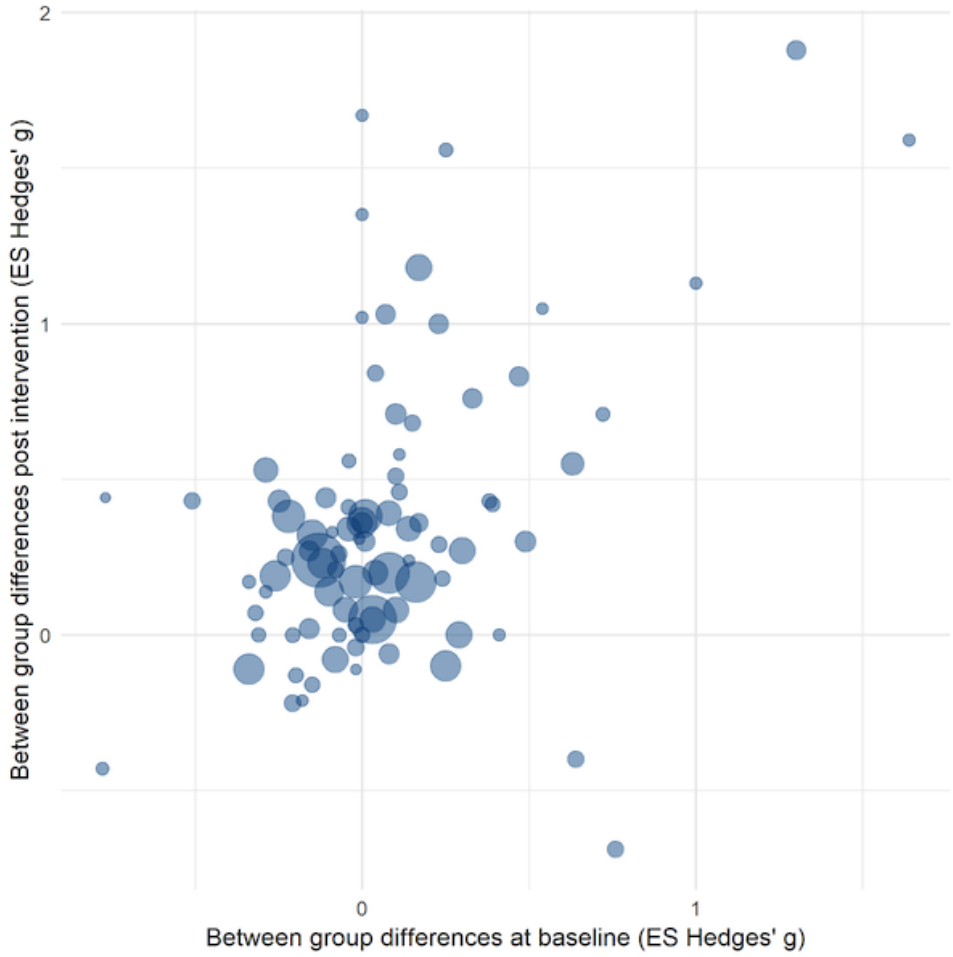
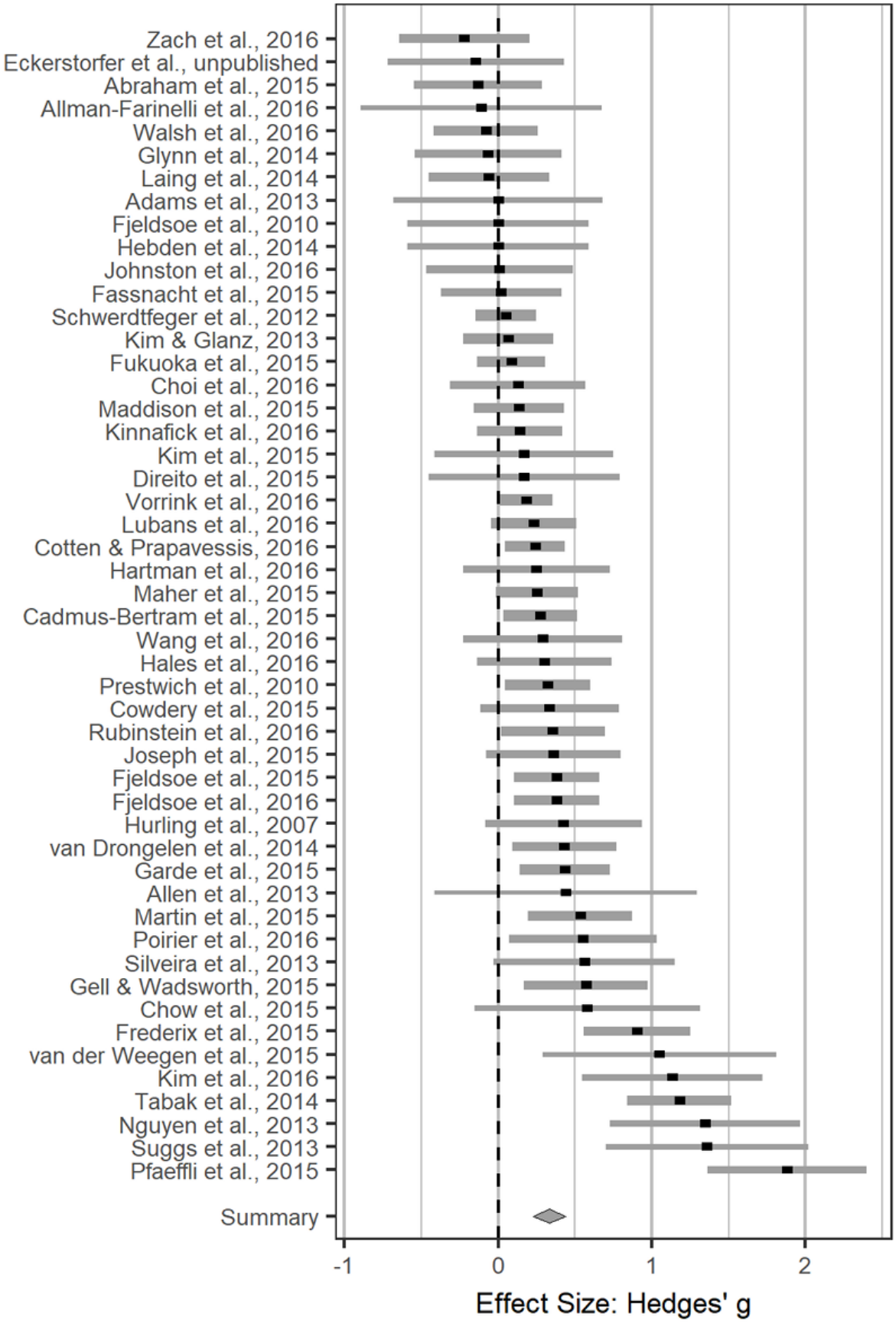


Figure 3. Forest plot for physical activity postintervention. The larger the values, the more active the intervention group was compared to the control group. Horizontal lines depict 95% CI and line thickness indicates each number's impact on the summary effect.



Also, we analyzed the effect of the five most frequently used BCTs. Two studies could not be coded for BCTs because they used very diverse interventions and from the text, it was not clear which participants received which intervention components. Thus, we conducted the analyses with a pool of 48 studies, using 84 outcomes. The moderator analyses are visually presented (see Figure 4). Using the BCTs behavioral goals and self-monitoring led to greater intervention success, but the other 3 tested BCTs (ie, general information, information on where and when, and instructions on how to) were not associated with greater intervention success (behavioral goals: Δg=0.20, 95% CI 0.02 to 0.38, P=.03; general information: Δg=–0.16, 95% CI –0.33 to 0.01, P=.07; self-monitoring: Δg=0.17 95% CI 0.01 to 0.34, P=.04; information on where and when: Δg=0.11, 95% CI –0.07 to 0.30, P=.22; instructions on how to: Δg=0.01, 95% CI –0.17 to 0.19, P=.88). We decided to explore the BCTs behavioral goals and self-monitoring further by combining them in a single moderator model and checking for additive effects. This analysis revealed that if neither behavioral goals nor self-monitoring are used this is associated with a lower intervention efficacy compared with the use of either or both BCTs ($\Delta g_{none\ versus\ either}$=0.31, 95% CI 0.13 to 0.49, P=.002; $\Delta g_{none\ versus\ both}$=0.36, 95% CI 0.15 to 0.56, P=.002; $\Delta g_{either\ versus\ both}$=0.05, 95% CI –0.19 to 0.29, P=.67). More precisely, interventions using neither goals nor rewards have an estimated effect of g=0.01, while interventions in which goals and rewards are used simultaneously have an estimated effect of g=0.36.

Figure 4. Differences in intervention efficacy depending on the use of the 5 most common behavior change techniques (BCTs). In each panel, “no” means that the BCT was not used and “yes” means that the BCT was used. The “n” next to “yes” and “no” indicates the number of studies in each group. The black line shows the mean intervention efficacy with a 95% CI in white. The curved areas depict density of data points (ie, fine-grained vertical histograms) for all included studies. The last panel shows intervention efficacy depending on a combination of behavioral goals and self-monitoring (n(none)=9, n(either)=22, n(both)=17). These depictions are not controlled for baseline group differences.

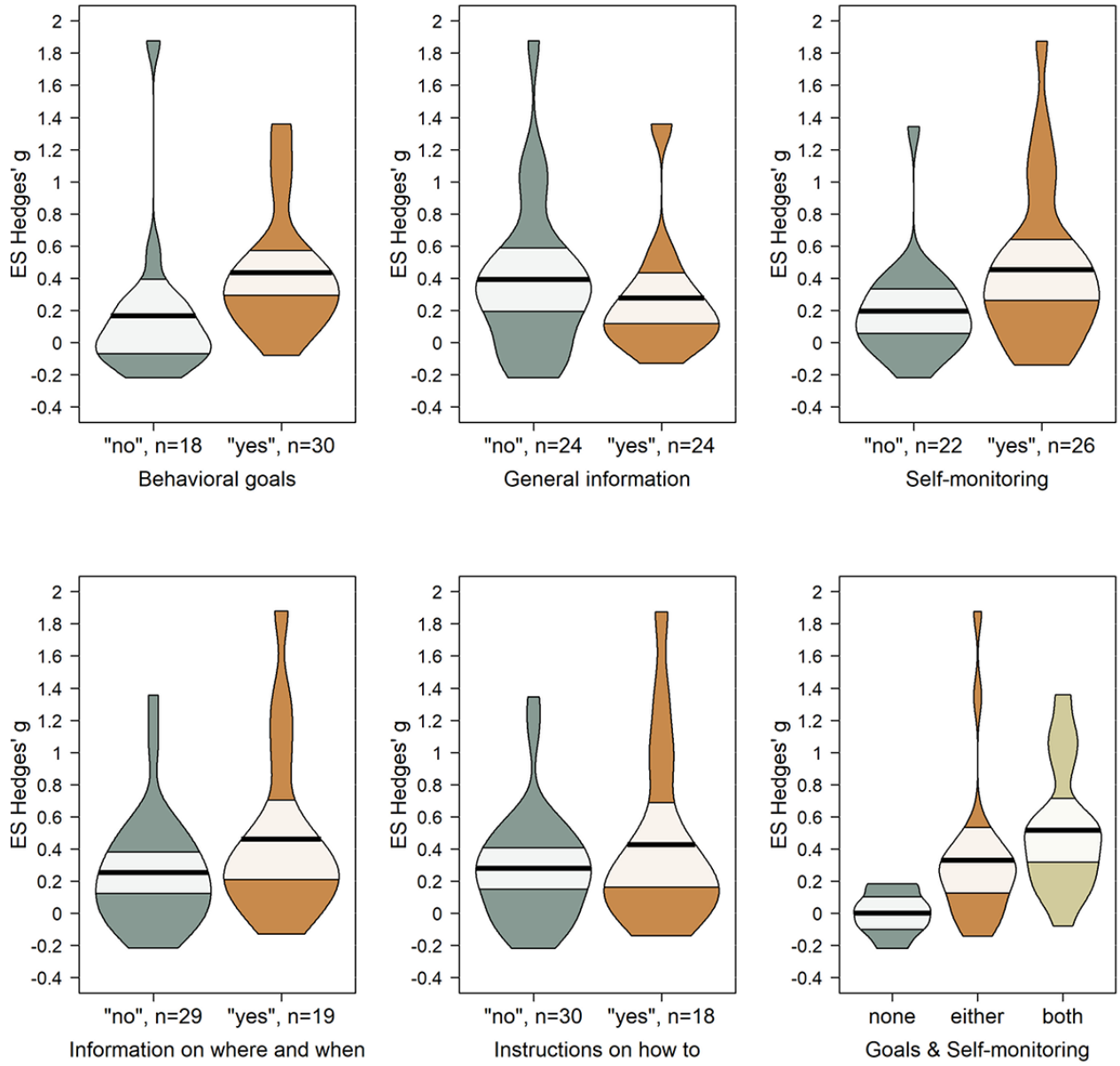
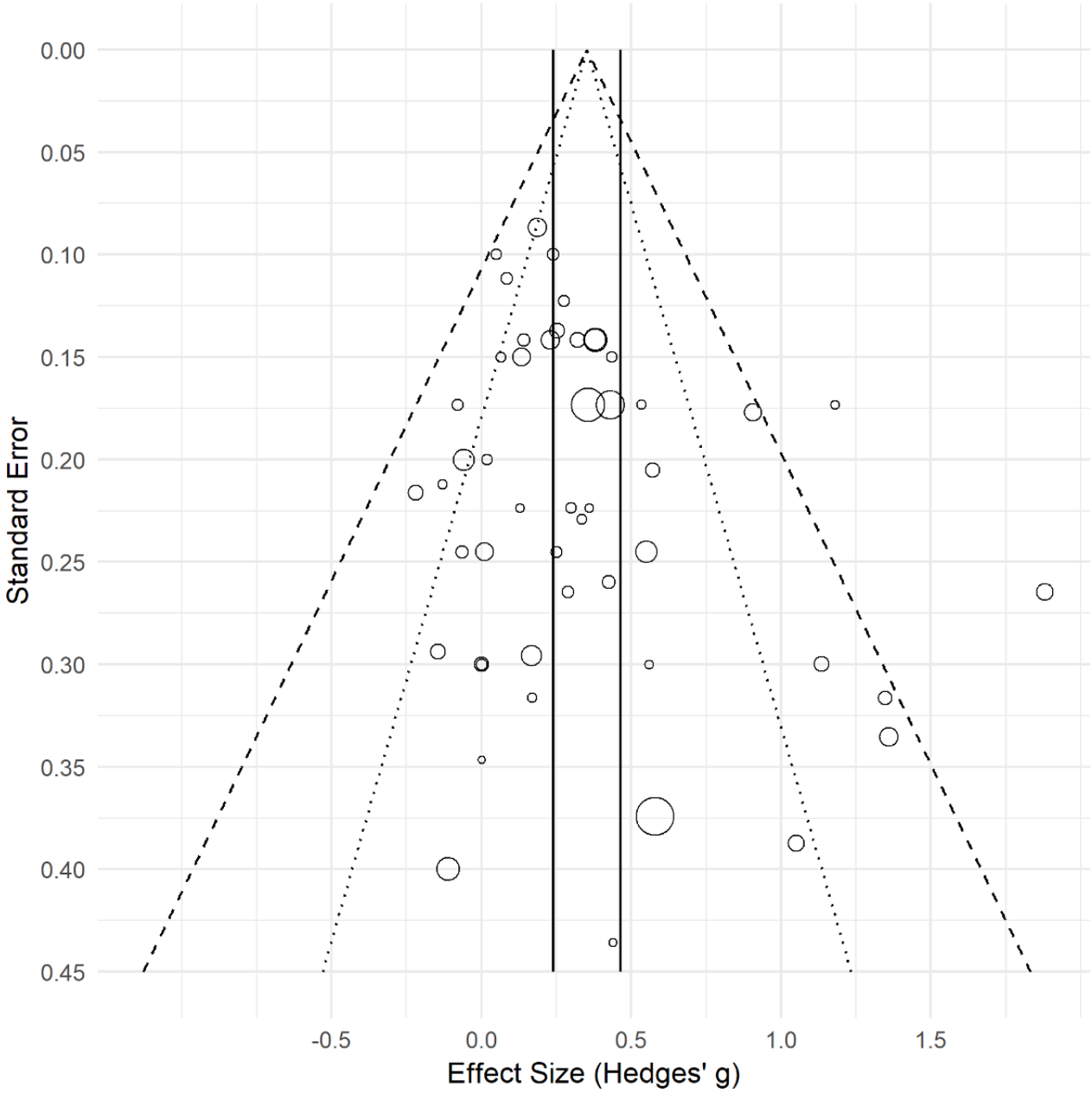


Figure 5. Funnel plot to assess publication bias. Point size indicates the number of participants in each study. The dotted and dashed lines show a 95% and 99% credibility region respectively and the full lines represent a 95% CI for the summary effect.



Publication Bias

To assess possible publication bias visually, we drew a funnel plot (see Figure 5). On the right side, there were 6 studies outside the 95% credibility region and 2 of those 6 were outside the 99% credibility region. The trim and fill method also suggested that 7 studies are missing on the left side (SE 4.00). These studies might lead to an overestimation of the overall effect. Filling in these "missing" studies led to an adjusted estimated overall effect size of g=0.22, meaning that we overestimated by Δg=0.07. The Egger asymmetry test was not significant (z=1.53, P=.12). Due to its low power in small pools of studies, this nonsignificance does not rule out that no studies are missing. However, the Orwin Failsafe-N test suggests that to obtain a point estimate half the size of the original effect (ie, g=0.14), we would need 76 more studies with a nil effect. One hundred and thirty-three studies with a nil effect would be needed to reduce the effect to a negligible effect size of g=0.10. Overall, publication bias seems to be acceptably low.

Discussion

This meta-analysis aimed to identify which BCTs contribute to behavior change in mHealth interventions targeting physical activity. We found an association between heightened efficacy and behavioral goals and the same for self-monitoring. Using either of these techniques pushed the effect from g=0.01 to g=0.32, but we found no further significant benefit from combining them in an intervention (g=0.36). Overall, we found a small to moderate effect of interventions on physical activity, which agrees with another recent meta-analysis of mHealth physical activity interventions [13]. Furthermore, targeting

physical activity directly instead of weight can lead to a greater increase of physical activity, but we did not find that the study quality or the tracked versus self-reported data contributed to the intervention efficacy. We also found that neglecting baseline group differences would lead to an overestimation of the efficacy of interventions in this pool of diverse RCTs with many small sample sizes.

To our knowledge, this is the most comprehensive assessment of mHealth interventions to increase physical activity so far. Further, this is the first one to address baseline group differences as covariates and to analyze moderating effects of the most frequently used BCTs while retaining sufficient power. However, 34/50 (68%) of all included studies were published in the years 2015 and 2016. This confirms the rapid growth of the mHealth literature concerned with increasing physical activity. To facilitate sequential meta-analyses and to increase transparency, we are sharing the information we coded and the annotated R script in Multimedia Appendix 2.

Limitations

There are 2 caveats we would like to address. Firstly, we did not look at actual changes within groups over time because the correlation between baseline and postintervention, which is necessary for this analysis, was generally not reported. Instead, we assessed group differences at baseline and postintervention. Using this approach, an increase of physical activity in the intervention group compared to the control group led us to assume intervention success. However, we also assumed intervention success if physical activity in the control group decreased during the intervention while being stable in the intervention group. Of course, the effects we analyzed might also have been a combination of both (ie, an increase of physical activity in the intervention group and a decrease of physical activity in the control group). Secondly, 4 studies in our pool of 50 (8%) studies used clustered randomization methods (for example by school) [41,54,56,67]. We did not differentiate those studies from fully RCTs, because we did not expect an interaction between the intervention effect and the type of unit randomized (ie, person versus school) [95].

Future Research

When more studies are available, we will have the statistical power for more sophisticated moderator analyses like meta-CART [96]. Our results suggest that behavioral goals and self-monitoring are especially beneficial, but it remains unclear which other BCTs work well and what interactions there are. Many BCTs from the taxonomy were not used at all, or only very little, and it was hard to code BCTs based on the available information. For the field as a whole it would be beneficial to report intervention content in a more structured way—if possible, based on a BCT taxonomy [24,97] to counteract the current confusion [20], and make our findings more replicable. Furthermore, not all BCTs might work equally well in each context and for each person. However, to assess this question, more research is needed.

In this meta-analysis, we only looked at BCTs, but of course, there is much more to think about when designing successful mHealth interventions. Intervention efficacy does not only depend on BCTs as intervention components, but also on the target population, the intervention design, and duration, as well as the intervention objectives. For example, participants who suffer from illness might be more motivated to use an mHealth intervention because they suffer more than healthy participants. Additionally, healthy participants might already be quite physically active and would therefore not gain a lot from an mHealth intervention targeting motivation to be physically active. Further, BCTs need to be carefully matched with intervention objectives. For example, when an intervention targets capability for physical activity, BCTs related to self-belief might have a greater impact on intervention success than behavioral goals and self-monitoring.

Conclusion

Despite a small to moderate overall success in increasing physical activity with mHealth interventions, setting behavioral goals or enabling self-monitoring, as well as a combination thereof, might be beneficial. With increasing technological possibilities, interventions will become ever more complex, and it is crucial to report their content thoroughly. However, let us not forget: BCTs are not everything. It is also important that people like and use the interventions. Elements of gamification and appealing visual presentation could be considered to address this issue.

Acknowledgments

This project was financed by the Austrian Science Fund (FWF) P 28393. The authors acknowledge the financial support from the University of Graz. Many thanks to Martin Aresin, Justus Baltzer and Robert Spörk for helping to code the studies. We would also like to thank Elizabeth Tipton for a fruitful discussion on how to handle baseline group differences.

Conflicts of Interest

None declared.

Authors' Contributions

The authors LVE, NKT, and KC conceived the presented idea. LVE developed the theory, supervised the data collection, performed the computations, and wrote the first draft of the manuscript. NKT verified the analytical methods. KC supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Multimedia Appendix 1

Description of included studies (population, country, duration, and outcomes).

[XLSX File (Microsoft Excel File), 63KB - mhealth_v6i11e10076_app1.xlsx]

Multimedia Appendix 2

Data spreadsheet and annotated analysis script.

[ZIP File (Zip Archive), 103KB - mhealth_v6i11e10076_app2.zip]

References

1. The Nielsen Company. http://www.nielsen.com/. 2017. Nielsen Total Audience Report - Q2 2017 URL: http://www.nielsen.com/us/en/insights/reports/2017/the-nielsen-total-audience-q2-2017.html [accessed 2017-12-30] [WebCite Cache ID 6wvWsI0WR]
2. European Commission. https://data.europa.eu/. 2014. Special Eurobarometer 412: Sport and Physical Activity 2014 URL: https://data.europa.eu/euodp/data/dataset/S1116_80_2_412 [accessed 2017-12-31] [WebCite Cache ID 6x4ssbQ6l]
3. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. The Lancet 2012 Jul;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1]
4. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
5. Patel AV, Hildebrand JS, Leach CR, Campbell PT, Doyle C, Shuval K, et al. Walking in Relation to Mortality in a Large Prospective Cohort of Older U.S. Adults. Am J Prev Med 2018 Jan;54(1):10-19. [doi: 10.1016/j.amepre.2017.08.019] [Medline: 29056372]
6. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Current Opinion in Psychiatry 2005;18(2):189-193. [doi: 10.1097/00001504-200503000-00013]
7. Fox KR. The influence of physical activity on mental well-being. PHN 2007 Jan 2;2(3a). [doi: 10.1017/S1368980099000567]
8. Statista. Number of smartphone users worldwide from 2014 to 2020 (in billions) 2017 WebCite original URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2018-02-02] [WebCite Cache ID 6wvXEk8pW]
9. Statista. Number of smartphone users in the United States from 2010 to 2022 (in millions) 2017 WebCite original URL: https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/ [accessed 2018-02-02] [WebCite Cache ID 6wvXX5zgJ]
10. Statista. Number of mobile phone users worldwide from 2013 to 2019 (in billions) 2018 WebCite original URL: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/ [accessed 2018-05-18] [WebCite Cache ID 6zVOXIY7A]
11. Miller G. The Smartphone Psychology Manifesto. Perspect Psychol Sci 2012 May;7(3):221-237 [FREE Full text] [doi: 10.1177/1745691612441215] [Medline: 26168460]
12. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
13. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to Influence Physical Activity and Sedentary Behaviors: Behavior Change Techniques, Systematic Review and Meta-Analysis of Randomized Controlled Trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
14. 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]
15. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
16. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
17. Monroe CM, Thompson DL, Bassett DR, Fitzhugh EC, Raynor HA. Usability of Mobile Phones in Physical Activity–Related Research: A Systematic Review. American Journal of Health Education 2015 Jul 06;46(4):196-206. [doi: 10.1080/19325037.2015.1044141]
18. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
19. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]

20. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
21. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002 Feb 28;21(4):589-624. [Medline: 11836738]
22. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
23. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
24. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
25. Abraham AA, Chow W, So H, Yip BH, Li AM, Kumta SM, et al. Lifestyle intervention using an internet-based curriculum with cell phone reminders for obese Chinese teens: a randomized controlled study. PLoS One 2015;10(5):e0125673 [FREE Full text] [doi: 10.1371/journal.pone.0125673] [Medline: 25946465]
26. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013 Dec;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
27. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
28. Allman-Farinelli M, Partridge SR, McGeechan K, Balestracci K, Hebden L, Wong A, et al. A Mobile Health Lifestyle Program for Prevention of Weight Gain in Young Adults (TXT2BFiT): Nine-Month Outcomes of a Randomized Controlled Trial. JMIR Mhealth Uhealth 2016 Jun 22;4(2):408-419 [FREE Full text] [doi: 10.2196/mhealth.5768] [Medline: 27335237]
29. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. Am J Prev Med 2015 Sep;49(3):414-418 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
30. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth Physical Activity Intervention: A Randomized Pilot Study in Physically Inactive Pregnant Women. Matern Child Health J 2016 May;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
31. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of Lifestyle-Focused Text Messaging on Risk Factor Modification in Patients With Coronary Heart Disease: A Randomized Clinical Trial. JAMA 2015 Sep;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
32. Cotten E, Prapavessis H. Increasing Nonsedentary Behaviors in University Students Using Text Messages: Randomized Controlled Trial. JMIR Mhealth Uhealth 2016 Aug 19;4(3):e99 [FREE Full text] [doi: 10.2196/mhealth.5411] [Medline: 27543317]
33. Cowdery J, Majeske P, Frank R, Brown D. Exergame Apps and Physical Activity: The Results of the ZOMBIE Trial. American Journal of Health Education 2015 Jul 06;46(4):216-222. [doi: 10.1080/19325037.2015.1043063]
34. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and Increasing Physical Activity Among Young People: The AIMFIT Pragmatic Randomized Controlled Trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
35. Fassnacht DB, Ali K, Silva C, Gonçalves S, Machado PPP. Use of text messaging services to promote health behaviors in children. J Nutr Educ Behav 2015;47(1):75-80. [doi: 10.1016/j.jneb.2014.08.006] [Medline: 25282200]
36. Fjeldsoe BS, Goode AD, Phongsavan P, Bauman A, Maher G, Winkler E, et al. Evaluating the Maintenance of Lifestyle Changes in a Randomized Controlled Trial of the 'Get Healthy, Stay Healthy' Program. JMIR Mhealth Uhealth 2016;4(2) [FREE Full text] [doi: 10.2196/mhealth.5280] [Medline: 27166643]
37. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]
38. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-Term Effectiveness of a Comprehensive Internet-Based and Patient-Specific Telerehabilitation Program With Text Messaging Support for Cardiac Patients: Randomized Controlled Trial. J Med Internet Res 2015;17(7) [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
39. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A Novel Diabetes Prevention Intervention Using a Mobile App: A Randomized Controlled Trial With Overweight Adults at Risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
40. Garde A, Umedaly A, Abulnaga SM, Robertson L, Junker A, Chanoine JP, et al. Assessment of a Mobile Game to Promote Physical Activity Among Children. Games Health J 2015 Apr 01;4(2):149-158. [doi: 10.1089/g4h.2014.0095] [Medline: 26181809]

41. Gell NM, Wadsworth DD. The Use of Text Messaging to Promote Physical Activity in Working Women: A Randomized Controlled Trial. J Phys Act Health 2015 Jun;12(6):756-763. [doi: 10.1123/jpah.2013-0144] [Medline: 25110303]
42. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
43. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: A randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Dec;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
44. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and Phone-Based Weight Loss Intervention: Pilot RCT in Women at Elevated Breast Cancer Risk. Am J Prev Med 2016 Nov;51(5):714-721 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.024] [Medline: 27593420]
45. Hebden L, Cook A, van der Ploeg HP, King L, Bauman A, Allman-Farinelli M. A mobile health intervention for weight management among young adults: a pilot randomised controlled trial. J Hum Nutr Diet 2013 Aug 29;27(4):322-332. [doi: 10.1111/jhn.12155]
46. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
47. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
48. Joseph RP, Keller C, Adams MA, Ainsworth BE. Print versus a culturally-relevant Facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
49. Kim BH, Glanz K. Text messaging to motivate walking in older African Americans: a randomized controlled trial. Am J Prev Med 2013 Jan;44(1):71-75. [doi: 10.1016/j.amepre.2012.09.050] [Medline: 23253653]
50. Kim J, Oh S, Steinhubl S, Kim S, Bae WK, Han JS, et al. Effectiveness of 6 months of tailored text message reminders for obese male participants in a worksite weight loss program: randomized controlled trial. JMIR Mhealth Uhealth 2015 Feb 03;3(1):e14 [FREE Full text] [doi: 10.2196/mhealth.3949] [Medline: 25648325]
51. Kim JY, Wineinger NE, Steinhubl SR. The Influence of Wireless Self-Monitoring Program on the Relationship Between Patient Activation and Health Behaviors, Medication Adherence, and Blood Pressure Levels in Hypertensive Patients: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
52. Kinnafick FE, Thøgersen-Ntoumani C, Duda J. The effect of need supportive text messages on motivation and physical activity behaviour. J Behav Med 2016 Dec;39(4):574-586 [FREE Full text] [doi: 10.1007/s10865-016-9722-1] [Medline: 26915963]
53. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12. [doi: 10.7326/M13-3005] [Medline: 25402403]
54. Lubans DR, Smith JJ, Plotnikoff RC, Dally KA, Okely AD, Salmon J, et al. Assessing the sustained impact of a school-based obesity prevention program for adolescent boys: the ATLAS cluster randomized controlled trial. Int J Behav Nutr Phys Act 2016 Aug 20;13:92 [FREE Full text] [doi: 10.1186/s12966-016-0420-8] [Medline: 27542825]
55. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: Results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
56. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij BI, Thomas S, et al. A Web-Based, Social Networking Physical Activity Intervention for Insufficiently Active Adults Delivered via Facebook App: Randomized Controlled Trial. J Med Internet Res 2015;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
57. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: A Randomized Clinical Trial of an Automated mHealth Intervention for Physical Activity Promotion. J Am Heart Assoc 2015 Nov 09;4(11) [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
58. Nguyen HQ, Donesky D, Reinke LF, Wolpin S, Chyall L, Benditt JO, et al. Internet-based dyspnea self-management support for patients with chronic obstructive pulmonary disease. J Pain Symptom Manage 2013 Jul;46(1):43-55 [FREE Full text] [doi: 10.1016/j.jpainsymman.2012.06.015] [Medline: 23073395]
59. Pfaeffli DL, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text Message and Internet Support for Coronary Heart Disease Self-Management: Results From the Text4Heart Randomized Controlled Trial. J Med Internet Res 2015;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
60. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an Activity Tracker- and Internet-Based Adaptive Walking Program for Adults: A Randomized Controlled Trial. J Med Internet Res 2016;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]

61. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychology 2010;29(1):40-49.
62. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, et al. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. The Lancet Diabetes & Endocrinology 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2]
63. Schwerdtfeger AR, Schmitz C, Warken M. Using text messages to bridge the intention-behavior gap? A pilot study on the use of text message reminders to increase objectively assessed physical activity in daily life. Front Psychol 2012;3:270 [FREE Full text] [doi: 10.3389/fpsyg.2012.00270] [Medline: 22876237]
64. Silveira P, van de Langenberg R, van Het Reve E, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase II preclinical exploratory trial. J Med Internet Res 2013;15(8):e159 [FREE Full text] [doi: 10.2196/jmir.2579] [Medline: 23939401]
65. Suggs S, Blake H, Bardus M, Lloyd S. Effects of text messaging in addition to emails on physical activity among university and college employees in the UK. J Health Serv Res Policy 2013 Apr;18(1 Suppl):56-64. [doi: 10.1177/1355819613478001] [Medline: 27552780]
66. Tabak M, op den Akker H, Hermens H. Motivational cues as real-time feedback for changing daily activity behavior of patients with COPD. Patient Educ Couns 2014 Mar;94(3):372-378. [doi: 10.1016/j.pec.2013.10.014] [Medline: 24332934]
67. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and Web-Based Monitoring and Feedback Tool Embedded in Primary Care Increases Physical Activity: A Cluster Randomized Controlled Trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
68. van Drongelen A, Boot CRL, Hlobil H, Twisk JWR, Smid T, van der Beek AJ. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. Scand J Work Environ Health 2014;40(6):557-568 [FREE Full text] [doi: 10.5271/sjweh.3447] [Medline: 25121620]
69. Vorrink SNW, Kort HSM, Troosters T, Zanen P, Lammers JJ. Efficacy of an mHealth intervention to stimulate physical activity in COPD patients after pulmonary rehabilitation. Eur Respir J 2016 Oct;48(4):1019-1029. [doi: 10.1183/13993003.00083-2016] [Medline: 27587557]
70. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth Intervention Using a Smartphone App to Increase Walking Behavior in Young Adults: A Pilot Study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
71. Wang JB, Cataldo JK, Ayala GX, Natarajan L, Cadmus-Bertram LA, White MM, et al. Mobile and Wearable Device Features that Matter in Promoting Physical Activity. J Mob Technol Med 2016 Jul;5(2):2-11. [doi: 10.7309/jmtm.5.2.2] [Medline: 27493694]
72. Zach S, Raviv T, Meckel Y. Using information communication technologies (ICTs) for motivating female adolescents to exercise/run in their leisure time. Computers in Human Behavior 2016;60:593-601.
73. Fjeldsoe BS, Miller YD, Graves N, Barnett AG, Marshall AL. Randomized Controlled Trial of an Improved Version of MobileMums, an Intervention for Increasing Physical Activity in Women with Young Children. Ann Behav Med 2015 Aug;49(4):487-499. [doi: 10.1007/s12160-014-9675-y] [Medline: 25582987]
74. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
75. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester: John Wiley & Sons Inc; 2009.
76. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
77. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: A critical review of current literature. Computers in Human Behavior 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]
78. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
79. Higgins JPT, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, et al. A tool to assess the quality of a meta-analysis. Res Synth Methods 2013;4(4):351-366. [doi: 10.1002/jrsm.1092] [Medline: 26053948]
80. R Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2017. R: A Language and Environment for Statistical Computing URL: http://www.r-project.org/ [accessed 2018-10-08] [WebCite Cache ID 731Ngy7bi]
81. Fisher Z, Tipton E, Zhipeng H. https://CRAN.R-project.org/package=robumeta. 2017. robumeta: Robust Variance Meta-Regression URL: https://cran.r-project.org/web/packages/robumeta/index.html [accessed 2018-02-08] [WebCite Cache ID 6x4rsiQGf]

82. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 2010;36:1-48 [FREE Full text] [doi: 10.18637/jss.v036.i03]
83. Del Re AC, Hoyt W. http://cran.r-project.org/web/packages/MAd. 2014. MAd: Meta-Analysis with Mean Differences URL: http://cran.r-project.org/web/packages/MAd [accessed 2018-02-08] [WebCite Cache ID 6x4rvDNZs]
84. Wickham H. https://CRAN.R-project.org/package=tidyverse. 2017. tidyverse: Easily Install and Load the 'Tidyverse' URL: https://CRAN.R-project.org/package=tidyverse [WebCite Cache ID 6x4rxbCsX]
85. Whickham H, Bryan J. https://CRAN.R-project.org/package=readxl. 2017. readxl: Read Excel Files URL: https://cran.r-project.org/web/packages/readxl/index.html [accessed 2018-02-08] [WebCite Cache ID 6x4s0Pqu6]
86. Del Re AC. http://cran.r-project.org/web/packages/compute.es. 2013. compute.es: Compute Effect Sizes URL: http://cran.r-project.org/web/packages/compute.es [accessed 2018-02-08] [WebCite Cache ID 6x4sMTYma]
87. Wickham H. ggplot2: Elegant Graphics for Data Analysis. In: ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2009.
88. Phillips N. https://CRAN.R-project.org/package=yarrr. 2017. YaRrr!: The Pirate's Guide to R URL: https://cran.r-project.org/web/packages/yarrr/index.html [accessed 2018-02-08] [WebCite Cache ID 6x4s4VbkG]
89. Ram K, Wickham H. https://CRAN.R-project.org/package=wesanderson. wesanderson: A Wes Anderson Palette Generator 2015 WebCite original URL: https://cran.r-project.org/web/packages/wesanderson/index.html [accessed 2018-02-08] [WebCite Cache ID 6x4s6wluZ]
90. Polanin JR, Hennessy EA, Tanner-Smith EE. A Review of Meta-Analysis Packages in R. Journal of Educational and Behavioral Statistics 2016;42(2):206-242. [doi: 10.3102/1076998616674315]
91. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods 2010 Jan;1(1):39-65. [doi: 10.1002/jrsm.5] [Medline: 26056092]
92. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634. [doi: 10.1136/bmj.315.7109.629]
93. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000;56(2):455-463. [Medline: 10877304]
94. Orwin RG. A Fail-Safe N for Effect Size in Meta-Analysis. Journal of Educational Statistics 2016 Nov 23;8(2):157-159. [doi: 10.3102/10769986008002157]
95. Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. Stat Med 2002 Oct 15;21(19):2971-2980. [doi: 10.1002/sim.1301] [Medline: 12325113]
96. Li X, Dusseldorp E, Meulman JJ. Meta-CART: A tool to identify interactions between moderators in meta-analysis. Br J Math Stat Psychol 2017 Feb;70(1):118-136. [doi: 10.1111/bmsp.12088] [Medline: 28130936]
97. Michie S, Wood C. Reporting Behaviour Change Interventions and Techniques. In: Benyamini Y, Johnston M, Karademas EC, editors. Assessment in health psychology. Goettingen: hogrefe; 2016:252-264.

Abbreviations

BCT: behavior change technique
EPHPP: Effective Public Health Practice Project
ES: effect size
mHealth: mobile health
PA: physical activity
RCT: randomized controlled trial

Edited by C Dias; submitted 08.02.18; peer-reviewed by M Conner, J Rawstorn, E Knight; comments to author 20.03.18; revised version received 29.06.18; accepted 26.07.18; published 12.11.18.

Please cite as:
Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, Corcoran K
Key Elements of mHealth Interventions to Successfully Increase Physical Activity: Meta-Regression
JMIR Mhealth Uhealth 2018;6(11):e10076
URL: https://mhealth.jmir.org/2018/11/e10076/
doi:10.2196/10076
PMID:30425028

Review

The Rise of Pregnancy Apps and the Implications for Culturally and Linguistically Diverse Women: Narrative Review

Jo-anne Patricia Hughson[1], BA (Hons), DEA, PhD; J Oliver Daly[2], BSC, MBBS, FRANZCOG; Robyn Woodward-Kron[3], BA, MA, MA (TESOL), PhD; John Hajek[1], BA (Hons), DPhil; David Story[4], MD, MBBS, BMedSc, FANZCA

[1]Research Unit for Multilingualism and Cross-Cultural Communication, University of Melbourne, Parkville, Australia

[2]Western Health, Sunshine Hospital, St Albans, Australia

[3]Department of Medical Education, University of Melbourne, Parkville, Australia

[4]Anaesthesia, Perioperative and Pain Medicine Unit, Melbourne Medical School, University of Melbourne, Parkville, Australia

Corresponding Author:
Jo-anne Patricia Hughson, BA (Hons), DEA, PhD
Research Unit for Multilingualism and Cross-Cultural Communication
University of Melbourne
Babel Building
Parkville, 3010
Australia
Phone: 61 425710396
Email: hughson@unimelb.edu.au

Abstract

Background: Pregnancy apps are a booming global industry, with most pregnant women in high-income countries now using them. From the perspective of health care and health information provision, this is both encouraging and unsettling; the demand indicates a clear direction for the development of future resources, but it also underscores the importance of processes ensuring access, reliability, and quality control.

Objective: This review provides an overview of current literature on pregnancy apps and aims at describing (1) the ways in which apps are used by women, in general, and by those of a culturally and linguistically diverse (CALD) background; (2) the utility and quality of information provided; and (3) areas where more research, development, and oversight are needed.

Methods: We chose a narrative review methodology for the study and performed a structured literature search including studies published between 2012 and 2017. Searches were performed using MEDLINE, EMBASE, and CINAHL databases. Studies were identified for inclusion using two separate search criteria and strategies: (1) studies on pregnancy apps and pregnant women's use of these apps and (2) studies on CALD pregnant women and their use of technology for accessing information on and services for pregnancy. Overall, we selected 38 studies.

Results: We found that pregnancy apps were principally used to access pregnancy health and fetal development information. Data storage capability, Web-based features or personalized tools, and social media features were also popular app features sought by women. Lower rates of the pregnancy app uptake were indicated among lower-income and non-English-speaking women. Preliminary evidence indicates that a combination of technological, health literacy, and language issues may result in lower uptake of pregnancy apps by these groups; however, further investigation is required. A marked limitation of the health app industry is lack of regulation in a commercially dominated field, making it difficult for users to assess the reliability of the information being presented. Health professionals and users alike indicate that given the choice, they would prefer using pregnancy apps that are relevant to their local health care context and come from a trusted source. Evidence indicates a need for greater health professional and institutional engagement in the app development, as well as awareness of and guidance for women's use of these resources.

Conclusions: This is the first review of pregnancy app use, types of information provided, and features preferred by pregnant women in general and by those of a CALD background in particular. It indicates the demand for access to accurate information that is relevant to users, their community, and their associated health services. Given the popularity of pregnancy apps, such apps have enormous potential to be used for the provision of accurate, evidence-based health information.

(JMIR Mhealth Uhealth 2018;6(11):e189) doi:10.2196/mhealth.9119

KEYWORDS

culture; emigrants and immigrants; health communication; information-seeking behavior; literacy; maternal health; mHealth; mobile phone; pregnancy; self-care; vulnerable populations

Introduction

Engaging with pregnancy apps appears to have become a routine part of the maternity experience for many expectant mothers. Globally, there are more pregnancy apps than for any other medical topic [1], which attests to their ever-increasing popularity. Internationally, the majority of smartphone owners are women [2], with indications that pregnancy itself is an incentive to buy a smartphone [3]. In advanced economies, like the United States and Australia, between 72% and 89% of the general population, and 92%-95% of 18- to 34-year olds, owned smartphones in 2015 [4,5].

There are several reasons for the high uptake of apps by women. Pregnancy is a normal life process influenced by many social and cultural practices and interactions with the health system. The types of information and resources sought by women during pregnancy span many domains such as health care, social, cultural, and material. Accordingly, pregnancy apps incorporate a range of platforms with diverse features catering to these diverse domains. We propose that the range of pregnancy apps can be categorized along a continuum from entertainment at one end through to health care at the other. Common features of pregnancy apps include the following: information provision, pregnancy tracking, record keeping, and gestation calculation. Of the plethora of pregnancy apps designed for entertainment purposes [6], games reportedly constitute the highest number of pregnancy apps [7], alongside Web-based shopping for pregnancy-related products, gender predictors, and baby name generators [8]. In addition, there are apps that are not pregnancy specific, such as for social networking, which may take on a specialized function during a woman's pregnancy, providing an important link to social connection and support [9,10]. Some apps offer multifunctionality while others are dedicated to a single function. While some apps clearly belong to either end of the continuum, many apps blend multiple functions that could place them at both ends of the same continuum.

While there is a growing popularity of pregnancy apps, it is unclear to what extent they actually address the needs of particular groups of women, such as culturally and linguistically diverse (CALD) populations. In this review, the term CALD refers to people living in an English language-dominant context who are of non-English-speaking background (NESB) and who may be bi-, tri-, or multilingual, but generally have low proficiency in English. CALD groups require particular consideration, given cultural variabilities in maternity practices of different ethnic groups [11], lower levels of literacy [12], health literacy [13], health care access [14], and concomitant poorer perinatal outcomes of some CALD groups in comparison to non-CALD women [15]. Thus, it is crucial to ascertain whether digital technologies improve outcomes for CALD and other disadvantaged groups or simply perpetuate the *status quo* [16].

This review provides an overview of women's preferences for and use of pregnancy apps as resources for pregnancy information, the effectiveness and acceptability of such resources, and their shortcomings, including unregulated development that may affect the quality of information provided. Based on the findings, recommendations are made regarding future research and features that need to be included in pregnancy apps to ensure that they provide useful, reliable information in a format that is accessible to the general population, including CALD users.

Methods

Research Question

The research question guiding this narrative review initially focused on studies examining the pregnancy app usage in CALD women. The review was motivated by the authors' involvement in providing digital health information to CALD pregnant women, in particular via smartphones. However, the scoping searches revealed a gap in the literature, prompting a broader review of women's use and uptake of pregnancy apps more generally, including issues related to their design and content.

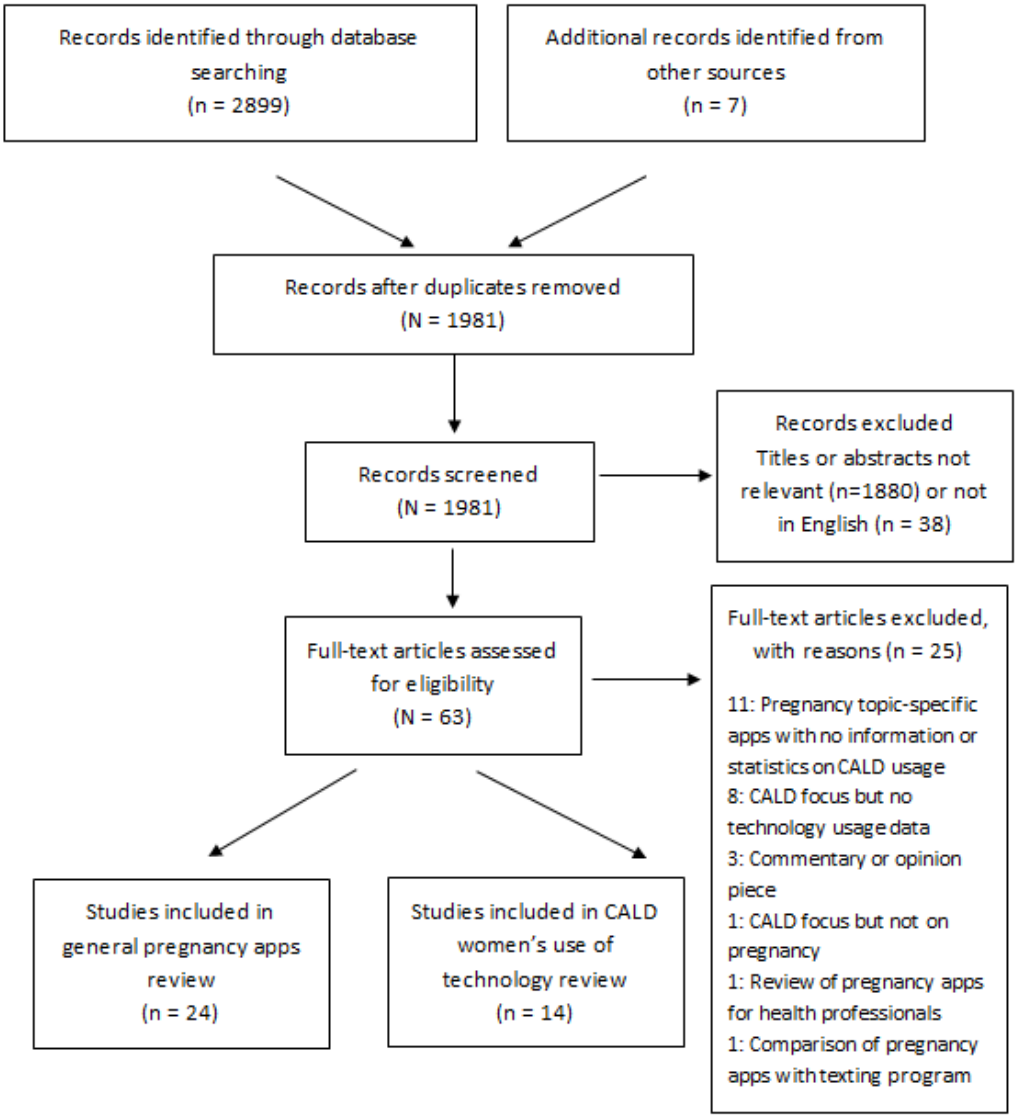
Search Strategy and Selection of Studies

A structured literature search was performed including studies published between 2012 and 2017. Searches were carried out using MEDLINE, EMBASE, and CINAHL databases. Given the varied nature of the publications under review, studies were identified for inclusion using two search strategies: (1) studies of pregnancy apps and pregnant women's use of these apps and (2) studies of CALD pregnant women and their use of technology for accessing information on and services for pregnancy. A sample search strategy used is reproduced in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

For studies reporting women's use of pregnancy apps, type 1 studies, the ones dedicated to a particular health topic in pregnancy (eg, nutrition education and breastfeeding), were excluded, although studies reporting on the general app usage to inform the design of pregnancy topic-specific resources were deemed suitable for inclusion. Studies of apps designed for use in clinical contexts and requiring the coparticipation of health professionals were also excluded unless they contained explicit content regarding CALD usage. These delimitations were set because (1) the authors were searching for information on designing an app on pregnancy, rather than an app on a specific pregnancy-related topic, and (2) the app was for pregnant women to use themselves, rather than an app for use with or by health professionals.

Figure 1. Literature search. CALD: culturally and linguistically diverse.



In the absence of substantial material on CALD women's use of apps, but seeking to gain an appreciation of factors that may help inform CALD usage of pregnancy apps, we additionally searched for studies that included information about CALD use of technology in pregnancy more broadly (including mobile phone usage, internet usage, general app or smartphone use). No exclusion criteria were applied to these studies provided they gave detailed information about the CALD women's usage and preferences with regard to digital technology. Further relevant articles were identified through manual searches of reference lists. In total, 38 studies were selected for inclusion in the review (Figure 1).

Study Analysis

Once accepted for inclusion, type 1 studies (n=24) were analyzed for any content pertaining to the pregnancy app usage, preferences in app features or design, or issues associated with app usage and type 2 studies (n=14) were analyzed for content concerning their attitudes toward and uptake or acceptance of digital technologies related to pregnancy care. These results were tabulated and are reproduced in Multimedia Appendices 2 and 3.

Results

How Pregnant Women Use Apps

The General Population

Information seeking is a key motivation for women using pregnancy apps. Certainly, the largest category of apps available on the market involves those providing information about pregnancy [1]. We found limited research providing evidence of the exponential rise in popularity of apps for accessing pregnancy information. A 2013 Australian study found that 40% of 35 participants had used at least one smartphone app to access pregnancy information [17]. In 2016, a larger Australian survey study of 410 women reported the proportion of pregnancy app users as 73%, the vast majority (92%) finding them to be useful [8,18]. The most valued feature was providing useful information (83%), mainly about fetal development (89%) or pregnancy-related changes in their bodies (71%) [8]. In a Korean

study [19] evaluating app use among 193 pregnant women, most participants used apps to obtain information, and the most frequently sought information was on risks and diseases during pregnancy (17%), physical changes related to a normal pregnancy (16%), prenatal education (15%), and breastfeeding and baby food (13%). An analysis of search engine use demonstrated pregnancy information seeking to be a primary function of women's digital pregnancy-related activities in several studies [6,19,20]. One US study found that Google searches were the most common mode of information seeking, particularly early in pregnancy [20]. An important factor for many women was the ability to obtain information instantly and easily [6,19,21,22].

The most popular apps have Web-based features [1]. When considering pregnancy apps, Web-based and mobile features include regular notifications, videos, and data storage capability, for example, for photos; taking notes; pregnancy tracking; and "personalized tools to assess nutrition, fitness, and weight" [18,23]. In addition, women value pregnancy apps and digital platforms "that are multifunctional and interact with each other" [6,24]. Participants in a recent German study strongly advocated for apps that had the functionality to provide individualized feedback, "communication platforms" for pregnant women and medical professionals, and indicated a desire for the integration of apps into routine clinical and pregnancy care [22]. Pregnant women interviewed to inform the design of a Dutch pregnancy app wanted all of the abovementioned features as well as a push notification system to remind them of tasks they needed to accomplish, an alert system to tell them if any monitored signs became irregular or dangerous, and a social media module for connecting with other pregnant women [25].

Social media use (eg, use of Facebook, discussion forums) has frequently featured in research on pregnancy information seeking and can be an important way for women to find support, develop networks, and foster emotional well-being [26,27]. Social media platforms, which are usually accessed via an app on women's phones, also provide women with a setting in which they can acquire and share (nonmedical) knowledge and expertise about motherhood and are a helpful way for women to get an idea of whether their symptoms and experiences during pregnancy are "normal" [21]. Apps with social media functionality may be particularly useful in late pregnancy and the postnatal period when needs change from seeking information to seeking support, being part of a community, and avoiding isolation [26]. One small US study reported 82% of pregnant women used Web-based social networking sites, at least, once a day [20], and discussion forums and social media were the most used websites in a large-scale Irish study [28]. However, the appeal of social media is not universal, with one study reporting low levels of interest by lower-income American women in using social networking tools to discuss their pregnancy—they preferred to seek support from their existing social circles [29]. It should be noted that some of the references above are general studies on social media use that did not disclose how women accessed social media, for example, via apps on their phones or other interfaces, and are, therefore, not included in the summary tables (Multimedia Appendices 2 and 3).

Pregnancy App and Other Technology Use Among Culturally and Linguistically Diverse and Socially Disadvantaged Women

This section discusses results of studies reporting CALD (or ethnic or racial minority) groups' use of technology during pregnancy and discusses whether or not, or to what degree, digital resources are acceptable and usable for these groups. We found only one study focusing on the pregnancy app use by CALD women [30], and most studies excluded women who were not proficient speakers of the given society's dominant language, for example, English, in English-speaking countries. Yet, CALD communities make up a proportion of the population in most countries. For example, in the Australian setting, almost one-quarter of women giving birth in 2013 were born outside Australia in non-English-speaking countries [31]. There appears to be little provision of multilingual apps or apps for use with migrants, with only 2 multilingual apps found for use in maternity care as part of our literature search; both were developed in Europe for health professional-patient interactions [32,33]. While these are much-needed resources, they fulfill a different role to the type of resources examined by this review.

There is evidence that CALD women use technology to source pregnancy information. An Australian study examining sources of information used by women in pregnancy did include CALD-background respondents, with more than one-third of respondents listing English as a second or other language [34]; it should be noted, however, that participants' English proficiency was sufficiently high to complete a written questionnaire in English. The study found that the internet was the most used resource by the NESB women in the study when seeking pregnancy-related information, although discussion with a midwife was deemed the most useful source of information [34]. While not reporting in any systematic way on app use, the study's findings indicate that CALD women are receptive and *au fait* with current technology. Evidence in favor of the acceptability and effectiveness of technology for increasing the knowledge in CALD groups has also been provided in a study testing an English language Web-based intervention in an Australian health service, to increase the knowledge of gestational diabetes mellitus in a multiethnic and low health literacy population [35]. Participants' knowledge scores increased across 3 domains from the pretest to the posttest. In the US context, a number of studies have indicated high ownership of smartphones and mobile phones in ethnic or racial minority groups and a preference for accessing Web-based content using these devices [36-39].

Nevertheless, others make the case that socially disadvantaged persons—including CALD people, especially those who have low English proficiency [40]—are further disadvantaged by digitization of health [16]. Some research suggests that pregnancy app engagement may be lower in areas of higher social disadvantage owing to factors such as less access to the internet and smartphones [41], as well as cultural and language barriers to app referral and content [42]. The existing research is inconclusive regarding whether, or to what extent, factors related to social disadvantage influence pregnancy app uptake, use, and utility. For instance, in the Australian study cited above

[34], while NESB women mostly used the internet to source pregnancy information, it was also found that the internet use was significantly higher among those who had tertiary education. Conversely, a Dublin study informed by survey data from a large representative sample of pregnant women (n=522) collected in 2012-13, found that there was extremely high general internet usage *across all social strata* (97%), high smartphone ownership (61%), and widespread usage of pregnancy apps among smartphone owners (48%) [28]. A limitation of this study is that it does not provide information about CALD or literacy status, and it must be assumed that respondents had sufficient literacy skills and English proficiency to fill out a written survey in English. Similarly, a 2015 US study examining the willingness of pregnant women to engage with computer- or mobile phone-delivered weight-loss interventions reported a high general use of and access to the internet via a computer or mobile phone, although there was slightly less access among nonwhite women and women who already had children [43]. Around half of the 100 women surveyed (which comprised 61% white, 26% black, 6% Hispanic, and 7% Asian women) were willing to participate in a mobile-based internet intervention (49%), while 83% were willing to participate in a computer-based intervention. This study's findings indicate that physical access to technology, then, may not be a significant limiting factor for marginalized groups in and of itself.

However, a person's ability to successfully engage with and retrieve required information from a resource may be a strong barrier. Baum et al [16] have demonstrated how fundamental literacy and health literacy issues can interact with digital literacy to exclude already marginalized groups from successfully taking up digital technologies, creating a "vicious cycle of disadvantage". The findings of Kraschnewski et al in their focus group study also support this contention [20]. The women in the study were of low socioeconomic status and all had smartphones and internet access; however, the women reported barriers to technology use, such as needing to conduct multiple searches to find the information they were seeking, and having concerns about receiving inaccurate information. It was suggested that these limitations might stem from a lower capacity to critically assess the source and accuracy of information. The authors expressed concern that the benefits of the internet and app-based tools may, thus, be limited by low electronic health (eHealth) literacy [20]. Low health literacy is proffered as a possible explanation for the presence of information "gaps" in women's knowledge in another study of predominantly white women, from a range of educational and income levels, who demonstrated active information-seeking behavior during pregnancy using both Web-based and other strategies [44]. Another study of low-income, predominantly black or Hispanic, American pregnant women and mothers of young children found low use of digital health management practices among almost three-quarters of their 92 study participants [45]. While around half of those who did not or rarely engaged with health-management practices indicated an interest in doing so in the future, several indicated a need for support or training to be able to manage existing digital resources. These findings highlight a key concern for CALD women, who are known to have lower health literacy levels than women in general [15,46,47].

A US study of pregnant women's use of information and communication technology by race and ethnicity ("white women," 23%; Latina women, 28%; African American women, 40%; and women of other ethnic groups, 9%) found that the vast majority owned mobile devices; however, those who had limited English proficiency (24%) and had not graduated from high school (24%), predominantly found in the Latina women group, showed lower usage patterns [48]. In addition, Latina women reported lower use of smartphones (55%), social networking sites (55%), and the internet (62%) than white or African American women; they also sought Web-based health information less (51%). The authors concluded that the uptake of Web-based apps was lower among low-income women, a group overrepresented in Latina and African American women, and that low English language proficiency and literacy were strong barriers to accessing the internet. They recommended that in these groups, alternatives such as paper or translations should be made available and interventions should be developed incorporating culturally and linguistically appropriate elements designed for people with low literacy in their first and second languages [48]. A recent New Zealand study on a short message service (SMS) text message-based maternal health information program reiterates these recommendations. In this study, the uptake and acceptability of a linguistically and culturally tailored program for Maori, Pacific, Asian, and South Asian families was high, and findings confirmed that women had greater access to information and felt supported by the regular messages they received [49].

Finally, previous studies focusing on CALD groups have recommended modifying existing health care procedures to facilitate service access for CALD pregnant women [50-52]. It is likely that the same recommendations should apply for pregnancy apps. CALD women have described struggling with incompatibility between the Western biomedical approaches to maternity practices and those of their own culture. Stapleton et al found that participants in their study were concerned that "the norms of the Australian maternity culture devalued, or indeed disregarded the intergenerational knowledge they viewed as precious, and which they understood offered the best protection for themselves and their infants" [11], a finding that is mirrored elsewhere [53]. The only study identified investigating CALD usage of pregnancy apps [30] advocated pregnant women's involvement in the app design to increase the effectiveness and usability and incorporated culturally sensitive components such as translations into CALD patients' languages, pictures of CALD women, and food items familiar to CALD users. During the design phase, participants indicated comprehension difficulties with the app content, which led the researchers to revise the language to make it more understandable. This app was yet to be evaluated, but initial prototype testing indicated participants' acceptance of the resource.

In this section, we have examined studies that present varying results regarding the uptake of digital resources by CALD women and also regarding the appropriateness or usefulness of these resources for CALD groups. It is possible that the extent to which language and culture create barriers for individual

CALD groups or people accounts for these divergent results. It seems reasonable that studies excluding non-English speakers will encounter fewer barriers to resource uptake than those with a representative proportion of participants with no or low English proficiency. This idea warrants further examination. Furthermore, it makes sense that CALD groups will be more inclined to accept and understand digital resources designed in consultation with them and using culturally sensitive methods.

Areas for Concern: Regulation and Validity and Reliability of Content

Previous studies have brought attention to the poor quality of some health-related apps [54,55] and even their potential to encourage harmful behaviors [56]. Although pregnancy apps offer many potential benefits, there are problems associated with their development and testing prior to release also. Researchers and health professionals have voiced concerns about the lack of regulation in the industry [1,57], as well as privacy and security issues [2,8,58]. In addition, pregnancy apps may provide inaccurate information [59], not include information that women are seeking [44], or have poor functionality [58]. Two separate studies examining the usefulness of pregnancy apps found that <6% (3.3% and 5.5%, respectively) were considered potentially useful by health care providers [1,59]. In a US study comparing 2 nationally endorsed apps [60], <20% of content explicitly addressed the recommended prenatal care content. In addition, examples of incomplete or confusing information were found, and the researchers reported significant gaps in the educational content, for example, postpartum contraception planning information was omitted, despite its inclusion in the American College of Obstetrics and Gynecology's guidelines for prenatal care.

Literature examining the perspectives of health professionals highlights other potential issues. A key concern for health professionals is the use of apps or other forms of eHealth as a replacement for members of the maternity health care team [24,61]. A number of other risks were identified by health professionals in qualitative studies of pregnancy mobile health interventions that included potential harm to "the personal or professional integrity of health professionals and health organizations (intellectual property, privacy, and legitimacy concerns)," the danger for misinterpretation if information was taken "out of context," and an undermining of professional legitimacy or control as women increase their reliance on health-related technology over health professionals "on the ground" [24]. In another study, challenges to professional legitimacy were a concern for health professionals in settings where professional and personal boundaries may become blurred, such as social media forums [62]. Finally, there is evidence that a self-perceived lack of technical skills can impinge on health professionals' willingness to engage with new technologies as a part of maternity care [62].

Patients may also be wary of the reliability of pregnancy apps or other eHealth content and have demonstrated concerns about personal data security [22,63]. A 2013 Australian study found that while the internet was the first-used source of information by most pregnant women, it was the least trusted [17]. In contrast, health care professionals and hospital print material were the most trusted sources. These are important concerns, particularly among certain groups of pregnant women. Pregnant adolescents, a high-risk group that is less likely to seek health professional advice and care, have been found to prefer apps as their information source for pregnancy education [64].

A lack of health professional guidance for women in their information seeking is an additional concern. Younger women with less experience of pregnancy [19] and first-time mothers [57] tend to use apps more but are particularly vulnerable to noncredible information sources as they are often active information seekers but less likely to know what to expect in pregnancy and childbirth. A German study found a statistically significant association between smartphone app use and younger maternal age, first pregnancy, lower self-rated health, and "influenceability" [65]. Another study found that young first-time mothers exhibited increased levels of anxiety and tension as a result of their information seeking [57]. Other studies have described how women's Web-based researching can turn into a stressful exercise where women may find it difficult to know when to stop searching [21], where unnecessary worry can be created from exposure to "horror stories" and "scare mongering" [66] and the net result can be more questions or confusion rather than answers [29]. These findings highlight a need for health professionals to have high levels of awareness of available information and resources, which has been demonstrated not to be the case in previous studies [62].

A Need for Expert, Reliable Advice From Local Sources

Pregnant women have expressed a need to be able to ascertain the credibility of information in apps. In one Australian study, some specified wanting "expert, credible, up-to-date advice," and others noted that they would like more "Australian-specific or locally based information" [8]. A 2013 Australian study found that women wanted "clinically endorsed" apps linked to trustworthy websites [23]. Another study found that women want apps and other digital media to give them "ready and instantaneous" access to expert professional information via real-time services such as Web-based messaging or video programs, for example, Skype [6]. The highest response rate for perceived weaknesses in pregnancy apps by respondents in a Korean study [19] was the "lack of credibility" (39%). Almost half of the participants (45%) expressed a need for expert opinions and opportunities for question-and-answer sessions on diet and medication administration during pregnancy. In the same study, the app evaluation component determined that disclosure of information sources in pregnancy apps scored lowest ("lower than average") on a credibility scale, signally the potentially low trustworthiness of pregnancy app information in general.

Currently, most app developers are commercial entities or internet portals [19]. Health professional and institution involvement in the creation or endorsement of digital information sources is very low [20], although improving [58]. In this scenario, the user becomes a commodity with marketable data, which can be sold to data-mining companies [7], raising concerns about the use, governance, and confidentiality of such data. Pregnancy app users are aware of this and indicate that

given the choice, they would prefer apps that are "not linked to the manufacturers of pregnancy or baby products" [8].

As well as calls for the creation of a regulatory framework [2], researchers have recommended that health professionals and health institutions take the initiative to provide both reliable information on maternal health through such mechanisms as apps, or advice about the reliability or validity of other available electronic information [2,57], a desire echoed by patients [23,24]. Health professional engagement with popular modes of information delivery is essential to ensure health services for women are responsive to the changes in information-seeking behaviors and attract rather than alienate women [61]. One reason women use digital media to find information is that they are not able to spend sufficient time with health professionals to have all their questions answered. It is, therefore, imperative that pregnant women can be confident that the information they are accessing is accurate. Johnson [67] has presented a compelling argument that pregnancy apps are effectively performative devices, exerting a strong influence on conceptions and enactments of motherhood in today's society, and the encouraging the "responsibilization" of health among users. Taking this perspective, the important role of health professionals and health institutions in the design and provision of pregnancy apps is clear and shifts the primary agenda from material gain for commercial entities to evidence-based, expert-endorsed health education for pregnant women.

Discussion

This review indicates that pregnancy apps are here to stay and confirms that women are increasingly using apps when seeking pregnancy-related information. The findings suggest that to ensure the uptake from women, future pregnancy app resources will need to include not only health services-endorsed information but also incorporate the Web-based and personalizable features that appeal to women currently using pregnancy apps. We have found substantial gaps in multilingual digital resources for the CALD population in Australia and elsewhere, as well as evidence of decreased uptake because of widespread lower literacy and health literacy status of this and other hard-to-reach populations. One of the very appealing aspects of digital and especially mobile technologies is to improve patient care for "unreachable populations" by overcoming the limitations imposed by cost and access [58]. However, this review has shown that there is conjecture regarding the extent to which such marginalized populations benefit from digital resources.

The provision of digital technologies such as apps is fast becoming standard practice in health services, but steps must be taken to ensure that these resources are fit for purpose. The studies discussed in this review have identified that addressing literacy and language barriers may be key components in ensuring the uptake of digital technologies such as apps among CALD and possibly other disadvantaged groups. Research exploring the relative effectiveness of different means of communicating information through the type of multimedia interface that an app offers, for example, graphs, pictures, audio, video, SMS text message lists, and more or less detailed textual information, is recommended to address knowledge gaps in this area. Consideration should be given to include different types of users, including those with higher or lower education levels and literacy, as well as to address cultural, social, and educational barriers to effective use of such resources. Addressing cultural components will be crucial to enhance the appeal and usability of resources for CALD women and will require sensitivity and awareness to ensure appropriateness. We note that one review of currently available apps comments on the ubiquity of white-skinned women and white or pink babies in depictions of pregnant women and their newborn infants [7], a feature that will need modification for such pregnancy apps to be culturally inclusive.

This review has highlighted the lack of reliable information provided by a high proportion of currently available apps—both in terms of content and functionality—as well as a clear demand for trustworthy, locally based, and professionally endorsed pregnancy information resources. It is essential to be able to ensure that information received by high-risk groups, such as adolescent and first-time mothers, is accurate and appropriate for their needs. The possible repercussions of not acting to ensure that the app content is reliable are potentially serious. This problem can be addressed through greater health professional engagement with and oversight of pregnancy app resources from a design perspective. It is important to mention that determining what information qualifies for inclusion in pregnancy information resources is not a straightforward matter. There are diverse views about how much information women should be given about pregnancy-related problems, for example, risks that may or may not become eventualities in individual cases [68]. This is an aspect of pregnancy app design that will require extensive consideration and necessitates health professional input, if not consensus.

In addition, it is suggested that health professionals guide women in their use of apps during pregnancy and that professionally endorsed apps be used adjunctively by health professionals as part of their maternity health care protocol. To ensure the accessibility and uptake of reliable app-based pregnancy information, the provision of free hospital-endorsed apps with complementary resources, such as supportive training and education to use them [16], is also recommended. To help achieve these objectives, more research into the effectiveness of apps is also required. Consideration of the issues raised in this review has markedly contributed to a pregnancy app resource developed by the authors [69], and it is hoped that similar consideration is given to the development, design, and testing of any pregnancy-related app resource in the future.

Acknowledgments

The authors wish to acknowledge funding from a Macedon Ranges North Western Melbourne Medicare Local Grant. We would also like to thank Evelyn Hutcheon at the Western Health Library Service for her valuable assistance.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample search strategy (MEDLINE search for type 1 studies).

[PNG File, 43KB - mhealth_v6i11e189_app1.png]

Multimedia Appendix 2

Papers examining women's use and preferred features in pregnancy apps, and issues in pregnancy apps use.

[PDF File (Adobe PDF File), 58KB - mhealth_v6i11e189_app2.pdf]

Multimedia Appendix 3

Papers including culturally and linguistically diverse women's use of digital technology during pregnancy.

[PDF File (Adobe PDF File), 66KB - mhealth_v6i11e189_app3.pdf]

References

1. 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]
2. Bert F, Passi S, Scaioli G, Gualano MR, Siliquini R. There comes a baby! What should I do? Smartphones' pregnancy-related applications: A web-based overview. Health Informatics J 2016 Sep;22(3):608-617. [doi: 10.1177/1460458215574120] [Medline: 25900813]
3. BabyCentre Internet. The rise of the mobile mum URL: http://www.babycentre.co.uk/a1052627/the-rise-of-the-mobile-mumA [WebCite Cache ID 6tvRCOcrl]
4. Maio Mackay M. AIMIA. Australian Mobile Phone Lifestyle Index 2014 URL: http://www.dtcollective.org.au/LiteratureRetrieve.aspx?ID=180479 [accessed 2017-10-03] [WebCite Cache ID 6tvbYul1O]
5. Poushter J. Pew Research Center. 2016 Feb 22. Smartphone ownership and Internet usage continues to climb in emerging economies URL: https://tinyurl.com/hpumdoz [WebCite Cache ID 6tvPjGtxs]
6. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
7. Lupton D, Thomas GM. Playing pregnancy: the ludification and gamification of expectant motherhood in smartphone apps. M/C Journal 2015;18(5) [FREE Full text]
8. Lupton D, Pedersen S. "What is happening with your body and your baby?": Australian women's use of pregnancy and parenting apps. Canberra: News & Media Research Centre, University of Canberra; 2015.
9. Hether HJ, Murphy ST, Valente TW. It's better to give than to receive: the role of social support, trust, and participation on health-related social networking sites. J Health Commun 2014 Dec;19(12):1424-1439. [doi: 10.1080/10810730.2014.894596] [Medline: 24766297]
10. Holtz B, Smock A, Reyes-Gastelum D. Connected Motherhood: Social Support for Moms and Moms-to-Be on Facebook. Telemed J E Health 2015 May;21(5):415-421. [doi: 10.1089/tmj.2014.0118] [Medline: 25665177]
11. Stapleton H, Murphy R, Correa-Velez I, Steel M, Kildea S. Women from refugee backgrounds and their experiences of attending a specialist antenatal clinic. Narratives from an Australian setting. Women Birth 2013 Dec;26(4):260-266. [doi: 10.1016/j.wombi.2013.07.004] [Medline: 23988383]
12. Floyd A, Sakellariou D. Healthcare access for refugee women with limited literacy: layers of disadvantage. Int J Equity Health 2017 Nov 10;16(1):195 [FREE Full text] [doi: 10.1186/s12939-017-0694-8] [Medline: 29126420]
13. An Investment Not an Expense: Enhancing Health Literacy in CALD Communities. Melbourne: Ethnic Communities Council of Victoria; 2012.
14. Hennegan J, Redshaw M, Kruske S. Another country, another language and a new baby: A quantitative study of the postnatal experiences of migrant women in Australia. Women Birth 2015 Dec;28(4):e124-e133. [doi: 10.1016/j.wombi.2015.07.001] [Medline: 26254551]
15. Almeida LM, Caldas J, Ayres-de-Campos D, Salcedo-Barrientos D, Dias S. Maternal healthcare in migrants: a systematic review. Matern Child Health J 2013 Oct;17(8):1346-1354. [doi: 10.1007/s10995-012-1149-x] [Medline: 23334357]

16. Baum F, Newman L, Biedrzycki K. Vicious cycles: digital technologies and determinants of health in Australia. Health Promot Int 2014 Jun;29(2):349-360. [doi: 10.1093/heapro/das062] [Medline: 23144236]
17. Rodger D, Skuse A, Wilmore M, Humphreys S, Dalton J, Flabouris M, et al. Pregnant women's use of information and communications technologies to access pregnancy-related health information in South Australia. Aust J Prim Health 2013;19(4):308-312. [doi: 10.1071/PY13029] [Medline: 24004661]
18. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
19. Lee Y, Moon M. Utilization and Content Evaluation of Mobile Applications for Pregnancy, Birth, and Child Care. Healthc Inform Res 2016 Apr;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
20. Kraschnewski JL, Chuang CH, Poole ES, Peyton T, Blubaugh I, Pauli J, et al. Paging "Dr. Google": does technology fill the gap created by the prenatal care visit structure? Qualitative focus group study with pregnant women. J Med Internet Res 2014;16(6):e147 [FREE Full text] [doi: 10.2196/jmir.3385] [Medline: 24892583]
21. Prescott J, Mackie L. "You Sort of Go Down a Rabbit Hole...You're Just Going to Keep on Searching": A Qualitative Study of Searching Online for Pregnancy-Related Information During Pregnancy. J Med Internet Res 2017 Jun 05;19(6):e194 [FREE Full text] [doi: 10.2196/jmir.6302] [Medline: 28583906]
22. Goetz M, Müller M, Matthies LM, Hansen J, Doster A, Szabo A, et al. Perceptions of Patient Engagement Applications During Pregnancy: A Qualitative Assessment of the Patient's Perspective. JMIR Mhealth Uhealth 2017 May 26;5(5):e73 [FREE Full text] [doi: 10.2196/mhealth.7040] [Medline: 28550005]
23. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [doi: 10.1071/PY13039] [Medline: 23899373]
24. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of Women and Health Professionals on mHealth Lifestyle Interventions in Pregnancy: A Qualitative Investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
25. Wierckx A, Shahid S, Al MA. Babywijzer: An Application to Support Women During Their Pregnancy. In: Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems Internet. New York, NY. USA: ACM; 2014 Presented at: ACM CHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto. [doi: 10.1145/2559206.2581179]
26. Newhouse N. "My Facebook is a bit of a multiple personality at the minute": social media and the transition to new motherhood. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction. New York: ACM; 2016 Presented at: NORDICHI'16; October 23-27, 2016; Gothenburg, Sweden URL: https://tinyurl.com/yd63zode
27. Fredriksen EH, Harris J, Moland KM. Web-based Discussion Forums on Pregnancy Complaints and Maternal Health Literacy in Norway: A Qualitative Study. J Med Internet Res 2016 May 26;18(5):e113 [FREE Full text] [doi: 10.2196/jmir.5270] [Medline: 27230094]
28. O'Higgins A, Murphy OC, Egan A, Mullaney L, Sheehan S, Turner MJ. The use of digital media by women using the maternity services in a developed country. Ir Med J 2014;107(10):313-315. [Medline: 25556255]
29. Peyton T, Poole E, Reddy M, Kraschnewski J, Chuang C. Every pregnancy is different: Designing mHealth interventions for the pregnancy ecology. In: Proceedings of the 2014 Conference on Designing interactive systems. New York: ACM; 2014 Presented at: ACM Conference on Designing Interactive Systems (DIS 2014); June 21, 2014; Vancouver, BC p. 577-586. [doi: 10.1145/2598510.2598572]
30. Garnweidner-Holme LM, Borgen I, Garitano I, Noll J, Lukasse M. Designing and Developing a Mobile Smartphone Application for Women with Gestational Diabetes Mellitus Followed-Up at Diabetes Outpatient Clinics in Norway. Healthcare (Basel) 2015 May 21;3(2):310-323 [FREE Full text] [doi: 10.3390/healthcare3020310] [Medline: 27417764]
31. Australian Institute of Health and Welfare (AIHW). Australia's Health 2016. Chapter 5 Health of population groups URL: https://www.aihw.gov.au:80/reports/aus/199/australias-health-2016/contents/population-groups [accessed 2018-10-01] [WebCite Cache ID 72qGO8HaI]
32. UniversalDoctor. 2013 Apr 13. UniversalWomen Speaker. Multilingual eHealth tool tailored for pregnancy, childbirth and maternal health Internet URL: https://tinyurl.com/y6u6e9su [WebCite Cache ID 6tvPjGty8]
33. Borsari L, Stancanelli G, Guarenti L, Grandi T, Leotta S, Barcellini L, et al. An Innovative Mobile Health System to Improve and Standardize Antenatal Care Among Underserved Communities: A Feasibility Study in an Italian Hosting Center for Asylum Seekers. J Immigr Minor Health 2017 Nov 16. [doi: 10.1007/s10903-017-0669-2] [Medline: 29143900]
34. Grimes HA, Forster DA, Newton MS. Sources of information used by women during pregnancy to meet their information needs. Midwifery 2014 Jan;30(1):e26-e33. [doi: 10.1016/j.midw.2013.10.007] [Medline: 24246969]
35. Carolan-Olah M, Steele C, Krenzin G. Development and initial testing of a GDM information website for multi-ethnic women with GDM. BMC Pregnancy Childbirth 2015 Jul 05;15:145 [FREE Full text] [doi: 10.1186/s12884-015-0578-0] [Medline: 26142482]
36. Bensley RJ, Hovis A, Horton KD, Loyo JJ, Bensley KM, Phillips D, et al. Accessibility and preferred use of online Web applications among WIC participants with Internet access. J Nutr Educ Behav 2014;46(3 Suppl):S87-S92. [doi: 10.1016/j.jneb.2014.02.007] [Medline: 24810002]

37. Brusk JJ, Bensley RJ. A Comparison of Mobile and Fixed Device Access on User Engagement Associated With Women, Infants, and Children (WIC) Online Nutrition Education. JMIR Res Protoc 2016 Nov 15;5(4):e216 [FREE Full text] [doi: 10.2196/resprot.6608] [Medline: 27847351]
38. Narasimhulu DM, Karakash S, Weedon J, Minkoff H. Patterns of Internet Use by Pregnant Women, and Reliability of Pregnancy-Related Searches. Matern Child Health J 2016 Dec;20(12):2502-2509. [doi: 10.1007/s10995-016-2075-0] [Medline: 27456311]
39. Waring ME, Moore STA, Xiao RS, Lombardini LM, Allison JJ, Rosal MC, et al. Pregnant women's interest in a website or mobile application for healthy gestational weight gain. Sex Reprod Healthc 2014 Dec;5(4):182-184. [doi: 10.1016/j.srhc.2014.05.002] [Medline: 25433828]
40. Commonwealth of Australia. Parliament of Australia. 2014. Bridging our growing divide: inequality in Australia URL: http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/Income_Inequality/Report [WebCite Cache ID 6tvPjGtyD]
41. Osma J, Barrera AZ, Ramphos E. Are Pregnant and Postpartum Women Interested in Health-Related Apps? Implications for the Prevention of Perinatal Depression. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):412-415. [doi: 10.1089/cyber.2015.0549] [Medline: 27327069]
42. Hearn L, Miller M, Lester L. Reaching perinatal women online: the Healthy You, Healthy Baby website and app. J Obes 2014;2014:573928 [FREE Full text] [doi: 10.1155/2014/573928] [Medline: 24872891]
43. Peragallo UR, Berger AA, Ivins AA, Beckham AJ, Thorp JM, Nicholson WK. Internet Use and Access Among Pregnant Women via Computer and Mobile Phone: Implications for Delivery of Perinatal Care. JMIR Mhealth Uhealth 2015 Mar 30;3(1):e25 [FREE Full text] [doi: 10.2196/mhealth.3347] [Medline: 25835744]
44. Guerra-Reyes L, Christie VM, Prabhakar A, Harris AL, Siek KA. Postpartum Health Information Seeking Using Mobile Phones: Experiences of Low-Income Mothers. Matern Child Health J 2016 Dec;20(Suppl 1):13-21 [FREE Full text] [doi: 10.1007/s10995-016-2185-8] [Medline: 27639571]
45. Guendelman S, Broderick A, Mlo H, Gemmill A, Lindeman D. Listening to Communities: Mixed-Method Study of the Engagement of Disadvantaged Mothers and Pregnant Women With Digital Health Technologies. J Med Internet Res 2017 Jul 05;19(7):e240 [FREE Full text] [doi: 10.2196/jmir.7736] [Medline: 28679489]
46. Peláez S, Hendricks KN, Merry LA, Gagnon AJ. Challenges newly-arrived migrant women in Montreal face when needing maternity care: Health care professionals' perspectives. Global Health 2017 Jan 25;13(1):5. [doi: 10.1186/s12992-016-0229-x] [Medline: 28122630]
47. Lauria L, Bonciani M, Spinelli A, Grandolfo ME. Inequalities in maternal care in Italy: the role of socioeconomic and migrant status. Ann Ist Super Sanita 2013;49(2):209-218 [FREE Full text] [Medline: 23771265]
48. Chilukuri N, West M, Henderson JL, Lawson S, Ehsanipoor R, Costigan K, et al. Information and Communication Technology Use Among Low-Income Pregnant and Postpartum Women by Race and Ethnicity: A Cross-Sectional Study. J Med Internet Res 2015;17(7):e163 [FREE Full text] [doi: 10.2196/jmir.3916] [Medline: 26142162]
49. Dobson R, Whittaker R, Bartley H, Connor A, Chen R, Ross M, et al. Development of a Culturally Tailored Text Message Maternal Health Program: TextMATCH. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e49 [FREE Full text] [doi: 10.2196/mhealth.7205] [Medline: 28428159]
50. Riggs E, Davis E, Gibbs L, Block K, Szwarc J, Casey S, et al. Accessing maternal and child health services in Melbourne, Australia: reflections from refugee families and service providers. BMC Health Serv Res 2012 May 15;12:117 [FREE Full text] [doi: 10.1186/1472-6963-12-117] [Medline: 22587587]
51. Reavy K, Hobbs J, Hereford M, Crosby K. A new clinic model for refugee health care: adaptation of cultural safety. Rural Remote Health 2012;12:1826 [FREE Full text] [Medline: 22263874]
52. Owens C, Dandy J, Hancock P. Perceptions of pregnancy experiences when using a community-based antenatal service: A qualitative study of refugee and migrant women in Perth, Western Australia. Women Birth 2016 Apr;29(2):128-137. [doi: 10.1016/j.wombi.2015.09.003] [Medline: 26410632]
53. Higginbottom GMA, Safipour J, Mumtaz Z, Chiu Y, Paton P, Pillay J. "I have to do what I believe": Sudanese women's beliefs and resistance to hegemonic practices at home and during experiences of maternity care in Canada. BMC Pregnancy Childbirth 2013 Feb 25;13:51 [FREE Full text] [doi: 10.1186/1471-2393-13-51] [Medline: 23442448]
54. Anthes G. Invasion of the mobile apps. Commun. ACM 2011 Sep 01;54(9):16. [doi: 10.1145/1995376.1995383]
55. Nasser F BinDhim, Trevena L. There's an App for That: A Guide for Healthcare Practitioners and Researchers on Smartphone Technology. Online J Public Health Inform 2015;7(2):e218 [FREE Full text] [doi: 10.5210/ojphi.v7i2.5522] [Medline: 26392848]
56. BinDhim NF, Trevena L. Health-related smartphone apps: regulations, safety, privacy and quality. BMJ Innov 2015 Mar 05;1(2):43-45. [doi: 10.1136/bmjinnov-2014-000019]
57. Fleming SE, Vandermause R, Shaw M. First-time mothers preparing for birthing in an electronic world: internet and mobile phone technology. Journal of Reproductive and Infant Psychology 2014 Mar 19;32(3):240-253. [doi: 10.1080/02646838.2014.886104]
58. Scott KM, Gome GA, Richards D, Caldwell PHY. How trustworthy are apps for maternal and child health? Health Technol 2015 Mar 12;4(4):329-336. [doi: 10.1007/s12553-015-0099-x]

59. Narasimhulu DM, Karakash S, Weedon J, Minkoff H. Patterns of Internet Use by Pregnant Women, and Reliability of Pregnancy-Related Searches. Matern Child Health J 2016 Dec;20(12):2502-2509. [doi: 10.1007/s10995-016-2075-0] [Medline: 27456311]
60. O'Donnell BE, Lewkowitz AK, Vargas JE, Zlatnik MG. Examining pregnancy-specific smartphone applications: what are patients being told? J Perinatol 2016 Oct;36(10):802-807. [doi: 10.1038/jp.2016.77] [Medline: 27195980]
61. Robinson F, Jones C. Women's engagement with mobile device applications in pregnancy and childbirth. Pract Midwife 2014 Jan;17(1):23-25. [Medline: 24520591]
62. Dalton JA, Rodger DL, Wilmore M, Skuse AJ, Humphreys S, Flabouris M, et al. "Who's afraid?": attitudes of midwives to the use of information and communication technologies (ICTs) for delivery of pregnancy-related health information. Women Birth 2014 Sep;27(3):168-173. [doi: 10.1016/j.wombi.2014.06.010] [Medline: 25070730]
63. Sundstrom B. Mothers "Google It Up": Extending Communication Channel Behavior in Diffusion of Innovations Theory. Health Commun 2016;31(1):91-101. [doi: 10.1080/10410236.2014.936339] [Medline: 26075413]
64. Ireson D. Edith Cowan University Research Online. Antenatal clinic: Using ethnographic methods to listen to the voices of pregnant adolescents URL: https://ro.ecu.edu.au/theses/1752/ [accessed 2018-10-01] [WebCite Cache ID 72qJXiVmh]
65. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using Web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Apr 15. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]
66. Lagan BM, Sinclair M, Kernohan WG. What is the impact of the Internet on decision-making in pregnancy? A global study. Birth 2011 Dec;38(4):336-345. [doi: 10.1111/j.1523-536X.2011.00488.x] [Medline: 22112334]
67. Johnson, SA. Maternal devices, social media and the self-management of pregnancy, mothering and child health. Societies 2014;4(2):330-350.
68. Callaghan S, Corderoy A. The Guardian. 2017 Sep 27. Women have the right to know about injuries of vaginal birth beforehand URL: https://tinyurl.com/yaeolbjo [WebCite Cache ID 6tvPjGtyN]
69. Smith W, Wadley G, Daly O, Webb M, Hughson J, Hajek J, et al. Designing an app for pregnancy care for a culturally and linguistically diverse community. In: Proceedings of the Australian Computer-Human Interaction Conference. New York: ACM; 2017 Presented at: OzCHI '17 29th Australian Conference on Human-Computer Interaction; November 29 - December 1, 2017; Brisbane, Australia p. 337-346. [doi: 10.1145/3152771.3152807]

Abbreviations

CALD: culturally and linguistically diverse
eHealth: electronic health
NESB: non-English-speaking background
SMS: short message service

Edited by G Eysenbach; submitted 05.10.17; peer-reviewed by S Guendelman, T Tamrat, M Wallwiener; comments to author 23.11.17; revised version received 10.02.18; accepted 18.07.18; published 16.11.18.

Please cite as:
Hughson JAP, Daly JO, Woodward-Kron R, Hajek J, Story D
The Rise of Pregnancy Apps and the Implications for Culturally and Linguistically Diverse Women: Narrative Review
JMIR Mhealth Uhealth 2018;6(11):e189
URL: https://mhealth.jmir.org/2018/11/e189/
doi:10.2196/mhealth.9119
PMID:30446483

Original Paper

Simulated Clinical Encounters Using Patient-Operated mHealth: Experimental Study to Investigate Patient-Provider Communication

Harry Tunnell[1], PhD; Anthony Faiola[2], PhD; Davide Bolchini[1], PhD; Rebecca Bartlett Ellis[3], RN, PhD

[1]School of Informatics and Computing, Department of Human-Centered Computing, Indiana University, Indianapolis, IN, United States

[2]Department of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL, United States

[3]School of Nursing, Indiana University, Indianapolis, IN, United States

Corresponding Author:
Harry Tunnell, PhD
School of Informatics and Computing
Department of Human-Centered Computing
Indiana University
535 West Michigan Street, IT 475
Indianapolis, IN, 46202
United States
Phone: 1 317 278 4636
Email: hadtunne@iupui.edu

Abstract

Background: This study investigates patient-centered mobile health (mHealth) technology in terms of the secondary user experience (UX). Specifically, it examines how personal mobile technology, under patient control, can be used to improve patient-provider communication about the patient's health care during their first visit to a provider. Common ground, a theory about language use, is used as the theoretical basis to examine interactions. A novel concept of this study is that it is one of the first empirical studies to explore the relative meaningfulness of a secondary UX for specific health care tasks.

Objective: The objective of this study was to investigate the extent that patient-operated mHealth technology can be designed to improve the communication between the patient and provider during an initial face-to-face encounter.

Methods: The experimental study was conducted in 2 large Midwestern cities from February 2016 to May 2016. A custom-designed smartphone app prototype was used as the study treatment. The experimental design was posttest-only control group and included video-recorded simulated face-to-face clinical encounters in which an actor role-played a patient. Experienced clinicians consisting of doctors (n=4) and nurses (n=8) were the study participants. A thematic analysis of qualitative data was performed. Quantitative data collected from time on task measurements were analyzed using descriptive statistics.

Results: Three themes that represent how grounding manifested during the encounter, what it meant for communication during the encounter, and how it influenced the provider's perception of the patient emerged from the qualitative analysis. The descriptive statistics were important for inferring evidence of efficiency and effectiveness of communication for providers. Overall, encounter and task times averaged slightly faster in almost every instance for the treatment group than that in the control group. Common ground clearly was better in the treatment group, indicating that the idea of designing for the secondary UX to improve provider outcomes has merit.

Conclusions: Combining the notions of common ground, human-computer interaction design, and smartphone technology resulted in a prototype that improved the efficiency and effectiveness of face-to-face collaboration for secondary users. The experimental study is one of the first studies to demonstrate that an investment in the secondary UX for high payoff tasks has value but that not all secondary UXs are meaningful for design. This observation is useful for prioritizing how resources should be applied when considering the secondary UX.

(JMIR Mhealth Uhealth 2018;6(11):e11131) doi:10.2196/11131

KEYWORDS

medical informatics; personal health record; medication reconciliation

Introduction

Statement of the Problem

The ubiquity of smartphones gives patients access to many apps that can potentially support health care needs. These types of technologies allow patients to track and trend various health behaviors that they may subsequently want to share with health care providers. In this context, health care providers become what is known as *secondary users*, a term from the human-computer interaction (HCI) discipline that reflects a type of user who is affected by the main or primary user's operation (ie, the patient) of a technology. [1-3]. Primary users, on the other hand, are the dominant operators that control the system and the dissemination of its information and are patients in this context [1]. Some researchers believe that the secondary user experience (UX) has the potential to improve collaboration and satisfaction in a variety of settings and, as such, advocate for its inclusion in interactive systems design [4-6]. However, little is known about the experience of the secondary user, especially in the context of health care provider encounters.

Recently, it has become common for more people to access the internet through mobile devices than personal computers [7]. The advent of always-internet-connected mobile technologies portends a wider set of UXs than previously envisioned. The lack of any practical geographic or temporal restrictions on the use of some types of patient-centered health information technology (HIT), such as a smartphone personal health record (PHR) app, is a recent phenomenon that impacts users and their experiences. Anytime someone acts with an interactive system through an interface in public, he potentially creates a UX for others; this is the secondary UX. With more than 165 million smartphone users in the United States as of 2014, secondary UXs have the potential to become routine [8].

The ubiquity of smartphones gives patients access to technologies that can support their health care needs in ways previously unavailable to lay people. Pew Research reported that within a 1-year period, 62% of US smartphone owners used their devices to look up health information [9]. The ability of smartphones to store and manage information allows patients to track and trend health data and share these data with health care providers in a manner that informs clinical decision making. It is in this context that we believe health care providers are interesting as secondary users of patient-controlled devices. In light of this assessment, we decided to investigate patient-centered mobile health (mHealth) technology in terms of the secondary UX. Specifically, this study examines how personal mobile technology, under patient control, can be used to improve patient-provider communication about the patient's health care during their first visit to a provider.

A novel concept of this study is that it is one of the first empirical studies to explore the relative meaningfulness of a secondary UX for a specific task. Not all UXs are likely to have value during task-oriented communication. In health care, patient health often relies on successful collaborations between patients and providers [10]. Because there are often common tasks for specific types of encounter, for example, a first face-to-face encounter between a provider and new patient, such tasks can be anticipated and designed for during a user-centered technology design process. However, before designing technology interventions for such tasks, it is important to understand which tasks benefit most from a technology solution. This study, through its emphasis on meaningfulness, provides a model for evaluating secondary UX value for a communication.

Secondary users have been identified in the medical informatics literature but operationalized in terms of patients. This study introduces the opposite phenomenon and looks at secondary users as providers. It is in this context that we explore how a patient as a primary user can manage technology in a way that improves outcomes for providers who are secondary users.

To meaningfully improve HCI design for secondary users, we must take into consideration the dynamic interplay between both types of users. This study does this by investigating secondary UXs according to common ground, a communication theory about language use. Overall, 2 assumptions underpinning this research are that (1) the creation of common ground is a key element of the secondary UX that contributes to improved communication and (2) interfaces for interactive systems can be designed to facilitate the creation of this key element.

The user type most often researched is the primary user, and investigations about secondary users are limited [11]. As the HCI literature about secondary users is sparse, technology is most often understood from the perspective of one user type. In fact, Inbar and Tractinsky [12] reported that secondary users are missing from both the theoretical and practical perspectives in HCI. Consequently, there is little research that exists to empirically demonstrate why secondary users should be included as a consideration in the design of technology. The currently available research merely indicates that secondary UXs exist and that secondary users are a relevant stakeholder group [3,11,13]. Health care, and the increasing emphasis on patient-centered technology, is an especially vital context in which to study the secondary UX because PHR apps are an example of HIT where one person's use creates UXs for others.

Another shortcoming in the HCI literature about secondary users is the lack of a theoretical basis to explain why secondary users are relevant and should be considered in technology design. We believe that common ground can fill this void. Common ground, a component of effective collaboration, is established when people have certain knowledge in common and know that they have this knowledge in common [14]. Our study revealed a gap in health care knowledge between patients and providers that can potentially be bridged with the aid of technology [15-17]. Such a knowledge disparity is not surprising, as patients are lay people and providers are experts regarding health care, which can make the attainment of common ground between them elusive during an encounter. The short duration of encounters can further constrain communication and information sharing, also making common ground difficult to obtain.

Although improving patient access to health information leads to increased participation of patients in health-related decision making, it is not merely the access to data that creates this impact [18]. The authors believe that it is an improvement in

common ground between patients and their providers that makes superior patient engagement possible. Technology designs that reduce ambiguity in communication between patients and providers have the potential to improve common ground.

Theoretical Model

Common ground theory provides the theoretical model for this study. Common ground is defined as "a proposition *p* is common ground if: all the people conversing know *p* and they all know they know *p* " [19]. The theory was developed by Clark [20] and explains how people achieve sufficient shared knowledge to successfully complete a communication. *Grounding* is the process to make communication effective and common ground is created because of the grounding process [21].

During face-to-face meetings, the ability to share content with someone can be limited by the lack of interactivity of typical physical tools (eg, pen and paper) used by team members [22]. One of the attractions of common ground for secondary UX research is the idea of external representation. External representation is a way to represent components of the communication in physical form. Clark [20] provided the example of a chess game with the board and pieces serving as external representations of the players. The position of the players on the board unambiguously shows the current state of the game.

In HCI, a smartphone interface that displays objects of interest to primary and secondary users is an external representation. Using the chess example once more, the board and chess pieces are digitally represented in the interface, again unambiguously showing the current state of the game. Direct manipulation interfaces (eg, graphical user interfaces) have provided excellent support for grounding for many years because of their capacity to continuously represent objects of interest and provide feedback about the effect of actions [23]. Thus, the relation between computer interfaces and grounding is already established and does not require additional explication for this study.

Constraints on Grounding

Actions are important components of common ground. A joint action occurs when people intend to do their parts in the communication and believe that the joint action includes their parts and the parts of the other participants [19]. Common ground is incrementally built based on the history of joint actions [24]. When joint actions are mediated by interactive systems, the technology places constraints on the establishment of common ground. Constraints are often considered as a negative attribute. However, in this context, constraints are *positive* for grounding because they reduce ambiguity [19]. The more constraints supported by a technology because of different combinations of devices and interfaces, the better [24]. There are 8 constraints for grounding (Table 1).

Constraints can be used to predict the problems that people will have with an information system by evaluating which constraints are present or absent when using the system [19]. The concept of constraints means that it is possible to anticipate what the UX will be with a product for a particular user type. As the secondary UX can be anticipated through the evaluation of the constraints, it can be designed for during product development. The ability to predict the experience means that common ground theory can be used to explain problems that people have with an information system in certain contexts [25].

Constraints and Smartphones

Unlike most communication technologies, smartphones can support all constraints in a face-to-face setting. Table 2 shows comparisons of communication mediums and constraints (X is a supported constraint). The fact that smartphones allow users to switch back and forth between functionalities (eg, email or short message service [SMS] text messaging) seamlessly is an exciting prospect and important for grounding. For example, smartphone users do not have to find and go to a fixed system or workstation to send and receive nontelephonic messages because they have smartphone apps that perform several communication functions.

Table 1. Constraints on grounding.

Constraint[a]	Description
Copresence	When A and B are colocated, such as in the same room
Visibility	When A and B can see each other
Audibility	When A and B talk to each other
Contemporality	When B receives messages at about the same time that A produces them
Simultaneity	When A and B can send and receive messages simultaneously
Sequentiality	When A's turn and B's turn cannot occur out of sequence
Reviewability	When B can review A's messages, as in written communication
Revisability	When A can revise messages for B before they are seen by others (outside of A and B)

[a]Adapted from Grounding in Communication by Clark and Brennan [21].

Table 2. Constraints on communication comparison chart. X refers to a supported constraint.

Medium	Copresence	Visibility	Audibility	Contemporality	Simultaneity	Sequentiality	Reviewability	Revisability
Face-to-face	X	X	X	X	X	X		
Telephone			X	X	X	X		
Video tele-conference		X	X	X	X	X		
Letters							X	X
Email or text							X	X
Mobile phone[a]	X	X	X	X	X	X	X	X

[a]Mobile phone as a medium added to table. Table adapted from Grounding in Communication by Clark and Brennan [21].

Table 3. Collaboration mechanics.

Category[a]	Description (mechanics)
Explicit communication	Planned and intentional communication (speaking, writing, gesturing, combining verbal and gestural, and manifesting actions)
Information gathering	Gathering information in shared workspaces from others and their activities (basic group awareness, feedthrough, consequential communication, visual evidence, and overhearing explicit communications)
Management of shared access	Managing group access to objects within the workspace (obtaining a resource, reserving a resource, and protecting work)
Transfer	The movement of objects and tools between people (handoff and deposit)

[a]Adapted from Task Analysis for Groupware Usability Evaluation: Modeling Shared-Workspace Tasks with the Mechanics of Collaboration, by Pinelle et al [27].

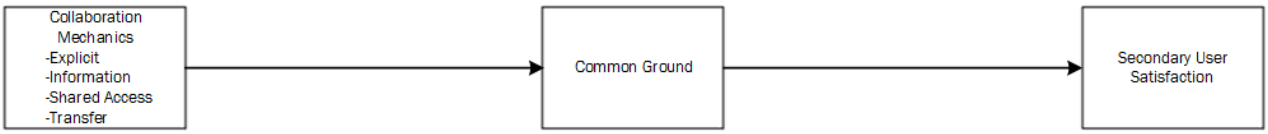
The potential to improve how language is used during the patient-provider encounter and thus increase the efficiency of communication between members of the dyad, makes the notion of common ground salient when examining primary and secondary user collaborations. Research suggests that well-designed collaborative technologies speed up the development of common ground by allowing teams to share knowledge, manage actions, and make decisions efficiently [26]. Fortunately, a model to define working with technology in the context of common ground exists. Collaboration mechanics (Table 3), the development of which was influenced by Clark [20] and his work on common ground theory [27], can be used to model collaboration with technology. Therefore, the relations among collaboration mechanics, common ground, and secondary user satisfaction are examined in the experimental study.

Research Questions

Although the HCI literature acknowledges that secondary UXs can occur, it does a poor job of revealing how to evaluate the importance of them for an interaction. It is the authors' contention that the value of secondary UXs in a unique context should be identified to assess its merit for design. Once this is understood, it becomes possible to make an informed decision about how a secondary UX may or may not improve communication. Common ground provides the vehicle in this research to evaluate the quality and importance of a secondary UX during a clinical encounter. The research questions emphasize how common ground is used to demonstrate the relation of language to digital systems, in a manner that results in better communication in an envisioned health care setting. The conceptual model for the research questions is shown in Figure 1. The research questions for this study are as follows:

1. To what extent do smartphone apps designed using collaboration mechanics support grounding (the process of creating common ground) between primary and secondary users during face-to-face collaborations?
2. To what extent do smartphone apps designed to support grounding impact the satisfaction of secondary users during face-to-face collaborations?

Figure 1. The conceptual model for the research questions.



Methods

Design

The experimental design was a two-group randomized experimental study with posttest measures. The study explored the effect of a prototype HIT compared with a control group (no HIT) on common ground and secondary user satisfaction in a simulated clinical encounter using both qualitative and quantitative methods. A custom-designed HIT was prototyped for use as the experimental treatment. Patient-provider communication was investigated during the performance of the following 3 tasks: problem identification, discussing a patient's medical history, and medication reconciliation.

Setting and Sample

The study was conducted between February and May 2016 in 2 large Midwestern cities. This study used simulated face-to-face clinical encounters using a trained patient actor. Experienced clinicians consisting of doctors and nurses were the study participants. Participants were recruited using snowball sampling and email solicitations. All participants received monetary compensation. An adult male, with physical characteristics closely matching the scenario and more than a decade of professional and community theater acting experience, was hired (with monetary compensation) to role-play the patient. The same actor was used for all simulations. See Figure 2 for a model of the experimental design.

Participants

Materials and Procedure

The custom-designed HIT prototype used as the treatment for the study was a smartphone PHR app. A PHR is a private, secure, electronic, and Web-based tool that people can use to communicate with their providers and access, manage, and share their personal health information [28-30]. PHRs are examples of HIT with primary (ie, the patient) and secondary (ie, the provider) users, the 2 stakeholders with different knowledge and priorities who need to collaborate to make treatment decisions. Extensive preliminary work before the study was performed to create the prototype as an mHealth platform to support both types of users [15,17].

Development of the Prototype

The prototype was developed as a WordPress website so that it could be accessed regardless of smartphone manufacturer and operating system. The personal smartphone of the trained patient actor, a Samsung Galaxy Prevail LTE Android Version 4.4.4, was used to access the prototype during all simulations. The trained patient actor used his own personal smartphone with the prototype installed to minimize training requirements for using the custom-designed HIT prototype during the simulation.

The prototype incorporated buttons for patients and providers at the bottom of each page (Figure 3) as a switching mechanism so that tailored views of information contained within the prototype were available for primary and secondary users. Before sharing information with providers, the trained patient actor selected the provider button, which changed the view from his patient view to a context tailored for providers. These different views of information were necessary to improve communication and enable grounding [15].

Simulation Scenarios

Both the control and experimental groups participated in a simulated clinical encounter. The scenario required that a provider see a patient for the first time (Multimedia Appendix 1). The reason for the patient's visit was that he had suffered a rash, which had healed and was no longer visible. As part of the scenario, the provider was required to perform the following 3 tasks with the trained patient actor: problem identification, take a medical history, and medication reconciliation. Overall, 4 medications (Table 4) were part of the patient profile and were expected to be reviewed during medication reconciliation.

Figure 2. Model of the experimental design.

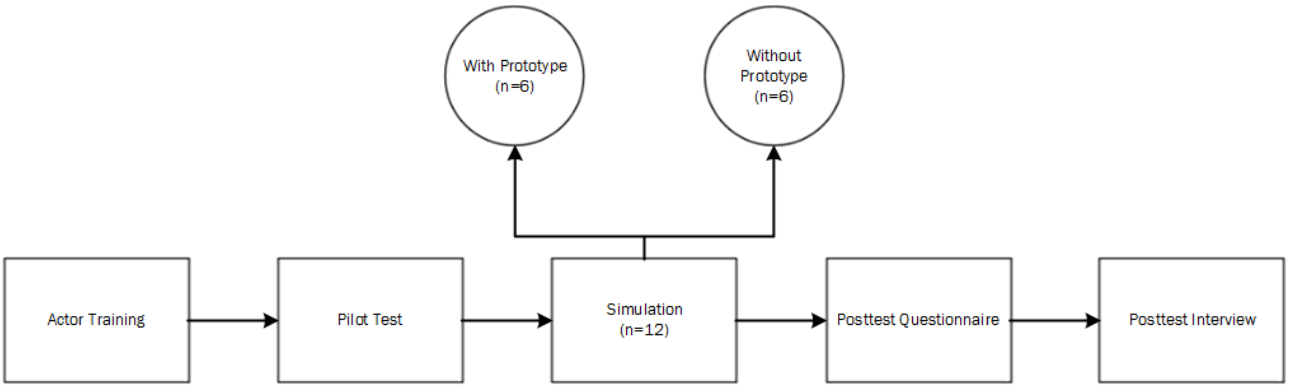


Figure 3. Example Samsung Galaxy Prevail LTE showing the patient and provider buttons. The screen background provides an additional visual cue to users about where they are in the interface. Primary user screens have a white background and secondary user screens a gray background.

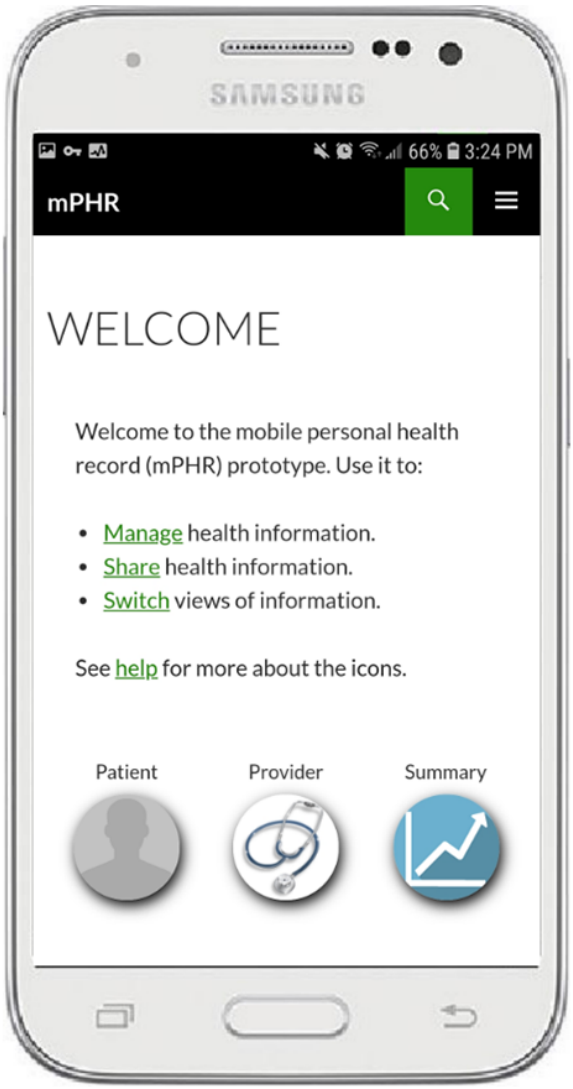


Table 4. Scenario patient medication profile.

Medication	Dose	Frequency	Reason
Metformin ER	500 mg	Twice/day	Diabetes
Lisinopril	10 mg	Once/day	Blood pressure
Atorvastatin	80 mg	Once/day (night)	Cholesterol
Glipizide	2.5 mg	Three/day with meals	Diabetes

Development of Simulation Scenarios

The orientation and training of the trained patient actor was extensive. The trained patient actor reviewed Web-based videos of simulated clinical encounters that included the tasks required for role-play. After reviewing the videos, a rehearsal training session with an experienced (13 years) registered nurse (RN), who has a PhD (Nursing Science) and is an assistant professor of nursing at a large Midwestern university, was conducted. After the session, the simulation documentation was revised and a checklist to assess treatment fidelity was created (Multimedia Appendix 2). The checklist enhanced the internal and external validity by ensuring that the trained patient actor addressed standardized areas with providers and ensured that the study could be replicated [31]. A second rehearsal using the treatment fidelity checklist as a guide was telephonically conducted. The simulation scenario was then pilot-tested for feasibility of study procedures. For this feasibility demonstration, we recruited 1 doctor in the control group and 1 advanced practice registered nurse (APRN) in the treatment group. Finally, a review of the pilot test simulation videos with the trained patient actor was completed before beginning the study.

Control Group Simulation

For the control group, the trained patient actor simulated real patient behavior during a typical encounter by relying on memory to share information with the study participant about the problem that sparked the visit (rash), medical history, and current medications. The medical history for the scenario included multiple ailments: type 2 diabetes mellitus, obesity, hypertension, and high cholesterol. See Figure 4 for an example of the interaction between the trained patient actor and study participant during a control group simulation.

Treatment Group Simulation

The treatment group was shown an image of the rash (Figure 5) by the trained patient actor. A verbal description of the ailment was also provided by the trained patient actor. The image used for the rash was of a patient with *bullosis diabeticorum*, a blistering condition that heals in a few weeks.

For medication reconciliation, details about the medications were shared with the treatment group from the prototype (Figure 6) by the trained patient actor. For the medical history tasks, the prototype was not offered to members of the treatment group unless they specifically made a request to look at the information on the device (there were no provider requests to review medical history on the prototype). Before sharing information from the prototype with participants, the trained patient actor switched the view of information from the primary user's view to the secondary user's view.

Figure 4. The patient attempts to recall from memory the details of specific medications during the medication reconciliation task.

Figure 5. Image used in the prototype to show the patient's past ailment. Note. Image adapted from Bullosis diabeticorum: Rare presentation in a common disease, by Gupta V, Gulati N, Bahl J, Bajwa J, and Dhawan N, 2014, Case Reports in Endocrinology, p. 2.

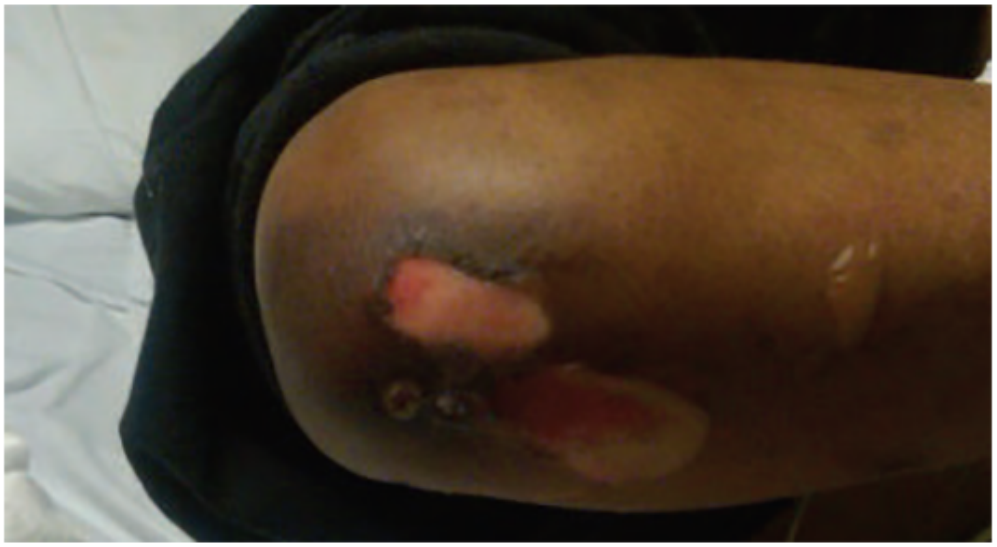
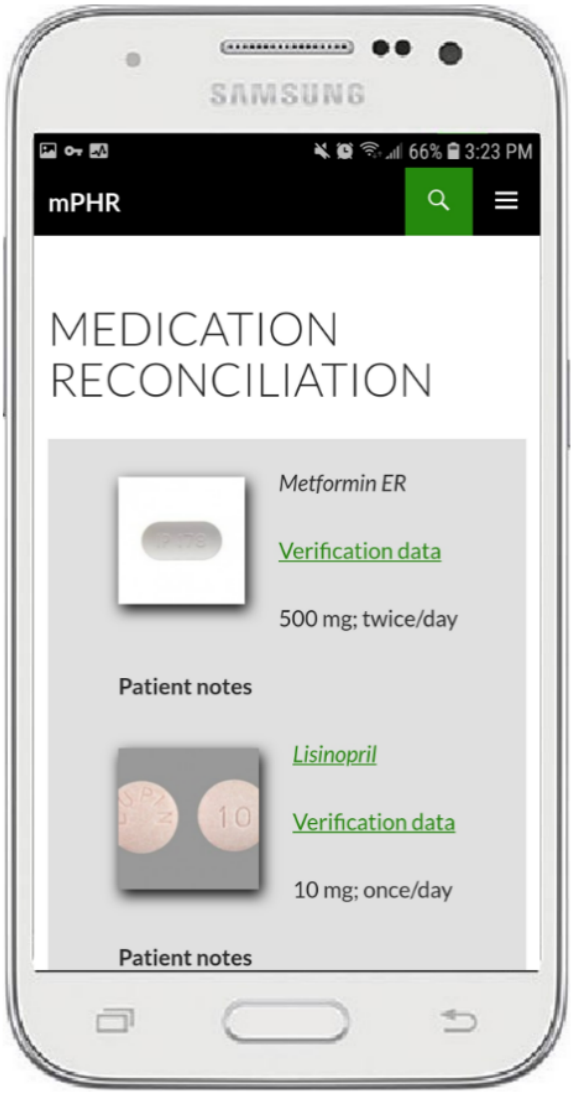


Figure 6. Example Samsung Galaxy Prevail LTE showing the provider view of the medication screen. The screen background provides an additional visual cue to the user about where they are in the interface. Primary user screens have a white background and secondary user screens a gray background.



Variables

The independent variable was the presence or absence of the prototype. The dependent variables were common ground and secondary user satisfaction. The participant's perception of performance was also measured because of its likely impact on satisfaction. Common ground and the remaining variables were measured by analyzing the simulation videos, Likert scale responses, and interview transcripts. The details of the 3 measures to evaluate the presence or absence of the treatment are as follows:

1. Recordings: The simulation was video recorded for later analysis with NVivo 11 (a qualitative data analysis software package).
2. Questionnaire: Upon the conclusion of the scenario, the participants completed a 7-point Likert scale questionnaire, which was composed of demographic questions and a set of 15 psychometric scales measuring satisfaction, common ground, and performance (Multimedia Appendix 3).
3. Interviews: Semistructured interviews were audio-recorded and conducted with each participant after the simulation (Multimedia Appendix 4). The results were transcribed using the *TranscribeMe!* service provided through NVivo 11. The transcriptions were *clean verbatim* in which filler words (eg, umm, ah, and you know) were removed. Irrelevant concluding remarks were deleted (eg, closing courtesies). The transcriptions were of good quality. When errors were discovered, they were reviewed, cross-checked, and corrected based upon the audio recordings.

During the study, an emphasis was placed on collecting rich qualitative data, which is gained when saturation is reached. The study sample size was considered adequate because saturation had been reached during earlier phases of this project with similar-sized groups. In addition, saturation was, in fact, achieved during the experimental study. A statistical power analysis indicated that the sample size was too small for hypothesis testing [32]. In lieu of this type of testing, descriptive statistics were used to compare interactions with and without the presence of the prototype. These factors, when taken together, ensured adequate investigation of the dependent variables.

Procedure

The study was performed at a location selected by the study participants. Hospital treatment rooms and administrative or public spaces in medical facilities and academic buildings were typical. At the beginning of each iteration, the participant was asked to review the study information sheet. The scenario was reviewed with the participant, who was then provided with a pen, clipboard, and paper for taking notes during the simulation. Subsequently, the simulation began, with the use of video recording.

Due to the different specialties and practices encountered during the study, it was impossible to anticipate the flow of the conversation during each simulation. The trained patient actor was authorized to incorporate real-life experiences as necessary to maintain realism during the simulation. For example, discussions about diet and exercise mimicked the trained patient actor's real-life experiences.

After each simulation, the trained patient actor's performance was reviewed and behaviors corrected as necessary. The potential effect of actor learning was controlled through training, the use of checklists for treatment fidelity, and counterbalancing between the control and treatment groups [31,33]. Consistency was maintained as much as possible. Minor errors completed by the actor did not disrupt the study as long as the actor was consistent throughout the duration of the simulation.

Qualitative Analysis

A thematic analysis of the videos and semistructured interview transcripts was performed using NVivo 11. This form of analysis was preferred because it is the most common approach for evaluating qualitative data in health care research [34]. The videos were coded by importing them into NVivo 11, reviewing each video within the software, selecting parts along the range of the media using the software tools, and assigning the range to an appropriate coding container. Summary transcriptions of the videos were also created within NVivo 11, showing general topics of conversation at selected points in the video (eg, medication reconciliation or problem identification). The audio-recorded, semistructured interviews were transcribed, as described earlier, and coded within NVivo 11.

There were 3 rounds of coding, with the first round being conducted manually within the software. During this round of coding, concepts were merged, revised, and discarded as patterns became more obvious. The second round relied on the analytical software tools of NVivo 11 to analyze results from the first round of coding and create a basic concept map. The map was used to sort the concepts that resulted from the first round of coding and aid in reorganizing them into a hierarchical system for analysis. The third round of coding scrutinized the data based on membership in the control or treatment group and evaluated how the content diverged thematically between the groups. A matrix query by assignment to experimental group was run to display the frequency of the coding in each major category of joint action. This round of coding was used to break out details of the data to satisfy the research questions. As a result, themes for the control and treatment groups were identified. The analysis revealed that both groups accomplished a substantial level of grounding for the medical history task. Consequently, it was not coded in detail because the analysis of the task, in and of itself, would not contribute to answering the research questions.

Quantitative Analysis

Descriptive statistics were employed because of their importance for inferring evidence of communication efficiency for the providers. Descriptive statistics of means, SDs, maximums, and minimums were calculated using SPSS 23. The posttest questionnaires were also evaluated using SPSS 23.

Each video was studied to assess the amount of time spent on the 3 scenario tasks (ie, problem identification, medical history, and medication reconciliation) and the percentage of time the tasks consumed for the entire encounter. The efficiency of the encounter was assessed based on the descriptive statistics for time spent on the 3 required tasks and the overall duration of encounters. The main effectiveness measure of encounters between the control and treatment groups also relied on descriptive statistics. In this case, they were used to evaluate the amount of time spent on patient education, which providers throughout the research identified as an important task to perform with patients. Finally, descriptive statistics were used to analyze the posttest questionnaires and compare responses between the control and treatment groups.

Results

Introduction

Overall, 13 nurses and doctors participated in the experimental study, with 12 included in the final analysis. One participant became frustrated during the simulation and requested that the video recording cease. The subsequent interview revealed that she normally used an electronic medical record (EMR), which is an electronic version of a patient chart, and followed a standardized routine when interviewing patients (an EMR was not provided as part of the scenario). The lack of an electronic tool made her feel as if she was not providing quality service to the trained patient actor, which caused her to become frustrated and culminated in her request to cease the video recording. Consequently, these results were excluded from the final analysis.

The backgrounds of the final 12 participants, by assignment to control group and treatment group, are shown in Table 5. Of these, 2 participants were educators and no longer interacted with patients; 1 had transitioned to teaching within the 6 months before the study and the other had 19 years of prior nursing experience. A third participant had shifted to case management within 6 months before the study and had more than 25 years of experience in nursing. The 3 participants had the necessary skills to perform the scenario tasks and were qualified for the study.

The professional experience of the participants varied from less than 1 year to more than 25 years (Table 6). All participants were familiar with the basic technologies used in the study. Each participant owned a smartphone and had 11 or more years of experience using computers. Overall, 9 of the 12 participants reported owning tablet computers.

Table 5. Background of participants (12 participants).

Group	Practice or specialty
Control group	
APRN[a]	Geriatric
APRN	Geriatric
Doctor	Internal medicine
Certified nurse midwife	Midwifery
Registered nurse	Intensive care or intravenous team
Doctor	Internal medicine
Treatment group	
Doctor	Pulmonary medicine
Registered nurse	Case management
Registered nurse	Clinical education
APRN	Internal medicine
Registered nurse	Perioperative nursing
Doctor	General surgery

[a]APRN: advanced practice registered nurse.

Table 6. Demographics of participants (12 participants).

Variable and attribute	Control group, n (%)	Treatment group, n (%)
Gender		
Male	2 (17)	1 (8)
Female	4 (33)	5 (42)
Age in years		
25-34	1 (8)	0 (0)
35-44	3 (25)	3 (25)
45-54	2 (17)	3 (25)
Years of experience		
<1	1 (17)	0 (0)
1-5	3 (25)	0 (0)
6-10	2 (17)	1 (8)
11-15	0 (0)	0 (0)
16-20	0 (0)	1 (8)
21-25	0 (0)	2 (17)
> 25	0 (0)	2 (17)

Themes

Overall, 3 themes emerged from the thematic analysis. The themes ultimately represent how grounding manifested during the encounter, what it meant for communication during the encounter, and how it influenced the provider's perception of the patient. The themes are discussed below.

- *The patient is engaged in his own health care*. If a patient is willing to take the time to manage his health information electronically, then it is an indication that he is engaged in his own health care.
- *The information is trustworthy*. The use of sophisticated technology by patients implies a higher trustworthiness of information—or, at least, the same level of trustworthiness as other traditionally accepted (but low-use) methods (eg, bringing medication bottles to an appointment).
- *There is enough information at an acceptable level of quality for some level of decision making to be obtained*. This amount of information, and its quality, does not mean that there is a diagnosis. It means that better planning is performed with the patient, even if that plan does nothing, because everything appears to be on track.

In addition to the basic scenario tasks, patient education emerged as a desired discussion topic with patients in the posttest interviews for most participants. As patient education emerged as a topic of broad concern to participants, patient education discussions about the rash, medical history, and medications were coded for the time analysis (see Quantitative Analysis section for these results). Time spent on general nutrition, health and fitness, and similar counseling was not evaluated.

Efficiency

In general, the trained patient actor's use of the smartphone seemed to improve provider's patient care efficiency during the encounter. The overall encounter and task times (Table 7) averaged slightly faster in almost every instance for the treatment group than that of the control group. Task times were faster in the control group for the medical history task, which can be explained by the fact that not all providers completed the task. As such, the overall mean was artificially reduced for the control group. Finally, SDs generally clustered more closely around the mean for tasks in the treatment group than that in the control group.

A benefit of patient care efficiency might be related to the fact that the providers had more time to conduct discussions on patient education. The mean for every educational task was higher for the treatment group than for the control group. This difference likely contributed to the higher percentage of the encounter time overall (ie, time spent on the 3 tasks by the treatment group).

The emphasis on task completion is also relevant for an analysis of patient education. Although efficient and effective patient care is not always correlated, we argue that time is another indicator of a more *effective* use of provider time in the study. The importance of patient education to providers underpins this analysis. If the efficient use of time leads to the discussion of additional important information (ie, patient education), then the encounter can be considered more effective.

Satisfaction, Common Ground, and Performance

The quantitative analysis of the posttest questionnaire examined secondary user perceptions of satisfaction, common ground, and performance. For the posttest questionnaire, a Cronbach alpha reliability analysis (Table 8) was performed on all the subscales using SPSS 23. To improve the reliability of the satisfaction subscale, 2 items were removed. This removal increased the reliability to ≥.80, which is good reliability [35]. Reliability for the common ground and performance subscales were ≥.80 and .90, respectively, which are good and excellent reliabilities, respectively [35].

Table 7. Descriptive statistics for encounter and task times for the 6 participants in the control group and 6 participants in the treatment group.

Task	Control group			Treatment group		
	Mean (SD)	Median	Min-Max	Mean (SD)	Median	Min-Max
Encounter[a]	0:12:45 (0:06:19)	0:11:09	0:06:35-0:25:05	0:12:22 (0:02:38)	0:12:30	0:08:47-0:15:40
Rash[b]	0:02:52 (0:02:33)	0:02:08	0:01:12-0:08:02	0:02:40[c] (0:01:36[c])	0:02:02[c]	0:01:44[c]-0:05:53[c]
Education[d]	0:00:18 (0:00:44)	0:00:00	0:00:00-0:01:50	0:00:19[c] (0:00:31[c])	0:00:00[c]	0:00:00[c]-0:01:13[c]
Med reconciliation[e]	0:03:03 (0:00:58)	0:02:37	0:02:33-0:05:02	0:02:24[f] (0:01:05[f])	0:02:11[f]	0:01:09[f]-0:04:14[f]
Education[d]	0:00:17 (0:00:29)	0:00:00	0:00:00-0:01:13	0:00:36[f] (0:00:58[f])	0:00:00[f]	0:00:00[f]-0:02:14[f]
History[g]	0:01:25[f] (0:01:43[f])	0:00:57[f]	0:00:00[f]-0:04:10[f]	0:02:46[f] (0:01:38[f])	0:02:23[f]	0:00:54[f]-0:05:04[f]
Education[d]	0:00:00[f] (0:00:00[f])	0:00:00[f]	0:00:00[f]-0:00:00[f]	0:00:23[f] (0:00:57[f])	0:00:00[f]	0:00:00[f]-0:02:21[f]
Total time spent on required tasks[h]	0:07:57 (0:02:39)	0:07:40	0:04:41-0:11:25	0:09:11 (0:01:49)	0:09:12	0:06:45-0:11:16

[a]The percentage of the total encounter time spent on the three tasks: control group=67%; treatment group=76%.

[b]Rash: time discussing the problem.

[c]Indicates a task for which partial common ground was achieved.

[d]The mean times for any patient education related to a specific task.

[e]Med reconciliation: time discussing medication reconciliation.

[f]Indicates a task for which common ground was achieved.

[g]History: time discussing medical history.

[h]The total time spent on the three tasks: Rash, med reconciliation, and history.

Table 8. Subscale reliability analysis.

Characteristics	Cronbach alpha	Item numbers
Satisfaction	.824	3
Common ground	.892	6
Performance	.981	4

Table 9. Descriptive statistics for the posttest questionnaire subscales.

Group	Satisfaction	Common ground	Performance
Control			
Mean (SD)	5.83 (1.44)	5.94 (1.25)	5.83 (1.54)
Minimum-maximum	3.00-7.00	3.83-7.00	3.00-7.00
Treatment			
Mean (SD)	5.67 (1.01)	5.75 (0.43)	5.25 (0.98)
Minimum-maximum	4.33-6.67	5.17-6.17	4.00-6.00

The descriptive statistics for each subscale (Table 9) show higher means for the control group than for the treatment group, indicating that members of the control group perceived that they had a higher level of satisfaction, common ground, and performance than members of the treatment group. SDs show better clustering around the mean for the treatment group on each subscale, which could indicate a better consensus among treatment group members, reflecting a more accurate evaluation of satisfaction, common ground, and performance in the treatment group than in the control group.

Discussion

Overview

The experimental study provided substantial insight on grounding in the context of a face-to-face clinical interaction. The research questions were comprehensively addressed during the study. Grounding was better in the treatment group, indicating that the idea of integrating collaboration mechanics into interactive technology designs with the intent to improve grounding has merit. The specific findings for each research question are listed below:

1. Research question 1 finding: The experimental study results indicated that smartphone apps designed using collaboration mechanics support grounding between primary and secondary users during face-to-face collaborations and can facilitate complete common ground. The success of grounding with them is task-dependent.
2. Research question 2 finding: The experimental study results indicated that smartphone apps designed to support grounding have the potential to positively impact secondary user satisfaction, performance, and perspective about the primary user's commitment to the collaboration.

Explanation of Outcomes

Similar Outcomes

Grounding occurred in both groups regarding the medical history task. However, the time analysis shows that the minimum time on task for the control group was 0 seconds, whereas the treatment group's minimum time was 54 seconds. Furthermore, none of the providers in the control group conducted patient education related to the medical history task, which was contrasted with a mean time in the treatment group of 23 seconds. Thus, even though grounding for this task occurred in both groups, it occurred throughout the treatment group with the additional benefit of including patient education. This aspect of the time analysis indicates that the encounters in the treatment group were more effective than those in the control group because all treatment group members accomplished the required task and made time for education, whereas those in the control group did not.

Control Group Outcomes

In the control group, grounding did not occur for problem identification or medication reconciliation. Participants were not able to confidently identify any aspect of the rash other than it had occurred and had healed. Although several participants were able to exclude some environmental causes during their discussions with the trained patient actor (eg, no recent changes to medications), the grounding that did occur involved future patient action if the rash recurred (ie, contact the office immediately). One participant described the difficulty of evaluating the no longer visible rash as:

> *The rash, that was difficult because once he described the rash, I thought of probably three things it could have been. So, maybe that was probably about an eight [out of 10] difficulty just because it's not there anymore. So, I can't treat something or even tell him what it is without having seen it. I can't treat something I can't see.* [Participant 3-C (APRN)]

Less grounding occurred in this instance, because the rash was gone, and the trained patient actor's verbal description was of limited utility.

Regarding medication reconciliation, there was no indication that the providers were able to glean enough information to be confident that they had correctly identified the patient's medication regime. On one hand, medication reconciliation was the most challenging task for participants in the scenario. The lack of patient medication knowledge inhibited decision making about the patient's health care. One participant commented:

> *He didn't know his medication doses and timing, so I was worried that there was a high risk for error in assuming [that] what he was telling me was right. It was difficult to make recommendations or a plan without knowing what those medications were.* [Participant 3-C (APRN)]

On the other hand, the preliminary work for the study indicated that providers typically have a substantial amount of experience with patients who lack detailed knowledge about their own medications. This is consistent with what was observed during the study.

Because participants were accustomed to dealing with a lack of medication information, they knew what questions to ask to devise workarounds. Consequently, they were able to create a plan of action with the trained patient actor that would result in getting the correct information. However, even with workarounds, significant treatment delays could be expected as 1 participant acknowledged:

> *What I didn't know was his medications, and I didn't want to guess. I needed that information, but I just have to find it from another source or ask him next time to bring his medications.* [Participant 2-C (APRN)]

The participants indicated that this situation—patient lack of knowledge about their medication details—was typical of encounters with new patients.

It was apparent that grounding did not occur in the control group for the medication reconciliation task. Although there were successful joint actions between the participants and trained patient actor that culminated in planning activities to get the correct information, these did not enhance the quality of the encounter. In fact, the lack of common ground made the encounters distinctly inefficient in the control group.

Treatment Group Outcomes

In the treatment group, grounding occurred for the problem identification and medication reconciliation tasks. Although participants could not determine the cause of the rash or swab it for testing, the picture sparked deeper engagement with the trained patient actor about the ailment and allowed participants to exclude some diagnoses. It also seemed to improve general confidence among participants about their interactions with the trained patient actor. The availability of an image of the rash was helpful, as highlighted by this participant:

> *I like that he took pictures of his wound. It would have been, like I told him, helpful in the future to actually be seen when the situation is acute versus resolved but taking that photo when it was an active rash was helpful—at least would be helpful—to the physician.* [Participant 6-T (RN)]

Although grounding was not complete for this task, treatment group participants were typically more willing to share detailed information about this type of wound with the patient. This is noteworthy because rashes are often a difficult clinical issue, so any improvement to communication about such a problem is important. The ability to view an image of the rash clearly aided the communication.

Use of the prototype resulted in common ground being obtained during the medication reconciliation task. Treatment group participants were confident at the end of the interaction that critical and accurate information about the medications had been relayed to them. This allowed them to probe deeper with specific questions or spend more time on patient education. See Figure 7 for an example of the interaction between the trained patient actor and a study participant during a treatment group simulation.

Treatment group members trusted the information provided from the prototype. For example, participant 5-T (APRN) remarked that:

> *...my gut reaction is that it's accurate and it's a tool that can be shared between the patient and the provider.* [Participant 5-T (APRN)]

The information obtained during the medication reconciliation task was of good enough quality to support medical decision making. In another example, Participant 6-T (RN) felt that the hospital could dispense medications to a patient admitted for an overnight stay with the level of detail provided by the trained patient actor.

Successful joint actions between the trained patient actor and participants, leading to common ground, occurred during all tasks for the treatment group. However, the completeness of common ground and relevance of the prototype did vary by task. The prototype supported successful joint action and partial common ground for the problem identification task and complete common ground for the medication reconciliation task. The prototype was unnecessary for accomplishing common ground during the medical history task.

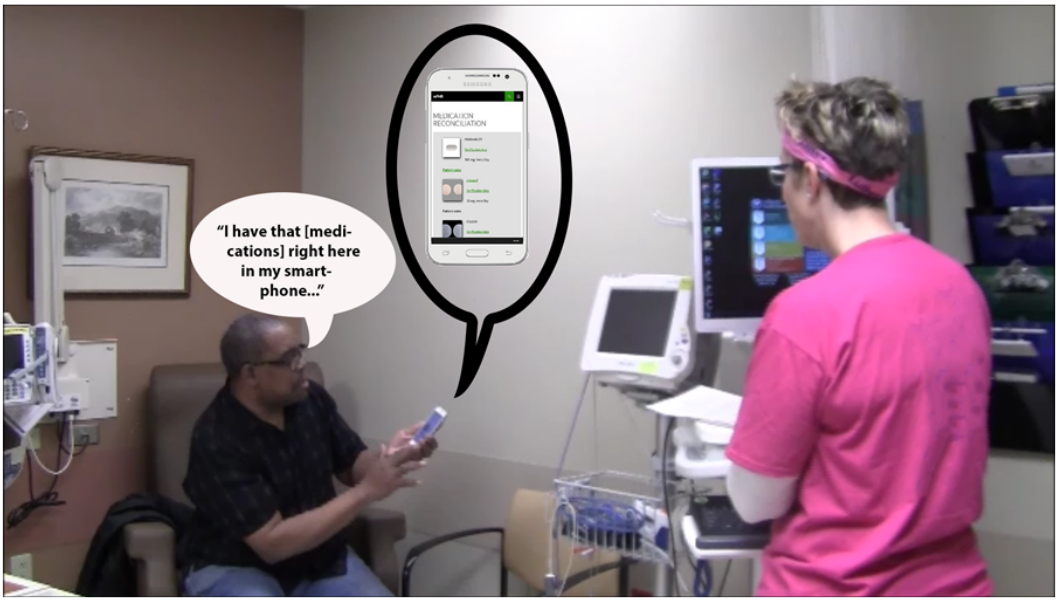
Overall in the treatment group, joint actions were comprehensive enough that participants could make plans for treatment based on the details obtained during the interview. For example, at the end of the encounter, the participants had enough accurate information about medications to maintain the current prescriptions (because they were the correct medicine at the proper dosage for the scenario). This was not the case in the control group. The lack of certainty about medications meant that participants could not evaluate if they were working for the ailments depicted in the scenario.

Although the posttest questionnaire implies less satisfaction among members of the treatment group, the richness and depth of the findings of the qualitative analysis indicate the opposite and provide a more comprehensive view of the data. The overall qualitative analysis clearly indicated that providers in the treatment group were more satisfied than their peers in the control group. The completeness of information and its contribution to a more successful encounter becomes obvious (because of the qualitative analysis) as reflected in the comments of 1 participant:

> *That information that you have to have to make decisions, we didn't have to spend a lot of time figuring that out [because of the prototype]. We were able to quickly get all of that and then move onto here's what we're going to do about the problems that you have, the issues that you have, and go from there. So, we had more time for that, rather than just trying to figure out the historical data.* [Participant 2-T (Doctor)]

For the treatment group, the analysis of satisfaction expressed by the providers combined with the descriptive statistics for the performance of the tasks, indicate that treatment group encounters were more efficient and effective than those in the control group.

Figure 7. The patient provides accurate medication information by using the prototype to augment his memory, making the attainment of common ground possible.



The qualitative analysis was a rich source of data and provided detailed insight for answering the research questions. The quantitative analysis of the task performance supported the conclusions of the qualitative analysis. According to these measures, the use of the mobile phone optimized common ground for members of the treatment group. In contrast, the results from the posttest questionnaire were a bit perplexing. For example, research about the impact of introducing examination room computers to the patient-provider encounter showed a higher degree of satisfaction among patients after the introduction of computers [36]. One would expect a similar outcome for clinicians in this research. It is unclear why the *perspectives* of the participants in each group, as assessed using the posttest questionnaire, were reversed.

Although the posttest questionnaire implied that the control group's perceptions were more favorable than that of the treatment group, this analysis was clearly not supported based on the strong positive responses of the treatment group members during the simulations and semistructured interviews. Poor clustering around the mean for satisfaction, common ground, and performance for the control group on the posttest questionnaire indicated less consensus among them about the variables than for the treatment group.

There are other factors, such as the distinction between the UX and usability, that may help explain how and why the differences in perception occurred. The UX is a person's emotional response to an interactive system, whereas usability is a technical aspect that emphasizes how well a person can use a system. The placement of a patient-managed technology between the provider and the trained patient actor may have subtly and negatively transformed provider perceptions in the treatment group. Although providers in the treatment group liked the usability of the system and found it helpful, they might have perceived it as creating a barrier between themselves and the trained patient actor in some way.

This idea of a barrier makes sense when considering the emotional nature of the UX and the technical nature of usability. Moreover, the control group was largely made up of primary care professionals who look at patients holistically and longitudinally. The treatment group was largely made up of specialists who may emphasize a procedural perspective because they look at patients for specific ailments and patient relationships tend to be much shorter than the relationships cultivated by their primary care peers. However, this distinction is a minor point because the overall analysis of the data clearly indicates that the performance was better and the UX was more satisfactory for the treatment group.

Unanticipated Outcomes

Finally, a few unanticipated outcomes of insight were gleaned. The first was the possible impact of the technology on the perspectives of the provider. For example, the semistructured interviews revealed that the providers in the treatment group were happy with the outcome of the encounter and felt that the information exchange was better than usual. However, they did not record a level of satisfaction higher than their peers in the control group on the posttest questionnaire. The disparity indicates the need for a closer look at the potential impact of patient-centered technology.

Another insight was the willingness of providers to become responsible for clinical data from nonclinical sources. An early assumption during the requirement-gathering process for designing the prototype was that providers might not be willing to help manage these type of data. Nothing in the following stages of preliminary work or the experimental study suggested any reluctance on the part of providers to interact with these type of data. Rather, the providers considered the data trustworthy or, at least, as trustworthy as other types of patient-provided information that they hold in high regard.

Implications of Results

The medication reconciliation task produced the strongest example of grounding in the research, which should not be surprising, as study participants indicated that 50% to 90% of their patients do not have accurate medication information with them during a real-world encounter. This lack of accurate information suggests a gap in knowledge and in patient-centered tools to aid patient recall. As the experimental study demonstrated, a tool that provides relevant medication information to providers can facilitate common ground. The prototype's interface provided external representations of medications (ie, images of each medication) that were clear objects of interest to secondary users. The representations allowed participants in the treatment group to quickly collaborate with the trained patient actor about his medications, regardless of the secondary user's interaction style with the prototype.

Limitations

The scope of the research was restricted to the relation between usability and common ground within an mHealth setting. The research initially relied on a thematic analysis according to common ground theory. The research culminated in a small sample experimental study using a simulation. Consequently, there are limitations for generalizability.

Form factor and device functionality were relevant to the experimental study regarding the simulated patient-provider interactions. In addition, the sample size was too small for hypothesis testing. Therefore, the results may not be generalizable to collaborations outside the scope of short-duration, face-to-face clinical encounters. Nonetheless, even with a small sample size, the groundbreaking nature of the study offers value to the health care community because it emphasizes the patient as a knowledgeable collaborator—and as one who (through mHealth) can share their personal health information in a manner that improves the overall efficiency and effectiveness of provider care.

Future Research and Recommendations

This study is the first that we know of to explore common ground using both primary and secondary UXs. The results indicate that there is a need to consider primary and secondary users when designing a single system for information sharing between those with expert and nonexpert levels of knowledge. To improve generalizability, the study should be replicated with a larger sample and providers who work in the same hospital or practice. It would be best to limit the study participants to 1 skill or specialty.

Limiting participants has several advantages for ecological validity. For example, EMRs and intake sheets with simulated patient data can be created using the systems common to the organization. This will increase realism in the simulation for participants. Furthermore, with the standardized procedures, training, and tools common to a single organization and specialty, confounding variables can be limited.

Finally, the prototype used for the simulation should be a mobile phone app running on the device rather than a responsive design website. There was some latency accessing the website based on location and network quality, which caused small delays. A native app running on a mobile phone mitigates the impact of a slow running website and will more accurately capture shorter encounter times because of the faster load times for a native app.

Summary

During the preliminary work, the priorities that emerged for providers in the type of encounter simulated were to conduct medication reconciliation, problem identification, and then medical history. The priorities, regarding the technology and its efficacy as a tool to facilitate communication, were supported during the experimental study. For example, peak usefulness of the prototype was demonstrated whenever participants attempted to glean highly clinical data (ie, detailed medication information) from the primary user, who is typically a nonclinical source for such information.

Furthermore, participants in the treatment group had improved confidence because the rash image was available. Finally, although technology did not improve or detract from grounding for members of the treatment group during the medical history task when compared with the control group, it seemed to create efficiencies for the overall encounter that allowed all members of the treatment group to complete the task (whereas all members of the control group did not).

The alignment of the perceptions of the relative importance of the respective tasks from providers to the actual creation of common ground during the experimental study should be interpreted as providing a level of awareness for design decisions regarding the allocation of time and other resources. If tasks are difficult to accomplish (ie, medication reconciliation), then attempts to promote grounding using technology may be a good use of resources, while it may be a poor use of resources for easy tasks (ie, medical history). This identification of importance and the relation to common ground is a vital insight for the overall body of secondary user research.

Regarding outcomes in the different experimental groups, it is not that grounding did not occur in the control group—it did. Rather, it is the relevance of the common ground achieved for solving the problem at the center of the need for communication that is at issue. Participants noted that the attainment of common ground during a first encounter with a new patient was frequently rare in real life. They also indicated that they would support patients' uses of smartphones as a tool by which to improve face-to-face communication during encounters. This participant highlights why smartphone use would be acceptable:

> *The easiest thing about sharing information with this patient was his ability to use technology to show me, so that we were both on the same page. As opposed to [a verbal] description, [where] I paint my own [picture]. Then, we're both on apples to apples, instead of me trying to paint in my head what he's describing.* [Participant 4-T (RN)]

This study indicated that well-designed systems that deepen the engagement of patients in their own health care while improving near-term communication with providers have a place in HIT.

The provider responses in the treatment group reinforced design decisions about the information that should be made available to secondary users. Overall, the utility of patient-controlled devices during a first encounter with a patient depends as much on the difficulty for the provider of normally obtaining the information intended to be shared, and its impact upon immediate decision making, as on the HCI design decisions.

Medication reconciliation is a difficult task that is necessary for successful treatment decisions. In real life, inconclusive reconciliation is so routine that it is expected among providers. The introduction of technology to mitigate a patient's personal lack of knowledge has the potential to create useful common ground within the dyad for this type of complex health care task. This observation is another critical insight for the overall body of secondary user research.

Conclusions

The experimental study is one of the first studies to directly demonstrate that not all secondary UXs are meaningful for design. For example, grounding occurred during the medical history task in the control and treatment groups. Any marginal improvement due to implementation of technology, where grounding occurs regardless of technology, is probably not worth the effort. The fact that common ground was only completed in the treatment group, and during medication reconciliation (an essential enabler for health care decision making), indicates that an investment in the secondary UX for high payoff tasks is valuable.

Combining the notions of common ground, HCI design, and smartphone technology resulted in a prototype that improved the efficiency and effectiveness of face-to-face collaboration for secondary users with the primary user. The prototype clearly facilitated a higher quality of information exchange than normal. Thus, the investigation substantiated the notion that properly designed interactive systems have the potential to facilitate common ground while providing a satisfactory secondary UX.

Acknowledgments

Majority support for the paper processing fee was provided by the IUPUI University Library Center for Digital Scholarship.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Experimental study scenario.

[PDF File (Adobe PDF File), 18KB - mhealth_v6i11e11131_app1.pdf]

Multimedia Appendix 2

Checklist to assess treatment fidelity.

[PDF File (Adobe PDF File), 70KB - mhealth_v6i11e11131_app2.pdf]

Multimedia Appendix 3

Posttest questionnaire.

[PDF File (Adobe PDF File), 55KB - mhealth_v6i11e11131_app3.pdf]

Multimedia Appendix 4

Posttest semistructured interview guide.

[PDF File (Adobe PDF File), 41KB - mhealth_v6i11e11131_app4.pdf]

References

1. Eason K. Information Technology and Organizational Change. Philadelphia, PA: Taylor and Francis Inc; 1988.
2. Alsos O, Svanæs D. Designing for the secondary user experience. In: Campos P, Graham N, Joaquim J, Nunes N, Palanque P, Winckler M, editors. Human-Computer Interaction – INTERACT 2011. Lecture Notes in Computer Science. New York: Springer, Berlin, Heidelberg; 2011:84-91.
3. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004:445-456.
4. Gonzales M, Riek L. A shared interface to improve oncologist-patient communication. 2012 Jan Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; 2012; San Diego, CA. [doi: 10.4108/icst.pervasivehealth.2012.248711]

5. Montague E, Xu J. Understanding active and passive users: the effects of an active user using normal, hard and unreliable technologies on user assessment of trust in technology and co-user. Appl Ergon 2012 Jul;43(4):702-712 [FREE Full text] [doi: 10.1016/j.apergo.2011.11.002] [Medline: 22192788]
6. Inbar O, Tractinsky N. Lowering the line of visibility: incidental users in service encounters. Behav Inf Technol 2012 Apr 27;31(3):245-260. [doi: 10.1080/0144929X.2011.563796]
7. Sathiaseelan A, Crowcroft J. Internet on the move: challenges and solutions. Comput Commun Rev 2013 Jan;43(1):51-55. [doi: 10.1145/2427036.2427046]
8. Reisinger D. Cnet. 2014 Dec 11. Over 2 billion smartphones will circulate in 2016, report predicts URL: https://www.cnet.com/news/over-2-billion-smartphones-to-be-in-use-by-end-of-2016-study-shows/ [accessed 2018-10-01] [WebCite Cache ID 72r5mn6fv]
9. Smith A. Pewinternet. 2015 Apr 1. US Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2018-10-01] [WebCite Cache ID 72r68p8zo]
10. Bardes CL. Defining "patient-centered medicine". N Engl J Med 2012 Mar 1;366(9):782-783. [doi: 10.1056/NEJMp1200070] [Medline: 22375968]
11. Ferneley E, Light B. Secondary user relations in emerging mobile computing environments. Eur J Inf Syst 2006 Jun;15(3):301-306.
12. Inbar O, Tractinsky N. The incidental user. Interactions 2009;16(4):56-59. [doi: 10.1145/1551986.1551998]
13. Sharp H, Finkelstein A, Galal G. Stakeholder identification in the requirements engineering process. IEEE. 1999 Presented at: 10th International Workshop on Database & Expert Systems Applications; September 1-3, 1999; Florence, Italy p. 387.
14. Kellogg WA, Erickson T, Wolf TV, Levy S, Christensen J, Sussman J, et al. Leveraging digital backchannels to enhance user experience in electronically mediated communication. 2006 Presented at: 20th Anniversary Conference on Computer Supported Cooperative Work; November 4-8, 2006; Banff, Alberta, Canada p. 451-454. [doi: 10.1145/1180875.1180943]
15. Tunnell H, Faiola A, Haggstrom D, Srinivas P. Clinicians as secondary users of patient-centered mobile technology in complex healthcare settings. 2015 Presented at: 17th International Conference on E-Health Networking, Application and Services, HealthCom 2015; October 13-17, 2015; Boston, MA p. 657-660.
16. Tunnell H, Faiola A, Bolchini D. Guidelines to incorporate a clinician user experience (UX) into the design of patient-operated mHealth. New York, US: ACM; 2017 Presented at: 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 385-388. [doi: 10.1145/3027063.3052966]
17. Tunnell HD, Pfaff M, Faiola A. Secondary users the personal mHealth record: designing tools to improve collaboration between patients providers. 2016 Presented at: IEEE International Conference on Healthcare Informatics. IEEE; October 4-7, 2016; Chicago, IL.
18. Prey JE, Woollen J, Wilcox L, Sackeim AD, Hripcsak G, Bakken S, et al. Patient engagement in the inpatient setting: a systematic review. J Am Med Inform Assoc 2014;21(4):742-750 [FREE Full text] [doi: 10.1136/amiajnl-2013-002141] [Medline: 24272163]
19. Monk A. Common ground in electronically mediated communication: Clark's theory of language use. In: Carroll JM, editor. Toward a Multidisciplinary Science of Human-Computer Interaction. Boston, MA: Morgan Kaufmann Publishers; 2003:265-289.
20. Clark HH. Using language. New York: Cambridge University Press; May 16, 1996:432.
21. Clark H, Brennan S. Grounding in communication. In: Resnick L, Levine J, Teasley S, editors. Perspectives on Socially Shared Cognition. Washington, DC: APA; 1991:127-149.
22. Convertino G, Mentis H, Rosson M, Slavkovic A, Carroll J. Supporting contentprocess common ground in computer-supported teamwork. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 4-9, 2009; Boston, MA p. 2339-2348.
23. Brennan S. The grounding problem conversations with and through computers. In: Fussell S, Kreuz R, editors. Social and Cognitive Psychological Approaches to Interpersonal Communication. Hillsdale, NJ: Lawrence Erlbaum; 1998:201-225.
24. Convertino G, Mentis H, Rosson M, Carroll J, Slavkovic A, Ganoe C. Articulating common ground in cooperative work: content and process. 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 5-8, 2008; Florence, Italy p. 1637-1646.
25. Monk A. Common ground in electronically mediated conversation. In: Carroll J, editor. Synthesis Lectures on Human-Centered Informatics. San Rafael, CA: Morgan & Claypool Publishers; 2009.
26. Convertino G, Carroll J. Toward experimental methods in collaborative computing research. In: Hesse-Biber S, editor. The Handbook of Emergent Technologies in Social Research. Oxford, UK: Oxford University Press; 2011:342-368.
27. Pinelle D, Gutwin C, Greenberg S. Task analysis for groupware usability evaluation: modeling shared-workspace tasks with the mechanics of collaboration. ACM Trans Comput Hum Interact 2003 Dec;10(4):281-311. [doi: 10.1145/966930.966932]
28. Señor IC, Fernández-Alemán J, Toval A. Personal health records: new means to safely handle health data? Computer 2012;45(11):27-33 [FREE Full text]
29. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]

30. Jones DA, Shipman JP, Plaut DA, Selden CR. Characteristics of personal health records: findings of the Medical Library Association/National Library of Medicine Joint Electronic Personal Health Record Task Force. J Med Libr Assoc 2010 Jul;98(3):243-249 [FREE Full text] [doi: 10.3163/1536-5050.98.3.013] [Medline: 20648259]
31. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(s1):S52-S63 [FREE Full text] [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21499543]
32. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
33. MacKenzie I. Human-Computer Interaction: An Empirical Research Perspective. Boston, MA: Morgan Kaufmann; 2013.
34. Pope C, Ziebland S, Mays N. Analysing qualitative data. In: Pope C, Mays N, editors. Qualitative Research in Health Care. Malden, MA: BMJ Books; Jan 8, 2006:63-81.
35. Gliem J, Gliem R. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. 2003 Presented at: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education; October 8-10, 2003; Columbus, OH.
36. Hsu J, Huang J, Fung V, Robertson N, Jimison H, Frankel R. Health information technology and physician-patient interactions: impact of computers on communication during outpatient primary care visits. J Am Med Inform Assoc 2005;12(4):474-480 [FREE Full text] [doi: 10.1197/jamia.M1741] [Medline: 15802484]

Abbreviations

APRN: advanced practice registered nurse
EMR: electronic medical record
HCI: human-computer interaction
HIT: health information technology
mHealth: mobile health
PHR: personal health record
RN: registered nurse
UX: user experience

Edited by G Eysenbach; submitted 28.05.18; peer-reviewed by R Lee, B Smaradottir; comments to author 11.07.18; accepted 17.07.18; published 01.11.18.

Please cite as:
Tunnell H, Faiola A, Bolchini D, Bartlett Ellis R
Simulated Clinical Encounters Using Patient-Operated mHealth: Experimental Study to Investigate Patient-Provider Communication
JMIR Mhealth Uhealth 2018;6(11):e11131
URL: https://mhealth.jmir.org/2018/11/e11131/
doi:10.2196/11131
PMID:30389652

Original Paper

A Tool to Measure Young Adults' Food Intake: Design and Development of an Australian Database of Foods for the Eat and Track Smartphone App

Lyndal Wellard-Cole[1], BHSci N&D, MHP; Melisa Potter[1], MND; Jisu (Joseph) Jung[2], BSE; Juliana Chen[1], MND; Judy Kay[2], PhD; Margaret Allman-Farinelli[1], PhD

[1]School of Life and Environmental Science, The University of Sydney, Sydney, Australia

[2]School of Information Technology, The University of Sydney, Sydney, Australia

Corresponding Author:
Lyndal Wellard-Cole, BHSci N&D, MHP
School of Life and Environmental Science
The University of Sydney
Level 4 East
Charles Perkins Centre D17
University of Sydney, 2006
Australia
Phone: 61 286274854
Email: lwel3754@uni.sydney.edu.au

Abstract

Background: Dietary assessment is reliant on the collection of accurate food and beverage consumption data. Technology has been harnessed to standardize recording and provide automatic nutritional analysis to reduce cost and researcher burden.

Objective: To better assess the diet of young adults, especially relating to the contribution of foods prepared outside the home, a database was needed to support a mobile phone data collection app. The app also required usability testing to assure ease of entry of foods and beverages. This paper describes the development of the Eat and Track app (EaT app) and the database underpinning it.

Methods: The Australian Food and Nutrient Database 2011-13, consisting of 5740 food items was modified. Four steps were undertaken: (1) foods not consumed by young adults were removed, (2) nutritionally similar foods were merged, (3) foods available from the 30 largest ready-to-eat food chains in Australia were added, and (4) long generic food names were shortened and simplified. This database was used to underpin the EaT app. Qualitative, iterative usability testing of the EaT app was conducted in three phases using the "Think Aloud" method. Responses were sorted and coded using content analysis. The System Usability Scale (SUS) was administered to measure the EaT app's perceived usability.

Results: In total, 1694 (29.51%) foods were removed from the Australian Food and Nutrient Database, including 608 (35.89%) ingredients, 81 (4.78%) foods already captured in the fast food chain information, 52 (3.07%) indigenous foods, 25 (1.48%) nutrients/dietary supplements, and 16 (0.94%) child-specific foods. The remaining 912 (53.84%) foods removed were not consumed by young adults in previous surveys or were "not defined" in the Australian Food and Nutrient Database. Another 220 (3.83%) nutritionally similar foods were combined. The final database consisted of 6274 foods. Fifteen participants completed usability testing. Issues identified by participants fell under six themes: keywords for searching, history list of entered foods, amounts and units, the keypad, food names, and search function. Suggestions for improvement were collected, incorporated, and tested in each iteration of the app. The SUS of the final version of the EaT app was rated 69.

Conclusions: A food and beverage database has been developed to underpin the EaT app, enabling data collection on the eating-out habits of 18- to 30-year-old Australians. The development process has resulted in a database with commonly used food names, extensive coverage of foods from ready-to-eat chains, and commonly eaten portion sizes. Feedback from app usability testing led to enhanced keyword searching and the addition of functions to enhance usability such as adding brief instructional screens. There is potential for the features of the EaT app to facilitate the collection of more accurate dietary intake data. The database and the app will be valuable dietary assessment resources for researchers.

(JMIR Mhealth Uhealth 2018;6(11):e12136) doi:10.2196/12136

KEYWORDS

diet surveys; smartphone; mobile phone; young adult

Introduction

Obesity is a global problem and although overweight and obesity is prevalent across all adult age groups in Australia, younger adults are a very vulnerable group. For example, Australian adults aged 18-30 years have had the greatest increase in Body Mass Index per year of any adult age group [1].

Young Australians are also the group who spend the highest proportion of their household income on fast foods and eating outside the home [2]. Frequent consumption of fast foods has been linked to poorer quality diets and weight gain [3,4]. However, there is little national data on the contribution these foods make to overall diets. While countries such as the United States have detailed data on the location of the purchase and consumption of foods, Australia does not. Therefore, there is no available consumption data for foods purchased from cafes, bars, restaurants, fast food outlets, takeout shops, and other food outlets. Research is needed to determine whether foods purchased and eaten outside the home are having an impact on young people's nutritional intake.

Dietary assessment is reliant on accurate unbiased collection of food and beverage consumption data. Typically, a participant must recall or record their dietary intake and provide portion sizes, which is a tedious process. In addition, to reveal the source of the meal preparation, further details must be recorded. The researcher must ensure the validity of the recording and convert dietary intake into nutrients. In recent years, technology has been harnessed to standardize recording and provide automatic nutritional analysis to reduce cost and researcher burden [5]. An example of technologies used to assist in recording intake is barcode scanning [6]. However, while scanning barcodes is useful in identifying foods purchased in packages, it is not applicable to most food prepared outside the home by the catering industry.

Recent studies have demonstrated that mobile phone apps can be valid measures of dietary intake at the population level [7-10]. However, a number of essential challenges must be addressed in the design of such apps. The starting point for creating an app for monitoring dietary intake must be the underpinning database of foods and beverages and their nutritional composition [11]. The usability of the interface for recording of foods is critical and equally important is the search functionality to support quick and easy locating of foods in the database. A comprehensive database of foods can be assembled, but it needs naming conventions that young adults readily recognize to enable better recording of the foods consumed [12]. Participants report that they never know if the food option they select is the appropriate one and they are confused by the large list of options to scroll through on an app like MyFitnessPal [12].

To better assess the diets of young adults in relation to the dietary contribution of foods prepared outside the home, a database was needed to support the development of a mobile phone app for dietary data collection. Assessment of potential participants' ability to search for food items was also required. Recording of the location of food purchase and tagging of food items would provide further insight into the outlets associated with food purchase and/or consumption.

This paper describes the development of the Eat and Track mobile phone app (EaT app), including the database underpinning it, and provides insights for other researchers seeking to develop apps for recording dietary intake.

Methods

Context

The development of the EaT app and the database that underpins it is part of the Measuring Young adults' Meals (MYMeals) Study, which aims to (1) determine how frequently young adults purchase and consume foods outside the home, and the types of foods they are purchasing and consuming, (2) the relative contributions of different food outlets to overall food and beverage intake, and (3) the extent that food and beverages consumed outside the home contribute to young adults' total energy and nutrient intakes [13]. The EaT app will collect 3 consecutive days of dietary intake data and information on where foods and drinks were obtained. This will allow analysis of the impact of foods eaten outside the home (for full methods, see the MYMeals Study protocol [13]).

The AUStralian Food and NUTrient Database (AUSNUT), 2011-13 [14], was used as a basis for the EaT app. The AUSNUT database was developed to analyze the foods and beverages that were consumed during the 2011-12 National Nutrition and Physical Activity Survey (NNPAS) and National Aboriginal and Torres Strait Islander Nutrition and Physical Activity Survey [15]. The food item names are mostly unbranded, and there was not always a distinction in the names between homemade foods and foods prepared ready-to-eat. Many meals and snacks from ready-to-eat food outlets cannot be readily identified using this database.

The EaT app database development (Stage 1) and formative usability testing of the EaT app (Stage 2) were conducted sequentially, but changes to the database were incorporated as a result of the usability testing. However, for clarity the Methods and Results for each part have been presented together.

Stage 1 Methods

Development of the Eat and Track App Database

The AUSNUT 2011-13 database [14] was used as the starting point of the EaT app database. The AUSNUT database contains the nutrition composition of 5740 foods. To modify the database, four steps were undertaken: (1) foods that would not be consumed by young adults were removed, (2) foods that were from the same subcategory food group and nutritionally very similar (±10% for energy and all 25 nutrients) and varieties of the same food were merged into one food item (eg, yellow and green apples), (3) foods available from the 30 largest ready-to-eat food chains in Australia were added, and (4) the

long string generic names were shortened and simplified to terms in common usage in the Australian community.

The database was reviewed to remove foods that were not likely to be eaten by potential participants. These include foods available only in certain states outside the EaT app's intended use (such as Indigenous foods available only in remote communities in the Northern Territory of Australia) [16], and food for specific groups outside of those who would be using the EaT app (eg, baby food and formula). The amounts of every food consumed by 18- to 30-year-olds during the NNPAS [17] were examined, and those that were not consumed were excluded. Examples of excluded foods included some offal meats and dietary supplements such as high-energy and high-protein formulas.

The AUSNUT food database organizes foods into three levels to assign each a unique ID: the major food group (eg, "Fruit products and dishes"), sub-major food group (eg, "Pome fruit"), and a minor category (eg, "apples") [18]. From that, a name is assigned to each food within the category. To identify nutritionally similar foods, the database was analyzed at the minor level. Any foods that were within ±10% for energy per 100 grams and all 25 different nutrients per 100 grams were identified and combined.

Due to the long character-length of many names in AUSNUT database, food names were shortened and simplified by an Accredited Practising Dietitian according to how foods are commonly referred to in Australia. For example, "Soft drink, lemonade, regular" was simplified to "Lemonade." Additionally, brand names were sourced from the AUSNUT 2011-13 Food Details File [19]. Some foods were renamed using multiple synonyms to increase likelihood of the end-user being able to find foods. For example, "Pizza, ham & pineapple" was replaced with "Ham & pineapple/Hawaiian pizza."

In addition to the generic foods from the AUSNUT database, an additional 2229 foods obtained from the 30 largest ready-to-eat food chains in Australia were included. Chains included in the database included traditional fast food (eg, burgers, fried chicken, pizza), fast casual chains, ice cream shops, bakery and salad chains, and beverage only or café chains.

Nutrition information for the ready-to-eat food chains was obtained from the companies, using methods detailed elsewhere [20]. However, this fast food nutrition information encompassed only the nutrients that are mandatory for Nutrition Information Panels in Australia (ie, energy, protein, total and saturated fats, carbohydrates, sugars, and sodium) [21] and not the comprehensive information from the AUSNUT database. Information was downloaded from chain websites, obtained in store or by request from company customer service enquiries. Data were collected in 2015 and checked for availability before being included in the EaT database. This was conducted by reviewing the menus of all the included chains and removing any foods that were no longer available. This was undertaken by one author (LWC) who was an Accredited Practising Dietitian in January 2017. All foods were named to include the chain name in the food name for the app, for example "Big Mac burger, McDonald's."

Portion Sizes and Measurements

The EaT app database contains information on portion sizes and measures for all foods. All foods were assigned either grams or millilitres as the unit of measurement. This was used in the EaT app for entry of the amount of the food. In addition, portion sizes were sourced from the companion AUSNUT food measures database [22]. This contains commonly eaten portion sizes reported in the NNPAS, often in metric household measures (eg, cup, tablespoon) [22]. To make recording of foods easier, we calculated and added portion sizes for each food using commonly eaten portions, such as standard glass sizes for alcoholic beverages. Where necessary, the portion sizes were calculated from the standard food weights and/or densities from the AUSNUT 2011-13 food measures database [22].

Quality Checking

Initial data quality checks were conducted on the additional foods added to AUSNUT by 2 independent researchers. Suspected errors in data were cross-checked from the original nutrition information and followed up with the chain, if necessary. Any fast foods with incomplete nutrient information were omitted from the database. Portion sizes for all foods were reviewed by an additional researcher independent of the database development process to ensure errors were detected and rectified. A final review of the entire database was conducted by 2 researchers before it was provided to the app development team for integration into the app.

Stage 1 Results

A total of 1694 foods (29.51%) were removed from the original AUSNUT database. Of the 1694 foods, 608 were ingredients that cannot be eaten without being made up (eg, dry soup mix or dried legumes and pulses) (35.89%), 81 were fast foods (4.78%) that were already captured in the fast food chain information, 52 were indigenous foods (3.07%), 25 were nutrients or dietary supplements (1.48%), and 16 were children's foods (including human breastmilk, infant and toddler formulas, and baby foods) (0.94%). The remaining 912 foods (53.84%) either were not consumed by young adults in the NNPAS or were included in AUSNUT as "not defined" or "not specified."

Once identified, foods that were within ±10% for energy and all 25 different nutrients were collapsed into single entries. For example, "apple, green" and "apple, golden" were combined into a single entry, "apple, green/golden." When there were small differences in nutrient composition, the collapsed items' nutrient contents were averaged. This process resulted in 220 foods (4%) being combined with other entries as they were nutritionally similar. The final database included in the EaT app consisted of 6274 foods, including 4046 foods (64%) from the AUSNUT database and 2229 branded ready-to-eat food chain items (36%).

Stage 2 Methods

The Eat and Track App

The EaT app was designed to draw on lessons from the previously validated electronic Dietary Intake Assessment (eDIA) app [7,8]. The app was developed using React-Native, a cross-platform app development platform that generates apps

for Android and iOS, which covers most mobile phone users. The EaT app stores every detailed interaction and food intake to a secure remote server in real time so that researchers are able to check and follow up logging. It was designed to reset the logging status every day at 3 a.m. to prevent participants from changing their data retrospectively.

The app interface (Figure 1) allows searches for common and brand names to increase potential for individuals to select appropriately matched foods. The interface was developed to include keyword functionality to improve food searches [13]. For example, when a user types "milk," a shortlist of all milk types appears. Keywords were added to the database for the most commonly eaten foods identified from the Australian Health Survey [17]. After selecting a food or drink, participants then record the amounts consumed and the location where the food was sourced. If participants cannot find a food listed in the app, they can manually enter it as a new food.

Iterative Usability Testing of the Eat and Track App

Qualitative, iterative usability testing was conducted in three phases during EaT app development. The "Think Aloud" usability testing method [23] was used to gain insights from participants on usability issues, such as the ease of finding specific foods using the search function and selecting the correct portion size.

Participants were recruited face-to-face and from posters on the University of Sydney campus, social media posts, and advertisements on the university's volunteer for a research study webpage. Potential participants were eligible if they were aged 18-30 years and could speak, write, and understand English. Participants were excluded if they had undertaken or were undertaking formal education in nutrition or information technology. Each participant was eligible to participate in only one phase of the study. Five participants were included in each phase and as an incentive they were entered in a prize draw for an Aus $50 gift voucher on completion of the study. Research has suggested that 80% of high severity usability issues can be determined with 5 participants, and 90% of issues with 10 participants [23-25]. Additionally, small samples are appropriate if there are to be multiple rounds of testing [24,25]. Therefore, 15 participants, 5 in each phase, were considered adequate for this study.

All participants provided consent. The usability studies were approved by the University of Sydney Human Ethics Research Committee (project number 2016/546).

In all phases, participants were provided with a device with the EaT app already installed to complete the testing. Participants were given information about the study and verbal instructions on how to use the EaT app and viewed a video demonstrating the Think Aloud method. All participants completed a demographic questionnaire. Figure 2 shows the three phases of the development.

In Phase 1, we tested as many of the features and functionalities of the app as possible, including how easily participants could find commonly consumed foods and enter correct portion sizes. Participants were provided with a list of 29 of the most commonly eaten foods for this age group from the latest Australian Health Survey [17] and relevant portion sizes and asked to search for the foods and enter these into the EaT app. Participants were instructed to verbalize their thoughts as they performed each task, but no assistance was provided. One researcher observed the participant, and another recorded the participants' comments. Comments were also audio-recorded for future analysis. All questions raised by the participants were answered after the testing was complete. Modifications to the app were made based on feedback from this phase, before further testing in Phase 2.

Phase 2 was designed to test the search functionality and keywords. Five different participants completed Phase 2. Participants were again given a mobile phone with the revised EaT app pre-installed and photographs of a hypothetical 2 days of food intake, including breakfast, a morning snack, lunch, an afternoon snack, dinner, and an evening snack, including drinks. Participants were instructed to enter these foods into the EaT app and verbalize their thought processes and questions. Further modifications were made based on this phase for summative testing in Phase 3.

Phase 3 tested participants' ability to estimate and enter portion sizes in a real-life situation and to test how long it would take to complete a day's worth of logging using the EaT app. We also tested the app's overall usability. In Phase 3, we presented 5 participants with 2 days of food intake using real foods that had been pre-weighed and measured. Each day consisted of breakfast, a morning snack, lunch, an afternoon snack, dinner, and an evening snack, including drinks (Figure 3). Participants were instructed to enter each food and drink into the EaT app as well as to estimate and enter the portion size for each food. The Australian Health Survey Food Model Booklet [26] containing to-scale images of different-sized foods and drinks was provided to help participants estimate portion sizes. The time it took participants to complete each task was recorded, to provide additional evidence about how difficult participants found the tasks.

An online version of the System Usability Scale (SUS) [27] was administered to participants in the third phase. The widely used SUS questionnaire measures a system/app's perceived usability through a series of 10 five-point Likert scale questions [27]. Additionally, participants were asked to rate the ease of estimating portion sizes of foods and beverages on a seven-point scale (extremely difficult to extremely easy). To collect feedback for further improvements to the EaT app, participants were asked two open-ended questions on their overall likes and dislikes about the app.

Figure 1. Eat and Track app screenshots.

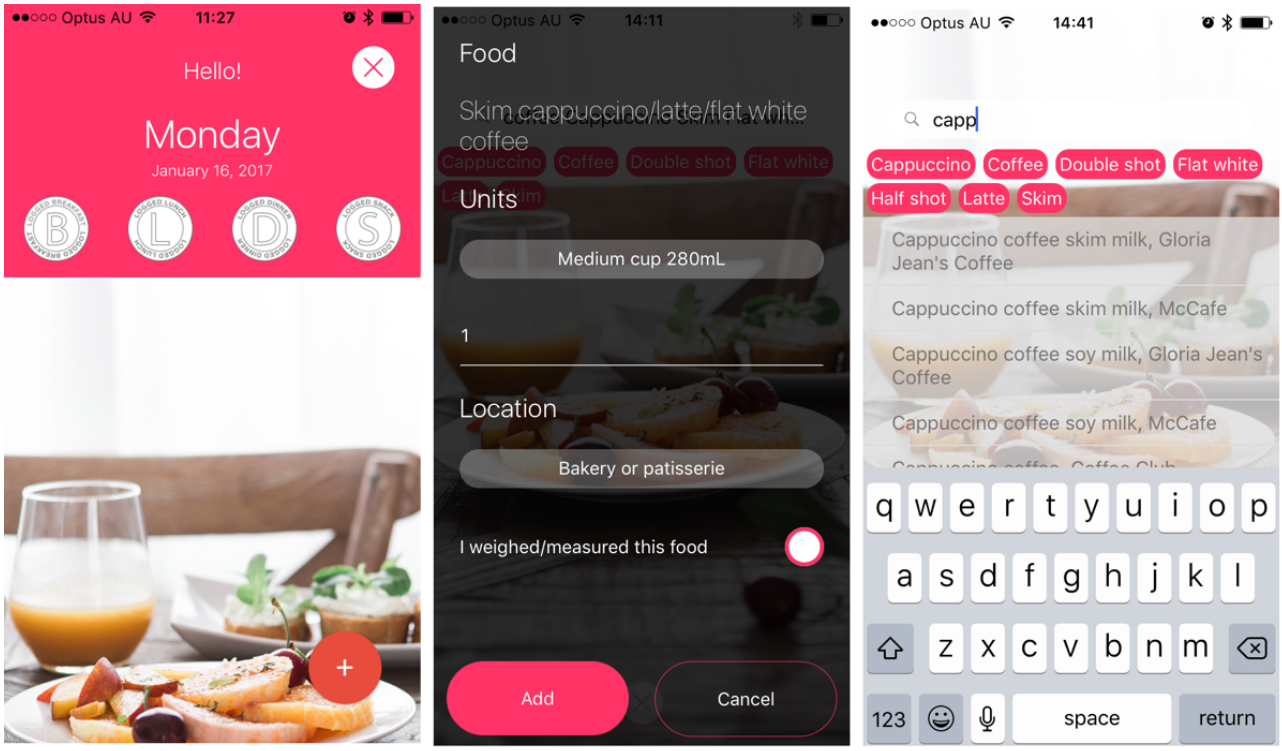


Figure 2. Iterative usability testing approach. EaT App: Eat and Track App.

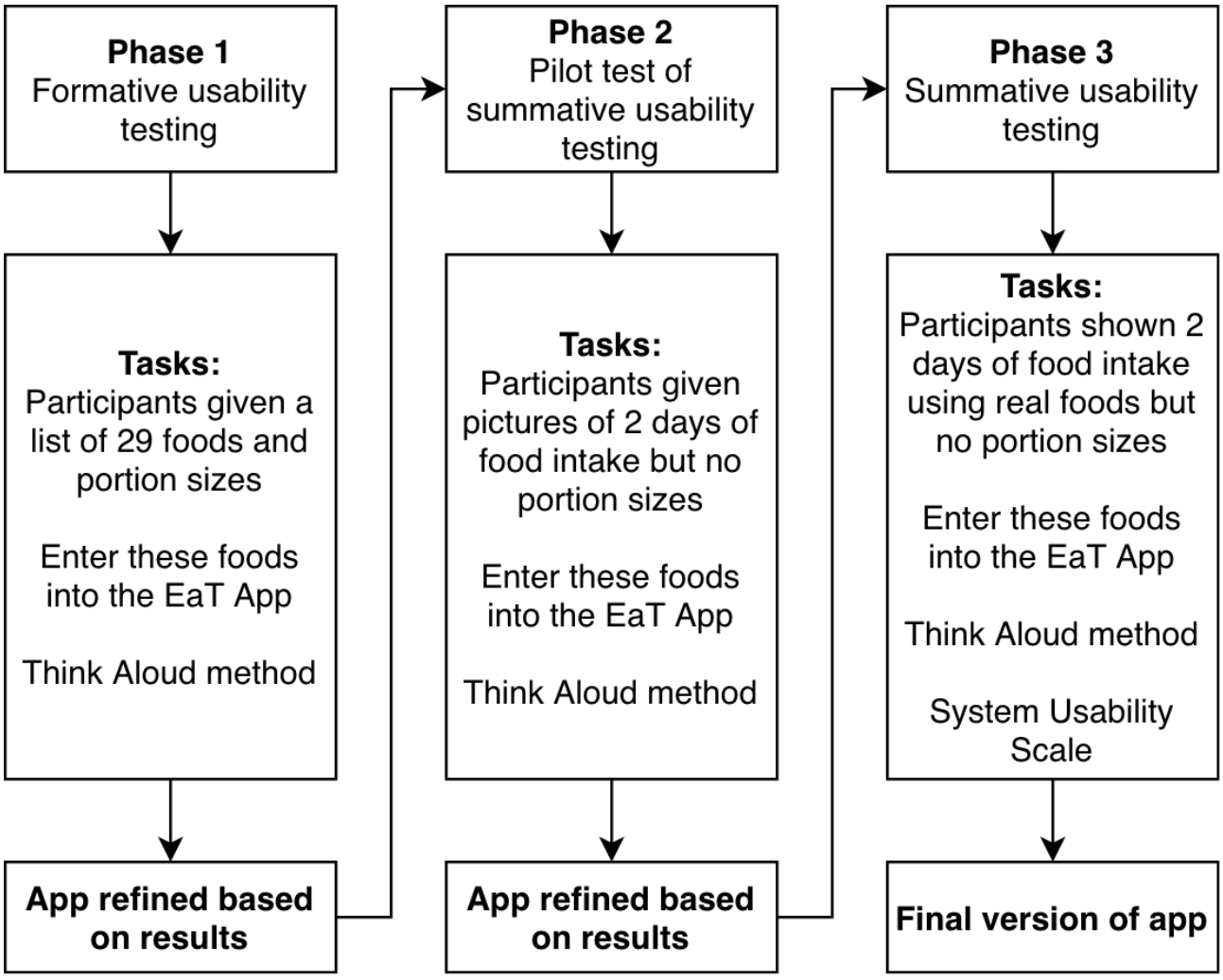


Figure 3. An example of the real foods presented to participants in Phase 3.

Data Analysis

Audio recordings of the usability testing and the researchers' observations were transcribed verbatim. Content analysis principles were used to sort and code data [28]. The transcription and observational data were reduced into themes that were identified a priori by the 2 researchers present in the usability testing sessions, in discussion with a third researcher. Once all the themes were identified, the data were aggregated to identify commonalities. Written data were then coded by a third researcher to ensure it was coded consistently.

The length of time taken for participants to complete each task, means, and standard deviations (SD) were calculated for each participant. Participants' scores for each item in the SUS were calculated to provide a score from 0 (very difficult to use) to 100 (very easy to use) [29]. A score above 68 is considered average [30]. Qualitative findings from the open-ended questions, including the question on portion size were organized into themes.

Stage 2 Results

A total of 15 participants aged 23-30 years completed the usability testing across the three phases. Eight participants were male. Nine participants had prior experience tracking health, nutrition, or physical activity behaviors.

Phases 1 and 2

Participants completed the tasks with varying levels of success (see Table 1). Common problems encountered by participants included incorrect spelling and/or punctuation despite being given the list of names with correct spelling (eg, omitting apostrophes or hyphens), foods being known as other names (eg, participants entering "hot chips" instead of "potato chips," or "Coke" instead of "Coca-Cola"), and words used in descriptions of many foods appearing in many searches (eg, searching for "milk" as opposed to "Cow's milk" resulted in "milk chocolate," "milk coffee," "coconut milk," and "milkshake," among others). Clear themes of usability problems emerged by the end of each iteration of testing, providing the basis for iterative refinement of the interface.

Table 1. Issues raised in Phases 1 and 2 of usability testing.

Success of participant and notes	Food item
All participants found these items easily and without help	
All participants found these items, though often the keywords participants entered were different. For 3 foods, they did not come up as the first item in the list.	• Black/green/chai tea no milk, regular/decaf; instant coffee, white • Light/mid-strength beer; red wine • Lemonade; lemon, lime and bitters; sports drink, bottled • Orange fruit drink 25% juice; apple fruit drink • Milo powder; drinking chocolate powder • Vegemite spread; peanut butter • Soy sauce; BBQ sauce; mayonnaise • Iceberg lettuce; mandarin; green cucumber; strawberry; watermelon, peeled • Cooked broccoli; cooked peas; cooked carrot, corn & pea/bean mix from frozen • Garden salad with cheese • Cooked white rice • Garlic/herb bread • Sweet plain biscuit, eg, Nice, Malt-o-Milk, Marie, Milk Arrowroot, plain Tiny Teddies, Morning Tea • Milk chocolate • Jelly lolly
All participants found these items with some searching difficulties	
Participants who entered only part of the search term (eg, “fried egg” instead of “fried chicken egg”) had to scroll through a long list to find the item. Often there were no keywords to narrow the search.	• Apple, red; orange • Orange juice; apple juice • Tomato sauce; honey; ground pepper • Coles regular margarine spread • Bacon middle rasher/shortcut fat trimmed, baked, roasted, fried, grilled or BBQed in butter/margarine • Processed ham and chicken luncheon meat • Fried chicken egg in butter • Wholemeal bread
Participants could not find the item if they entered “Coke.” One participant needed help as they misspelled “Coca” (“cocoa”).	• Coca-Cola
Some participants stated they would not think to enter “raw.” Entering only the keyword meant that participants had to scroll through a long list.	• Raw banana • Raw onion • Raw carrot • Raw avocado • Peeled Desiree/Coliban/red skin potato, raw • Raw common/Roma tomato
Some participants had difficulty finding these items	
One participant suggested that a keyword for “sugar” would make the search easier.	• Raw sugar
Some participants could not find this term as they omitted the apostrophe. Other participants tried searching with “milk” and had to scroll through a long list to find the item.	• Skim cow’s milk
Some participants entered “ham” and had to scroll through a long list to find the item.	• Leg ham
If “weetbix” was entered without the hyphen, participants could not find the item and needed help.	• Sanitarium Weet-Bix Original
If “ice cream” was entered, participants had to scroll through a long list as the list was not in alphabetical order. One participant entered “icecream,” which returned no items.	• Ice cream, all flavors
Many participants had difficulty finding these items without help	
Most participants had trouble with this task. There was confusion with entering the “amount” and “unit.”	• Tap water

Success of participant and notes	Food item
Some participants were confused by the many different spreads in the app, but eventually found the item after assistance from a researcher. Some participants entered what they would enter in a real-life situation. This included "butter," which required extensive scrolling, and "margarine," which did not yield the item at all.	• Regular fat dairy blend spread, eg, Western Star spreadable, Coles Spreadable Dairy Blend, Beautifully Butterfully Premium
Most participants had difficulty with this task. Some participants tried entering "hot chips" and were confused when the search returned no items.	• Deep fried potato chips, from restaurant/takeout

Table 2. Issues and improvements identified in Phase 3 of usability testing.

Function	Participant quote	Issue
Keywords	"There is inconsistency in that some products have keywords and some don't, for example, it would be helpful to have keywords for chips, such as 'potato,' 'crisps,' and 'plain' to help narrow the search down"	• Not all foods have keywords • Some participants did not notice pink keyword buttons • Other participants did not know what pink keyword buttons were for
History of previously added foods	"I didn't realize what the history list was, it wasn't obvious and the list of foods looked exactly the same as the list of foods provided when you are searching for a food"	• Several participants did not notice or use the history list
Amounts and units	"I find this function very unintuitive, I entered 600 under amount and then chose 600 mL bottle under unit, which actually meant that I had 600 x 600 mL bottles, which is not accurate"	• Several participants found the order of the Amount and Unit fields confusing
Keypad	"I have entered orange juice into the search bar and it has given me a list of options to choose from, but I can't figure out how to get the keypad out of the way so I can see the rest of the list"	• Several participants did not know how to minimize the keypad
Food names	"It would be helpful if the app recognized different synonyms of for food names, eg, chips/crisps, Coca Cola/Coke and chips/fries"	• Descriptions of foods inconsistent and confusing, making it harder to find the right item • App does not recognize some common synonyms of food names (eg, chips/fries) • App does not recognize the word "and" and the symbol "&" as the same thing
Search function	"Some items don't appear at the top of the list even when you've typed the exact phrase as the item on the list appears (eg, orange juice—the item is half way down the list)"	• App does not return 2-word matches as it does for single word matches

Modifications from Phase 1 included updating keywords to assist with searching for foods and drinks, and improving the search function to return single word matches first followed by all foods with the word entered appearing in alphabetical order. The feedback from Phase 2 led to several further updates to the app including improving the search functionality to ignore hyphens. Additionally, further improvements to the usability of the interface were also incorporated, including the addition of a sign-out button, changes to the look of the app, fixing app bugs, and allowing users to minimize the keypad so the shortlist was visible on the entire screen.

Phase 3

The mean time taken to enter 2 days of dietary intake into the EaT app in Phase 3 was 22 minutes, 13 seconds (SD 4.0). The slowest completion time was 30 minutes, 10 seconds, and the fastest completion time was 16 minutes, 22 seconds. Participants spent the least amount of time completing the task when they entered a fast food beverage and the most time completing the task was entering a pasta dish. Single food items were entered more quickly than mixed dishes or composite foods.

The Think Aloud protocol provided participants with the opportunity to provide insights into the difficulties they had and the reasons for these difficulties when using the app. They were also invited to make suggestions for improvements as they completed the usability testing. This highlighted several issues and improvements that would make using the app simpler. These are shown in Table 2.

Issues in this phase broadly fell under six themes: keywords for searching, the history list of previously entered foods, fields for amounts and units, keypad, food names, and search function. Some participants did not notice or understand the keyword buttons or the list of previously entered foods. The participants were unclear that there were two fields for amounts of food—one for the number and a separate one for the unit of measurement. Some participants could not minimize the keypad. Some still struggled with the food names, and even when a food was entered, the exact match did not necessarily show as number one in the options.

The EaT app obtained a mean SUS score of 69 (range 45-90), rating it about average for usability. Ten of the 15 participants (67%) rated it as above average for usability. After the third

phase, we concluded that remaining usability problems were minor and the app was ready for use.

Final Eat and Track App

The final EaT app incorporated the Phase 3 feedback. Specifically, more refinements were made to food names, for example adding the term "hot chips" to the name of "potato chips" to reflect what participants searched for in the usability testing, or adding alternative spellings of food names to enhance searchability. Additional keywords were added to aid narrowing lists of foods. The "unit" and "amount" fields order were swapped, to enable participants to enter the correct amounts, and pop-up instructional screens were introduced at sign-in to provide participants with brief information on how to use the app.

Discussion

Principal Considerations

A food and beverage database, including comprehensive ready-to-eat food chain data, has been developed to underpin the EaT app. This will support data collection on the eating-out habits of 18- to 30-year-olds in New South Wales, Australia's most populous state. The database development process has resulted in a database with commonly used food names and extensive coverage of foods from ready-to-eat chains. The usability testing allowed search and food naming issues to be identified and resolved, and refinement of app to improve usability.

Carter et al designed a database with inclusion of an extensive number of commercial foods available in the United Kingdom to support their online dietary assessment platform, myfood24 [11]. Including brand names is a way that database developers can minimize misinterpretation of foods [31]. The EaT app is different from other dietary data collection apps in that it includes a large number of branded ready-to-eat chain items that contribute a large proportion of foods consumed by our target group of young adults. Our database provides access to food names that the public understands, which may improve accuracy beyond that when generic names are used. While nutrition researchers may have less difficulty matching commercial foods to generic foods in a database, this may not be the case for the general public. Newer technology approaches transfer the burden of correctly identifying the food consumed from the researcher to the participant, hence the need for increased "user friendly" food and beverage names in an electronic dietary assessment tool. Similar to previous Australian research on an earlier version of AUSNUT, we renamed most foods and provided branded examples, which increased face validity of the final food names [31].

A database of packaged food and beverages has been developed in the United States [32]. Consistent with the current study, they found some limitations in the US Food Composition database for assessing population trends in the nutrients provided from packaged foods that were continually changing and being updated [32]. In our case, the deficiency is with the inclusion of extensive data on foods prepared outside the home, which appears to be dominant in young adults' food intake. The US research group was able to demonstrate that the use of the modified database identified changes in the energy density of foods from stores and vending machines that the generic database could not when applied to the same National Health and Nutrition Examination Survey dietary data [32]. This may be important in studying trends in food consumption over time and explanation of changes in population prevalence of overweight and obesity.

The EaT app includes a function to record where participants sourced their foods [13]. The app has been designed specifically with the intention of collecting data on eating food prepared outside the home and is unique in this aspect. Identifying the source of foods may help direct policies. In the state in Australia where the app will be used to collect data, menu labeling legislation is enforced where any chain food outlet with more than 20 stores in the state or 50 in Australia must display energy information for all menu items [33]. However, it is not known if other food outlets exempt from the regulation may make a larger contribution to overall intake of energy and deleterious nutrients. The EaT app will enable the examination of young adults' eating habits in relation to foods eaten outside the home. This is a research gap here in Australia, and results of the MYMeals Study will be used to shape health promotion messages for improving young adults' food choices when eating outside the home.

Strengths and Limitations

As suggested by other researchers, a limitation of food composition databases is that they are correct only at one time-point, and there is an ongoing challenge in updating them [11]. In particular, the ready-to-eat chains constantly offer new, limited-time-only menu items [34], meaning that the app will not include the most recent items. Similarly, the EaT app database is based on the foods consumed as part of the NNPAS, which was collected in 2011-2012 [35]. In 2015, there were 4143 new grocery items introduced in Australian supermarkets [36]. Considering this large number of new items, there will be many items consumed by future participants and these will not be in the database. However, if participants cannot find a food, such as a new ready-to-eat chain or grocery item, the app allows them to manually enter it as a new food. This flags the food to the research team for follow-up. New grocery items will also be identified this way.

Other researchers developing databases to underpin dietary assessment methods can learn from our study. Naming conventions for database entries must contain brand and colloquial names for foods to improve participants' ability to find foods. Approximate spelling matching functionality or alternative spelling of food names should be incorporated to account for spelling errors. The EaT app is a research tool that could be used to investigate other aspects of young adults' dietary consumption, for example foods prepared at home, or further tested in other population groups.

In the future, the EaT app will be validated by comparing dietary data collected using the app with 24-hour recalls and used to assess the eating-out habits of 1000 young adults [13]. If finances permit, biomarkers such as doubly labeled water and urinary nitrogens could be used. The EaT app could be further

developed for other future purposes, including behavior change programs by incorporating feedback to participants. By validating the app, it can be determined whether an updated database will be required in future versions and uses of the app.

Conclusions

This paper has described the development of the EaT mobile phone dietary assessment app and the database that underpins it. The database contains comprehensive nutrition information about foods from many large Australian ready-to-eat chains, including commonly eaten portion sizes. Feedback from app usability testing led to updated database naming, enhanced keyword searching, and the addition of functions to enhance usability, such as adding brief instructional screens. There is potential for the features of the EaT app to facilitate the collection of more accurate dietary intake data. The database and the app will be valuable dietary assessment resources for researchers.

Conflicts of Interest

None declared.

References

1. Hayes A, Gearon E, Backholer K, Bauman A, Peeters A. Age-specific changes in BMI and BMI distribution among Australian adults using cross-sectional surveys from 1980 to 2008. Int J Obes (Lond) 2015 Aug;39(8):1209-1216. [doi: 10.1038/ijo.2015.50] [Medline: 25869604]
2. Australian Bureau of Statistics. 6530.0 - Household Expenditure Survey, Australia: Summary of Results, 2015-16 Canberra. 2017. URL: http://www.abs.gov.au/Ausstats/abs@.nsf/0/16D1D54EA86276D0CA2581990017282E?OpenDocument [accessed 2018-09-04] [WebCite Cache ID 72AkC0CBc]
3. Prentice A, Jebb S. Fast foods, energy density and obesity: a possible mechanistic link. Obes Rev 2003 Nov;4(4):187-194. [Medline: 14649369]
4. Lachat C, Nago E, Verstraeten R, Roberfroid D, Van Camp J, Kolsteren P. Eating out of home and its association with dietary intake: a systematic review of the evidence. Obes Rev 2012 Apr;13(4):329-346. [doi: 10.1111/j.1467-789X.2011.00953.x] [Medline: 22106948]
5. Rollo M, Williams R, Burrows T, Kirkpatrick S, Bucher T, Collins C. What Are They Really Eating? A Review on New Approaches to Dietary Intake Assessment and Validation. Curr Nutr Rep 2016 Oct 15;5(4):307-314. [doi: 10.1007/s13668-016-0182-6]
6. Maringer M, Wisse-Voorwinden N, Veer P, Geelen A. Food identification by barcode scanning in the Netherlands: a quality assessment of labelled food product databases underlying popular nutrition applications. Public Health Nutr 2018 Jul 02:1-8. [doi: 10.1017/S136898001800157X] [Medline: 29962361]
7. Rangan AM, O'Connor S, Giannelli V, Yap ML, Tang LM, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): Comparison of a Mobile Phone Digital Entry App for Dietary Data Collection With 24-Hour Dietary Recalls. JMIR Mhealth Uhealth 2015 Oct 27;3(4):e98 [FREE Full text] [doi: 10.2196/mhealth.4613] [Medline: 26508282]
8. Rangan A, Tieleman L, Louie J, Tang L, Hebden L, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): relative validity of a mobile phone application to measure intake of food groups. Br J Nutr 2016 Dec;115(12):2219-2226. [doi: 10.1017/S0007114516001525] [Medline: 27121045]
9. Bucher Della Torre S, Carrard I, Farina E, Danuser B, Kruseman M. Development and Evaluation of e-CA, an Electronic Mobile-Based Food Record. Nutrients 2017 Jan 18;9(1) [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
10. Pendergast F, Ridgers N, Worsley A, McNaughton S. Evaluation of a smartphone food diary application using objectively measured energy expenditure. Int J Behav Nutr Phys Act 2017 Dec 14;14(1):30 [FREE Full text] [doi: 10.1186/s12966-017-0488-9] [Medline: 28288657]
11. Carter CM, Hancock N, Albar AS, Brown H, Greenwood CD, Hardie JL, et al. Development of a New Branded UK Food Composition Database for an Online Dietary Assessment Tool. Nutrients 2016 Aug 05;8(8) [FREE Full text] [doi: 10.3390/nu8080480] [Medline: 27527214]
12. Chen J, Berkman W, Bardouh M, Ng CYK, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2018 May 25;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]
13. Wellard-Cole L, Jung J, Kay J, Rangan A, Chapman K, Watson WL, et al. Examining the Frequency and Contribution of Foods Eaten Away From Home in the Diets of 18- to 30-Year-Old Australians Using Smartphone Dietary Assessment (MYMeals): Protocol for a Cross-Sectional Study. JMIR Res Protoc 2018 Jan 26;7(1):e24 [FREE Full text] [doi: 10.2196/resprot.9038] [Medline: 29374002]
14. Food Standards Australia New Zealand. AUSNUT 2011-2013. Canberra, Australia; 2016 Apr. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/foodnutrient/Pages/default.aspx [accessed 2018-03-06] [WebCite Cache ID 6v9Z4xkkv]

15. Food Standards Australia New Zealand. Food Nutrient Database. Canberra, Australia; 2018. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/ausnutdatafiles/Pages/foodnutrient.aspx [accessed 2018-03-06] [WebCite Cache ID 1536034484718722]
16. Food Standards Australia New Zealand. Development of additional nutrient profiles for foods and beverages consumed in the NATSINPAS. Canberra, Australia; 2018. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/foodnutrient/Pages/devnutrientsNATSINPAS.aspx [accessed 2018-03-06] [WebCite Cache ID 1536034654895700]
17. Australian Bureau of Statistics. 4364.0.55.007 - Australian Health Survey: Nutrition First Results - Foods and Nutrients, 2011-12. Canberra, Australia; 2014 May 09. URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4364.0.55.007Appendix22011-12 [accessed 2018-09-04] [WebCite Cache ID 1536035260315739]
18. Food Standards Australia New Zealand. AUSNUT 2011-13 food and dietary supplement classification system. Canberra, Australia; 2018. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/ausnutdatafiles/Pages/foodclassification.aspx [accessed 2018-08-21] [WebCite Cache ID 1536035544815132]
19. Food Standards Australia New Zealand. AUSNUT 2011-13 Food Details File. Canberra, Australia; 2014. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/ausnutdatafiles/Pages/fooddetails.aspx [accessed 2018-10-28] [WebCite Cache ID 73V2OxOtN]
20. Wellard-Cole L, Goldsbury D, Havill M, Hughes C, Watson W, Dunford E, et al. Monitoring the changes to the nutrient composition of fast foods following the introduction of menu labelling in New South Wales, Australia: an observational study. Public Health Nutr 2018 Apr;21(6):1194-1199. [doi: 10.1017/S1368980017003706] [Medline: 29262878]
21. Australian Government. Australia New Zealand Food Standards Code - Standard 1.2.8 Nutrition Requirements. Canberra, Australia; 2017 Apr 13. URL: https://www.legislation.gov.au/Series/F2015L00395 [accessed 2018-09-04] [WebCite Cache ID 1536035697851003]
22. Food Standards Australia New Zealand. AUSNUT 2011-13 Food Measures Database. Canberra, Australia; 2014. URL: http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/ausnutdatafiles/Pages/foodmeasures.aspx [accessed 2018-10-28] [WebCite Cache ID 73V2reNp4]
23. Nielsen J. Usability Engineering. Fremont, CA: Elsevier Science; 1994.
24. Davids M, Harvey J, Halperin M, Chikte U. Determining the number of participants needed for the usability evaluation of e-learning resources: A Monte Carlo simulation. Br J Educ Technol 2015 Aug 13;46(5):1051-1055. [doi: 10.1111/bjet.12336]
25. Turner C, Lewis J, Nielsen J. Determining Usability Test Sample Size. International Encyclopedia of Ergonomics and Human Facotrs 2006;2(3):3084-3088.
26. Australian Bureau of Statistics. Australian Health Survey. Food Model Booklet. Canberra, Australia; 2010. URL: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/05E75E65AD98B1C0CA257CD20014B24B/$File/food%20model%20booklet.pdf [accessed 2018-10-29] [WebCite Cache ID 73XBD1vpF]
27. US Department of Health and Human Services. System Usability Scale (SUS). 2016. URL: https://www.usability.gov/node/368 [accessed 2018-09-04] [WebCite Cache ID 1536035814791735]
28. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
29. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Scales 2009;4(3):114-123.
30. Sauro J. A Practical Guide to the System Usability Scale. [online]: CreateSpace Independent Publishing Platform; 2011.
31. Probst Y, Tapsell L. What to ask in a self-administered dietary assessment website: The role of professional judgement. Journal of Food Composition and Analysis 2007;20(8):696-703.
32. Poti JM, Yoon E, Hollingsworth B, Ostrowski J, Wandell J, Miles DR, et al. Development of a food composition database to monitor changes in packaged foods and beverages. J Food Compost Anal 2017 Dec;64(Pt 1):18-26. [doi: 10.1016/j.jfca.2017.07.024] [Medline: 29230079]
33. New South Wales Government. Food Act 2003 No 43. URL: https://www.legislation.nsw.gov.au/inforce/21dd14e3-569d-eb35-e54b-cd410db63d52/2003-43.pdf [WebCite Cache ID 1536036422701791]
34. Glanz K, Resnicow K, Seymour J, Hoy K, Stewart H, Lyons M, et al. How major restaurant chains plan their menus: the role of profit, demand, and health. Am J Prev Med 2007 May;32(5):383-388. [doi: 10.1016/j.amepre.2007.01.003] [Medline: 17478263]
35. Australian Bureau of Statistics. 4363.0.55.001 - Australian Health Survey: Users' Guide, 2011-13. Canberra, Australia; 2014. URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4363.0.55.001Chapter6502011-13 [accessed 2017-11-22] [WebCite Cache ID 6v9ZFIq3t]
36. Spiteri S, Olstad DL, Woods J. Nutritional quality of new food products released into the Australian retail food market in 2015 - is the food industry part of the solution? BMC Public Health 2018 Dec 07;18(1):222 [FREE Full text] [doi: 10.1186/s12889-018-5127-0] [Medline: 29415698]

Abbreviations

AUSNUT: AUStralian Food and NUTrient Database

EaT app: Eat and Track app
MYMeals Study: Measuring Young adults' Meals Study
NNPAS: National Nutrition and Physical Activity Survey
SUS: System Usability Scale

Edited by G Eysenbach; submitted 06.09.18; peer-reviewed by Y Probst, K Appleton; comments to author 08.10.18; revised version received 11.10.18; accepted 17.10.18; published 07.11.18.

Please cite as:
Wellard-Cole L, Potter M, Jung J(, Chen J, Kay J, Allman-Farinelli M
A Tool to Measure Young Adults' Food Intake: Design and Development of an Australian Database of Foods for the Eat and Track Smartphone App
JMIR Mhealth Uhealth 2018;6(11):e12136
URL: http://mhealth.jmir.org/2018/11/e12136/
doi:10.2196/12136
PMID:30404768

Original Paper

A Cardiopulmonary Monitoring System for Patient Transport Within Hospitals Using Mobile Internet of Things Technology: Observational Validation Study

Jang Ho Lee[1], MD; Yu Rang Park[2], PhD; Solbi Kweon[2], MS; Seulgi Kim[1], MS; Wonjun Ji[1], MD; Chang-Min Choi[1,3], MD, PhD

[1]Department of Pulmonology and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[2]Department of Biomedical System Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[3]Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Chang-Min Choi, MD, PhD
Department of Pulmonology and Critical Care Medicine
Asan Medical Center
University of Ulsan College of Medicine
88 Olympic-ro, 43-gil
Songpa-gu
Seoul, 05505
Republic of Korea
Phone: 82 2 3010 5902
Email: ccm@amc.seoul.kr

Related Article:

Correction of: http://mhealth.jmir.org/2018/12/e12864/

Abstract

Background: During intrahospital transport, adverse events are inevitable. Real-time monitoring can be helpful for preventing these events during intrahospital transport.

Objective: We attempted to determine the viability of risk signal detection using wearable devices and mobile apps during intrahospital transport. An alarm was sent to clinicians in the event of oxygen saturation below 90%, heart rate above 140 or below 60 beats per minute (bpm), and network errors. We validated the reliability of the risk signal transmitted over the network.

Methods: We used two wearable devices to monitor oxygen saturation and heart rate for 23 patients during intrahospital transport for diagnostic workup or rehabilitation. To determine the agreement between the devices, records collected every 4 seconds were matched and imputation was performed if no records were collected at the same time by both devices. We used intraclass correlation coefficients (ICC) to evaluate the relationships between the two devices.

Results: Data for 21 patients were delivered to the cloud over LTE, and data for two patients were delivered over Wi-Fi. Monitoring devices were used for 20 patients during intrahospital transport for diagnostic work up and for three patients during rehabilitation. Three patients using supplemental oxygen before the study were included. In our study, the ICC for the heart rate between the two devices was 0.940 (95% CI 0.939-0.942) and that of oxygen saturation was 0.719 (95% CI 0.711-0.727). Systemic error analyzed with Bland-Altman analysis was 0.428 for heart rate and –1.404 for oxygen saturation. During the study, 14 patients had 20 risk signals: nine signals for eight patients with less than 90% oxygen saturation, four for four patients with a heart rate of 60 bpm or less, and seven for five patients due to network error.

Conclusions: We developed a system that notifies the health care provider of the risk level of a patient during transportation using a wearable device and a mobile app. Although there were some problems such as missing values and network errors, this paper is meaningful in that the previously mentioned risk detection system was validated with actual patients.

(JMIR Mhealth Uhealth 2018;6(11):e12048) doi:10.2196/12048

KEYWORDS

wearable device; patient safety; intrahospital transport; oxygen saturation; heart rate; mobile application; real-time monitoring

Introduction

As medicine continues to progress, patient safety has become more important. Intrahospital transport (IHT) is necessary during clinical practice. Pulmonary rehabilitation is regarded as one of the most important interventions for chronic pulmonary disease patients [1-4]. Nevertheless, adverse events during IHT and rehabilitation are inevitable. One study reported that 1.7% of critically ill patients suffered adverse events during IHT, defined as life-threatening events. Another study reported that adverse events, from equipment problems to life-threatening situations, occurred in 79.8% of patients during IHT. [2,5-7]. Although it is desirable for clinicians to accompany and observe their patients during IHT to reduce these events, this is impossible in the real world due to medical resource limitations.

Recent changes in the mobile and internet environment have an effect on the medical field [8-11]. Due to progress in high-speed data transmission capabilities and the ability to wirelessly connect to external devices, several telemonitoring techniques have been developed in various fields [12-14]. These techniques offer immediate information about patients to clinicians and facilitate adequate management of patients.

There are several telemonitoring solutions for oxygen saturation and heart rate, but most of them were developed for long term rather than immediate management [15,16]. The Prince 100-H wrist oximeter and SpO_2 monitor version 0.23 were developed for simultaneous supervision of oxygen saturation and heart rate during patient transport or rehabilitation and can notify clinicians of adverse events.

In this study, we aimed to answer two questions: (1) Are wearable devices and mobile applications suitable for recognizing risk during transit? and (2) Is transmission of the risk signal over the network reliable? To address these questions, we collected biometrics data during the transport of respiratory medicine inpatients using wearable devices and a mobile app. We developed a risk detection algorithm to analyze the collected data and notified the health care provider when necessary.

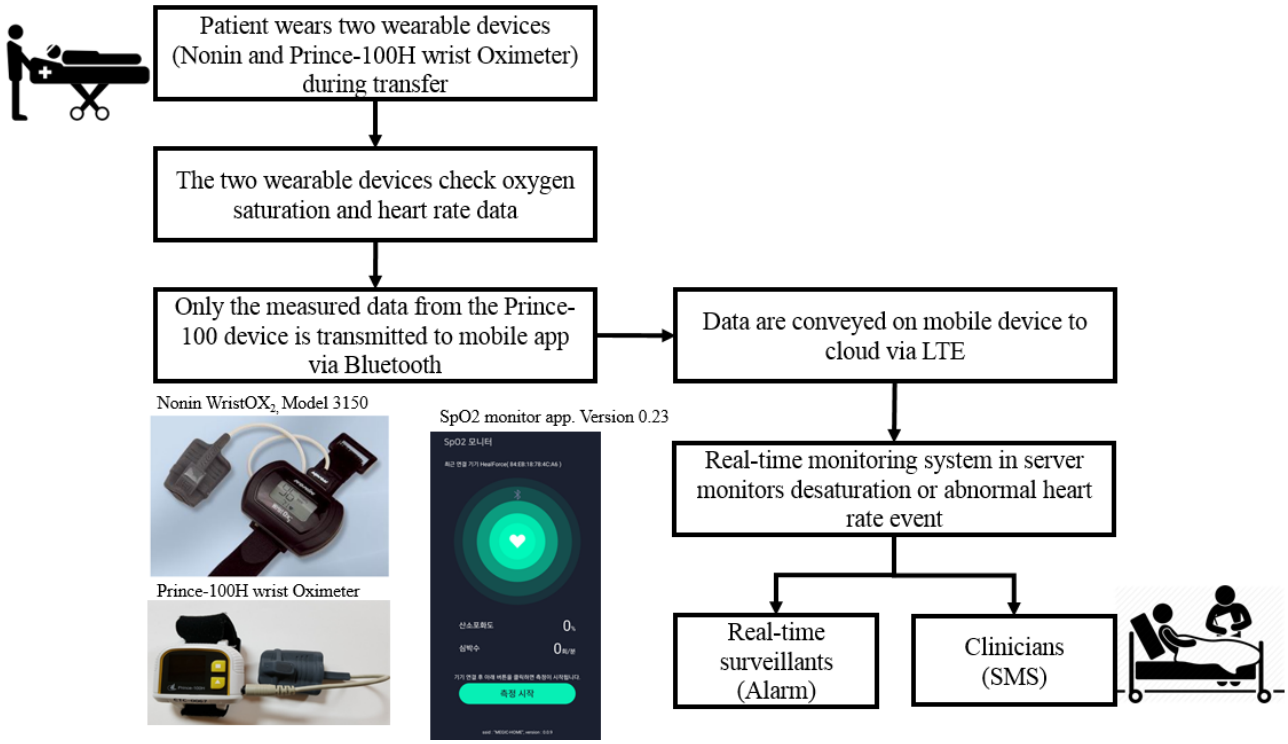
Methods

Study Design

We performed a single-center study at Asan Medical Center in South Korea. Patient screening was based on admission to the pulmonology ward between May 16, 2018 and May 31, 2018. Exclusion criteria were as follows: (1) patients in the acute phase based on the clinicians' judgment, and (2) patients not transported for work up or rehabilitation.

Oxygen saturation and pulsation were measured in real time using portable oxygen saturation measurement equipment for patients with risk factors that reduce oxygen saturation and were measured in the hospital for diagnostic purposes or rehabilitation. The measured data were collected in real time on a connected mobile phone and transmitted to a real-time monitoring system via the Internet of Things (IoT). The transmitted data were checked in real time on monitoring equipment on the ward where the patient was hospitalized by clinicians on the study team. When there was a risk, an alarm system was activated, allowing the staff to immediately (within 1 minute) identify and respond to the danger. Using a wearable sensing device and mobile phone app, we collected data from 23 patients at Asan Medical Center in Seoul and analyzed the risk factors during patient transport (Figure 1).

Figure 1. Data flow for risk signal detection system during patient transfer through wearable device and a mobile app.



Device and Mobile App

When patients needed to be transported for rehabilitation or study, we employed the Prince-100H wrist oximeter and the Nonin for monitoring. Both devices check oxygen saturation and heart rate in real time. The Prince-100H wrist oximeter generates SpO_2 and pulse data at 1 record per second, while the Nonin, which is approved by the US Food and Drug Administration (FDA), generates one record every 4 seconds. Only measured data from the Prince-100H wrist oximeter were transmitted to the mobile app, SpO_2 monitor version 0.23, via Bluetooth, and delivered to the cloud over the network. When the monitoring system detected desaturation or abnormal heart rate events, a notification was transmitted to real-time surveillants by displaying an alarm on the monitor and to clinicians by short message service (SMS) text messaging. We selected SMS text messaging as tools for notification to clinicians because our judgment was that SMS text messaging was relatively less affected by the mobile environment. Notification of a risk signal occurred only in the following four cases: (1) oxygen saturation less than 90%, (2) heart rate greater than 140 beats per minute (bpm), (3) heart rate less than 60 bpm, and (4) network error.

The device and mobile phone app were used for IHT during the hospitalization of 23 patients. We investigated the possible risk factors and disease severity during transport to determine the efficiency of the monitoring and transmission system. Furthermore, the risk factors during transport were identified through comparative analysis of the patients with and without the hazards identified during transport.

This study was approved by the institutional review board of the Asan Medical Center (IRB no. 2018-0480). We obtained informed consent from all study participants.

Data Description

In this study, we collected three types of data: (1) clinical information of patients collected on admission to Asan Medical Center, (2) the data collected during the clinical trial through the case report form for participants, and (3) the patient's real-time pulse and oxygen saturation by pulse oximetry (SpO_2) values were measured using two wearable devices (Nonin and Prince-100H wrist oximeters). The clinical information for the patients included age, sex, body mass index, smoking history, oxygen use before trial, and underlying disease (including diabetes mellitus, hypertension, tuberculosis history, respiratory disease, arrhythmia, lung cancer, and other malignancies). The data collected during the clinical trial included the reason for study device application, vital signs before the trial, and the results of a pulmonary function test given to available patients. We only employed the study device for monitoring during IHT or rehabilitation.

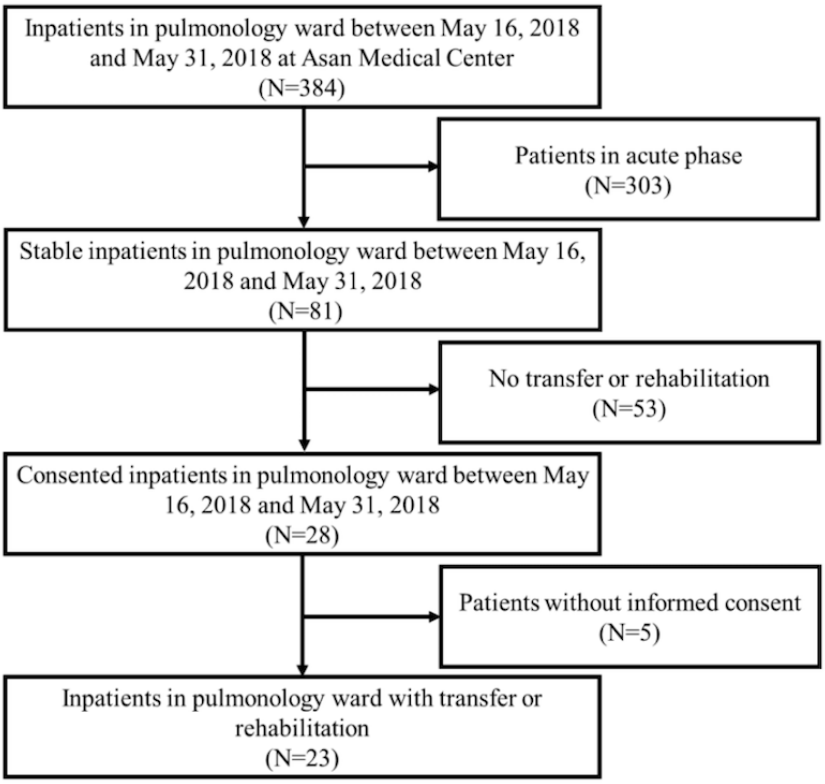
Data Analysis

Figure 2 shows the patient selection flowchart for the study. Among 384 patients admitted to the pulmonology ward between May 16, 2018 and May 31, 2018, we excluded the following patients: (1) 303 patients in the acute phase, (2) 53 patients without in-hospital transfer or a rehabilitation schedule, and (3) five patients without informed consent. Twenty-three patients were finally selected for this study.

To analyze the interdevice agreement rate, the Nonin and Prince-100H data collected every 4 seconds, which is the data generation criterion for the Nonin. If the Prince-100H did not have a value at exactly the same time, imputation was performed at an average of ±1 second. Based on matching data, the Pearson correlation coefficient (r) between the different wearable devices was determined with linear regression. The intraclass correlation coefficients (ICCs) were used to assess the relationships between

the two wearable devices [17,18]. Point estimates of the ICCs were interpreted as follows: excellent (0.75-1), modest (0.4-0.74), or poor (0-0.39). The Bland-Altman method was also used to measure the agreement of the pulse and SpO_2 values for the two wearable devices. The Student *t* test was used to compare differences in the groups for abnormal and unmatched signals. All reported *P* values were two-sided, and *P* values less than .05 were considered significant. All statistical analyses were performed using R version 3.5.0. We expressed the categorical variables as numbers with the proportion of subjects and continuous variables presented as means with standard deviations.

Figure 2. Recruitment for patient safety study related to patient transfer within hospital.



Results

Overall Characteristics

Twenty-three patients consented to inclusion in this study. The baseline characteristics of the enrolled patients are summarized in Table 1. Categorical variables are expressed as a number with the proportion of participants. Continuous variables are expressed as means with standard deviations. Pulmonary function test results were extracted from 18 patients, because five did not take a pulmonary function test before the study. The mean age was 64.4 (SD 11.1) years, and 12 men were included. The mean body mass index was 23.7 (SD 2.4) kg/m^2. Seven patients had underlying respiratory disease and three of them used an oxygen supply before the study. Nineteen patients had malignancies, and none had underlying arrhythmia. Twenty patients used the study device and app during IHT for diagnostic purpose, and the other three patients used the device for rehabilitation. In the baseline pulmonary function test of 18 patients, the mean forced expiratory volume in first second of expiration (FEV_1) was 2.1 (SD 0.7) L, mean forced vital capacity (FVC) was 3.1 (SD 0.9) L, mean FEV_1/FVC was 70.4% (SD 11.0%), and mean diffusing capacity for carbon monoxide (DLCO) was 67.1% (SD 20.3%).

Difference in Data Transmission by Network Type

Of the 23 patients, only two used the Wi-Fi network and 21 used the LTE network to transmit data generated from the Prince-100H device to the cloud (Multimedia Appendix 1). The total measurement time of this study was 875.1 minutes (mean 38.0, SD 29.4 minutes). During this time, the Nonin produced 14,161 records and the Prince-100 produced 49,282 records. Since the Prince-100H measures SpO_2 and pulses data in 1 second units, this was matched with one record every 4 seconds from the Nonin. The mean value for 109.34 records before the imputation was “not available” and the mean value of 24.21 records after imputation was the not available value. The remaining mean not available ratio was 3.79%. Patients with the highest not available ratios were in the following order: P05 (28.02%), P07 (14.56%), and P03 (12.74%). Two patients with Wi-Fi transmission had higher not available ratios before imputation and their not available ratio values after imputation were lower than those for LTE patients.

Correlation Analysis of Two Wearable Devices

The SpO_2 and pulse variables measured by the Prince-100H and Nonin are given in Table 2. The pulse ICC between the two devices was 0.940 (95% CI 0.939-0.942), which indicated “excellent” agreement, and the SpO_2 ICC was 0.719 (95% CI 0.711-0.727), which indicated “good” agreement (Table 2). In

addition, Bland-Altman analysis of the pulse revealed that the systematic error was low at 0.428 compared to –1.404 for SpO_2 (Figure 3). The 95% limit of agreement was in the –9.344 to 10.201 range for the pulse and –5.496 to 2.688 for the SpO_2.

Table 1. Basic characteristics of patients (N=23).

Variable	Total
Age (years), mean (SD)	64.4 (11.1)
Male, n (%)	12 (52.2)
Body mass index (kg/m^2), mean (SD)	23.7 (2.4)
Smoking history, n (%)	13 (56.5)
Oxygen use before trial, n (%)	3 (13.0)
Vital sign before trial, mean (SD)	
Systolic blood pressure (mm Hg)	117.1 (13.5)
Diastolic blood pressure (mm Hg)	73.0 (7.3)
Heart rate	75.0 (11.3)
Respiratory rate	18.3 (1.4)
Underlying disease, n (%)	
Diabetes mellitus	5 (21.7)
Hypertension	9 (39.1)
Tuberculosis history	3 (13.0)
Respiratory disease	7 (30.4)
Arrhythmia	0 (0.0)
Lung cancer	17 (73.9)
Other malignancy	2 (8.7)
Reason for study device application, n (%)	
Intrahospital transport	20 (87.0)
Rehabilitation	3 (13.0)
Pulmonary function test (N=18), mean (SD)	
FVC[a] (L)	3.1 (0.9)
FEV_1[b] (L)	2.1 (0.7)
FEV_1/FVC (%)	70.4 (11.0)
DLCO[c] (%)	67.1 (20.3)

[a]FVC: forced vital capacity.

[b]FEV_1: forced expiratory volume in 1 sec.

[c]DLCO: diffusing capacity for carbon monoxide.

Table 2. Characteristics of the patients with abnormal signals ("yes") versus those without ("no") during transport (N=23).

Variables and categories	Yes (n=14)	No (n=9)	*P* value[a]
Age (years), mean (SD)	62.07 (12.91)	68 (6.75)	.17
Sex, n (%)			.40
Male	6 (43.86)	6 (66.67)	
Female	8 (57.14)	3 (33.33)	
Weight, mean (SD)	62.71 (11.76)	61.11 (7.47)	.69
Body mass index, mean (SD)	24.01 (2.71)	23.33 (1.91)	.49
Oxygen use before study n (%)	3 (21.43)	0 (0)	.25
Study device application: rehabilitation n (%)	3 (21.43)	0 (0)	.25
Smoking n (%)			.22
Nonsmoker	8 (57.14)	2 (22.22)	
Ex-smoker	3 (21.43)	5 (55.56)	
Current-smoker	3 (21.43)	2 (22.22)	
Underlying disease n (%)			
Diabetes mellitus	3 (21.41)	2 (22.22)	>.99
Hypertension	4 (28.57)	5 (55.56)	.38
History of tuberculosis	0 (0)	3 (33.33)	.05
Pulmonary disease	4 (28.57)	3 (33.33)	>.99
Arrhythmia	0 (0)	0 (0)	>.99
Lung cancer	10 (71.43)	7 (77.78)	>.99
Other malignancy	1 (7.14)	1 (11.11)	>.99
Pulmonary function test, mean (SD)			
FVC[b]	2.99 (0.97)	3.22 (0.73)	.57
FEV_1[c]	2.13 (0.77)	2.17 (0.48)	.88
FEV_1/FVC	0.72 (0.13)	0.68 (0.04)	.35
DLCO[d]	63.42 (20.32)	74.33 (19.92)	.30
Vital sign before trial, mean (SD)			
Systolic blood pressure (mmHg)	115.79 (11.81)	119.11 (16.22)	.60
Diastolic blood pressure (mmHg)	73.36 (4.97)	72.44 (10.33)	.81
Heart rate	73.07 (11.09)	78.11 (11.56)	.31
Respiratory rate	17.86 (1.46)	18.89 (1.05)	.06
Temperature	36.60 (0.27)	36.64 (0.44)	.79

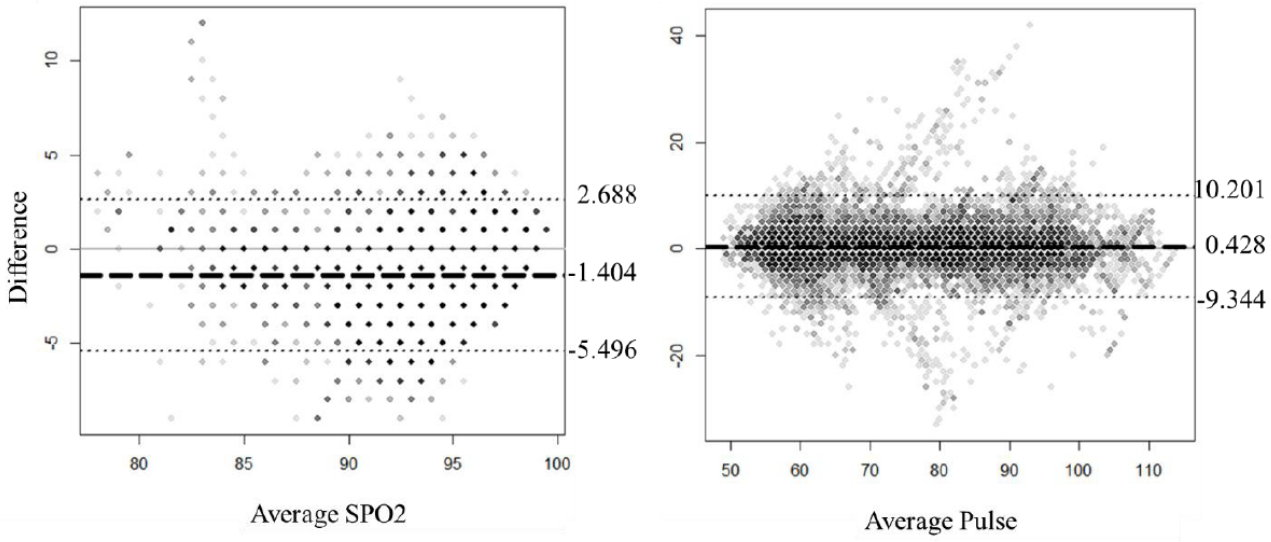
[a]Student *t* test for continuous variables and Fisher exact test for categorical variables.

[b]FVC: forced vital capacity; FEV_1: forced expiratory volume in 1 sec; DLCO: diffusing capacity for carbon monoxide.

[c]FEV_1: forced expiratory volume in 1 sec.

[d]DLCO: diffusing capacity for carbon monoxide.

Figure 3. Bland-Altman plots representing comparisons between the Nonin and Prince-100H devices for SpO_2 (left) and pulse (right). The dashed line represents the mean difference between the devices, with the upper and lower lines (dotted lines) representing the limits of agreement (±2SD).



Risk Signal Detection Using Prince-100H Wrist Oximeter and SpO $_2$ Monitor

Among the 23 patients, 14 had 20 risk signals during transport within the hospital, and none of the patients had a heart rate above 140. The risk signals occurred nine times for eight patients with less than 90% saturation, four times for four patients with heart rates below 60 bpm, and seven times for five patients due to network errors. Except for the risk signals for network error, we compared the characteristics of the patients with and without risk signals (Table 2). Although most variables were not statistically significant between the alarm and nonalarm groups, three patients in the alarm group used oxygen supplement devices before monitoring device use compared to none in the nonalarm group. Patients in the alarm group had lower pulmonary function test results than those in the nonalarm group, especially the DLCO value. After notification of risk signal, clinicians visited patients at risk and properly managed them. For example, for patients with hypoxemia, clinicians applied oxygen therapy until the patients stabilized without hypoxemia.

Discussion

In this study, we confirmed that a wearable device and mobile app can detect risk signals effectively during the transport or rehabilitation of a patient within the hospital. In addition, a real-time risk signal was sent to the health care provider in a message to ensure patient safety. To our knowledge, this is the first study involving simultaneous monitoring of oxygen saturation and heart rate in patients during transport or rehabilitation. In our study, the Prince-100H wrist oximeter and a mobile app, SpO_2 monitor, showed comparable results for oxygen saturation and heart rate with the Nonin, which is approved by the FDA for patient monitoring. This means that oxygen saturation and heart rate monitoring using a wearable device and mobile app is stable and reliable for patients.

Telemetry monitoring is a well-known helpful technique for real-time monitoring. The American Heart Association recommends the use of real-time electrocardiographic monitoring for patients with underlying cardiac disease or high-risk patients [19]. After introducing telemetry monitoring, cardiologists can early detect abnormal cardiac rhythm 24 hours per day in real time. Practical applications of real-time electrocardiographic monitoring have led to the expectation that real-time monitoring systems can be utilized in more situations with various parameters. Sala and colleagues [20] suggested that monitoring oxygen saturation and heart rate during rehabilitation after cardiac surgery can be helpful to ensure patient safety. However, the limitation of this suggestion is that physiotherapists have to directly monitor the oxygen saturation and heart rate of patients. To address these problems, wearable devices connected to a network and algorithms to detect risk signals are needed. In our study, we developed a real-time risk monitoring system that included cloud transmission, a mobile app, and a wearable device connected to the network through Wi-Fi or LTE. Because clinicians received the risk signal from the device through a Wi-Fi or LTE network in real time, they did not need to stand by the patients. In addition, they could monitor several patients simultaneously.

In this study, we selected patients admitted to the pulmonology ward, because patients with pulmonary disease have a relatively high risk of desaturation. Most of the patients enrolled after applying the exclusion criteria were patients admitted for diagnostic workup for lung cancer. Two patients were admitted for acute exacerbation of chronic obstructive pulmonary disease and another was admitted for an acute phase of interstitial lung disease.

In our study, there was no significant intergroup differences in the baseline characteristics between the alarm and nonalarm groups. However, all patients who used oxygen supplement devices before the study and one patient who used the monitoring device for rehabilitation were included in the alarm group. In addition, patients in the alarm group had poor pulmonary function test results, especially DLCO. This observation was reasonable because poor pulmonary function, demand for oxygen supply, or rehabilitation imply a higher risk of desaturation or elevated heart rate than normal. Although the

nonalarm group had more patients with a history of tuberculosis with statistical significance, this did not appear to have an effect on alarm events. Tuberculosis can cause bronchiectasis or airflow obstruction [21,22], which can lead to chronic airflow obstruction. However, all patients with obstructive patterns had only mild pulmonary dysfunction and there was no difference between the two groups in FEV_1/FVC, which represents the degree of airflow obstruction [23]. These results indicate that the device sent a risk signal to clinicians regardless of the patient's history.

We initially used Wi-Fi to transfer the data from the mobile app to the cloud. Because there were a lot of "not applicable" records, we used only LTE for the transfer after the second patient. For the two patients whose data were transmitted using Wi-Fi, most of the values were recorded as not applicable during the imputation process. We concluded that transmission of the measured data to the cloud was delayed because of network traffic. In addition to the two Wi-Fi patients, three LTE patients had exceptionally high not applicable ratios: P05 (28.02%), P07 (14.56%), and P03 (12.74%). Most of the not applicable signals occurred in the elevator or corridor during transport. There is a possibility that a poor mobile app signal delayed data delivery to the cloud. In this situation, we recorded the signal as not applicable and used an imputation value for matching and analysis.

During the trial, there were 20 risk signals for 14 patients. Except for network errors, risk signals for oxygen saturation below 90% were the most common. After clinicians received the risk signals, they applied oxygen supplements and confirmed improved oxygen saturation. In our study, there was no significant difference between risk signal detected and not detected patients. Nevertheless, if we selected patients at high risk and applied the study device to them, it may be more adequate. Patients who were expected to have desaturation events due to the use of an oxygen supplement device before monitoring and IHT for rehabilitation were in the alarm group. In addition, the results of the pulmonary function test for the nonalarm group were better than those for the alarm group. Risk signals for a heart rate of 60 bpm or less occurred for four patients. Among them, three patients had heart rates of 60 bpm or less measured before the study. This suggests that the target heart rate should be personalized before device application. One patient presented decreased heart rate after device application. Clinicians visited this patient and concluded that the decrease was a side effect of pethidine, which had been injected before the diagnostic procedure. The patient recovered after adequate hydration. This is a representative example of the usefulness of real-time monitoring. If clinicians are able to detect early signs of risk in patients, it can reduce the task of managing critically ill patients. Because this study was a feasibility study, we defined risk signal based on nonindividualized criteria rather than real-world situations. For this reason, we believe that there were relatively frequent risk signals. For applying the study device to a real-world setting, we plan further study based on individualized criteria.

This study had several limitations. First, this study was a single-center study with a small number of patients. We wanted to determine the feasibility of real-time monitoring with wearable devices and mobile apps. For this reason, we conducted the study with a small number of patients in a single center. Owing to favorable results in this study, our research team is planning a multicenter study with more patients. Secondly, we measured only oxygen saturation and heart rate for real-time monitoring. We selected these parameters because oxygen saturation and heart rate can be measured simply with only a pulse oximeter and are some of the earliest risk indicators for patients. Third, the patients were not transported through the same route. However, we performed this study to validate wearable devices and mobile apps in a variety of environments. Therefore, we tried to apply them without restrictions. Consequently, we were able to identify issues such as increased not applicable ratio in the elevator or corridor in transit to the room for bronchoscopy. Finally, we applied the study device to only two patients through hospital Wi-Fi, and the others through LTE. In this case, there was a possibility of security problem regarding patients' medical information. Nevertheless, we changed to LTE due to instability of Wi-Fi and planned to further study for Wi-Fi performance in real-time monitoring.

Despite these limitations, this study demonstrated favorable validation of telemonitoring application with LTE and Wi-Fi during patient transport. During IHT and rehabilitation, utilization of real-time monitoring can help clinicians with early risk detection or decisions such as prescribing oxygen supplements. Further study is needed for generalization to critically ill patients and other applications.

New techniques have been developed to ensure patient safety during transit. In this study, we constructed a system that notifies the health care provider by detecting the risk signal for the patient during transport based on a wearable device and a mobile app. Although there were some problems such as missing values and network errors, this paper is meaningful because the previously mentioned risk detection system was verified on actual patients.

Acknowledgments

This study was supported by a grant from Korea Health Industry Development Institute (grant number HI17C0178).

Authors' Contributions

Study concept and design: all authors; data acquisition: JHL and WJ; data preprocessing: JHL and YRP; statistical analysis and interpretation: JHL, YRP, and SK; discussion: JHL, YRP, and CMC; drafting of the manuscript: JHL and YRP; and study supervision: CMC and WJ. JHL and YRP should be considered as co-first authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Number of raw, matched, and imputation records from two wearable devices for twenty-three patients.

[PDF File (Adobe PDF File), 43KB - mhealth_v6i11e12048_app1.pdf]

References

1. Bolton CE, Bevan-Smith EF, Blakey JD, Crowe P, Elkin SL, Garrod R, et al. British Thoracic Society guideline on pulmonary rehabilitation in adults. Thorax 2013 Sep;68 Suppl 2:ii1-i30. [doi: 10.1136/thoraxjnl-2013-203808] [Medline: 23880483]
2. Jia L, Wang H, Gao Y, Liu H, Yu K. High incidence of adverse events during intra-hospital transport of critically ill patients and new related risk factors: a prospective, multicenter study in China. Crit Care 2016 Jan 18;20:12 [FREE Full text] [doi: 10.1186/s13054-016-1183-y] [Medline: 26781179]
3. Casaburi R, ZuWallack R. Pulmonary rehabilitation for management of chronic obstructive pulmonary disease. N Engl J Med 2009 Mar 26;360(13):1329-1335. [doi: 10.1056/NEJMct0804632] [Medline: 19321869]
4. Lee Y, Jung M, Shin GW, Bahn S, Park T, Cho I, et al. Safety and usability guidelines of clinical information systems integrating clinical workflow: a systematic review. Healthc Inform Res 2018 Jul;24(3):157-169 [FREE Full text] [doi: 10.4258/hir.2018.24.3.157] [Medline: 30109149]
5. Kue R, Brown P, Ness C, Scheulen J. Adverse clinical events during intrahospital transport by a specialized team: a preliminary report. Am J Crit Care 2011 Mar;20(2):153-161; quiz 162 [FREE Full text] [doi: 10.4037/ajcc2011478] [Medline: 21362719]
6. Lahner D, Nikolic A, Marhofer P, Koinig H, Germann P, Weinstabl C, et al. Incidence of complications in intrahospital transport of critically ill patients--experience in an Austrian university hospital. Wien Klin Wochenschr 2007;119(13-14):412-416. [doi: 10.1007/s00508-007-0813-4] [Medline: 17671822]
7. Parmentier-Decrucq E, Poissy J, Favory R, Nseir S, Onimus T, Guerry M, et al. Adverse events during intrahospital transport of critically ill patients: incidence and risk factors. Ann Intensive Care 2013 Apr 12;3(1):10 [FREE Full text] [doi: 10.1186/2110-5820-3-10] [Medline: 23587445]
8. Kim J, Kam HJ, Park YR, Yoo S, Oh JS, Kim Y, et al. Enchanted life space: adding value to smart health by integrating human desires. Healthc Inform Res 2018 Jan;24(1):3-11 [FREE Full text] [doi: 10.4258/hir.2018.24.1.3] [Medline: 29503747]
9. Il-Young JH, Lee E, Jung HW, Park H, Cheon S, Lee YS, et al. Wearable device-based walking programs in rural older adults can improve physical activity and health outcome: a feasibility study. JMIR Mhealth Uhealth 2018:e1 (forthcoming).
10. Loveday A, Sherar LB, Sanders JP, Sanderson PW, Esliger DW. Technologies that assess the location of physical activity and sedentary behavior: a systematic review. J Med Internet Res 2015 Aug 05;17(8):e192 [FREE Full text] [doi: 10.2196/jmir.4761] [Medline: 26245157]
11. Chai PR, Zhang H, Baugh CW, Jambaulikar GD, McCabe JC, Gorman JM, et al. Internet of Things buttons for real-time notifications in hospital operations: proposal for hospital implementation. J Med Internet Res 2018 Aug 10;20(8):e251 [FREE Full text] [doi: 10.2196/jmir.9454] [Medline: 30097420]
12. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
13. Vandenberk T, Stans J, Mortelmans C, Van Haelst R, Van Schelvergem G, Pelckmans C, et al. Clinical validation of heart rate apps: mixed-methods evaluation study. JMIR Mhealth Uhealth 2017 Dec 25;5(8):e129 [FREE Full text] [doi: 10.2196/mhealth.7254] [Medline: 28842392]
14. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. npj Digital Med 2018 Jan 15;1(1). [doi: 10.1038/s41746-017-0002-4]
15. Buekers J, De Boever P, Vaes AW, Aerts J, Wouters EF, Spruit MA, et al. Oxygen saturation measurements in telemonitoring of patients with COPD: a systematic review. Expert Rev Respir Med 2018 Feb;12(2):113-123. [doi: 10.1080/17476348.2018.1417842] [Medline: 29241369]
16. Vitacca M. Telemonitoring in patients with chronic respiratory insufficiency: expectations deluded? Thorax 2016 Apr;71(4):299-301. [doi: 10.1136/thoraxjnl-2015-208211] [Medline: 26962014]
17. Haas M. Statistical methodology for reliability studies. J Manipulative Physiol Ther 1991 Feb;14(2):119-132. [Medline: 2019821]
18. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
19. Sandau KE, Funk M, Auerbach A, Barsness GW, Blum K, Cvach M, American Heart Association Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology;Council on Cardiovascular Disease in the Young. Update to practice standards for electrocardiographic monitoring in hospital settings: a scientific statement from the American Heart Association. Circulation 2017 Nov 07;136(19):e273-e344. [doi: 10.1161/CIR.0000000000000527] [Medline: 28974521]

20. Sala V, Petrucci L, Monteleone S, Dall'Angelo A, Miracca S, Conte T, et al. Oxygen saturation and heart rate monitoring during a single session of early rehabilitation after cardiac surgery. Eur J Phys Rehabil Med 2016 Feb;52(1):12-19 [FREE Full text] [Medline: 26220328]
21. Chakrabarti B, Calverley PM, Davies PD. Tuberculosis and its incidence, special nature, and relationship with chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 2007;2(3):263-272 [FREE Full text] [Medline: 18229564]
22. Polverino E, Goeminne PC, McDonnell MJ, Aliberti S, Marshall SE, Loebinger MR, et al. European Respiratory Society guidelines for the management of adult bronchiectasis. Eur Respir J 2017 Sep;50(3). [doi: 10.1183/13993003.00629-2017] [Medline: 28889110]
23. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J 2005 Nov;26(5):948-968 [FREE Full text] [doi: 10.1183/09031936.05.00035205] [Medline: 16264058]

Abbreviation

bpm: beats per minute
DLCO: diffusing capacity for carbon monoxide
FEV $_1$: forced expiratory volume in 1 sec
FVC: forced vital capacity
ICC: intraclass correlation coefficient
IHT: intrahospital transport
IoT: Internet of Things
SMS: short message service
SpO $_2$: oxygen saturation by pulse oximetry

Edited by G Eysenbach; submitted 28.08.18; peer-reviewed by H Zhang, J Hefner; comments to author 07.10.18; revised version received 19.10.18; accepted 19.10.18; published 14.11.18.

Please cite as:
Lee JH, Park YR, Kweon S, Kim S, Ji W, Choi CM
A Cardiopulmonary Monitoring System for Patient Transport Within Hospitals Using Mobile Internet of Things Technology: Observational Validation Study
JMIR Mhealth Uhealth 2018;6(11):e12048
URL: http://mhealth.jmir.org/2018/11/e12048/
doi:10.2196/12048
PMID:30429115

Original Paper

Experience of Emergency Department Patients With Using the Talking Pole Device: Prospective Interventional Descriptive Study

Junsang Yoo[1*], RN; Ji Yeong Soh[1*], RN, MSc; Wan Hyoung Lee[2]; Dong Kyung Chang[1,3,4], MD, PhD; Se Uk Lee[5], MD; Won Chul Cha[1,4,6], MD

[1]Department of Digital Health, Samsung Advanced Institute of Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea
[2]Creative Laboratory, Samsung Electronics, Suwon, Republic of Korea
[3]Department of Gastroenterology, Samsung Medical Center, Seoul, Republic of Korea
[4]Health Information Center, Samsung Medical Center, Seoul, Republic of Korea
[5]Department of Emergency Medicine, Seoul National University Hospital, Seoul, Republic of Korea
[6]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Won Chul Cha, MD
Department of Emergency Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81, Irwon-ro
Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 234102053
Fax: 82 234100012
Email: wc.cha@samsung.com

Abstract

Background: Patient engagement is important. However, it can be difficult in emergency departments (EDs).

Objective: The aim of this study was to evaluate the satisfaction of ED patients using a patient-friendly health information technology (HIT) device, the "Talking Pole," and to assess the factors relevant to their satisfaction.

Methods: This study was conducted in May 2017 at the ED of a tertiary hospital. The "Talking Pole" is a smartphone-based device attached to a intravenous infusion pole with sensors. It is capable of sensing patient movement and fluid dynamics. In addition, it provides clinical information from electronic medical records to patients and serves as a wireless communication tool between patients and nurses. Patients and caregivers who entered the observation room of the ED were selected for the study. The "Talking Pole" devices were provided to all participants, regardless of their need for an intravenous pole upon admittance to the ED. After 2 hours, each participant was given an 18-item questionnaire created for this research, measured on a 5-point Likert scale, regarding their satisfaction with "Talking Pole."

Results: Among 52 participants recruited, 54% (28/52) were patients and the remaining were caregivers. In total, 38% (20/52) were male participants; the average age was 54.6 (SD 12.9) years, and 63% (33/52) of the participants were oncology patients and their caregivers. The overall satisfaction rate was 4.17 (SD 0.79) points. Spearman correlation coefficient showed a strong association of "overall satisfaction" with "comparison to the previous visit" (ρ=.73), "perceived benefit" (ρ=.73), "information satisfaction" (ρ=.70), and "efficiency" (ρ=.70).

Conclusions: In this study, we introduced a patient-friendly HIT device, the "Talking Pole." Its architecture focused on enhancing information delivery, which is regarded as a bottleneck toward achieving patient engagement in EDs. Patient and caregiver satisfaction with the "Talking Pole" was positive in the ED environment. In particular, correlation coefficient results improved our understanding about patients' satisfaction, HIT devices, and services used in the ED.

(JMIR Mhealth Uhealth 2018;6(11):e191) doi:10.2196/mhealth.9676

KEYWORDS

emergency department; health information technology; Internet of Things; mobile phone; patient engagement

Introduction

The needs of patients and importance of patient engagement are increasing; therefore, informed decision making in emergency departments (EDs) is critical [1,2]. Shared decision intervention, which is based on patients' proper understanding of their treatment, not only improves patient outcomes but also increases satisfaction and reduces health care utilization [3,4]. One of the important prerequisites for shared decision interventions is that patients have sufficient knowledge about their health care plan [5].

Accomplishing patient engagement in the ED is regarded as difficult because information delivery to patients is disrupted by the hostile and confusing circumstances of the ED [6-8], even if most patients wish to know about their illness and treatment [9-11]. Delivering information is an essential first step in patient engagement [3,4,12]. Unfortunately, the rapid pace of the process and the volume of information required often exceed an individual's comprehension [1,13]; moreover, interrelated factors, including the uncertainty of diagnosis and treatment, further complicate this situation [14]. However, information transfer does not have to rely solely on the relationship between a patient and their health care provider [15].

Health information technology (HIT) has the potential to improve patient engagement in EDs. The Society for Academic Emergency Medicine has presented strategies to accomplish patient engagement in the ED, including "using HIT to enhance patient communication" [8]. With the systemic constraints of health systems and advancement of the health information technology infrastructure, HIT ranks as the most efficient candidate for improving patient engagement. At present, however, the potential of HIT has not been fully reached owing to deficiencies in design and implementation issues [16,17].

The aims of this study were to evaluate the satisfaction of a patient-friendly HIT device, the "Talking Pole," in the ED environment and to assess the factors relevant to patient satisfaction.

Methods

Introduction to the "Talking Pole"

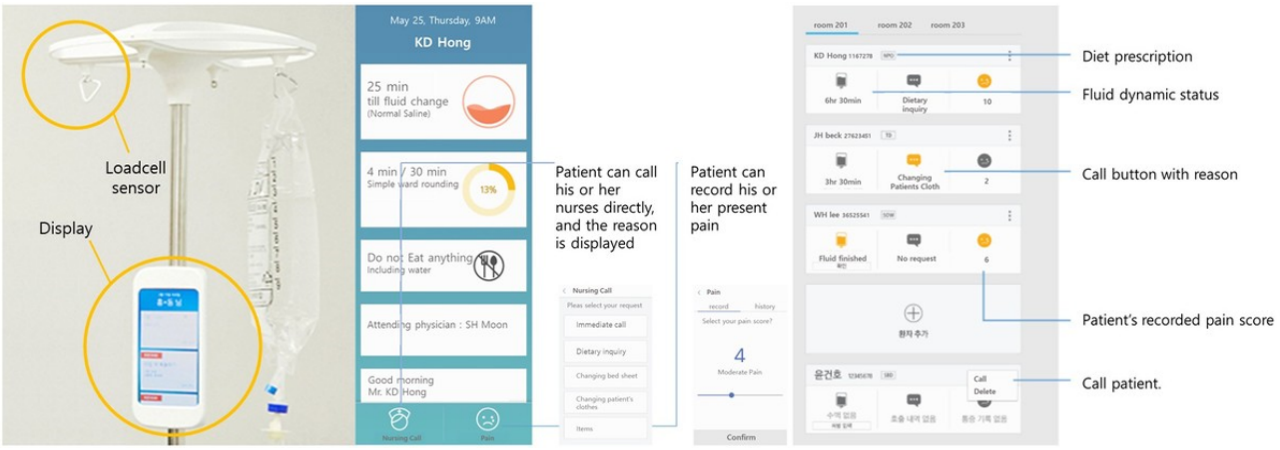
The "Talking Pole" is a patient-friendly HIT device that was developed to provide smart care to in-patients through the Internet of Things technology. The development team was comprised of both clinical and technical domain experts, and the device consisted of a weight sensor and a smartphone-based display (Figure 1).

The device has the following capacities. First, the device can deliver medical information from the electronic medical records system to the patient. Patients can check their medical schedule, information on medications, vital sign records, diet order information, and so on. Second, the "Talking Pole" enhances communication with medical staff in a subtle way. Two buttons at the bottom of the display, the "Call" and "Pain" buttons, allow patients to enter information for medical staff to see in real-time. Third, the "Talking Pole" checks fluid infusion status in real-time, so patients need not be concerned about receiving the appropriate amount of fluids and nurses can more conveniently monitor the flow. Finally, the "Talking Pole" provides patients with appropriate exercise goals according to their treatment plan and measures actual time spent exercising.

Study Setting

This is a prospective interventional descriptive study. The study was undertaken at an ED with an annual visit volume of 76,000 in a tertiary academic teaching hospital in Seoul. Participants were recruited from May 1, 2017 to May 31, 2017. The study was approved by the institution's ethics committee (Institutional Review Board File # SMC 2017-03-034-002).

Figure 1. General features of the "Talking Pole" system, from left to right: the "Talking Pole" device, patient display, and dashboard for medical staff.



Selection of Participants

Inclusion criteria were patients and their caregivers who entered the observation room in the ED and agreed to participate in this study. Usually, patients entering the observation room are distinguished from patients treated in other ED areas in the following ways: those who had finished initial assessment or treatment and those waiting for admission or determined to need monitoring of their condition over time. Such patients receive relatively planned and static treatment. By determining patients with these inclusion criteria, we intended to minimize the extent to which the study affected the subjects' treatment process. A caregiver was defined as a family member or friend who visited the ED with the patient. The criteria excluded patients declaring a "do not resuscitate", those younger than 18 years, those whose mental state was not alert, those with a critical medical device, and those at level 1 on the Korean Triage and Acuity Scale (KTAS). The KTAS was developed based on the Canadian Triage and Acuity Scale; level 1 indicates the highest acuity or severity of distress and level 5 indicates the lowest [18]. Patients with higher severity levels were excluded because it was clinically infeasible for them to use the "Talking Pole" device or because would require an amount of information that would exceed the capability of the device.

Interventions

After obtaining their consent, participants received the "Talking Pole" devices regardless of their need for an intravenous pole and were encouraged to use the device. After 2 hours, each participant was given an 18-item questionnaire developed for this study containing a 5-point Likert scale; questions regarded patients' satisfaction with the "Talking Pole." We prepared a total of 5 devices, including 2 extras for use in case of device failure, and the hardware and software of all devices were identical.

Outcome Measures

The main outcome was the determination of overall patient satisfaction with the "Talking Pole" system. Secondary outcome was a Spearman rank correlation coefficient between overall satisfaction and other questionnaire items. Secondary analysis was performed to determine where to focus and improve in the next iteration of our development process and to identify the factors affecting the users' satisfaction with HIT devices or services. Identifying these factors is crucial for future researchers and developers to find the best methods of applying HIT services in the clinical setting.

Data Analysis

Controversy exists among scholars about whether Likert scales should be analyzed in a parametric or nonparametric way. Likert scales are generally considered to be ordinal scales, but they are also used as interval scales [19,20]. In this study, we considered each item of the questionnaire on an interval scale.

We performed subgroup analyses of overall satisfaction and reported the means and SDs. We also examined Spearman rank correlation coefficient to examine associations between overall satisfaction and other factors. Correlation coefficients of |0.7-1| are considered to be strong, |0.5-0.7| are moderate, and less than |0.5| indicate weak relationships [21]. R version 3.4.3 was used for statistical analysis [22].

Results

Characteristics of Study Participants

A total of 52 participants were recruited, and there were no cases of dropout. The general characteristics of participants are presented in Table 1.

Evaluation Outcomes

Overall satisfaction with the "Talking Pole" system was measured at 4.17 points. The overall satisfaction score was higher for the male group (4.25 points) than for the female group (4.12 points), the caregiver group (4.25 points) than for the patient group (4.11 points), the moderate-severity group (4.23 points) than for the low-severity group (4.09 points), and the general patients group (4.37 points) than for the oncology patients group (4.06 points), but statistical significance of each subgroup was not verified (Figure 2).

The bars represent SEs of means. "Moderate severity" is indicated by KTAS levels 2 and 3, while "low" includes levels 4 and 5. Under "department," "general" includes cardiology, gastroenterology, infection, neurology, and nephrology patients.

The mean of participant responses was at least 4.0 points for all items. Items that evaluated interactions with medical staff, such as "call button" and "input pain score," were rated higher than those that evaluated the information display, such as "display information of username," "medical staff," "fluid infusion," and "dietary prescription" (Table 2).

Spearman correlation coefficient showed a strong correlation of "overall satisfaction" with "comparison to the previous visit" (ρ=.73), "perceived benefit" (ρ=.73), "information satisfaction" (ρ=.70), and "efficiency" (ρ=.70); on the other hand, items related to function were low (ρ=.29-.48; see Figure 3). All correlation coefficients were significant at P=.05.

Table 1. General characteristics of study participants.

Characteristics	Patient (n=28)	Caregiver (n=24)	Total (n=52)
Age (years), mean (SD)	57.7 (13.9)	50.9 (10.9)	54.6 (12.9)
Sex, n (%)			
Male	16 (57)	4 (17)	20 (38)
Female	12 (43)	20 (83)	32 (62)
Diagnosis category, n (%)			
Cardiology	1 (4)	0 (0)	1 (2)
Gastroenterology	2 (7)	1 (4)	3 (6)
Infectious	5 (18)	4 (17)	9 (17)
Neurology	1 (4)	1 (4)	2 (4)
Oncology	17 (61)	16 (67)	33 (63)
Nephrology	2 (7)	2 (8)	4 (8)
Severity (Korean Triage and Acuity Scale), n (%)			
1	0 (0)	0 (0)	0 (0)
2	1 (4)	0 (0)	1 (2)
3	15 (54)	14 (58)	29 (56)
4	11 (39)	9 (38)	20 (38)
5	1 (4)	1 (4)	2 (4)

Table 2. Mean score for each question

Questions	Score, mean (SD)
Perceived benefit	4.00 (0.74)
Learnability	4.27 (0.69)
Efficiency	4.25 (0.81)
Feeling safe	4.19 (0.86)
Overall satisfaction	4.17 (0.79)
Information satisfaction	4.17 (0.76)
Intention to reuse	4.33 (0.79)
Impact on hospital image	4.38 (0.69)
Comparison to the previous visit	4.38 (0.75)
Display information about user name	4.31 (0.78)
Display information about medical staff	4.25 (0.62)
Display information about fluid infusion	4.27 (0.72)
Display information about dietary prescription	4.38 (0.57)
Call button	4.48 (0.64)
Input pain score	4.56 (0.57)
Exercise measurement	4.46 (0.58)
Expectation of information use by medical staff	4.08 (0.90)
Service method evaluation	4.33 (0.62)

Figure 2. Overall satisfaction score and subgroup scores.

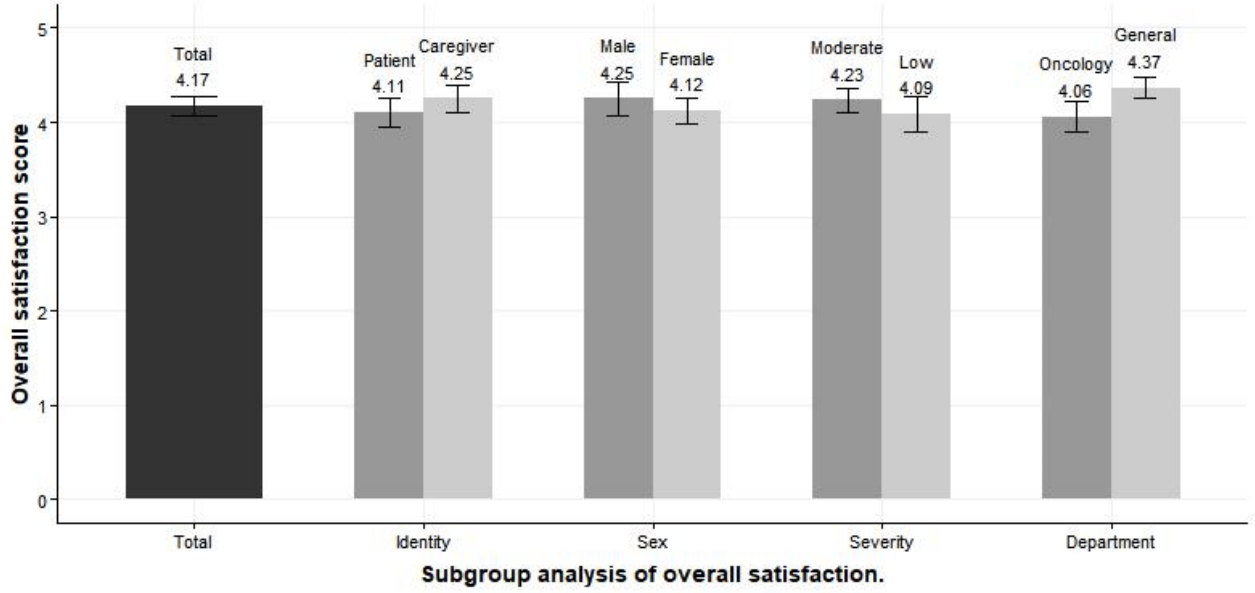
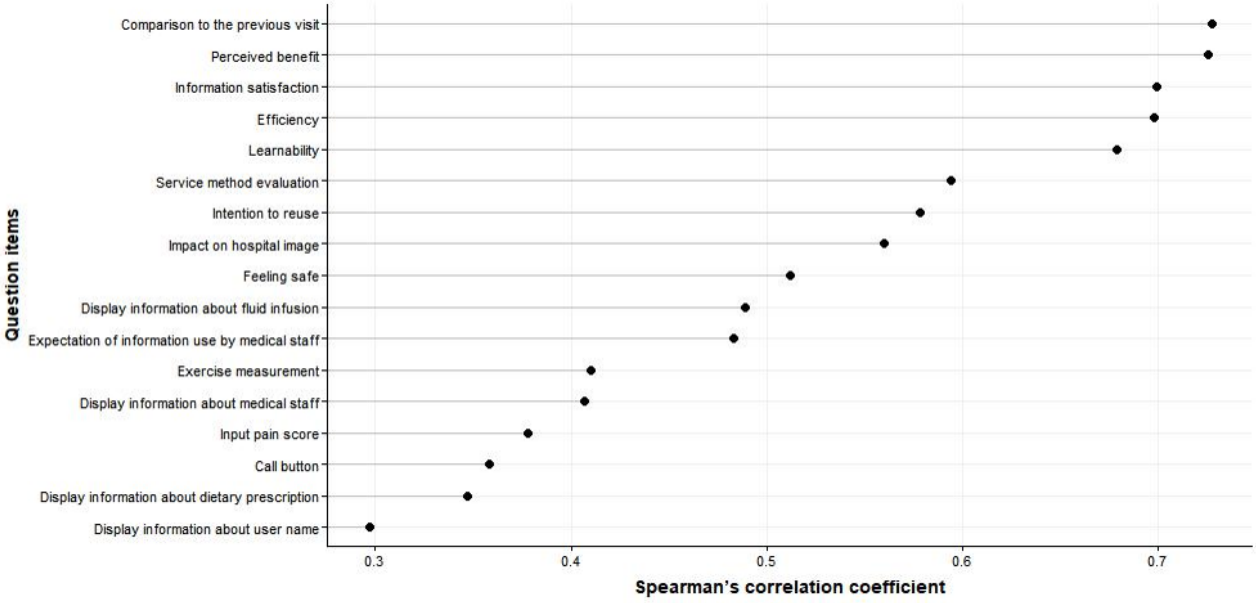


Figure 3. Correlation between overall satisfaction and each question.



Discussion

Principal Findings

In this study, we investigated the overall satisfaction of a patient-friendly HIT device by patients and caregivers in a real clinical environment as well as the correlation between overall satisfaction and other factors based on user surveys. The overall satisfaction score of the "Talking Pole" system was high. In addition, we found a high correlation of overall satisfaction with "comparison to the previous visit," "perceived benefit," "information satisfaction," and "efficiency." Under the TURF framework, "perceived benefit" and "efficiency" in our survey correspond to "useful" and "learnability" corresponds to "usable" [23]. The findings from the correlation coefficient may be consistent with this framework.

The participants also reported "information satisfaction" and "feeling safe" with scores of 4.17, with high correlation to overall satisfaction, and 4.19, with moderate correlation, respectively. The high satisfaction rate could be partially due to strong informational needs of ED patients. Thus, having personal information transferred directly to them would be satisfying. Based on the study settings and results, we can conclude that it is feasible to use our device to deliver information in a real ED environment.

Sharing information is an essential first step, as well as being a significant barrier, of patient engagement. Prior literature has reported that patients who visit the ED experience severe anxiety

and concern due to uncertainty about their disease, diagnosis, treatment, admission, and even medical error [14,24]. Thus, we need to decrease that uncertainty by providing information that patients and caregivers wish to have in a timely and personalized manner. However, until now, satisfying patients' informational needs has been regarded as difficult, especially in EDs. Delivering information has relied considerably on interpersonal communication between medical staff and patients, but this communication is often disturbed due to a high workload as well as a confusing and complex ED environment [6,7,25].

Under these circumstances, HIT can be a good solution for information delivery as well as for cost and quality of health care [26]. With this in mind, we developed our product, the "Talking Pole," with the expectation that it would improve patients' and caregivers' information-seeking behaviors by delivering medical information directly from the electronic medical records to the patient. However, incorporating HIT into EDs does not guarantee patient engagement. There are numerous unintended consequences associated with imprudent implementation, some even harmful to patients [27-30]. For this reason, we investigated patient and caregiver satisfaction with our devices by testing the "Talking Pole" with ED patients and caregivers in the actual ED environment. Satisfaction, as opposed to other usability factors, could be readily assessed in our study setting. The International Standard for Organization has defined satisfaction as the "extent to which the user's physical, cognitive, and emotional responses that result from the use of a system, product, or service meet the user's needs and expectations" [31]. A further explanation states that "a system is satisfying to use if the users have a good subjective impression of how useful, usable, and likable the system is" [23]. Although satisfaction is only one aspect of usability, it is associated with various factors, including intuitive design, ease of learning, efficiency of use, error frequency, and severity [32].

Prior literature has reported that multidisciplinary collaboration involving health care professionals is a factor in the successful application of HIT [33,34], and our experience is consistent with this. In our development process, clinical experts participated in the team from the ideation stage throughout. We thought that this active conjunction between both medical and technical domain experts may help the "Talking Pole" become more feasible by reflecting domain specificity of the clinical environment as well as by uncovering patients' unmet needs.

Finally, we routinely use the phrase "patient engagement," but this is an abbreviation of "patient and family engagement" [35]. It is a common phenomenon for a patient to bring a family member, friend, or accompanying person with them when they come to the hospital. Therefore, when we measured the satisfaction of the "Talking Pole," which is a system designed to improve "patient engagement," the research team thought it would be appropriate to include caregivers in the participant group.

Limitations

First, this research was conducted in an ED of a single tertiary academic hospital; readers must be careful when extending their interpretations of these results to other departments or hospitals. However, considering that the need for information is a common phenomenon for patients under a variety of circumstances [25,36-40], the "Talking Pole" has potential applicability to other departments, such as wards. Second, we assessed the satisfaction of the "Talking Pole," which is only one aspect of usability, so this research cannot conclude that our product is usable. Third, we used a questionnaire that we developed ourselves and that has not been validated; there is a possibility that it contains response biases, such as an acquiescence bias. Further researchers may consider using inversely coded questionnaires to overcome this kind of bias. However, it is not a fundamental solution to this problem, since the acquiescence response style itself tends to produce positive responses regardless of content [41]. Fourth, we did not investigate the patients' clinical outcomes and usability. Further research is needed to evaluate usability with a validated tool and when the device is implemented in other hospitals or other departments.

Conclusion

The overall satisfaction of the "Talking Pole" was high, and it highly correlated with "comparison to the previous visit," "perceived benefit," "information satisfaction," and "efficiency." Through this study, we were able to verify that the "Talking Pole" was able to help meet the needs of patients' and caregivers' information-seeking behaviors, which are regarded as the primary barrier of patient engagement in an ED environment.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (Grant #2018R1C1B6002877) and funded by Samsung Electronics. The results, discussion, and conclusion of this paper are independent of the funding source.

Conflicts of Interest

WHL is employed by Samsung Electronics.

References

1. Flynn D, Knoedler MA, Hess EP, Murad MH, Erwin PJ, Montori VM, et al. Engaging patients in health care decisions in the emergency department through shared decision-making: a systematic review. Acad Emerg Med 2012 Aug;19(8):959-967 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01414.x] [Medline: 22853804]

2. Charmel PA, Frampton SB. Building the business case for patient-centered care. Healthc Financ Manage 2008 Mar;62(3):80-85. [Medline: 19097611]
3. Hess EP, Grudzen CR, Thomson R, Raja AS, Carpenter CR. Shared Decision-making in the Emergency Department: Respecting Patient Autonomy When Seconds Count. Acad Emerg Med 2015 Jul;22(7):856-864 [FREE Full text] [doi: 10.1111/acem.12703] [Medline: 26112797]
4. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
5. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
6. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, et al. The effect of emergency department crowding on clinically oriented outcomes. Acad Emerg Med 2009 Jan;16(1):1-10 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00295.x] [Medline: 19007346]
7. Cha WC, Shin SD, Song KJ, Jung SK, Suh GJ. Effect of an independent-capacity protocol on overcrowding in an urban emergency department. Acad Emerg Med 2009 Dec;16(12):1277-1283 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00526.x] [Medline: 19912131]
8. Pham JC, Trueger NS, Hilton J, Khare RK, Smith JP, Bernstein SL. Interventions to improve patient-centered care during times of emergency department crowding. Acad Emerg Med 2011 Dec;18(12):1289-1294 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01224.x] [Medline: 22168193]
9. Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys. Behav Med 1998;24(2):81-88. [doi: 10.1080/08964289809596384] [Medline: 9695899]
10. Shiloh S, Gerad L, Goldman B. Patients' information needs and decision-making processes: what can be learned from genetic counselees? Health Psychol 2006 Mar;25(2):211-219. [doi: 10.1037/0278-6133.25.2.211] [Medline: 16569113]
11. Scott JT, Thompson DR. Assessing the information needs of post-myocardial infarction patients: a systematic review. Patient Educ Couns 2003 Jun;50(2):167-177. [Medline: 12781932]
12. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
13. Kovacs G, Croskerry P. Clinical decision making: an emergency medicine perspective. Acad Emerg Med 1999 Sep;6(9):947-952 [FREE Full text] [Medline: 10490259]
14. Sun BC, Adams JG, Burstin HR. Validating a model of patient satisfaction with emergency care. Ann Emerg Med 2001 Nov;38(5):527-532. [doi: 10.1067/mem.2001.119250] [Medline: 11679864]
15. Boudreaux ED, O'Hea EL. Patient satisfaction in the Emergency Department: a review of the literature and implications for practice. J Emerg Med 2004 Jan;26(1):13-26. [Medline: 14751474]
16. Kellermann AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Aff (Millwood) 2013 Jan;32(1):63-68. [doi: 10.1377/hlthaff.2012.0693] [Medline: 23297272]
17. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
18. Kim JH, Kim JW, Kim SY, Hong DY, Park SO, Baek KJ. Validation of the Korean Triage and Acuity Scale Compare to Triage by Emergency Severity Index for Emergency Adult Patient: Preliminary Study in a Tertiary Hospital Emergency Medical Center. Journal of The Korean Society of Emergency Medicine 2016 Oct 30;27(5):436-441 [FREE Full text]
19. Clason DL, Dormody TJ. Analyzing Data Measured By Individual Likert-Type Items. JAE 1994 Dec;35(4):31-35. [doi: 10.5032/jae.1994.04031]
20. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
21. Rumsey DJ. Statistics II for Dummies. Hoboken, New Jersey: Wiley Publishing Inc; 2009.
22. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. URL: https://www.r-project.org/ [accessed 2018-10-08] [WebCite Cache ID 730iEFRnW]
23. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011 Dec;44(6):1056-1067 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.005] [Medline: 21867774]
24. Burroughs TE, Waterman AD, Gallagher TH, Waterman B, Adams D, Jeffe DB, et al. Patient concerns about medical errors in emergency departments. Acad Emerg Med 2005 Jan;12(1):57-64 [FREE Full text] [doi: 10.1197/j.aem.2004.08.052] [Medline: 15635139]
25. Yu M, Chair SY, Chan CW, Choi KC. Information needs of patients with heart failure: Health professionals' perspectives. Int J Nurs Pract 2016 Aug;22(4):348-355. [doi: 10.1111/ijn.12442] [Medline: 27245373]
26. Bardhan I, Thouin M. Health information technology and its impact on the quality and cost of healthcare delivery. Decision Support Systems 2013 May;55(2):438-449 [FREE Full text] [doi: 10.1016/j.dss.2012.10.003] [Medline: 25904163]

27. Bloomrosen M, Starren J, Lorenzi NM, Ash JS, Patel VL, Shortliffe EH. Anticipating and addressing the unintended consequences of health IT and policy: a report from the AMIA 2009 Health Policy Meeting. J Am Med Inform Assoc 2011;18(1):82-90 [FREE Full text] [doi: 10.1136/jamia.2010.007567] [Medline: 21169620]
28. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Apr;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
29. Ranji SR, Rennke S, Wachter RM. Computerised provider order entry combined with clinical decision support systems to improve medication safety: a narrative review. BMJ Qual Saf 2014 Sep;23(9):773-780. [doi: 10.1136/bmjqs-2013-002165] [Medline: 24728888]
30. Sittig DF, Singh H. Defining health information technology-related errors: new developments since to err is human. Arch Intern Med 2011 Jul 25;171(14):1281-1284 [FREE Full text] [doi: 10.1001/archinternmed.2011.327] [Medline: 21788544]
31. International Organization for Standardization. ISO 9241-11Ergonomics of human-system interaction. Geneva, Switzerland; 2018. Part 11: Usability: Definitions and concepts URL: https://www.iso.org/standard/63500.html [accessed 2018-10-08] [WebCite Cache ID 730igdzZb]
32. Berner ES. Clinical Decision Support Systems: Theory and Practice. 3rd edition. New York City: Springer International Publishing; 2016:79.
33. Devlin AM, McGee-Lennon M, O'Donnell CA, Bouamrane M, Agbakoba R, O'Connor S, et al. Delivering Digital Health and Well-Being at Scale: Lessons Learned during the Implementation of the dallas Program in the United Kingdom. J Am Med Inform Assoc 2015 Aug 8 [FREE Full text] [doi: 10.1093/jamia/ocv097] [Medline: 26254480]
34. Soh JY, Cha WC, Chang DK, Hwang JH, Kim K, Rha M, et al. Development and Validation of a Multidisciplinary Mobile Care System for Patients With Advanced Gastrointestinal Cancer: Interventional Observation Study. JMIR Mhealth Uhealth 2018 May 07;6(5):e115 [FREE Full text] [doi: 10.2196/mhealth.9363] [Medline: 29735478]
35. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb;32(2):223-231 [FREE Full text] [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
36. Stavropoulou C. Perceived information needs and non-adherence: evidence from Greek patients with hypertension. Health Expect 2012 Jun;15(2):187-196 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00679.x] [Medline: 21496190]
37. Rutten LJF, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns 2005 Jun;57(3):250-261. [doi: 10.1016/j.pec.2004.06.006] [Medline: 15893206]
38. Søndergaard EG, Grøne BH, Wulff CN, Larsen PV, Søndergaard J. A survey of cancer patients' unmet information and coordination needs in handovers--a cross-sectional study. BMC Res Notes 2013 Sep 25;6:378 [FREE Full text] [doi: 10.1186/1756-0500-6-378] [Medline: 24066725]
39. Abt SA, Pablo HS, Serrano AP, Fernández VE, Martín FR. [Information needs and internet use in patients with breast cancer in Spain]. Gac Sanit 2013;27(3):241-247 [FREE Full text] [doi: 10.1016/j.gaceta.2012.06.014] [Medline: 22943976]
40. Lambert V, Glacken M, McCarron M. Meeting the information needs of children in hospital. J Child Health Care 2013 Dec;17(4):338-353. [doi: 10.1177/1367493512462155] [Medline: 23411658]
41. Kim SK, Shin IC, Jeong JK. Personality Traits and Response Styles. Survey Research 2011;12(2):51-76.

Abbreviations

ED: emergency department
HIT: health Information technology
KTAS: Korean Triage and Acuity Scale

Edited by G Eysenbach; submitted 23.12.17; peer-reviewed by YR Park, G Constantinescu; comments to author 20.03.18; revised version received 18.06.18; accepted 21.08.18; published 22.11.18.

Please cite as:
Yoo J, Soh JY, Lee WH, Chang DK, Lee SU, Cha WC
Experience of Emergency Department Patients With Using the Talking Pole Device: Prospective Interventional Descriptive Study
JMIR Mhealth Uhealth 2018;6(11):e191
URL: http://mhealth.jmir.org/2018/11/e191/
doi:10.2196/mhealth.9676
PMID:30467105

Original Paper

A Real-Time Autonomous Dashboard for the Emergency Department: 5-Year Case Study

Junsang Yoo[1], RN; Kwang Yul Jung[2], MD; Taerim Kim[2], MD; Taerim Lee[3], MD; Sung Yeon Hwang[2], MD; Hee Yoon[2], MD; Tae Gun Shin[2], MD; Min Seob Sim[2], MD; Ik Joon Jo[2], MD; Hansol Paeng[4]; Jong Soo Choi[1,5], PhD; Won Chul Cha[2,5,6], MD

[1]SAIHST, Department of Digital Health, Sungkyunkwan University, Seoul, Republic of Korea

[2]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[3]Department of Emergency Medicine, Chamjoeun Hospital, Gwangju, Republic of Korea

[4]Human Understanding Design Center (HUDC), Seoul Medical Center, Seoul, Republic of Korea

[5]Health Information Center, Samsung Medical Center, Seoul, Republic of Korea

[6]Department of Digital Health, Samsung Advanced Institute of Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Won Chul Cha, MD
Department of Emergency Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81, Irwon-ro
Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 2053
Fax: 82 2 3410 0012
Email: wc.cha@samsung.com

Abstract

Background: The task of monitoring and managing the entire emergency department (ED) is becoming more important due to increasing pressure on the ED. Recently, dashboards have received the spotlight as health information technology to support these tasks.

Objective: This study aimed to describe the development of a real-time autonomous dashboard for the ED and to evaluate perspectives of clinical staff on its usability.

Methods: We developed a dashboard based on three principles—"anytime, anywhere, at a glance;" "minimal interruption to workflow;" and "protect patient privacy"—and 3 design features—"geographical layout," "patient-level alert," and "real-time summary data." Items to evaluate the dashboard were selected based on the throughput factor of the conceptual model of ED crowding. Moreover, ED physicians and nurses were surveyed using the system usability scale (SUS) and situation awareness index as well as a questionnaire we created on the basis of the construct of the Situation Awareness Rating Technique.

Results: The first version of the ED dashboard was successfully launched in 2013, and it has undergone 3 major revisions since then because of geographical changes in ED and modifications to improve usability. A total of 52 ED staff members participated in the survey. The average SUS score of the dashboard was 67.6 points, which indicates "OK-to-Good" usability. The participants also reported that the dashboard provided efficient "concentration support" (4.15 points), "complexity representation" (4.02 points), "variability representation" (3.96 points), "information quality" (3.94 points), and "familiarity" (3.94 points). However, the "division of attention" was rated at 2.25 points.

Conclusions: We developed a real-time autonomous ED dashboard and successfully used it for 5 years with good evaluation from users.

(JMIR Mhealth Uhealth 2018;6(11):e10666) doi:10.2196/10666

KEYWORDS

dashboard; development; emergency department; evaluation; health information technology; situation awareness; usability

Introduction

An emergency department (ED) is a complex system designed to treat patients with various conditions simultaneously. Even though ED providers try to triage patients according to their clinical needs while managing scarce resources, these activities are often overwhelmed by the complexity and large volume of data. Emergency physicians are required to treat multiple patients while maintaining situational awareness of the ED surroundings [1]. This task is very challenging because it requires acquisition, processing, integration, and archiving of large data at multiple levels [2]. Thus, physicians frequently feel like they are losing control over the ED, which aggravates burnout and affects performance [3,4].

Moreover, many EDs are already overcrowded, thus increasing the complexity [5-7]. While ED overcrowding remains a major health care issue, it significantly and adversely affects quality of care by influencing major quality factors, such as timeliness, effectiveness, efficiency, safety, and patient-centeredness, resulting in increased mortality and morbidity [5,8-10]. Multiple studies have suggested the use of system engineering and science to improve ED performance, streamline the process, and improve the throughput [11-13]. However, a strategic approach to identify process delays and supply-demand mismatch using traditional hospital information systems (HISs) is not feasible. It is important to monitor and manage the ED as a whole.

The dashboard is "a visual display of the most important information needed to achieve one or more objectives" [14]. Because the situation in the ED affects the quality of care and patient outcomes, clinicians use different diagnostic and treatment strategies depending on the situation [15]. Therefore, recognizing the correct situation is becoming increasingly important to emergency medical clinicians [16].

While a quality dashboard helps decision making at the organization level, a clinical dashboard helps decision making regarding patient care [17]. A dashboard that fits the changing situations in the ED in the real time must have the characteristics of both quality and clinical dashboards [17]. Recent studies on the ED dashboard system have reported its potential to improve patient safety, situation awareness, and workflow [15,18], but such studies have not addressed long-term experiences and association between the introduction of the dashboard and mental workload of the users. In particular, ensuring real-time availability is challenging and important because nonreal–time dashboards cannot support decision making [15].

This study aimed to describe the development of a real-time organizational dashboard for the ED and to evaluate its usability.

Methods

This study was approved by the Institutional Review Board of the study site (Institutional Review Board File #SMC 2018-01-040-001).

Study Setting

The study was carried out in a metropolis: Seoul. It was undertaken at an ED with an annual visit volume of 79,000 patients in a tertiary teaching hospital. The hospital has about 2000 inpatient beds. This ED is one of the most overcrowded EDs in the country [19].

The HIS was in use in this ED since 1994, supported by electronic medical records and a picture archiving and communication system. Although the history of this HIS is long and its technical quality was one of the most advanced in the country, no ED-specific dashboard system was ever used in this hospital.

Development

The Happinovation and the Happy Emergency Room Team

In 2012, an institution-wide project, the "Happinovation" (Happy innovation), was initiated to enhance patient and provider happiness through process and hardware innovations. The "Happy Emergency Room Team" was formed as a satellite team for the overall project. The team focused on visualizing the ED process for providers and patients.

Specifically, 2 subprojects were implemented. One was a visualization project for providers—an electronic dashboard to develop visualization of ED performance status on wall-mounted monitors and PCs. The other was a visualization project for patients and their families—a wall-mounted electronic dashboard, kiosks, and tablets.

The multidisciplinary Happy Emergency Room Team included 6 physicians, 4 nurses, 1 administrator, 2 quality improvement team members, 2 consultants, and 2 designers. While hospital staff provided inputs, consultants and designers tried to see things from patients' perspective, thus balancing the conclusion. Several rounds of discussions and debates took place before the first dashboard principle and design feature was created.

Dashboard Principle

The team had agreed on the following 3 dashboard principles that define the characteristics of the dashboard.

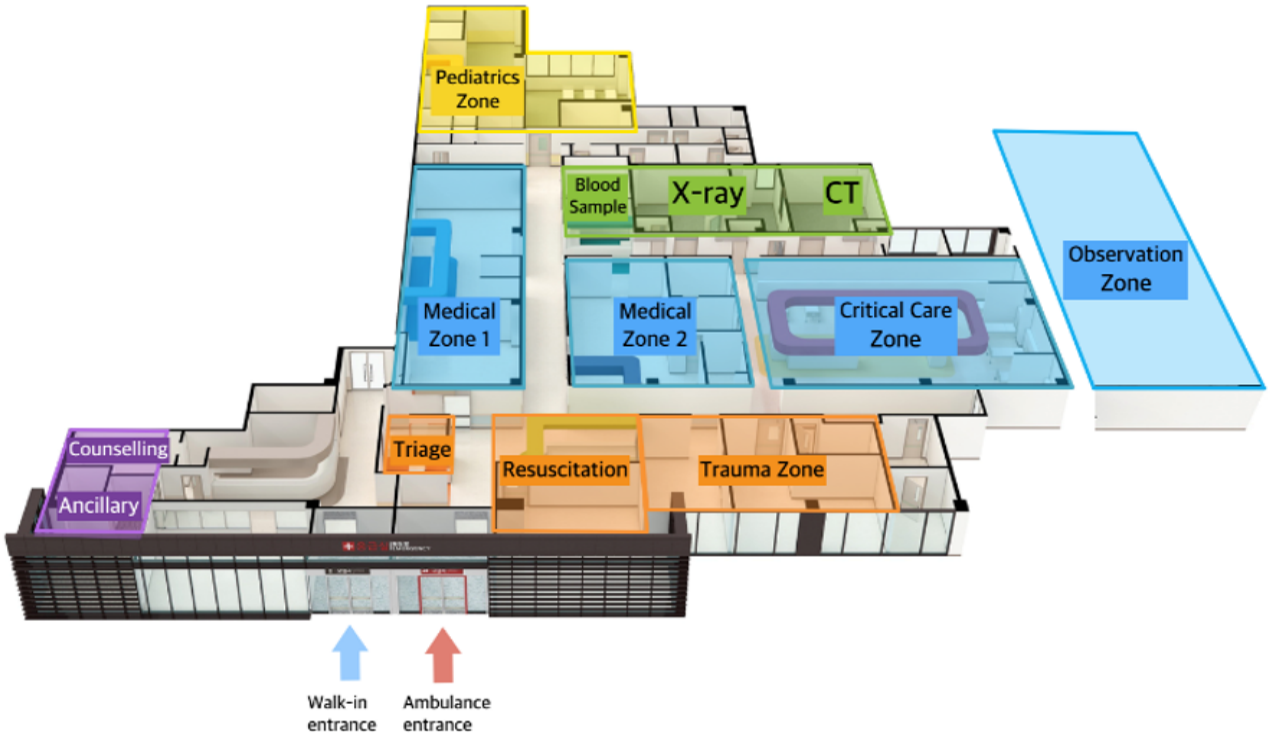
Anytime, Anywhere, at a Glance

The dashboard was designed to be like a traditional whiteboard for the ED, which means that it should be mounted on a wall and be visible to providers from about 2-5 m away. It should be in a static mode, without flipping the screen, such that providers would not waste any time to find the desired information [20]. Since multiple providers should be able to access the information, interactivity was not feasible. We did not include a writing and communication function from the dashboard because it was designed to only offer providers the overall ED status at a glance [15,21].

Minimal Interruption to Workflow

The dashboard should be integrated with the clinical workflow. An unfavorable clinical workflow often changes after implementing health information technology [21-23]. However, the Happy Emergency Room Team worried that the dashboard could be a source of botheration. Therefore, we designed it such that it would not require any additional inputs. This means that all data should come from the legacy system.

Figure 1. Geographical layout of the emergency department.



All information on the dashboard is automatically generated from previous columns of HIS. For example, the moment at which a physician's order or barcode of a clinical sample is input by a nurse, it is linked to the dashboard system such that the whole operating process could be automatically marked and shown without additional inputting processes. This is also different from traditional dry-erase whiteboards [24].

Protect Patient Privacy

It is important that patients' privacy be secured [25]. Because the screen is physically accessible to patients and families, most information should be deidentified and symbolized to prevent unnecessary misunderstandings and debates.

Design Features

The team adopted the following 3 design features for the dashboard:

Geographical Layout

The dashboard should indicate the ED floor plan. A geographical layout would give intuitive information to ED providers on what is going on where in the ED. It is also very common that ED providers recognize and communicate about patients with their bed locations and not with their numbers or names, which made it more intuitive and effective to use a geographical layout rather than a patient list (Figure 1).

Patient-Level Alert

All beds and chairs were symbolized to stand for a patient to provide patient-specific information to ED clinicians. Additional patient-level information was included to provide a real-time alert to providers through encoded colors and symbols, especially regarding process delays. The main objective of this concept was to provide ED staff with patient-level alerts such that a provider could immediately notice delays in catering to patients.

Real-Time Summary Data

The dashboard would show summary statistics regarding ED performance in the real time. While in-depth information for each patient is available in the pre-existing HIS, a real-time summary depends mostly on subjective feeling of individual providers, which varies significantly due to the lack of information on summary information over ED state. With summary data, providers would be able to reschedule clinical processes for their patients' efficient journey. For example, a nurse could direct a patient to the X-ray if there is a long queue for a co-ordered computed tomography.

Prototyping

Initially, multiple measures were suggested for the dashboard. Measures were chosen and categorized based on a "conceptual model of ED crowding." The model consisted of 3 factors, among which the throughput factors, which reflect the internal process of ED care, were mostly demonstrated with the dashboard [13]. Among input factors, which are components that contribute to the demand for ED services, patient severity and measure of visits were included. Cautions on infection information were also included since they were well correlated

with severity and patient allocation within the ED and hospital. Throughput factors were divided into 2 sections: structures and functions. Output factors, such as boarding and discharge data, were also included. Boarding pertains to the inability to admit a patient to a ward due to the lack of inpatient beds even through the patient is determined to be admitted.

We used a Windows server as the ED visualization platform to support various devices in the ED, including the dashboard and kiosks for patients and mobile devices. Our entire platform was developed and deployed on 2 Windows 2008 servers, each with a 1200-GB hard drive and 20-GB memory, and 2 4-core Intel Xeon 2.4 GHz processors. The servers queried the measurements mentioned above from the electronic medical records and picture archiving and communication system servers. A Windows communication foundation was used as a visualization tool (Figure 2).

Evolution of the Emergency Department Dashboard

We developed and updated the dashboard over 5 years. During the observation, the ED had gone through a Middle East Respiratory Syndrome outbreak, followed by a major structural and functional renovation [26]. The ED dashboard adopted such changes and evolved through the process.

Evaluation

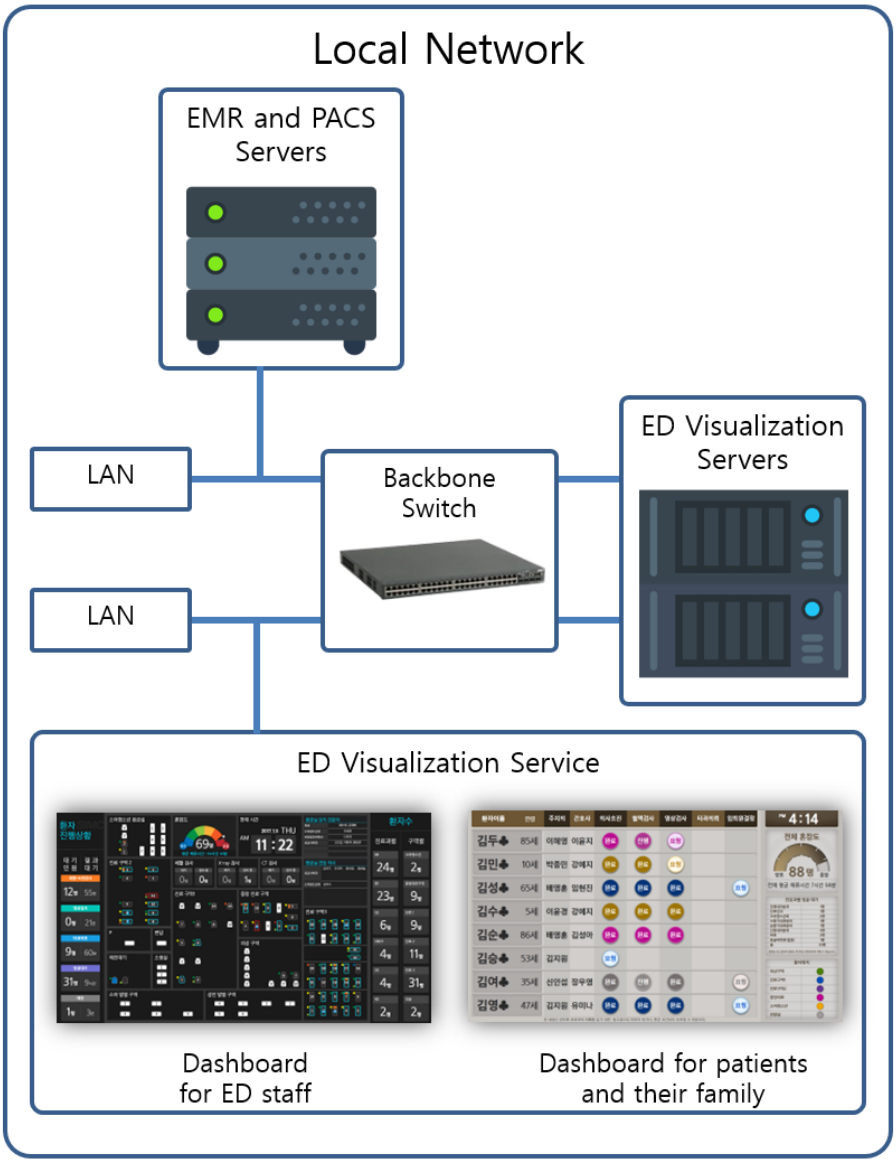
Selection of Participants

Inclusion criteria were those currently working as ED physicians and triage or charge nurses and those using the dashboard. Participants were recruited from January 1 to February 10, 2018.

Intervention

Participants responded to 20 questions on a 5-point Likert scale after completing a consent form. The first 10 items of the questionnaire were derived from the system usability scale (SUS) to investigate the usability of the dashboard [27]. The last 10 items were derived from the situation awareness index (SAI), which we composed based on the Situation Awareness Rating Technique (SART) [28]. The SAI aimed to assess whether ED physicians and nurses were using the dashboard to help them recognize the situation. The SAI includes the constructs of the SART as a whole, and small modifications were made to adjust it to fit the dashboard.

Figure 2. Transfer of electronic medical record (EMR) and picture archiving and communication system (PACS) data to the emergency department (ED) visual architecture platform. LAN: Local Area Network.



After completing the questionnaire, the participants received about US $8 as compensation for participating in the survey.

Outcome Analysis

The SUS scores were interpreted using an adjective rating scale [29].

The SAI score was calculated using the following formula:

> SAI = {Q11 + Q12 + Q13 + Q14 + Q15 + Q16 + (6 - Q17) + Q18 + Q19 + Q20} / 10, where Q: question number.

The SAI scores were also examined using descriptive analyses.

Results

Design and Structure

Introduction to the First Version of the Emergency Department Dashboard

The semicircle-shaped indicator in the upper central area represents the complexity of the ED and presents the expected mean length of stay of a current patient in the ED. The semicircle borrows the scheme of the traffic light so that the user intuitively grasps the current situation of the ED.

A semitransparent colored square represents each section of the ED and matches the geographical layout presented in Figure 1. The squares are not visible on the actual dashboard. Small icons, such as that shown in the red circle, indicate patients, and they intuitively inform the user about the patient's journey through the color corresponding to the patient's process on the left.

The left side of the dashboard presents a summary of the patient process. This information allows the physician to set up a patient

diagnosis strategy and the nurse to determine the order of various tests. The central area of the dashboard reflects the geographical layout of the ED. Here, individual patients' specific information is displayed, such as real-time clinical processes. The right side of the dashboard presents the number of patients by zone, thereby enabling efficient distribution of medical labor (Figure 3).

Evolution of the Emergency Department Dashboard

Major difference between versions 1 and 2 of the dashboard is that the latter reflects a geographical change. In 2015, the ED of the study site underwent a Middle East Respiratory Syndrome outbreak, and a respiratory isolation area was established and operated. This structural change to the ED was reflected on the dashboard (Figure 4).

In version 3, a revision was made to improve usability. We updated the dashboard to change the overall color coding. By doing so, we were able to reflect the users' suggestions, such as "difficulty to identify bed status" (Figure 4).

A major change in version 4 is the improvement of patient-specific information.

The circle next to the patient indicates the mapping of the Korean Triage and Acuity Scale score from 1 to 5, presented from red to green. Additionally, the rectangle next to the patient reflects "infection caution," such as "air caution" and "blood caution" (Figure 4).

Figure 3. The first version of the emergency department dashboard.

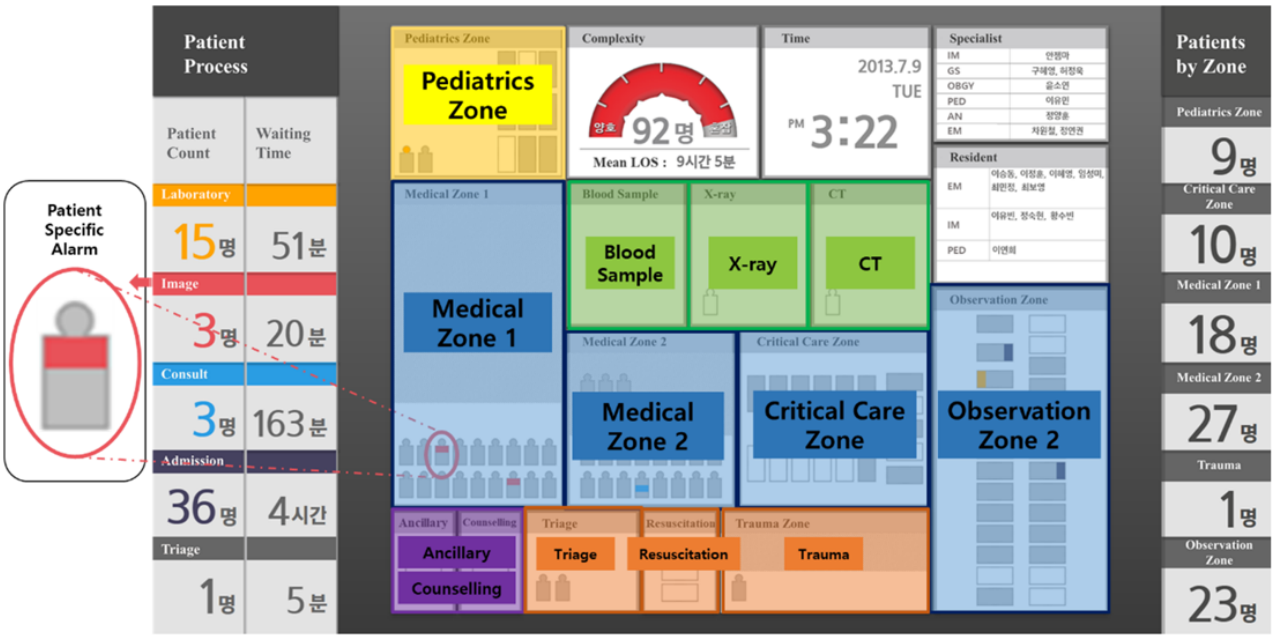
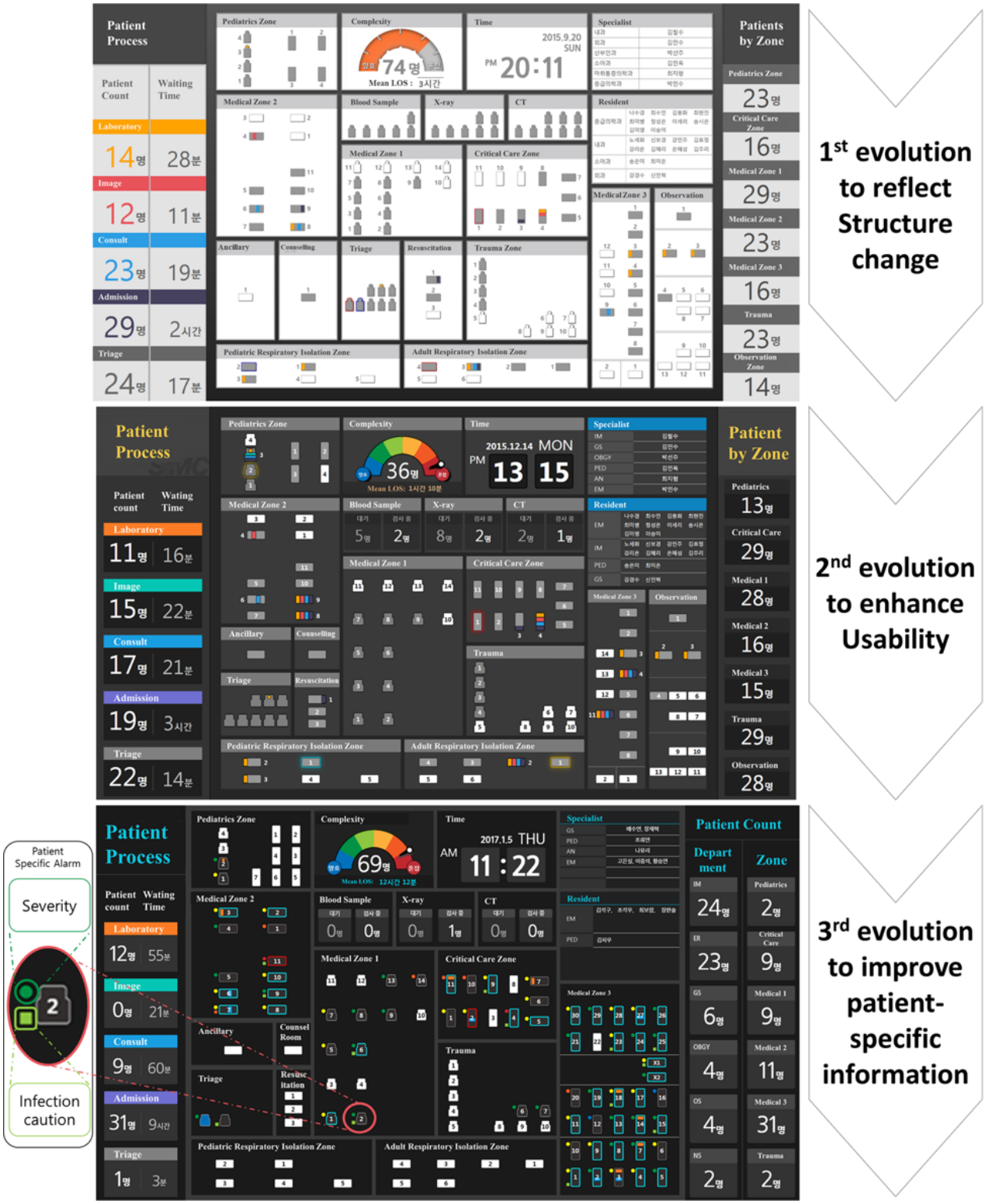


Figure 4. Evolution of the emergency department dashboard, showing the second, third, and final versions.



Evaluation

Participant Characteristics

A total of 52 participants were recruited; 25 were physicians and 27 were nurses. In the physician group, 17 were males and 8 were females; in the nurse group, only 2 were males and 25 were females. The participants' years of experience also varied; 8 had worked for less than 3 years, 10 had worked for 3-5 years, 19 had worked for 5-10 years, and 15 had worked for over 10 years (Table 1).

System Usability Scale

The SUS score of the ED dashboard system was 67.6 points (Table 2). The SUS score showed a slight difference between the 2 groups; the physician and nurse groups had scores of 67.5 and 67.69 points, respectively. We could interpret our result as follows: acceptability of the dashboard was "marginally high,"

and the adjective rating was "OK-to-Good" [29]. These findings indicate that the participants used it very frequently. They also felt that this system was easy to learn and use.

Situation Awareness Index

The overall SAI score was 3.87 points, and the score of the physician group (3.95) was higher than that of the nurse group (3.80). The top 5 rated items were "concentration support" (4.15), "complexity representation" (4.02 points), "variability representation" (3.96 points), "information quality provided" (3.94 points), and "familiarity of dashboard" (3.94 points). However, the score for "division of attention" was 2.25 points (Table 3).

Table 1. Characteristics of participants.

Characteristic	Physician (n=25)	Nurse (n=27)	Total (N=52)
Age group, n (%)			
20s	2 (8)	6 (22)	8 (15)
30s	19 (76)	19 (70)	38 (73)
40s	1 (4)	2 (7)	3 (6)
50s	3 (12)	0 (0)	3 (6)
Sex, n (%)			
Male	17 (68)	2 (7)	19 (37)
Female	8 (32)	25 (93)	33 (64)
Years of experience, n (%)			
0-3	8 (32)	0 (0)	8 (15)
3-5	6 (24)	4 (15)	10 (19)
5-10	4 (16)	15 (56)	19 (37)
>10	7 (28)	8 (30)	15 (29)

Table 2. System usability scale scores.

Item	Physician, mean (SD)	Nurse, mean (SD)	Total, mean (SD)
Q1. I think that I would like to use this dashboard frequently.	4.52 (0.7)	4.15 (0.5)	4.33 (0.6)
Q2. I found the dashboard unnecessarily complex.	2.44 (0.9)	2.52 (1.0)	2.48 (0.9)
Q3. I thought the dashboard was easy to use.	3.76 (0.7)	3.85 (0.6)	3.81 (0.7)
Q4. I think that I would need the support of a technician to be able to use this dashboard.	3.08 (1.1)	2.89 (1.1)	2.98 (1.1)
Q5. I found that the various functions in this dashboard were well integrated.	3.56 (0.8)	3.78 (0.7)	3.67 (0.7)
Q6. I thought there was too much inconsistency in this dashboard.	2.24 (0.9)	2.07 (0.7)	2.15 (0.8)
Q7. I would imagine that most people would learn to use this dashboard very quickly.	3.72 (0.9)	3.93 (0.6)	3.83 (0.8)
Q8. I found the dashboard very cumbersome to use.	2.12 (0.7)	2.11 (0.6)	2.12 (0.7)
Q9. I felt very confident using the dashboard.	3.84 (0.9)	3.67 (0.8)	3.75 (0.9)
Q10. I needed to learn a lot of things before I could get going with this dashboard.	2.52 (1.1)	2.70 (1.1)	2.62 (1.1)
System usability scale score	67.50 (12.0)	67.69 (11.0)	67.60 (11.4)

Table 3. Situation awareness and dashboard results.

Construct	Item	Physician, mean (SD)	Nurse, mean (SD)	Total, mean (SD)
Instability representation	Q11. The dashboard adequately represents the instability of the ED[a].	3.96 (1.0)	3.78 (0.7)	3.87 (0.8)
Complexity representation	Q12. The dashboard adequately represents the complexity of the ED.	4.12 (0.9)	3.93 (0.6)	4.02 (0.8)
Variability representation	Q13. The dashboard contains key elements that are changing in the ED.	4.08 (0.6)	3.85 (0.8)	3.96 (0.7)
Arousal support	Q14. The dashboard helps me be alert and clearer.	3.92 (0.9)	3.74 (0.7)	3.83 (0.8)
Concentration support	Q15. The dashboard helps me focus on the situation in the ED.	4.20 (0.6)	4.11 (0.6)	4.15 (0.6)
Spare mental capacity support	Q16. I can acquire additional mental capacity in a pressing ED situation.	3.60 (1.2)	3.48 (0.9)	3.54 (1.0)
Division of attention	Q17. The dashboard distracts attention from important tasks of the ED.	2.12 (0.7)	2.37 (0.9)	2.25 (0.8)
Information quantity provided	Q18. The quantity of information provided by the dashboard is appropriate for performing ED tasks.	3.84 (0.9)	3.67 (0.8)	3.75 (0.8)
Information quality provided	Q19. The quality of information provided by the dashboard is appropriate for performing ED tasks.	3.92 (1.0)	3.96 (0.6)	3.94 (0.8)
Familiarity of dashboard	Q20. I can perform ED tasks more proficiently using the dashboard.	4.00 (0.9)	3.89 (0.8)	3.94 (0.9)
Situation awareness index	—	3.95 (0.6)	3.80 (0.5)	3.87 (0.6)

[a]ED: emergency department.

Discussion

Principal Findings

The ED dashboard was successfully developed and implemented. The system is independent of manual input and is fully connected to the legacy HIS. The graphical and statistical concepts were determined during the developmental period, and they were upgraded gradually. Though clinical dashboards for EDs have been developed in other studies [15,30-32], this study is the first to examine its long-term use and conduct serial upgrades.

We used SUS, a formal, highly validated usability test, and obtained a score of 67.6 points from physicians and nurses. This score could be interpreted as indicating "marginally high acceptability" with "OK-to-Good usability." Additionally, the ED staff responded that the dashboard presented the situation in the ED effectively and that they could better focus on the changing situation in the ED by using it. The quality of the information provided by the dashboard was rated high; however, the quantity of information was rated relatively low. There is a need for a systematic investigation to establish the information that ED staff seek and a subsequent improvement plan to reflect the same in the system.

Clinical Aspects

It has been one of the major responsibilities of ED staff to keep patients' processes on track and ensure timely results [33]. To do so, ED providers had to call numerous departments and browse through multiple windows repeatedly during their duty time. As the tasks and volume of ED work-ups grew, this timekeeping function had become significantly heavy. The ED dashboard described in this study focused on providing a visual representation of this "hidden" information pertaining to the ED process, such that ED providers could plan and carry out their tasks proactively.

The visualized information not only pertained to individual patients' processes but also reported the department's performance status. As the government and insurance companies focus on performance as a group of patients, it has become an essential job for the system administrator to be able to assess real-time statistics. ED providers' responsibility as administrators demands tools like our dashboard during their work.

Our ED dashboard affects the workflow of various medical personnel in the ED in a variety of ways, ranging from simple information delivery to clinical decision-making support. If the dashboard indicates that the ED is extremely overcrowded, the ED chief could contact the national acute care system to control the transfer of new patients. Additionally, physicians could grasp the timely information of each patient using this dashboard. The charge nurse could improve the efficacy of the ED by reassigning each nurse to more appropriate zones based on the information obtained from the dashboard. Triage nurses could also use the dashboard to allocate patients to appropriate treatment zones.

Comparison With Previous Work

Recent review articles have indicated that an ED dashboard could be useful for "saving time and reducing the risk of errors or delay" [18]. However, the association between the introduction of an electronic dashboard and the mental workload of ED staff is controversial. Further, there is little evidence to support that it improves clinical outcomes [18].

A major difference between our system and other electronic boards is that our dashboard does not allow manual or direct input of patient-specific information. We have fully synchronized our dashboard with the legacy system such that physicians' and nurses' additional inputs are not required to use the dashboard. By achieving this goal, our system was found to provide effective arousal support, concentration support, and spare mental capacity (3.83 points, 4.15 points, and 3.54 points, respectively) even in the busy clinical setting.

Limitations

First, this is a single-center case study with its unique HIS. Its feasibility and usability should be validated in other institutions. Considering that the implementation of this system in other institutions has been discussed, subsequent investigation on ED dashboard utilization is expected in the near future.

The measures used for the dashboard are not universally agreed upon. They have mainly been used in a highly crowded EDs of a teaching hospital. When used in smaller EDs, the measures of interest would be different. Additionally, the measures were not compared with the national standard, which requires further research.

The SAI has not been validated. We searched for questionnaires to investigate the association between dashboards and situation awareness, but we could not find any that suited our purpose. Therefore, we developed a questionnaire based on the SART and applied it in this study. However, this questionnaire is not validated, and therefore, its interpretation value is limited.

Finally, we could not investigate the association between the effectiveness of the ED dashboard and clinical outcomes. However, it is difficult to identify this association because the ED is one of the most complex systems affected by several uncontrollable factors. Therefore, multicenter comparative studies need to be conducted to examine this association.

Conclusions

We developed a real-time autonomous ED dashboard and successfully used it for 5 years. ED physicians and nurses rated the usability of the ED dashboard as "OK-to-Good." We realize that continuous maintenance is important because the dashboard should reflect the situation of the ED.

Acknowledgments

The authors acknowledge the members of the Department of Quality Improvement and of the Health Information Center for their contribution to the design and development of the emergency department dashboard. This research was supported by Basic Science Research Program through the National Research Foundation of Korea and funded by the Ministry of Education (Grant #2017R1D1A1B03034690).

Conflicts of Interest

None declared.

References

1. Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. hum factors 1995 Mar 01;37(1):32-64. [doi: 10.1518/001872095779049543]
2. Wickens CD. Situation Awareness and Workload in Aviation. Curr Dir Psychol Sci 2016 Jun 24;11(4):128-133 [FREE Full text] [doi: 10.1111/1467-8721.00184]
3. De Stefano C, Philippon A, Krastinova E, Hausfater P, Riou B, Adnet F, et al. Effect of emergency physician burnout on patient waiting times. Intern Emerg Med 2018 Dec;13(3):421-428. [doi: 10.1007/s11739-017-1706-9] [Medline: 28677043]
4. Garrouste-Orgeas M, Perrin M, Soufir L, Vesin A, Blot F, Maxime V, et al. The Iatroref study: medical errors are associated with symptoms of depression in ICU staff but not burnout or safety culture. Intensive Care Med 2015 Feb;41(2):273-284 [FREE Full text] [doi: 10.1007/s00134-014-3601-4] [Medline: 25576157]
5. Pines JM, Iyer S, Disbot M, Hollander JE, Shofer FS, Datner EM. The effect of emergency department crowding on patient satisfaction for admitted patients. Acad Emerg Med 2008 Sep;15(9):825-831. [Medline: 19244633]
6. Pines JM, McCarthy ML. The crowding-effectiveness link: it doesn't matter how fast we deliver care if we don't deliver it right. Ann Emerg Med 2011 Mar;57(3):201-202. [doi: 10.1016/j.annemergmed.2010.12.005] [Medline: 21232815]
7. Warden G, Griffin RB, Erickson SM, Mchugh M, Wheatley B, Dharshi A, et al. Hospital-based emergency care: at the breaking point. Washington, DC: National Academies Press; Jun 2006.
8. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, et al. The effect of emergency department crowding on clinically oriented outcomes. Acad Emerg Med 2009 Jan;16(1):1-10 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00295.x] [Medline: 19007346]
9. Kang J, Kim J, Jo YH, Kim K, Lee JH, Kim T, et al. ED crowding and the outcomes of out-of-hospital cardiac arrest. Am J Emerg Med 2015 Nov;33(11):1659-1664. [doi: 10.1016/j.ajem.2015.08.002] [Medline: 26324002]

10. Cha WC, Cho JS, Shin SD, Lee EJ, Ro YS. The impact of prolonged boarding of successfully resuscitated out-of-hospital cardiac arrest patients on survival-to-discharge rates. Resuscitation 2015 May;90:25-29. [doi: 10.1016/j.resuscitation.2015.02.004] [Medline: 25680821]
11. Storrow AB, Zhou C, Gaddis G, Han JH, Miller K, Klubert D, et al. Decreasing lab turnaround time improves emergency department throughput and decreases emergency medical services diversion: a simulation model. Acad Emerg Med 2008 Nov;15(11):1130-1135 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00181.x] [Medline: 18638034]
12. White BA, Baron JM, Dighe AS, Camargo CA, Brown DFM. Applying Lean methodologies reduces ED laboratory turnaround times. Am J Emerg Med 2015 Nov;33(11):1572-1576 [FREE Full text] [doi: 10.1016/j.ajem.2015.06.013] [Medline: 26145581]
13. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of emergency department crowding. Ann Emerg Med 2003 Aug;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
14. Few S. Information Dashboard Design: The Effective Visual Communication of Data. Sebastopol, California: O'Reilly Media, Incorporated; 2006.
15. Franklin A, Gantela S, Shifarraw S, Johnson TR, Robinson DJ, King BR, et al. Dashboard visualizations: Supporting real-time throughput decision-making. J Biomed Inform 2017 Dec;71:211-221. [doi: 10.1016/j.jbi.2017.05.024] [Medline: 28579532]
16. Jarvis PRE. Improving emergency department patient flow. Clin Exp Emerg Med 2016 Jun;3(2):63-68 [FREE Full text] [doi: 10.15441/ceem.16.127] [Medline: 27752619]
17. Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015 Feb;84(2):87-100. [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
18. Randell R, Greenhalgh J, Wyatt J, Gardner P, Pearman A, Honey S, et al. Electronic whiteboards: review of the literature. Stud Health Technol Inform 2015;210:389-393. [Medline: 25991172]
19. Cha WC, Ahn KO, Shin SD, Park JH, Cho JS. Emergency Department Crowding Disparity: a Nationwide Cross-Sectional Study. J Korean Med Sci 2016 Aug;31(8):1331-1336 [FREE Full text] [doi: 10.3346/jkms.2016.31.8.1331] [Medline: 27478347]
20. Bjørn P, Hertzum M. Artefactual Multiplicity: A Study of Emergency-Department Whiteboards. Comput Supported Coop Work 2010 Oct 27;20(1-2):93-121. [doi: 10.1007/s10606-010-9126-7]
21. Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. J Am Med Inform Assoc 2018 May 01;25(5):585-592. [doi: 10.1093/jamia/ocx118] [Medline: 29126196]
22. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Apr;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
23. Bloomrosen M, Starren J, Lorenzi NM, Ash JS, Patel VL, Shortliffe EH. Anticipating and addressing the unintended consequences of health IT and policy: a report from the AMIA 2009 Health Policy Meeting. J Am Med Inform Assoc 2011;18(1):82-90 [FREE Full text] [doi: 10.1136/jamia.2010.007567] [Medline: 21169620]
24. Zhang AL, Chen RX, Kang MF, Fan HL, Wang WL. Study of the regulatory effect of acupuncture on rotation-induced gastric dysrhythmia in rabbits. World J Gastroenterol 1997 Mar 15;3(1):54-55 [FREE Full text] [doi: 10.3748/wjg.v3.i1.54] [Medline: 27006588]
25. Tsou M, Jung C, Allen C, Yang J, Gawron J, Spitzberg B, et al. Social media analytics and research test-bed (SMART dashboard). In: Proceedings of the 2015 International Conference on Social Media & Society Article No. 2. 2015 Presented at: The 2015 International Conference on Social Media & Society; July 27 - 29, 2015; Toronto, Canada. [doi: 10.1145/2789187.2789196]
26. Park GE, Ko J, Peck KR, Lee JY, Lee JY, Cho SY, et al. Control of an Outbreak of Middle East Respiratory Syndrome in a Tertiary Hospital in Korea. Ann Intern Med 2016 Jul 19;165(2):87-93. [doi: 10.7326/M15-2495] [Medline: 27272273]
27. Jordan P, Jordan J, McClelland I. SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry. Boca Raton, Florida: Taylor & Francis, Inc; 1996:189-194.
28. Selcon S, Taylor R. Evaluation of the Situational Awareness Rating Technique(SART) as a tool for aircrew systems design. In: AGARD, Situational Awareness in Aerospace Operations. 1989 Presented at: Advisory Group For Airospace Research & Development; october 02 1989; Copenhagen, Denmark URL: http://www.dtic.mil/dtic/tr/fulltext/u2/a223939.pdf
29. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies 2009 May;4(3):114-123 [FREE Full text]
30. Staib A, Sullivan C, Jones M, Griffin B, Bell A, Scott I. The ED-inpatient dashboard: Uniting emergency and inpatient clinicians to improve the efficiency and quality of care for patients requiring emergency admission to hospital. Emerg Med Australas 2017 Jun;29(3):363-366. [doi: 10.1111/1742-6723.12661] [Medline: 27592365]
31. Stone-Griffith S, Englebright JD, Cheung D, Korwek KM, Perlin JB. Data-driven process and operational improvement in the emergency department: the ED Dashboard and Reporting Application. J Healthc Manag 2012;57(3):167-80; discussion 180. [Medline: 22724375]

32. Batley NJ, Osman HO, Kazzi AA, Musallam KM. Implementation of an emergency department computer system: design features that users value. J Emerg Med 2011 Dec;41(6):693-700. [doi: 10.1016/j.jemermed.2010.05.014] [Medline: 20619572]
33. Randell R, Keen J, Gates C, Ferguson E, Long A, Ginn C, et al. Managing Quality and Safety in Real Time? Evidence from an Interview Study. Stud Health Technol Inform 2016;228:23-27. [Medline: 27577334]

Abbreviations

ED: emergency department
HIS: hospital information system
LAN: Local Area Network
SAI: situation awareness index
SART: Situation Awareness Rating Technique
SUS: system usability scale

Edited by C Dias; submitted 02.04.18; peer-reviewed by R Randell, A Franklin, A Staib, J Jones; comments to author 14.05.18; revised version received 02.07.18; accepted 10.08.18; published 22.11.18.

Please cite as:
Yoo J, Jung KY, Kim T, Lee T, Hwang SY, Yoon H, Shin TG, Sim MS, Jo IJ, Paeng H, Choi JS, Cha WC
A Real-Time Autonomous Dashboard for the Emergency Department: 5-Year Case Study
JMIR Mhealth Uhealth 2018;6(11):e10666
URL: http://mhealth.jmir.org/2018/11/e10666/
doi:10.2196/10666
PMID:30467100

Review

App Features for Type 1 Diabetes Support and Patient Empowerment: Systematic Literature Review and Benchmark Comparison

Antonio Martinez-Millana[1*], PhD; Elena Jarones[1*], BSc Biomedical Engineering; Carlos Fernandez-Llatas[1*], PhD; Gunnar Hartvigsen[2*], Prof Dr; Vicente Traver[1*], PhD

[1]ITACA, Universitat Politècnica de València, Valencia, Spain

[2]Department of Computer Science, University of Tromsø, Tromsø, Norway

[*]all authors contributed equally

Corresponding Author:
Antonio Martinez-Millana, PhD
ITACA
Universitat Politècnica de València
Camino de Vera s/n
Valencia, 46022
Spain
Phone: 34 963877606
Email: anmarmil@itaca.upv.es

Abstract

Background: Research in type 1 diabetes management has increased exponentially since the irruption of mobile health apps for its remote and self-management. Despite this fact, the features affect in the disease management and patient empowerment are adopted by app makers and provided to the general population remain unexplored.

Objective: To study the gap between literature and available apps for type 1 diabetes self-management and patient empowerment and to discover the features that an ideal app should provide to people with diabetes.

Methods: The methodology comprises systematic reviews in the scientific literature and app marketplaces. We included articles describing interventions that demonstrated an effect on diabetes management with particular clinical endpoints through the use of mobile technologies. The features of these apps were gathered in a taxonomy of what an ideal app should look like to then assess which of these features are available in the market.

Results: The literature search resulted in 231 matches. Of these, 55 met the inclusion criteria. A taxonomy featuring 3 levels of characteristics was designed based on 5 papers which were selected for the synthesis. Level 1 includes 10 general features (Personalization, Family support, Agenda, Data record, Insulin bolus calculator, Data management, Interaction, Tips and support, Reminders, and Rewards) Level 2 and Level 3 included features providing a descriptive detail of Level 1 features. Eighty apps matching the inclusion criteria were analyzed. None of the assessed apps fulfilled the features of the taxonomy of an ideal app. Personalization (70/80, 87.5%) and Data record (64/80, 80.0%) were the 2 top prevalent features, whereas Agenda (5/80, 6.3%) and Rewards (3/80, 3.8%) where the less predominant. The operating system was not associated with the number of features (P=.42, F=.81) nor the type of feature (P=.20, χ^2=11.7). Apps were classified according to the number of level 1 features and sorted into quartiles. First quartile apps had a regular distribution of the ten features in the taxonomy whereas the other 3 quartiles had an irregular distribution.

Conclusions: There are significant gaps between research and the market in mobile health for type 1 diabetes management. While the literature focuses on aspects related to gamification, rewarding, and social communities, the available apps are focused on disease management aspects such as data record and appointments. Personalized and tailored empowerment features should be included in commercial apps for large-scale assessment of potential in the self-management of the disease.

(JMIR Mhealth Uhealth 2018;6(11):e12237) doi:10.2196/12237

KEYWORDS

mHealth; type 1 diabetes mellitus; patient empowerment; apps; diabetes self-management

Introduction

Diabetes mellitus is a metabolic syndrome, which comprises an impaired insulin production and action [1]. Type 1 diabetes mellitus (T1DM) prevalence is rapidly rising throughout the world [2]. Supporting T1DM patients is a major health care challenge, as it involves many aspects of daily routine activities (eg, food intake, physical activity, motivation) and specific knowledge of disease mechanisms (eg, blood glucose regulation, insulin intake) [3,4].

Despite the promise of mobile health (mHealth) in the specific field of diabetes [5,6] and the explosion of diabetes-related apps in markets, T1DM management is yet undertaken on a routine care basis, in which glycemic levels (blood glucose and hemoglobin A_{1c} (HbA_{1c}) among others) and other health-related outcomes are supervised by general practitioners, endocrinologists, and nurses. This type of care has shown limited effects on empowering patients to control blood sugar levels [7].

In their review from 2011, Chomutare et al [8] compared the recommendations from evidence-based guidelines and the features of mobile apps. The evaluation of features was analyzed for 6 types of functionalities: (1) self-monitoring, (2) education, (3) alerts, (4) reminders, (5) social media, and (6) personal health records synchronization. The authors concluded that apps did not include education structural elements from the empowerment of patients other than including functionalities for routine management (record of lifestyle and measurements). The lack of core recommendations compromised the effectiveness of mobile health in diabetes clinical management [9].

Since then, several studies have been conducted to test the effectiveness of apps in reduced samples of patients. In such studies, researchers evaluate apps with different approaches and patient groups, yielding conflicting conclusions due to the different methodology of the interventions [10].

This study aimed to assess whether app manufacturers adopted the findings from mHealth evidence-based interventions in diabetes. The rationale is to mind the gap between research and the market to identify the features that are not available in commercial apps. The methodology comprises 2 systematic reviews using the PRISMA methodology, one in the scientific literature (without meta-analysis) and the other in app marketplaces. The systematic review of scientific literature is focused on interventions with apps on patients with T1DM into randomized controlled trials (RCTs). In the first stage, we built the schema of what could be an ideal diabetes app (including all the features which have shown a positive effect on diabetes management) and to then analyze the characteristics of available apps and their distance to the ideal app.

Methods

Selection Criteria

The primary objective of this study was to review scientific literature using the PRISMA methodology to enumerate evidence-based features that have demonstrated a positive effect in the management of T1DM in RCTs. Inclusion criteria were defined as (1) a mHealth intervention on T1DM patients using an app for remote and self-management of the disease and (2) the intervention was performed in an RCT and reports on clinical outcomes (HbA_{1c}, in-range time or self-monitoring blood glucose). Exclusion criteria included (1) gray literature, (2) studies not reporting RCTs on T1DM (eg, type 2 diabetes, gestational diabetes), and (3) studies not using an app (eg, text messages, manual notes) or not describing the app's functionality. The secondary objective was to compare the evidence-based features with the characteristics of the available apps.

Recruitment Strategy

The source of the literature review were online journal databases and indexers (PubMed, Medline, Google Scholar, and Cochrane Trials). We searched a combination of keywords including type 1 diabetes mellitus, mobile health, RCTs and self-management. The complete search strategy, combination of keywords in the queries and results for PubMed are described in Multimedia Appendix 1. The search included all the parts of the manuscript (title, abstract, keywords, and text). The review was conducted in July 2018, and we focused our review on recently published papers (January 2012-July 2018). Only publications in English were considered for inclusion.

Subsequently, the approach for recruiting apps was twofold. First, a web search was conducted by using keywords of 3 different groups: (1) "diabetes mellitus 1 AND apps AND (android OR iPhone) AND self-management," (2) "diabetes mellitus," and (3) "diabetes apps AND self-management." A second search was conducted in Google Play and the App Store to recruit a greater number of apps. In this case, we introduced 2 keywords: "diabetes AND management." After collecting all the available apps, the screening and selection were done in the same way as with the publications (using a PRISMA flow diagram). Subsequently, the selected apps were downloaded and tested to know which of the characteristics obtained with the systematic literature review were available in the apps.

Data from the literature review and the apps assessment was extracted by 2 of the authors (AMM and EJP) using a structured data form. For the literature review, we extracted data related to the study (year of publication, sample size, the age of participants, methodology, intervention, clinical endpoints, features, usability, and satisfaction). Studies were assessed using the Cochrane Collaboration's tool to assess the risk of bias of included (selection, reporting, performance, and attrition) [11]. For the app review, we extracted the type of features each app offered, the operative system, language, icon, and link.

Statistical Analysis

A descriptive analysis of the features in the apps was done before association analytics. Association of the type of features and the type of operative system was evaluated with a two-way analysis of variance, in which we assessed the *P* value and the F statistic. Association of the number of features and the operative system was assessed using Kruskal-Wallis tests at 95% CI, in which we calculated the *P* value and the chi-square value. A value of $P<.05$ was considered as significant The

Feature Factor (FF) was defined as the polynomial formula (Equation 1) FF= $N_{L1}^3+N_{L2}^2+N_{L3}$ where N_{Lx} is the number of features in level 1, level 2, and level 3 respectively. We used MATLAB 2017Ra framework under Academic License to conduct the statistical analysis.

Results

Systematic Literature Review of Features

Figure 1 shows the cascade flow of search process for the literature review. A total of 449 scientific publications were found in the selected search engines with the chosen keywords and the date constraints. After excluding 216 duplicated entries, we obtained a total of 233 original publications. In the second screening of these 233 articles, 176 were not related to the objective of this study and were eliminated. A big part of the remaining 51 articles was not matching the inclusion criteria. Ten of the 51 (19.6%) did not refer to T1DM, 5/51 (9.8%) were not related to mHealth, 11/51 (21.6%) were reviews, 17/51 (33.3%) were not in the scope of our study, 3/51 (5.9%) did not allow full-text publication accessibility, and 5/51 (9.8%) did not provide relevant data for the synthesis. Finally, once read and analyzed, 6/57 (11.8%) scientific publications were included for the synthesis [12-17].

These 6 studies reported on the results evaluating the benefits of apps in the management of T1DM (Table 1). These studies consisted of 266 participants (as we only considered intervention groups) in RCTs and lasted from 3 to 12 months. Only studies with a clear clinical endpoint were considered (eg, HbA_{1c} improvement, self-monitoring blood glucose (SMBG) daily rate).

Figure 1. Selection of the literature for evidence-based features of mobile apps for type 1 diabetes management and empowerment.

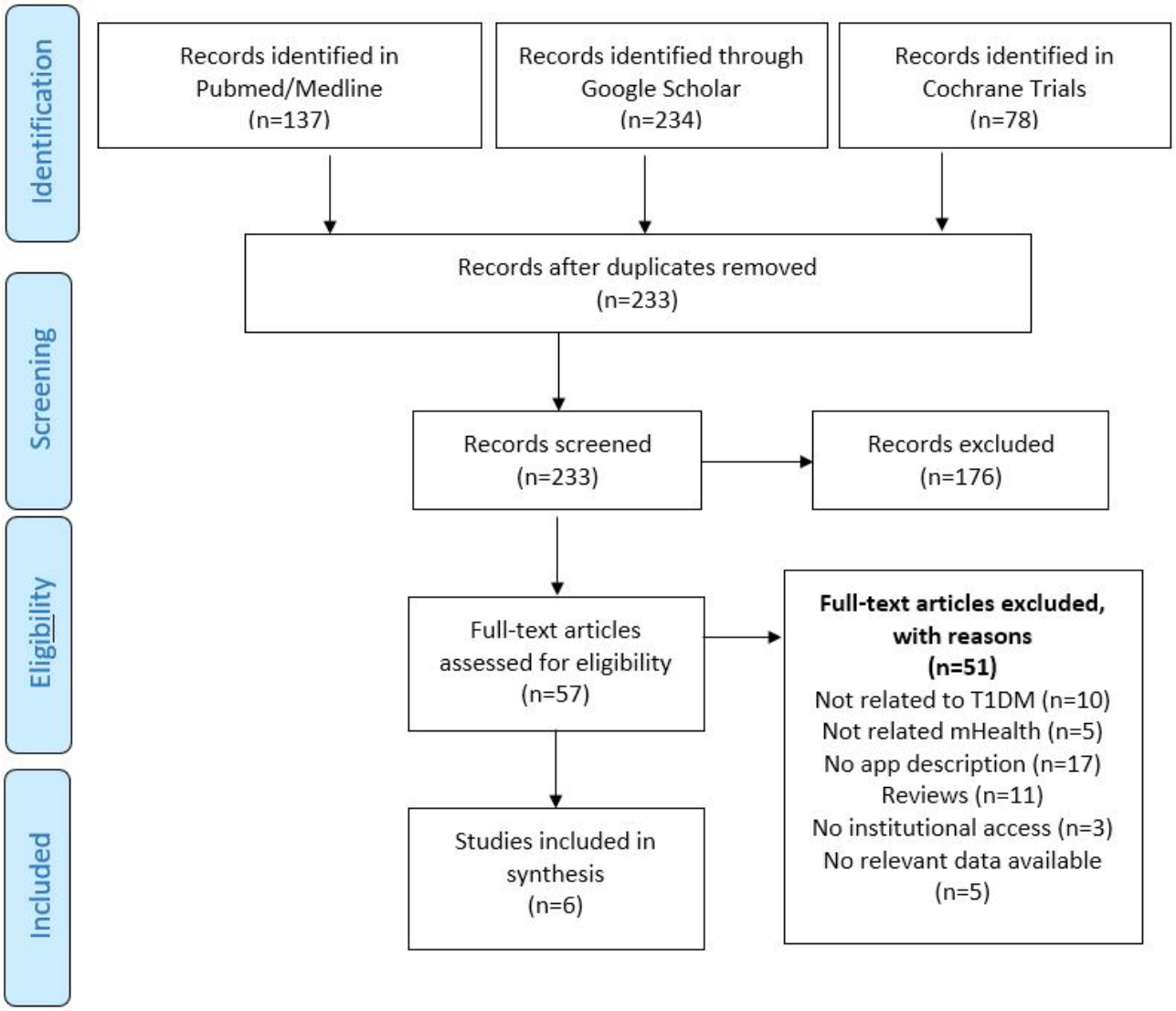


Table 1. Characteristics of the studies evaluating mobile apps for type 1 diabetes mellitus management and empowerment.

Characteristics	Castensøe-Seidenfaden [12]	Cafazzo [13]	Goyal [14]	Kirwan [15]	Clemens [16]	Ryan [17]
Publication year	2018	2012	2017	2013	2017	2017
Intervention, n	76	20	46	25	81	18
Age (years), mean (SD)	17.6 (2.6)	14.9 (1.3)	14.1 (1.7)	35.9 (10.6)	14.0 (10.4-15.9)[a]	40 (13.9)
Time[b] (years), mean (SD)	8.0 (4.5)	NS[c]	7.1 (3.2)	19.7 (9.6)	4.9 (2.7-7.5)[a]	27.3 (14.9)
Duration (months)	12	3	12	6	3.6	4
Intervention type	Usual care App	App	Usual care App	App Feedback	Retrospective analysis	Usual care App
App name	Young with Diabetes	Bant	Bant	Glucose Buddy	NS	NS
HbA_{1c}[d] outcome	No significant change	No significant change	Decrease by 0.58% (*P*=.02)	Decrease in mean (SD) from 9.08% (1.18%) to 7.8% (0.75%)	No significant change	Decrease in median (9.1% to 7.8%)
SMBG[e] outcome	—	Increased mean daily frequency (2.4 to 3.6, *P*=.006, n=12)	—	No significant change	Increased 2.3 times	—
App perceived usefulness[f]	Chat Room (among young people)	Reminders, blood glucose regulation, insulin and food regulation, emergency readiness, exercise	Trending feature, logbook, and home menu (statistics)	NA[g]; texting extensively used	Data synchronization	Bolus calculator and glucose control. Badges used by 17%
User satisfaction	>80% would recommend	88% would continue to use	76% "satisfied/very satisfied" 96% would continue using app	NA	NA	NA

[a]Median and interquartile range.

[b]Since diagnosis.

[c]NS: not specified.

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]SMBG: self-monitoring blood glucose.

[f]Either a 5-point or 10-point Likert scale was used to score.

[g]NA: not assessed.

Overall, the included trials adequately achieved a low level of risk of bias (Table 2). Three of the studies did not report on random sequence generation, and 2 did not adequately report concealing the allocation sequence to determine if the group could have been foreseen. Only 2 trials blinded the outcome assessment. Besides, all trials except 1 adequately described data completeness (including exclusion and drops out) and did not perform selective outcome description.

The majority of these studies assessed the features of the app which had a higher perceived usefulness or a good adoption among study subjects. Castensøe-Seidenfaden et al [12] found the Chat Room as the most rated feature, a virtual space for user communication in which users posted comments about alcohol and sex. Cafazzo et al [13] report the user-centered design and evaluation in 12 subjects of the Bant app; a mobile app that's focused on the simple and automated transfer of glucometer readings, a social community and rewards for healthy behavior adherence (average 8 rewards per user in a three-month trial). An updated version of the Bant app was tested by Goyal et al [14] in a 12-months trial involving 46 subjects, pointing out the trends, logbook, and homepage as the preferred features of the app. Kirwan et al [15] examined the effectiveness of Glucose Buddy, a free app combined with texting from a certified diabetes educator in 25 subjects.

Table 2. Cochrane Collaboration's tool risk-of-bias assessment for the clinical outcomes (hemoglobin A_{1c} and self-monitoring blood glucose changes) of the included study papers.

	Castensøe-Seidenfaden [12]	Cafazzo [13]	Goyal [14]	Kirwan [15]	Clemens [16]	Ryan [17]
Random sequence generation	Low	High	Low	Low	High	High
Allocation concealment	Low	High	Low	Low	High	Low
Blinding of participants and personnel	N/A[a]	N/A	N/A	N/A	N/A	N/A
Blinding of outcome assessment	Low	High	High	Low	High	Unclear
Incomplete outcome data	Low	Low	Low	High	Low	Low
Selective reporting	Low	Low	Low	High	Low	Low
Other bias	Low	Unclear	Low	Unclear	Unclear	Low

[a]N/A: not applicable (both patient and doctors know the group they are allocated).

Clemens and Staggs [16] reported a retrospective study on 81 youth T1DM patients who used a mobile app connected to a glucose meter, showing that data synchronizations were associated with an increased rate of SMBG but not with HbA_{1c} or mean glucose values. Beyond data synchronization, no more app features were provided by authors in this paper. Finally, Ryan et al [17] developed and evaluated Intelligent Diabetes Management, an app for insulin bolus calculation and an electronic diabetes diary in clinical visits. The app included a "badge" feature for motivational accolades based on the reported measurements. The 18 subjects, who completed the study, rated the insulin calculator with 8 in a 10-point Likert scale and only 17% of subjects used the "badges" feature regularly.

Three of these studies evaluated the user satisfaction in terms of willingness to use the app after the trial and willingness to recommend its use to peers. Each showed a high percentage of subjects willing to continue using the app and willing to recommend the app to peers [12-14]. In our study, we analyzed the features available in the mobile apps (rated or not) and gathered in a hierarchical map of functionalities.

Features Taxonomy of an Ideal App

The qualitative study of the apps used in the studies allowed us to extract the characteristics of the apps that users rated as key and researchers considered of value. These features were sorted into functional areas beyond the traditional 4 clinical areas of diabetes management taxonomy (glycemic control, carbohydrate intake, insulin and exercise) [18]. The new proposed taxonomy (Figure 2) had 3 hierarchical levels, the first of which had 10 areas: (1) Personalization, (2) Family support, (3) Agenda, (4) Data record, (5) Insulin bolus calculation, (6) Data management, (7) Interaction, (8) Tips and support, (9) Reminders, and (10) Rewards. In this area, we included all the functions and settings that users can adjust to customize the app interface, the behavior, and calculations (if any). The basic feature was building a user profile for the social community of app users, with the capability of adding a picture or avatar. The second feature in Personalization was the possibility of configuring the blood glucose diary (or record), adjusting personalized thresholds for supporting the patient, and product recommendation. The third feature was the configuration of the user mood. The fourth feature was the self-adjustment of goals of any type (glucose goals, carbs intake, weight loss). The fifth feature was setting up the type of insulin, the brand, and the intake method. The sixth feature was configuring tailored advices (eg, topic selection, when to show the advice). The seventh feature was the personalization of the measurements reporting, instead of having a default form for data introduction, for all the relevant variables (Figure 2). The last feature of personalization was the customization of physical activity reporting.

Family support was a feature in which relatives can perform a follow-up of the data introduced by the patient in which we have identified read access and read-and-write access. This feature also contains literature support for relatives in the management of T1DM through web links and books. The Agenda feature was a common factor in the apps analyzed in the review, and it was mainly devoted for scheduling medical and personal appointments. Another recurrent feature was the storage of measurements and reports related to food intake, insulin intake, and physical activity, which are under the Data record feature. In this feature, we made a distinction between the apps that had manual data entry (eg, forms, pictures, voice recognition, avatar) and the automatic entry of data using wireless sensors, mobile phone sensors (such as an accelerometer for physical activity) and smartwatches and wearables. The last feature (right side in Figure 2) was the Insulin bolus calculator, which provided calculations for exogenous insulin injection for fast action (bolus) or slow action (basal), and for the insulin pump device (bolus and basal).

The Data management feature involved the capability of the app of exporting, storing, and analyzing the data collected in the app. In the review, one of the top-rated features was the graphical representation of measurements and the calculation of statistics based on these measurements. The second feature in the right side of Figure 2 refers to the ability of the app to interact with other subjects beyond the T1DM user (clinicians, parents, and other app users). This interaction was generic and involves social communities, participation in forums, and data sharing. All the analyzed apps had a feature for Tips and support, which involved a user manual of the app. Some of them also included generic information related to diet and physical activity. Only 1 had online 24/7 medical support and emergency management. The Reminders feature enabled users to set up alerts for measuring and recording blood glucose tests, insulin intakes, insulin and glucometer strip purchasing, and attending medical or personal appointments scheduled in the Agenda.

Finally, the Reward feature contained the motivation mechanisms that the app implements for control and therapy adherence. Rewarding strategies in the review were related for providing credit in virtual markets, premium access to the app (rewards for enhanced functionalities) and other types of credit for online stores (Amazon, eBay vouchers).

Systematic Review of Commercial Diabetes Management Apps

We evaluated the available apps through an online search, the App Store (Apple) and Google Play (Android) for diabetes management and support. Following the PRISMA methodology approach (Figure 3), we obtained 93 single apps devoted to T1DM (20/93 (21.5%) from the online search and 73/93 (78.5%) from the previously mentioned app markets). Five of the 93 apps (5.4%) were excluded because the links for downloading the software were unavailable. Of the remainder, 8/87 (91.9%) apps were excluded from the detailed analysis: 2/87 (2.3%) because they were not tailored to T1DM, 2/87 (2.3%) because they focused on healthcare professionals, 2/87 (2.3%) because the access was secured with a password (corporate use), 1/87 (1.1%) because it was not free and 1 (1.1%) because it was a copy of another app with different name and icon. Finally, 80 apps (Figure 4) were assessed to discover the extent to which these apps fulfilled the feature schema of an ideal app taxonomy.

Android and iOS apps were tested during a month since many apps needed control over a longer period to provide data (graphics, statistics) to the user. A mHealth expert analyzed the apps and collected all the features which matched the taxonomy of the ideal diabetes management app. Features were flagged with a Boolean value depending on whether the app provided the analyzed feature. Occasionally, the mHealth expert was provided with clarification to ensure that the analysis was more precise.

Figure 2. Taxonomy of the features of an ideal app according to the evidence-based effectiveness of mobile health in diabetes support and empowerment.

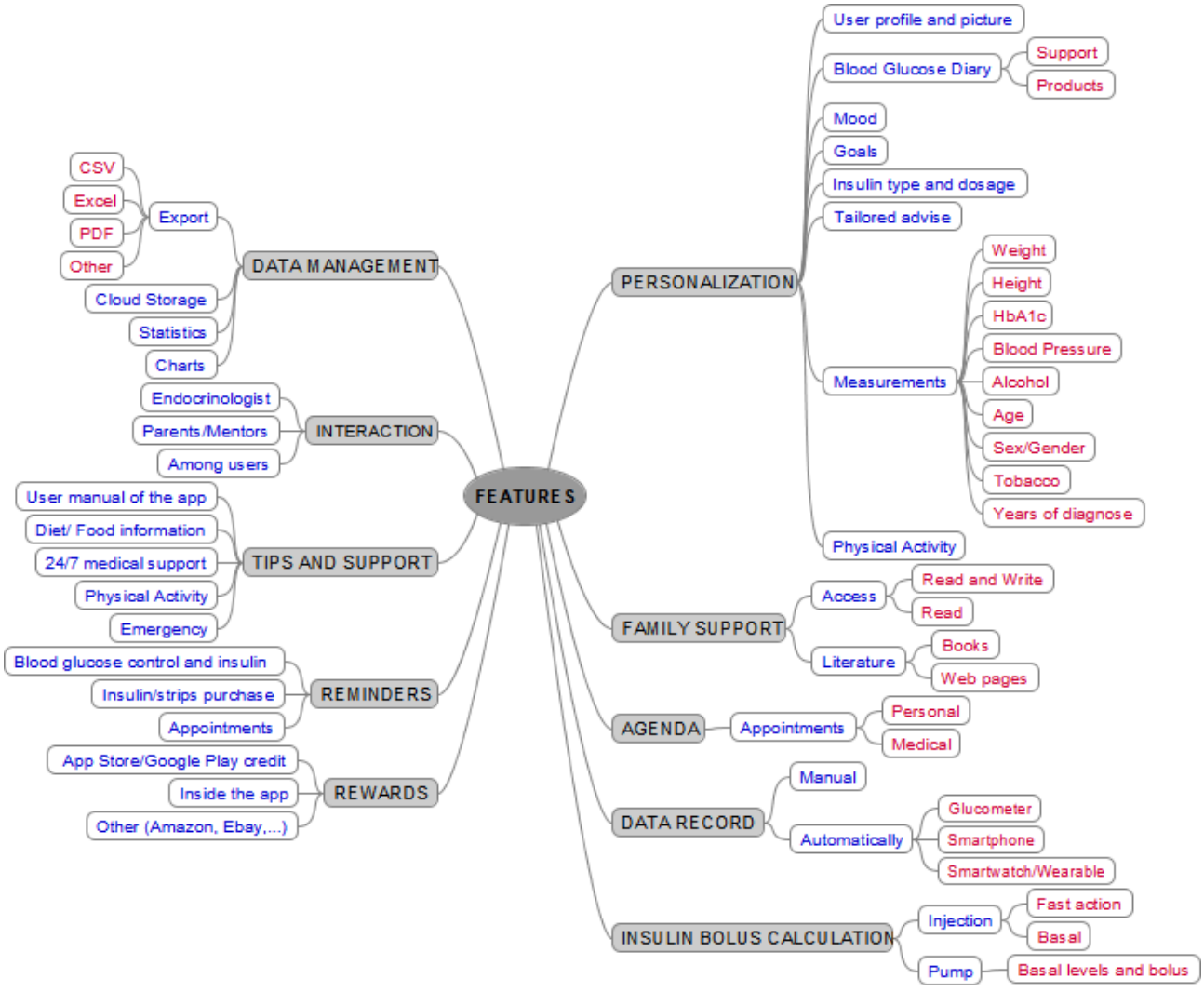


Figure 3. Selection of the available apps for type 1 diabetes management and empowerment.

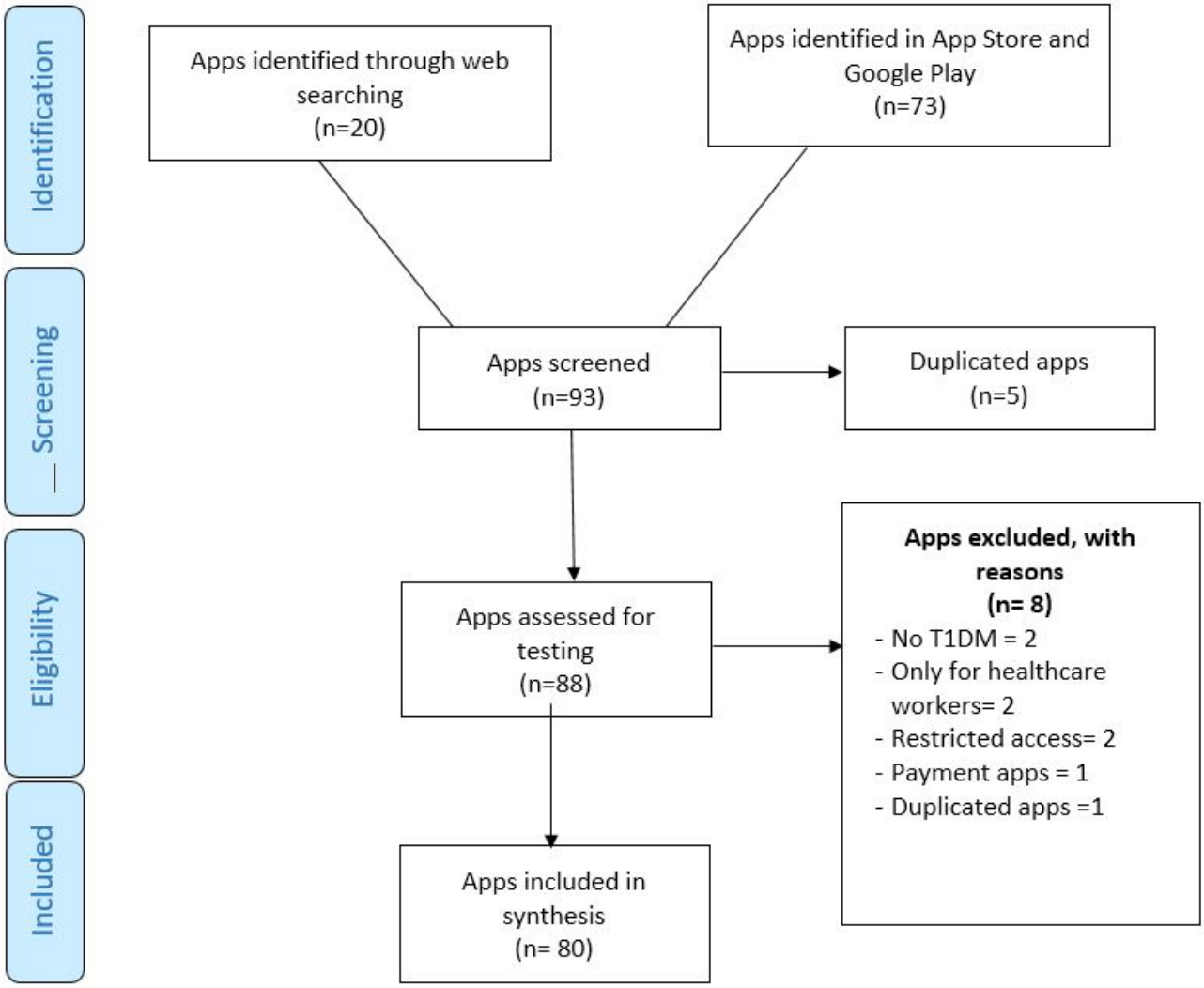
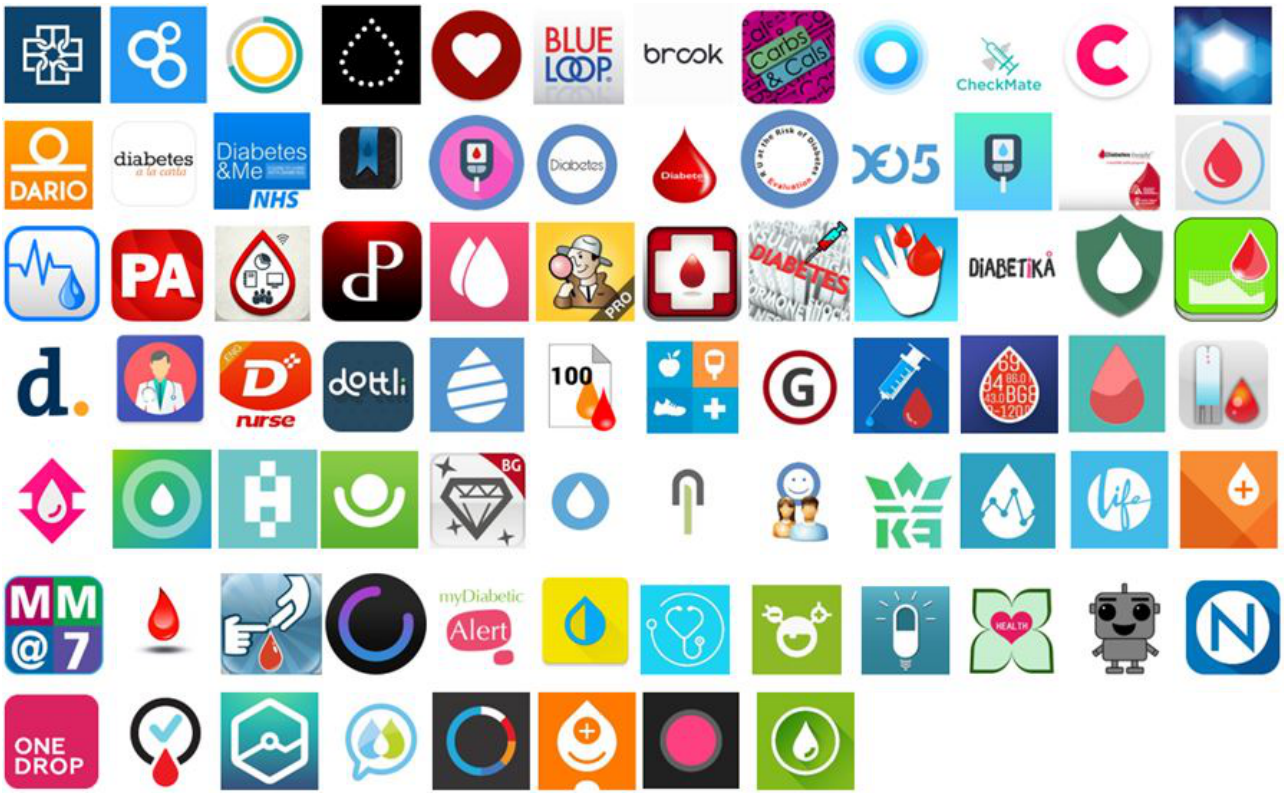


Figure 4. Apps included in the analysis. From top-down left-right: Balansio, Bant, BeatO, Beyond type 1 diabetes, Blood glucose tracker, BlueLoop, Brook, Carbs & Cals, Center health, Checkmate diabetes, Chron, Contour diabetes, Dario, Diabetes a la carta, Diabetes & Me, Diabetes connect, Diabetes diary, Diabetes diet and management, Diabetes digest, Diabetes evaluation, Diabetes experience day, Diabetes ID, Diabetes insight, Diabetes kit blood, Diabetes metrics, Diabetes PA, Diabetes pal, Diabetes parent, Diabetes passport, Diabetes pilot pro, Diabetes plus, Diabetes treatment, Diabetes vue, Diabetika, Diaguard, Diario de sangre, Diasend, DMI from zero to hero, Dnurse, Dottli, Dr. Diabetes, Easy diabetes, Glooko, Glucool diabetes, Glucosa compañero, Glucose buddy, Glucose wiz, Uright, Glucosio, GluQUO, Health2sync, Helparound, iFora, Inrange, Insulclock, Kids and teens diabetes, Kingfit, La diabetes M, Life in control, MedM diabetes, Meet me, Mi glucemia, Monitor de glucosa, Mumoactive, My diabetic alert, Nagbot, Neptun, One drop, Ontrack diabetes, PredictBGL, Social diabetes, SOS diabetes, Sugar sense, Sugarmate, Track3lite.



The Gap Between Literature and Apps in Diabetes Management

None from the 80 analyzed apps fulfilled the criteria of the taxonomy of an ideal app. Only BlueLoop had 9 out of the 10 ideal features in level 1. Table 3 describes the distribution of features amongst the analyzed apps. Sixteen of the 80 (19.7%) apps were only for Android, 15/80 (18.5%) only for iOS, and 50/80 (62.7%) for both operative systems. The mean for the number of level 1 features in the analyzed apps was 4.6 (SD 1.7) with interquartile range 3-6 and range 2-9. Based on the taxonomy of an ideal app, the predominant category of features was Personalization (70/80, 87.5%), followed by Data record (64/80, 80.0%), Tips and support and Data management (60/80, 75.0%). Reminders were featured in less than a half (33/80, 40.7%), Family support in 37.5% (30/80), and Agenda in 36.3% (29/80). Less than a third (23/80, 28.7%) featured Interaction and 25.0% (20/80) had Insulin bolus calculation. The least predominant feature was Rewards (3/80, 3.8%). The number of level 1 features was not associated with the operative system (P=.42, F=.81). In the same way, the level 1 type of feature was not associated with the operative system (P=.20, X^2_{29}=11.7).

Table 3. Feature frequency in the reviewed diabetes management apps for the 3 hierarchy levels.

Level hierarchy	Apps containing the feature, n (%)
Personalization (n=70)	70 (100.0)
User profile and picture	40 (57.1)
Blood glucose diary	64 (91.4)
Mood	14 (20.0)
Goals	33 (47.1)
Insulin type and dosage	21 (30.0)
Tailored advice	8 (11.4)
Measurements	24 (34.3)
Weight	24 (34.3)
Height	14 (20.0)
Hemoglobin A_{1c}	17 (24.3)
Blood pressure	15 (21.4)
Alcohol	1 (0.1)
Age	6 (8.6)
Gender	14 (0.2)
Tobacco	1 (0.1)
Years with type 1 diabetes mellitus	12 (17.1)
Physical activity	35 (50.0)
Family support (n=30)	30 (100.0)
Access	30 (100.0)
Read and write	8 (26.7)
Read	22 (73.3)
Literature	7 (23.3)
Books	12 (40.0)
Web pages	7 (23.3)
Agenda (n=29)	29 (100.0)
Appointments	29 (100.0)
Data record (n=64)	64 (100.0)
Manual	62 (96.9)
Automatically	30 (46.9)
Glucometer	27 (42.2)
Mobile phone	17 (26.6)
Smartwatch/wearable	1 (1.6)
Insulin bolus calculation (n=20)	20 (100.0)
Injection	17 (85.0)
Bolus	9 (45.0)
Pump	7 (35.0)
Basal	9 (45.0)
Data management (n=60)	60 (100.0)
Export	34 (56.7)
Comma-separated values	16 (26.7)
Excel	6 (10.0)

Level hierarchy	Apps containing the feature, n (%)
PDF	10 (16.7)
Other	20 (33.3)
Cloud storage	28 (46.7)
Statistics	18 (30.0)
Charts	51 (85.0)
Interaction (n=23)	23 (100.0)
Endocrinologists	16 (69.6)
Parents/mentors	14 (60.9)
Among users	9 (39.1)
Tips and support (n=60)	60 (100.0)
User manual of the app	46 (76.7)
Diet/food information	28 (46.7)
24/7 medical support	7 (11.7)
Emergency	11 (18.3)
Reminders (n=33)	33 (100.0)
Blood glucose control and insulin	33 (100.0)
Insulin/strips purchase	28 (84.8)
Appointments	29 (87.9)
Rewards (n=3)	3 (100.0)
App Store/Google Play credit	0 (0.0)
Inside the app	3 (100.0)
Other (Amazon, eBay, etc.)	1 (33.3)

Level 2 features in Personalization were dominated by the Blood glucose diary (64/70, 91.4%), whereas Tailored advice was only present in 11.4% (8/70). If we look into level 3 Measurements in Personalization, only 0.1% (1/70) of the apps included Tobacco and Alcohol, and 17.1% (12/70) included a field for Years of diagnosis. With regards to Family support, 100.0% (30/30) featured Read access, and 26.7% (8/30) included Writing access to relatives.

Fifty-one of 60 apps (85.0%) featuring Data management had the possibility of drawing charts with the stored measurements, and more than a half (34/60, 56.7%), offered the possibility of data exportation in several formats. Nearly a half (28/60, 46.7%) offered the possibility of storing data in the Cloud, and only 30.0% (18/60) had the option of calculating statistics.

Level 2 features of Tips and support showed that not all the apps have a user manual (46/60, 76.7%), less than a half (28/60, 46.7%) provided Diet/food information, only 11.67% (7/60) had 24/7 Medical support specific for diabetes, and 18.3% (11/60) had Emergency support.

All the Reminders (33/33, 100%) were for blood glucose and insulin schedule. A majority were for purchasing fungibles (28/33, 84.8%) and Appointments (29/33, 87.9%). Three of 3 (100%) provided Rewards inside the app, and only 1/3 (33.3%) gave Rewards in online markets.

Figure 5 shows the relationship between level 1 features and the FF as defined in Equation 1. The size and color of the point indicates the level of feature completeness according to the ideal app taxonomy. The yellow point refers to the app which had 9 of the 10 level 1 features, which is surpassed by 2 apps with 8 level 1 features, but which also implement more level 2 features.

Apps were classified according to the number of level 1 features they offered. Figure 6 shows the distribution of these features in a quartile classification of the number of level 1 features variable. First quartile (>75% in Figure 6) shows a regular distribution of the features, except to Agenda and Rewards. Second and third quartile apps (25%-75% in Figure 6) have a heterogeneous distribution of features, with an increased presence of Reminders, Personalization and Data management, and a lower presence of Insulin calculation, Agenda, Family support, and Rewards. First quartile (<25% in Figure 6) shows fewer features (4 versus 10 in the other quartiles), which are dominated by Family support and Tips and support features, completed by Data record and Personalization.

Figure 5. The relationship among the number of level 1 features and the Feature Factor.

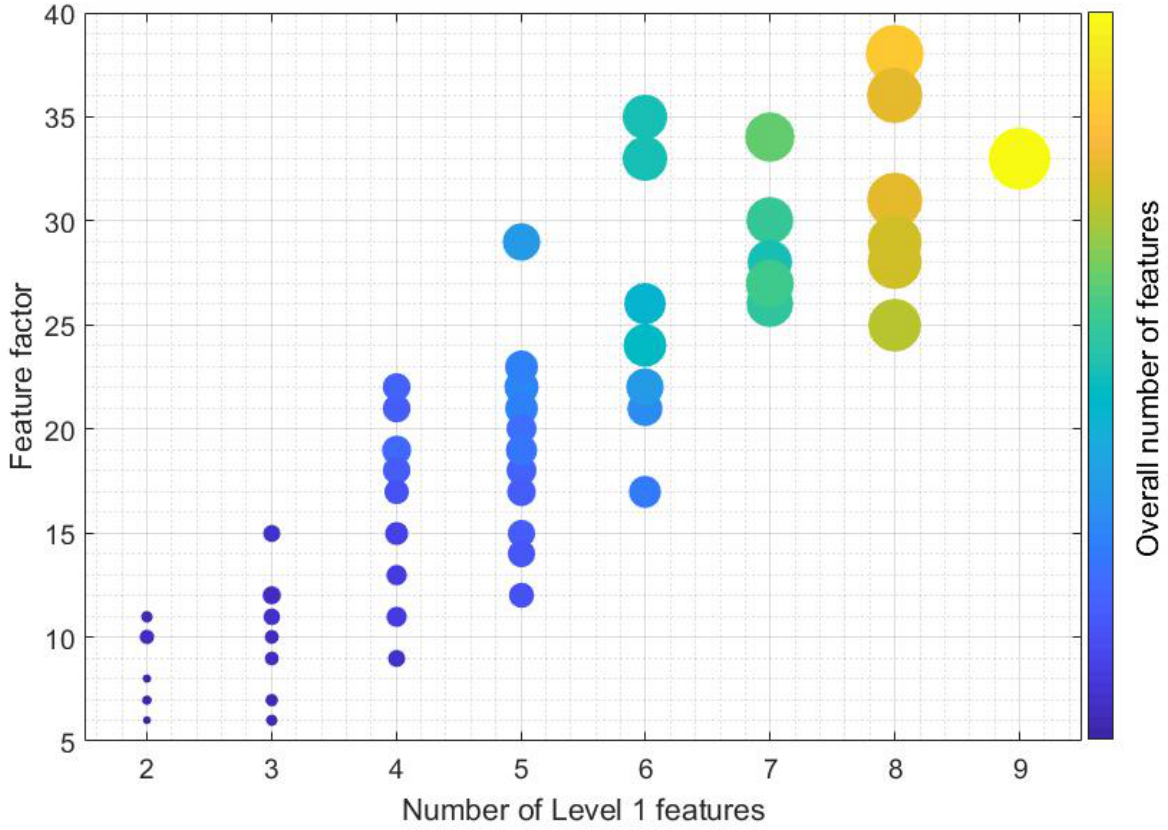
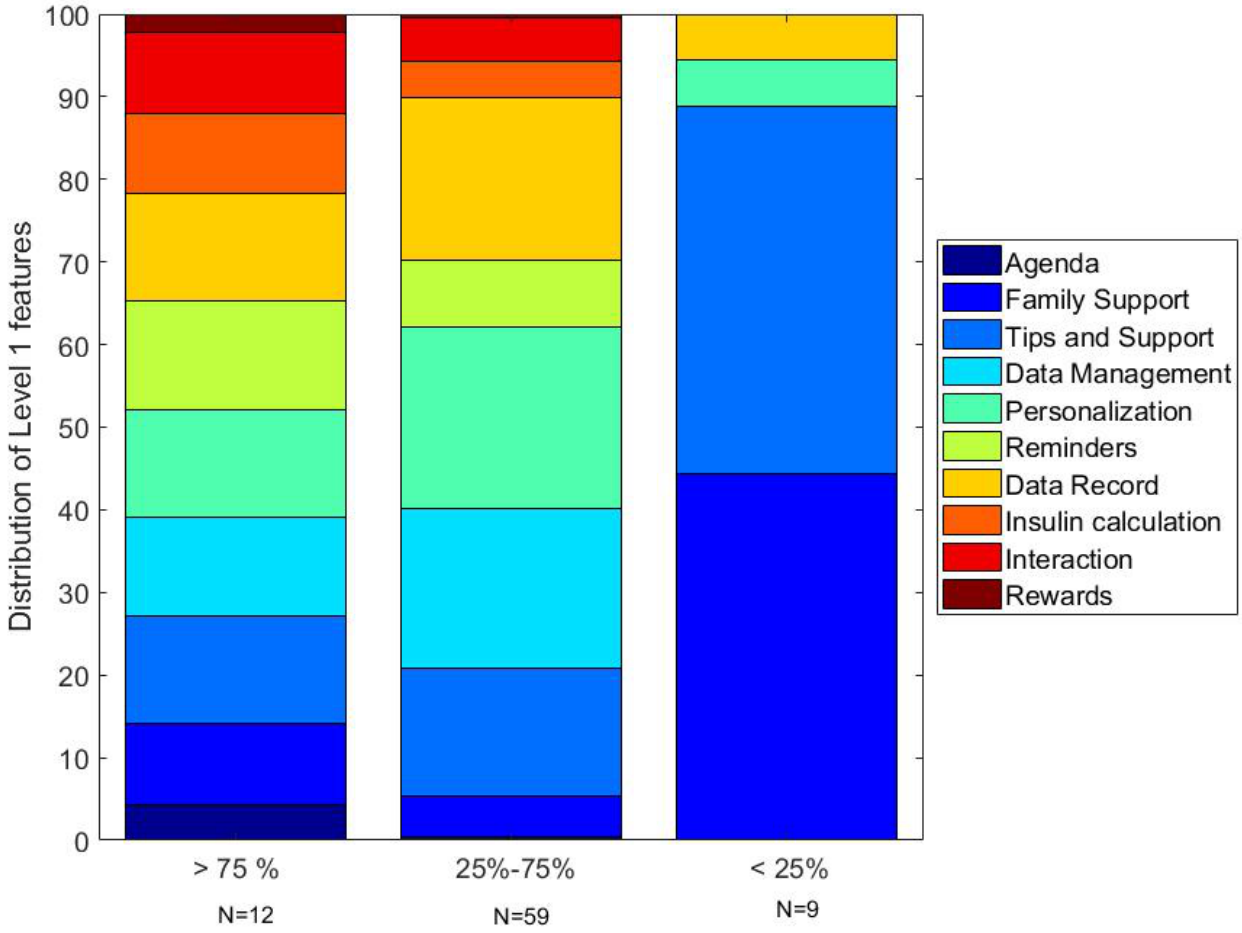


Figure 6. Distribution of level 1 features classified according to the number of level 1 features in the apps divided into quartiles.



Discussion

Principal Results

We conducted a systematic review to discover the features of the apps that had shown an effect in T1DM management. We found a big gap between research and market in the apps for supporting and empowering T1DM patients. While research is currently testing the effectiveness of mHealth in the improvement of clinical outcomes related to T1DM and therapy adherence, the characteristics of such apps are heterogeneous and not consistently justified. The systematic search discovered 6 studies consisting of mHealth interventions on 266 participants with a study duration ranging from 3 to 12 months. Studies described the app and assessed the user perceived usefulness of the app characteristics. Three of the 6 (50%) studies also reported on user satisfaction. Features of the apps were categorized and merged into a taxonomy of what would be an ideal app for T1DM management and patient empowerment (Figure 2).

The newly proposed taxonomy featured 3 hierarchical levels, the first of which has 10 areas. Subsequent level 2 and level 3 features are embedded into level 1 features, enabling us to detail what type of feature is offered to the app user (eg, Personalization: setting up a user profile). Regarding the interaction of patients and health care professionals, we discovered apps including contact to and support from endocrinologists and diabetes educators (Figure 2). We think that a general practitioner should also be considered as a reference contact point for some cases. Moreover, information about current trends might also be of interest (eg, blood glucose is increasing or decreasing) for patients having a continuous blood glucose monitoring system, or by interpolating self-reported measurements.

Moreover, this study explored systematically the features that are present in apps available at zero cost for users on the internet and mobile apps markets. Following the PRISMA approach, we found 80 apps which were analyzed in detail for 30 days. None of the assessed apps fulfilled the features of the taxonomy of an ideal app, but 1 featured 9 characteristics (BlueLoop). Personalization (70/80, 87.5%) and Data record (64/80, 80.0%) were the 2 top prevalent features, whereas Rewards (3/80, 3.8%) was the less predominant. We did not find an association on the number of features (P=.42, F=.81) nor the type of feature (P=.20, X^2_{29}=11.7) with the app operative system. Level 2 features were highly heterogeneous, but we highlight the Blood glucose diary featured in 91.4% (64/70) of apps and the low rate of Tailored advice only present in 11.4% (8/70). Level 2 features of Tips and support showed that not all the apps included a user manual (46/60, 76.7%), less than a half (28/60, 46.7%) provided Diet/food information. Only 11.67% (7/60) offered 24/7 medical support specific for diabetes, and 18.3% (11/60) offered Emergency support. To our knowledge, this is the first comprehensive taxonomy specific definition of features and the first review exploring the availability of such features in commercial apps for mHealth interventions in T1DM.

In a secondary analysis apps were classified according to the number of level 1 features and sorted into quartiles (Figure 5). First quartile apps had a regular distribution of the 10 features in the taxonomy, expect Agenda and Rewards. Second and third quartile did not follow this distribution, having a greater frequency of Data record, Personalization, and Data management. The fourth quartile featured Family support and Tips and support.

Comparison with Prior Work

Patient empowerment is essential for T1DM management and control [19]. Interventions for T1DM with mobile applications are mainly used for managing measurements, reminders and charting data, instead of promoting the self-management of the disease with a comprehensive strategy of skills development. Gamification and social communities have been observed as a key factor for empowering patients, though in our review we confirm these to features as a testimonial.

This research stresses the fact that there is a need to consider the key features to be included in an app for T1DM. This consideration is straightforward related to the ultimate objective of the app. Is it for recording and storing measurements? Is it for managing other aspects as the disposables or the appointments? Is it targeted to empower the user? Is it to build a social community? Goyal and colleagues argued that the design of an app for T1DM young patients had to consider 3 factors: (1) relationship to technology, (2) how this relationship might make a difference to users, and (3) considering when it might not be a suitable mechanism to use [14]. There is a ring to rule them all, but there should not be a T1DM app to rule them all. Patients and relatives will experience an intense evolution since disease onset until blood glucose control. This means that there can be an app for each type of T1DM patient depending on their momentum with the disease and their own context.

The taxonomy of features was designed based on previous clinical interventions [12-17] which have shown a clinical effect on the management of diabetes with mHealth. However, more evidence is needed to correlate specific improvements with the types of features. Many studies revealed the clinical effectiveness of diabetes management based on the definition of the American Association of Diabetes Educators 7 Self-Care Behaviors (AADE7) [20]. The AADE7defines 7 key areas for conducting and effective management of diabetes: (1) monitoring, (2) healthy eating, (3) taking medication, (4) being active, (5) problem solving, (6) healthy coping, and (7) reducing risks. Recently Ye et al [21] reviewed 137 apps for diabetes management based on AADE7 guidelines, concluding that many of the apps were designed for supporting healthy behaviors and a few were for supporting patients at problem-solving, risk reduction, and facing health from a holistic perspective. Our findings support that the current apps are suited for basic management instead of promoting the long-term empowerment of patients.

In 2017, Holtz et al [22] reported the development of a T1DM app with a patient-centered approach. One of the desirable app features that the adolescents mentioned in the evaluation of this study was to be part of a community. Although researchers and clinicians are focusing on intervention based on social networks and mobile apps [23], we do not see such progress when these

apps are used for improving clinical endpoints nor in the apps already available to users. Gamification and rewarding are the least prevalent of the features, a reality that should encourage researchers to design new gamification strategies to engage T1DM patients on the importance of self-management.

Gamification and coaching techniques are also a promising feature of mobile health apps for diabetes management. Sannino et al [24] introduced the concept of a constant follow-up of the patient's performance along with continuous feedback and reward system according to the user behavior and disease control. Although this app was only tested in a pilot set up for assessing usability and satisfaction, some of the features could be introduced in future RCTs to discern which gamification approaches fit better for each type of T1DM user.

The review of apps allowed us to conclude that level 1 features in apps have a balanced distribution (Figure 6). However, rather than focusing on the number of features in an app, research should investigate which of the features contribute to achieving goals effectively. A better understanding of the disease mechanism and treatment effects lead to an improvement of health outcomes [25]. This suggests that small changes in patient environments can have more significant effects on behavior and can be utilized in the self-management of diseases. The mHealth solutions are a promising alternative enabling context-awareness and personalization, but these solutions must be designed integrating evidence-based behavioral change theoretical foundations to be effective [26].

In this study, we have discovered Data record and Personalization as the most prevalent features in mHealth diabetes apps. This finding should be further explored to know how many of these apps that also offer a dashboard for professional management. A recent study has discovered a decrease of the consultation time in type 2 diabetes management by using artificial intelligence and predictive modeling [27,28], and moreover, a review has proved these methods are being progressively established as suitable for use in daily clinical practice [29]. Future research should tap into the application of these methods for supporting both T1DM patients and healthcare professionals on the follow-up and control of this complex disease.

Finally, for the main purpose of disease management apps, rather than investigating the number of app features, we should investigate what kind of features could achieve the goal effectively. In our study, we were not able to distinguish which feature (or combination of features) was helping patients to achieve their goals. Acceptability and usability studies may help to identify the features that have a higher impact on the self-management of the disease, but further research should be conducted to critically identify the sets of features valuable for patients.

Limitations

A limitation of this study is the set of papers and apps selected for the literature review and the app review. The authors may have omitted significant contributions for both searches due to the incompleteness of the query commands and mismatch in the searches. We have focused our research on apps tested in RCTs with significant clinical end-points. Authors are aware of the relevant research done in the past and conducted in the present in the design of T1DM apps, which do not involve clinical endpoints nor RCTs, which may also contribute to the taxonomy of an ideal app defined in this paper. Another limitation is that the graphical user interface (GUI) of the apps are not evaluated or studied. If the GUI is not appropriate, many features might be less accessible and thus less used.

Conclusions

This study assessed the existing gap between research and market in mobile health apps for the management of T1DM and the empowerment of patients. The mHealth has potential to support the management of T1DM, to catalyze the information exchange between patients, parents, and caregivers, and to empower and educate patients in the management of T1DM. The current landscape of apps for T1DM does not seem to be close to what researchers promote from RCTs and user-centered design. A majority of the apps mainly support the collection of measurements, and only a few of them offer a wide range of features for a personalize self-management. Rewards and social communities are not yet well adopted in market environments.

Acknowledgments

The authors wish to acknowledge the ITACA Institute (Universitat Politècnica de València) for making possible the publication of this paper through the Excellence Support program for the publication in high-impact international journals.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search Strategy and results in PubMed (MEDLINE).

[PDF File (Adobe PDF File), 35KB - mhealth_v6i11e12237_app1.pdf]

References

1. American Diabetes Association. Classification and Diagnosis of Diabetes. Diabetes Care 2017 Dec;40(Suppl 1):S11-S24. [doi: 10.2337/dc17-S005] [Medline: 27979889]
2. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014 Feb;103(2):137-149. [doi: 10.1016/j.diabres.2013.11.002] [Medline: 24630390]
3. Diabetes ControlComplications Trial/Epidemiology of Diabetes InterventionsComplications (DCCT/EDIC) Research Group, Nathan DM, Zinman B, Cleary PA, Backlund JC, Genuth S, et al. Modern-day clinical course of type 1 diabetes mellitus after 30 years' duration: the diabetes control and complications trial/epidemiology of diabetes interventions and complications and Pittsburgh epidemiology of diabetes complications experience (1983-2005). Arch Intern Med 2009 Jul 27;169(14):1307-1316 [FREE Full text] [doi: 10.1001/archinternmed.2009.193] [Medline: 19636033]
4. Martinez-Millana A, Fico G, Fernández-Llatas C, Traver V. Performance assessment of a closed-loop system for diabetes management. Med Biol Eng Comput 2015 Dec;53(12):1295-1303. [doi: 10.1007/s11517-015-1245-3] [Medline: 25667016]
5. Lim S, Kang SM, Shin H, Lee HJ, Won Yoon J, Yu SH, et al. Improved glycemic control without hypoglycemia in elderly diabetic patients using the ubiquitous healthcare service, a new medical information system. Diabetes Care 2011 Feb 19;34(2):308-313 [FREE Full text] [doi: 10.2337/dc10-1447] [Medline: 21270188]
6. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
7. Ashurst EJ, Jones RB, Abraham C, Jenner M, Boddy K, Besser RE, et al. The diabetes app challenge: user-led development and piloting of internet applications enabling young people with diabetes to set the focus for their diabetes consultations. Med 2 0 2014 Nov 07;3(2):e5 [FREE Full text] [doi: 10.2196/med20.3032] [Medline: 25654312]
8. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011 Sep 22;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
9. Chavez S, Fedele D, Guo Y, Bernier A, Smith M, Warnick J, et al. Mobile Apps for the Management of Diabetes. Diabetes Care 2017 Dec;40(10):e145-e146 [FREE Full text] [doi: 10.2337/dc17-0853] [Medline: 28774944]
10. Castensøe-Seidenfaden P, Reventlov Husted G, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a Self-Management App for Young People With Type 1 Diabetes: Methodological Challenges, Experiences, and Recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
11. Higgins J. Cochrane Handbook for Systematic Reviews of Interventions Version 5.: The Cochrane Collaboration; 2011. URL: https://training.cochrane.org/handbook [accessed 2018-10-29] [WebCite Cache ID 73XCJzC9A]
12. Castensøe-Seidenfaden P, Husted GR, Jensen AK, Hommel E, Olsen B, Pedersen-Bjergaard U, et al. Testing a Smartphone App (Young with Diabetes) to Improve Self-Management of Diabetes Over 12 Months: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Jun 26;6(6):e141 [FREE Full text] [doi: 10.2196/mhealth.9487] [Medline: 29945861]
13. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
14. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A Mobile App for the Self-Management of Type 1 Diabetes Among Adolescents: A Randomized Controlled Trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82 [FREE Full text] [doi: 10.2196/mhealth.7336] [Medline: 28630037]
15. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
16. Clements MA, Staggs VS. A Mobile App for Synchronizing Glucometer Data: Impact on Adherence and Glycemic Control Among Youths With Type 1 Diabetes in Routine Care. J Diabetes Sci Technol 2017 Dec;11(3):461-467 [FREE Full text] [doi: 10.1177/1932296817691302] [Medline: 28745097]
17. Ryan EA, Holland J, Stroulia E, Bazelli B, Babwik SA, Li H, et al. Improved A1C Levels in Type 1 Diabetes with Smartphone App Use. Can J Diabetes 2017 Feb;41(1):33-40. [doi: 10.1016/j.jcjd.2016.06.001] [Medline: 27570203]
18. Sun C, Malcolm JC, Wong B, Shorr R, Doyle M. Improving Glycemic Control in Adults and Children With Type 1 Diabetes With the Use of Smartphone-Based Mobile Applications: A Systematic Review. Can J Diabetes 2018 Mar 27. [doi: 10.1016/j.jcjd.2018.03.010] [Medline: 30026048]
19. Chen L, Chuang L, Chang C, Wang C, Wang I, Chung Y, et al. Evaluating self-management behaviors of diabetic patients in a telehealthcare program: longitudinal study over 18 months. J Med Internet Res 2013 Dec 09;15(12):e266 [FREE Full text] [doi: 10.2196/jmir.2699] [Medline: 24323283]
20. Tomky D, Cypress M, Dang D, Maryniuk M, Peyrot M, Mensing C. AADE7 Self-Care Behaviors. In: Diabetes Educ. Diabetes Educ SAGE Publications Sage CA: Los Angeles, CA; 2008:445-449.
21. Ye Q, Khan U, Boren SA, Simoes EJ, Kim MS. An Analysis of Diabetes Mobile Applications Features Compared to AADE7: Addressing Self-Management Behaviors in People With Diabetes. J Diabetes Sci Technol 2018 Jul;12(4):808-816. [doi: 10.1177/1932296818754907] [Medline: 29390917]

22. Holtz BE, Murray KM, Hershey DD, Dunneback JK, Cotten SR, Holmstrom AJ, et al. Developing a Patient-Centered mHealth App: A Tool for Adolescents With Type 1 Diabetes and Their Parents. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e53 [FREE Full text] [doi: 10.2196/mhealth.6654] [Medline: 28428167]
23. Gabarron E, Årsand E, Wynn R. Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/10303] [Medline: 30097421]
24. Sannino G, Forastiere M, De Pietro G. A Wellness Mobile Application for Smart Health: Pilot Study Design and Results. Sensors (Basel) 2017 Mar 17;17(3) [FREE Full text] [doi: 10.3390/s17030611] [Medline: 28304332]
25. Adams RJ. Improving health outcomes with better patient understanding and education. Risk Manag Healthc Policy 2010;3:61-72 [FREE Full text] [doi: 10.2147/RMHP.S7500] [Medline: 22312219]
26. Giunti G. 3MD for Chronic Conditions, a Model for Motivational mHealth Design: Embedded Case Study. JMIR Serious Games 2018 Aug 24;6(3):e11631 [FREE Full text] [doi: 10.2196/11631] [Medline: 30143476]
27. Dagliati A, Sacchi L, Tibollo V, Cogni G, Teliti M, Martinez-Millana A, et al. A dashboard-based system for supporting diabetes care. J Am Med Inform Assoc 2018 May 01;25(5):538-547. [doi: 10.1093/jamia/ocx159] [Medline: 29409033]
28. Martinez-Millana A, Bayo-Monton J, Argente-Pla M, Fernandez-Llatas C, Merino-Torres JF, Traver-Salcedo V. Integration of Distributed Services and Hybrid Models Based on Process Choreography to Predict and Detect Type 2 Diabetes. Sensors (Basel) 2017 Dec 29;18(1) [FREE Full text] [doi: 10.3390/s18010079] [Medline: 29286314]
29. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]

Abbreviations

AADE7: American Association of Diabetes Educators 7 Self-Care Behaviors
CSV: comma-separated values
FF: Feature Factor
GUI: graphical user interface
HbA $_{1c}$: hemoglobin A_{1c}
mHealth: mobile health
RCT: randomized controlled trial
SMBG: self-monitoring blood glucose
T1DM: type 1 diabetes mellitus

Edited by G Eysenbach; submitted 17.09.18; peer-reviewed by A Fioravanti, TW Ho; comments to author 07.10.18; revised version received 12.10.18; accepted 01.11.18; published 21.11.18.

Please cite as:
Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V
App Features for Type 1 Diabetes Support and Patient Empowerment: Systematic Literature Review and Benchmark Comparison
JMIR Mhealth Uhealth 2018;6(11):e12237
URL: http://mhealth.jmir.org/2018/11/e12237/
doi:10.2196/12237
PMID:30463839

Original Paper

Implementation Tells Us More Beyond Pooled Estimates: Secondary Analysis of a Multicountry mHealth Trial to Reduce Blood Pressure

Rodrigo M Carrillo-Larco[1,2], MD; Safia S Jiwani[1], MSPH; Francisco Diez-Canseco[1], MPH; Rebecca Kanter[3,4], PhD; Andrea Beratarrechea[5], MD; Vilma Irazola[5], MD; Manuel Ramirez-Zea[3], PhD; Adolfo Rubinstein[5], PhD; Homero Martinez[6,7], PhD; J Jaime Miranda[1,8], PhD; GISMAL Group[1]

[1]CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru

[2]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom

[3]INCAP Research Center for the Prevention of Chronic Diseases, Institute of Nutrition of Central America and Panama, Guatemala, Guatemala

[4]Department of Nutrition, Faculty of Medicine, University of Chile, Santiago de Chile, Chile

[5]South American Center of Excellence for Cardiovascular Health, Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina

[6]Nutrition International, Ottawa, ON, Canada

[7]Hospital Infantil de Mexico Federico Gomez, Mexico DF, Mexico

[8]Department of Medicine, School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

Corresponding Author:
Manuel Ramirez-Zea, PhD
INCAP Research Center for the Prevention of Chronic Diseases
Institute of Nutrition of Central America and Panama
Calzada Roosevelt 6-25 Zona 11
Guatemala, Apartado Postal 1188
Guatemala
Phone: 502 2472 3762
Fax: 502 2472 3762
Email: mramirez@incap.int

Abstract

Background: The uptake of an intervention aimed at improving health-related lifestyles may be influenced by the participant's stage of readiness to change behaviors.

Objective: We conducted secondary analysis of the Grupo de Investigación en Salud Móvil en América Latina (GISMAL) trial according to levels of uptake of intervention (dose-response) to explore outcomes by country, in order to verify the consistency of the trial's pooled results, and by each participant's stage of readiness to change a given lifestyle at baseline. The rationale for this secondary analysis is motivated by the original design of the GISMAL study that was independently powered for the primary outcome—blood pressure—for each country.

Methods: We conducted a secondary analysis of a mobile health (mHealth) multicountry trial conducted in Argentina, Guatemala, and Peru. The intervention consisted of monthly motivational phone calls by a trained nutritionist and weekly tailored text messages (short message service), over a 12-month period, aimed to enact change on 4 health-related behaviors: salt added to foods when cooking, consumption of high-fat and high-sugar foods, consumption of fruits or vegetables, and practice of physical activity. Results were stratified by country and by participants' stage of readiness to change (precontemplation or contemplation; preparation or action; or maintenance) at baseline. Exposure (intervention uptake) was the level of intervention (<50%, 50%-74%, and ≥75%) received by the participant in terms of phone calls. Linear regressions were performed to model the outcomes of interest, presented as standardized mean values of the following: blood pressure, body weight, body mass index, waist circumference, physical activity, and the 4 health-related behaviors.

Results: For each outcome of interest, considering the intervention uptake, the magnitude and direction of the intervention effect differed by country and by participants' stage of readiness to change at baseline. Among those in the high intervention uptake category, reductions in systolic blood pressure were only achieved in Peru, whereas fruit and vegetable consumption also showed reductions among those who were at the maintenance stage at baseline in Argentina and Guatemala.

Conclusions: Designing interventions oriented toward improving health-related lifestyle behaviors may benefit from recognizing baseline readiness to change and issues in implementation uptake.

Trial Registration: ClinicalTrials.gov NCT01295216; http://clinicaltrials.gov/ct2/show/NCT01295216 (Archived by WebCite at http://www.webcitation.org/72tMF0B7B).

(JMIR Mhealth Uhealth 2018;6(11):e10226) doi:10.2196/10226

KEYWORDS

Argentina; behavior; clinical trial; Guatemala; health risk behaviors, lifestyle risk reduction; mHealth; Peru

Introduction

Mobile health (mHealth) strategies have been increasingly used in public health research studies, some of them showing effective results in improving the profiles of traditional risk factors for noncommunicable diseases [1,2], including in developing countries [3-6]. Most mHealth projects involve multifaceted complex interventions, where the interplay of their components is the key to determining an effect, thus requiring many more angles for their evaluation rather than a single indicator of effectiveness at the end of the trial [7-10].

Multisite trials are efficient in expanding enrolment targets [11,12]. Multicountry studies, in addition, can provide heterogeneity in the diversity of settings, where an intervention is being tested. It is well known that several individual and contextual factors may be directly related with the uptake and implementation of the desired intervention [13,14]. Yet, the effect of any given intervention conducted in a multisite or multicountry study may differ by site or country, and thus a single pooled effect estimate can cloud the true range of the intervention impact.

In addition to context, another major factor affecting the success of an intervention relates to the profile of participants. In biomedical and mechanistic research, the same chain of events applies to all participants. In behavioral sciences, however, it is known that targeting certain behaviors may not be the same for each person, often requiring the acknowledgment and identification of the stage of readiness to change certain habits [15,16]. The applicability of such readiness to incorporate changes toward healthier habits has been signaled for smoking cessation [17,18]. An intervention may not have the same effect on people merely thinking of quitting an unhealthy habit compared with that on those already committed to quitting.

This Grupo de Investigación en Salud Móvil en América Latina (GISMAL) trial aimed to reduce blood pressure and prevent the shift to hypertension in adult subjects with prehypertension in resource-constrained urban settings in Argentina, Guatemala, and Peru [19]. This trial used customized weekly short message service (SMS) text messages and motivational monthly phone calls aimed at the adoption of healthy lifestyles. Pooled results of the intervention showed reductions in weight but not in blood pressure after 12 months [19]. However, further scrutiny regarding the implementation of the intervention is needed to better understand what works regarding mHealth and its capability to support behavior change in real-world conditions [20]. Consequently, we aimed to conduct a secondary analysis of the GISMAL trial according to levels of uptake of the mHealth intervention to explore outcomes by country, in order to verify the consistency of the trial's pooled results, and by each participant's stage of readiness to change a given lifestyle at baseline. The rationale for this secondary analysis is motivated by the original design of the GISMAL study that was independently powered for the primary outcome, blood pressure, for each country.

Methods

Study Design

This is a secondary analysis using data collected from an mHealth randomized controlled trial (RCT) known as GISMAL (NCT01295216). The results of the trial have been reported elsewhere [19]. GISMAL was conducted in Argentina, Guatemala, and Peru by targeting adult individuals with prehypertension with the primary aim of reducing blood pressure. The intervention arm received tailored phone calls and associated SMS text messages to foster the adoption of a healthy diet and improve physical activity profiles. Phone calls using the motivational interview method were delivered monthly by nutritionists in conjunction with weekly tailored SMS text messages during a 12-month period. These communications were tailored according to the participants' stage of readiness to change regarding the 4 hypertension-related risk factors the RCT aimed to improve, that is, physical activity, fruit and vegetable consumption, decreased consumption of high-fat and high-sugar foods, and reduced salt intake. The control arm received usual care without any mHealth components.

Study Population

Subjects enrolled in the GISMAL trial were 30- to 60-year-old males and females with prehypertension (systolic blood pressure between 120 and 139 mm Hg, diastolic blood pressure between 80 and 89 mm Hg, or both). Further inclusion criteria were (1) those not receiving medication for hypertension; (2) those owning a mobile telephone; and (3) those able to read and understand SMS text messages in Spanish. Pregnant women and people who reported having ever been diagnosed with hypertension, diabetes, or cardiovascular disease were excluded [19].

Variable Assessment

In total, 3 blood pressure measurements were taken in a sitting position after a 5-minute resting period using calibrated digital devices (Omron HEM-742INT, Omron Healthcare, Lake Forest, IL, United States); the mean of the last 2 readings was herein used. A digital scale (Omron SC-100/SECA 803) was used to measure bodyweight. Height was measured with the participant

shoeless using a stadiometer, and waist circumference was assessed with a flexible nonelastic measuring tape. All other variables were assessed using standardized and validated questionnaires including the food frequency questionnaire [21] and the International Physical Activity Questionnaire to assess diet and physical activity, respectively [22].

Exposure Variable

For this study, the exposure variable was intervention uptake, defined as the intervention receiving <50%, 50%-74%, or ≥75% out of the 12 planned phone calls. The reference category was the control group of the original RCT, that is, those who did not receive the mHealth intervention. The mHealth intervention also included weekly SMS text messages related to the theme of the phone call, but it is difficult to ascertain SMS text message reception. Therefore, it was decided not to consider the SMS text messages as part of the mHealth intervention in the exposure categories. This was not considered a weakness in the ascertainment of exposure as SMS text messages were only enacted following the completion of a phone call.

Outcome Variables

The original trial measured the following biological and behavioral risk factors before and after the intervention: (1) systolic blood pressure (mm Hg); (2) diastolic blood pressure (mm Hg); (3) weight (kg); (4) body mass index (kg/m^2); (5) waist circumference (cm); (6) physical activity (metabolic equivalents of task/minute per week); (7) fruit and vegetable consumption (number of daily servings); (8) high-sodium food intake (number of daily servings); and (9) consumption of high-fat and high-sugar foods (number of daily servings). For analytical purposes, these variables were treated as continuous and mean standardized (ie, the mean was subtracted from the observed values) and then divided by SD.

Regarding the analysis of these 9 outcomes, we also performed stratified analyses by country. However, when our analysis was stratified by baseline readiness to change, we chose only 4 outcomes for illustrative purposes: 2 outcomes that were expected to increase following the mHealth intervention—fruit and vegetable consumption and physical activity—and 2 outcomes that were expected to decrease following the mHealth intervention—salt added when cooking and high-fat and high-sugar food consumption.

Variables Used to Perform Stratified Analyses

Analyses were stratified by country (Argentina, Guatemala, and Peru) and by participants' stage of readiness to change regarding the improvement of a particular health-related lifestyle at baseline. The 3 stages of readiness to change were constructed based on the transtheoretical model of health behavior change [23]: (1) precontemplation or contemplation; (2) preparation or action; and (3) maintenance. For the second aim of our study, based on stage of readiness to change, we used the 4 previously described health-related lifestyles, with the first 2 expected to increase and the latter 2 expected to decrease postintervention: (1) consumption of fruits and vegetables; (2) physical activity profile; (3) salt added while cooking; and (4) intake of high-fat and high-sugar foods. Of note, the transtheoretical model of health behavior change appears to be effective when implemented in culturally diverse populations [24,25].

Ethics Approval and Consent to Participate

The protocol for the original RCT was independently reviewed and approved by Institutional Review Boards in the 3 participating countries: Hospital Italiano de Buenos Aires (Argentina), Institute of Nutrition of Central America and Panama (Guatemala), and Universidad Peruana Cayetano Heredia (Peru). The original RCT protocol was also reviewed and approved by the RAND Corporation, Santa Monica, CA, United States. Written informed consent was provided by all participants. The trial is registered at ClinicalTrials.gov (NCT01295216).

Statistical Analysis

Numeric variables were described using means and SDs, while categorical variables were summarized using frequencies (%). Comparisons between numeric variables were assessed using analysis of variance test. Linear regression models were used, with and without stratification by country and stage of change; mHealth intervention uptake was the independent variable, and biological and behavioral risk factors (mean standardized) were the dependent variables. The regression models without any stratification were adjusted by country. The regression models stratified by country or stage of readiness did not include variables other than the exposure and outcome. Results from the regression models are presented as coefficients and 95% CIs. Analyses were conducted using STATA 13.0 (StataCorp, College Station, TX, United States).

Results

Principal Results

At the baseline, there were 637 participants (321 in the control group and 316 in the intervention group), 53.7% (342/637) were women, and the mean age was 43.4 (SD 8.4) years. At the end of the study, 89.4% (287/321) subjects remained in the control group and 84.2% (266/316) in the intervention group. Among those who were initially in the intervention group, 40.5% (128/316) received <50%, 36.1% (114/316) received between 50% and 74%, and 23.4% (74/316) received ≥75% of the planned intervention phone calls. There were no differences in intervention dose by sex (P=.05) or country (P=.34). However, there were more young subjects (≤45 years) among those who received either <50% or between 50% and 74% of the intervention compared with those who received ≥75% of the intervention phone calls (P=.001).

Multimedia Appendix 1 shows the means and SDs for each outcome of interest so that the estimates of the figures can be computed in their original units (eg, blood pressure in mm Hg). The overall unstandardized estimates are also presented in footnotes to all figures.

Results by Country

Figure 1 shows the intervention effect for each of the biological and behavioral risk factors assessed by country among subjects who received a higher dose (≥75%) of the planned intervention phone calls. Regarding all 9 metabolic risk factors, the

magnitude of the intervention effect varied across countries. The intervention had an effect in the opposite direction than expected whereby systolic blood pressure increased in Argentina and Guatemala. Those in the intervention group in Peru showed a 4-fold greater fruit and vegetable consumption than those in Guatemala and almost 2-fold the consumption of those in Argentina. Moreover, the reduction of high-fat and high-sugar foods was almost 2-fold higher in Guatemala and Argentina than in Peru. Moreover, Peru appeared to be the only country where a reduction in systolic blood pressure was achieved. Overall, varying directions and magnitudes of effect were also observed among those who received <50% and 50%-74% of the intervention phone calls (Multimedia Appendices 2 and 3).

To compute each estimate in their respective units (eg, blood pressure in mm Hg), multiply SD with the reported value in the figures and then divide by the mean (Multimedia Appendix 1). The overall estimates were as follows: systolic blood pressure=123.45 mm Hg; diastolic blood pressure=74.87 mm Hg; weight=74.36 Kg; body mass index=28.85 Kg/m^2; waist circumference=95.50 cm; salt consumption=0.40 servings/day; high-fat and high-sugar foods consumption=3.24 servings/day; fruits and vegetables consumption=2.70 servings/day; and physical activity=583.62 metabolic equivalents of task /min per week.

Outcomes by Participants' Stage of Readiness at Baseline

After assessing the health-related lifestyle outcome of salt added when cooking (Figure 2) among those who received ≥75% of the intervention phone calls, the magnitude of the intervention effect varied according to stage of readiness to change at baseline. Specifically, the magnitude of the intervention effect was greater in those in the precontemplation or contemplation stage than in those in the maintenance stage in Guatemala and Peru.

Regarding fruit and vegetable consumption as a health-related lifestyle outcome (Figure 3), among those who received ≥75% of the intervention phone calls, the magnitude and direction of the intervention effect differed by country. In Peru, the direction of the intervention was the same according to stages of change, with those in the maintenance stage showing the largest effect. Overall, as well as in Guatemala and Argentina, those in the maintenance stage of readiness to change at baseline had an intervention effect that was in the opposite direction than expected, whereby fruit and vegetable consumption decreased.

The results of the intervention exposure on the other 2 health-related lifestyle factors are shown in Multimedia Appendices 4 and 5, respectively. Regarding the consumption of high-fat and high-sugar foods (Multimedia Appendix 4), Argentinians in the precontemplation or contemplation stage of readiness to change at baseline had an increased consumption that was in the opposite direction than expected. Finally, regarding the health-related lifestyle outcome of physical activity (Multimedia Appendix 5), those in the maintenance stage of readiness to change at baseline in Guatemala and Peru had decreased their physical activity profile, which was also in the opposite direction than expected.

Figure 1. Assessed outcomes of subjects who received ≥75% of the intervention overall and by country. BMI: body mass index.

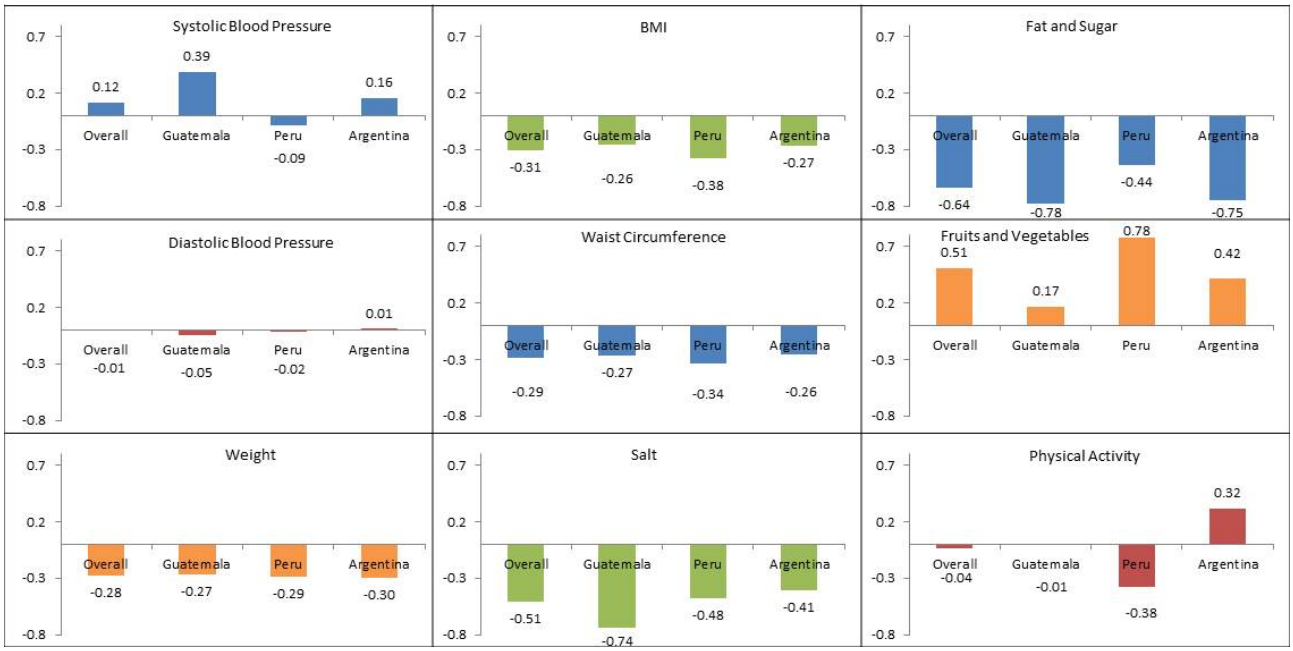


Figure 2. Intervention effect on salt added when cooking according to participant baseline stage of readiness status, overall and by country. Pre-Con/Con: precontemplation or contemplation; Prep/Act: preparation or action; Maint: maintenance.

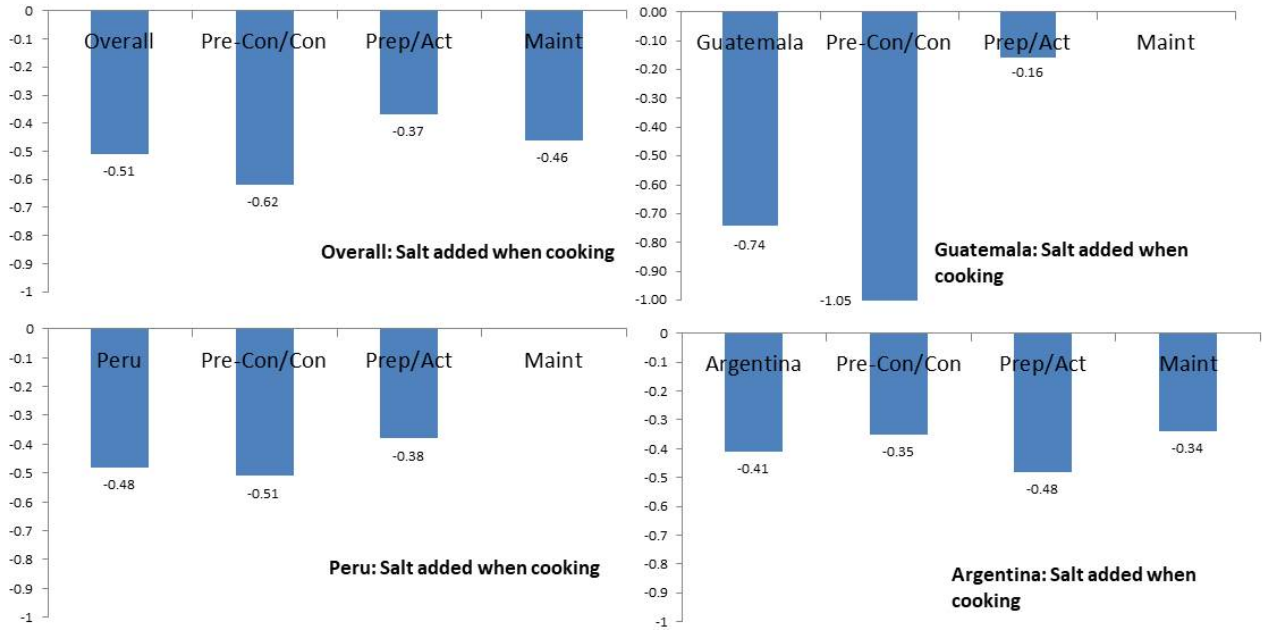
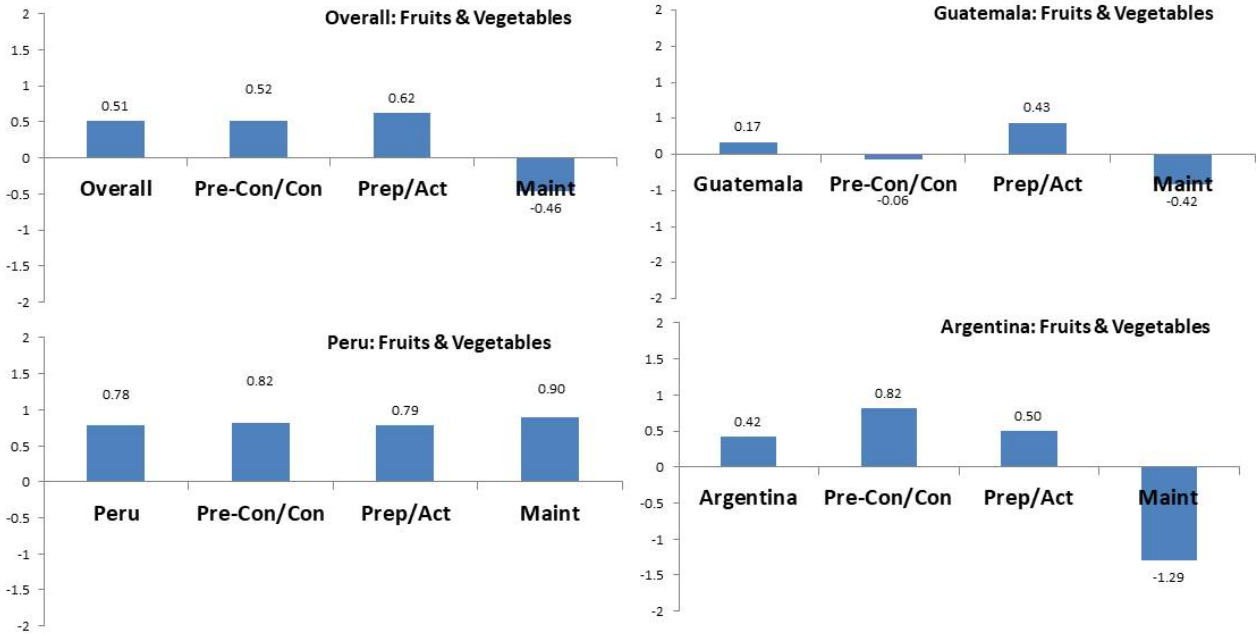


Figure 3. Intervention effect on fruit and vegetable consumption according to the participant baseline readiness to change status, overall and by country. Pre-Con/Con: precontemplation or contemplation; Prep/Act: preparation or action; Maint: maintenance.



Discussion

Principal Findings

Our secondary analysis of the GISMAL trial shows variations in the effect by country and stage of readiness to change at baseline. Acknowledging that the original trial was independently powered for the primary outcome—blood pressure—for each country, we found that the trial only had a positive effect on systolic blood pressure among those in Peru. Considering intervention uptake, the magnitude and direction of the intervention effect differed by country and stage of readiness to change at baseline. Among those in the higher category of intervention uptake, reductions in systolic blood pressure were also only achieved in Peru. However, fruit or vegetable intake declined among those who were at the maintenance stage at baseline in Argentina and Guatemala, respectively. These findings call for additional considerations when conducting complex multicountry or multisite behavioral interventions. For example, when planning future interventions, readiness to change could be a parameter to account for in sample size calculations among other considerations during the design stage of a study.

We reported a greater magnitude of the intervention effect among those in the precontemplation or contemplation stage of readiness to change at baseline or, in other words, those who appeared to have the least predisposition to uptake healthy

lifestyles in the study sample. Those in the precontemplation or contemplation stage may have been unaware of the health risks of certain lifestyle or how to make their lifestyles healthier. However, we can speculate that the mHealth intervention provided them with the necessary amount of interaction, information, and tools to improve their lifestyles. In doing so, those in the precontemplation or contemplation stage may have been keener to engage with and uptake healthier habits compared to those in the other stages of readiness to change. Moreover, people in the precontemplation or contemplation stage had more room for improvement in terms of lifestyle behaviors; thus, even small changes in adopting healthier lifestyles were likely to have had a greater health impact than similar changes observed among those in other stages of readiness to change.

Our results from this secondary analysis align with recommendations for the need to better understand the role of technology in enacting and sustaining behavioral change [26]. The results observed in our study suggest tailoring future mHealth study interventions to specific stages of readiness to change. The duration and multipronged design of the GISMAL mHealth trial, conducted over a 12-month period targeting multiple behaviors, provides more insights into short-term action and long-term behavioral change. We did not explore predictors of engagement or habituation, which opens additional venues to understand short- and long-term effects of future mHealth behavior-oriented studies. Additional caution should be placed on the mode of mHealth delivery as it should not be assumed that all different platforms to deliver technology-based interventions would have the same adoption, engagement over time, and expected effects in the same order of magnitude [27].

From an implementation perspective, the varying results observed may have been due to differences in a number of implementation research-related indicators [20]. Among the strengths of our study lies the multicountry deployment of the same intervention allowing this mHealth intervention to operate in different the “real-world” settings. Additional strengths rely on the objective ascertainment of intervention uptake through completed phone calls made by the nutritionists, thus providing a pragmatic approach to evaluate the implementation of mHealth strategies without the need to rely on SMS text messaging (short message service, SMS) alone as a means to deliver mHealth. Fidelity was enacted before and during the conduction of the trial through a standardized approach using the same training and supervision procedure for nurse calls, the same algorithm to enact SMS text message delivery, and training of fieldworkers [19]. Appropriateness was anticipated through a qualitative study and a theory-driven development of the SMS text messages involving health communication and psychology experts before the intervention [28]. We do not have information regarding the acceptability of the intervention, which may have provided a richer picture of implementation issues. However, acceptability is in part implied by the intervention uptake over its 12-month duration, which permits a partial picture of acceptability from the end user’s point of view.

Limitations

Some limitations include the fact that some outcomes were self-reported, such as physical activity or food consumption; specifically, salt, fruit, and vegetable consumption could have been particularly affected by self-reported information. People with higher health awareness were more likely to change their dietary behaviors but also underestimate their salt intake or overestimate their fruit and vegetable consumption. Our results could support the first hypothesis because salt intake was consistently reduced. Although fruit and vegetable consumption improved for those in different stages of readiness to change, the magnitude was relatively small and, in some cases, there was a reduction. This suggests that participants tend to underestimate unhealthy lifestyles (eg, salt consumption) rather than to overestimate the frequency of healthy lifestyles (eg, fruit and vegetable consumption). Yet, our results warrant cautious interpretation because of the lack of statistical power for subgroup analyses for all the outcomes. As the overall goal of this study was to identify signals in the effect of the intervention across countries and stage of readiness to change at baseline, we did not aim to assess such potential variations in clinically relevant units, that is, transforming standardized mean values into mm Hg for blood pressure. Lastly, a caveat of this study is the theory used to inform the methodology of the original trial [23], which has been challenged regarding its interventions to reduce smoking. Although the available evidence is inconclusive [29], some reports have suggested that the transtheoretical model could have positive results in improving weight loss and healthy diets [30].

Conclusions

In summary, the results of this multicountry mHealth intervention trial, originally aimed at reducing the progression of prehypertension to hypertension by improving health-related lifestyles, show that outcomes vary by country and according to the participants’ stage of readiness to change a specific behavior at baseline. This information will be of utmost utility when designing future studies and provides important pragmatic lessons regarding implementation issues of mHealth interventions, emphasizing indicators amenable to be monitored.

Acknowledgments

National Heart, Lung, and Blood Institute (National Institutes of Health, United States) and the Medtronic Foundation. Subcontract number HHSN268200900028C. There are no other financial disclosures. The sponsor played no role in this work.

Authors' Contributions

JJM conceived the research idea. RMC-L, SSJ, and JJM designed the analysis plan and outline of the study. RMC-L conducted the analysis and wrote the first draft of the paper. All authors gave major comments and edited the paper. All authors approved the final version. Additional members of the GISMAL group are Adrían Alasino (FunPRECAL, Mar del Plata, Prov de Buenos

Aires, Argentina), Berneth Nuris Budiel Moscoso (Universidad Peruana Cayetano Heredia, Lima, Peru), Carolina Carrara (Hospital Italiano de Buenos Aires, Buenos Aires, Argentina), Jackelyn Espinoza Surichaqui (Universidad Peruana Cayetano Heredia, Lima, Peru), Gimena Giardini (Hospital Italiano de Buenos Aires, Buenos Aires, Argentina), Jesica Guevara (Institute of Nutrition of Central America and Panama, Guatemala City, Guatemala), Analí Morales Juárez (Institute of Nutrition of Central America and Panama, Guatemala City, Guatemala), Lorena Lázaro Cuesta (FunPRECAL, Mar del Plata, Prov de Buenos Aires, Argentina), Dalia Lewitan (Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina), Lita Palomares Estrada (Universidad Peruana Cayetano Heredia, Lima, Peru), Carla Martínez Ramírez (Universidad Peruana Cayetano Heredia, Lima, Peru), Gloria Robles de la Cruz (Universidad Peruana Cayetano Heredia, Lima, Peru), Julissa Salguero (Institute of Nutrition of Central America and Panama, Guatemala City, Guatemala), Juan Carlos Saravia Drago (Universidad Peruana Cayetano Heredia, Lima, Peru), María Urtasún (Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina), and José Alfredo Zavala Loayza (Universidad Peruana Cayetano Heredia, Lima, Peru).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Means and SDs for each outcome variable at 12 months of intervention, overall and by country.

[PDF File (Adobe PDF File), 416KB - mhealth_v6i11e10226_app1.pdf]

Multimedia Appendix 2

Assessed outcomes for subjects who received <50% of the planned intervention phone calls, overall and by country. Results are. BMI: body mass index.

[JPG File, 64KB - mhealth_v6i11e10226_app2.jpg]

Multimedia Appendix 3

Assessed outcomes of subjects who received 50%-74% of the planned intervention phone calls, overall and by country. BMI: body mass index.

[JPG File, 64KB - mhealth_v6i11e10226_app3.jpg]

Multimedia Appendix 4

Intervention's effect on high-sugar and high-fat food consumption according to the participants' baseline readiness status, overall and by country. Precon/Con: precontemplation or contemplation; Prep/Act: preparation or action; Maint: maintenance.

[JPG File, 46KB - mhealth_v6i11e10226_app4.jpg]

Multimedia Appendix 5

Intervention's effect on physical activity according to the participant's baseline readiness status, overall and by country. Precon/Con: precontemplation or contemplation; Prep/Act: preparation or action; Maint: maintenance.

[JPG File, 46KB - mhealth_v6i11e10226_app5.jpg]

References

1. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A Systematic Review of the mHealth Interventions to Prevent Alcohol and Substance Abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
2. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to Influence Physical Activity and Sedentary Behaviors: Behavior Change Techniques, Systematic Review and Meta-Analysis of Randomized Controlled Trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
3. Beratarrechea A, Moyano D, Irazola V, Rubinstein A. mHealth Interventions to Counter Noncommunicable Diseases in Developing Countries: Still an Uncertain Promise. Cardiol Clin 2017 Feb;35(1):13-30. [doi: 10.1016/j.ccl.2016.08.009] [Medline: 27886783]
4. Müller AM, Alley S, Schoeppe S, Vandelanotte C. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: A systematic review. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]

5. Stephani V, Opoku D, Quentin W. A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. BMC Public Health 2016 Dec 15;16:572 [FREE Full text] [doi: 10.1186/s12889-016-3226-3] [Medline: 27417513]
6. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014;10:49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]
7. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015;350:h1258 [FREE Full text] [Medline: 25791983]
8. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [Medline: 24609605]
9. 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]
10. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]
11. Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, Giobbie-Hurder A, et al. Multisite randomized controlled trials in health services research: scientific challenges and operational issues. Med Care 2001 Jun;39(6):627-634. [Medline: 11404645]
12. Zweben A, Donovan DM, Randall CL, Barrett D, Dermen K, Kabela E, et al. Issues in the development of subject recruitment strategies and eligibility criteria in multisite trials of matching. J Stud Alcohol Suppl 1994 Dec;12:62-69. [Medline: 7723000]
13. Eisenstein EL, Lobach DF, Montgomery P, Kawamoto K, Anstrom KJ. Evaluating implementation fidelity in health information technology interventions. AMIA Annu Symp Proc 2007 Oct 11:211-215 [FREE Full text] [Medline: 18693828]
14. Medical Research Council (MRC). Developing and evaluating complex interventions URL: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/ [accessed 2018-10-08] [WebCite Cache ID 731dQc51x]
15. Grier S, Bryant CA. Social marketing in public health. Annu Rev Public Health 2005;26:319-339. [doi: 10.1146/annurev.publhealth.26.021304.144610] [Medline: 15760292]
16. Harris JR, Cheadle A, Hannon PA, Forehand M, Lichiello P, Mahoney E, et al. A framework for disseminating evidence-based health promotion practices. Prev Chronic Dis 2012;9:E22 [FREE Full text] [Medline: 22172189]
17. Cahill K, Lancaster T, Green N. Stage-based interventions for smoking cessation. Cochrane Database Syst Rev 2010 Nov 10(11):CD004492. [doi: 10.1002/14651858.CD004492.pub4] [Medline: 21069681]
18. Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, et al. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. BMJ 2003 May 31;326(7400):1175-1177 [FREE Full text] [doi: 10.1136/bmj.326.7400.1175] [Medline: 12775617]
19. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, et al. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2] [Medline: 26653067]
20. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. BMJ 2013 Nov 20;347:f6753. [Medline: 24259324]
21. Kabagambe EK, Baylin A, Allan DA, Siles X, Spiegelman D, Campos H. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. Am J Epidemiol 2001 Dec 15;154(12):1126-1135. [Medline: 11744518]
22. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
23. Prochaska J, DiCle MC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51:95.
24. Callaghan RC, Hathaway A, Cunningham JA, Vettese LC, Wyatt S, Taylor L. Does stage-of-change predict dropout in a culturally diverse sample of adolescents admitted to inpatient substance-abuse treatment? A test of the Transtheoretical Model. Addict Behav 2005 Oct;30(9):1834-1847. [doi: 10.1016/j.addbeh.2005.07.015] [Medline: 16111832]
25. Kelaher M, Gillespie AG, Allotey P, Manderson L, Potts H, Sheldrake M, et al. The Transtheoretical Model and cervical screening: its application among culturally diverse communities in Queensland, Australia. Ethn Health 1999 Nov;4(4):259-276. [doi: 10.1080/13557859998047] [Medline: 10705563]
26. Muench F, Baumel A. More Than a Text Message: Dismantling Digital Triggers to Curate Behavior Change in Patient-Centered Health Interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]

27. Gomez QS, Walthouwer MJL, Schulz DN, de Vries H. mHealth or eHealth? Efficacy, Use, and Appreciation of a Web-Based Computer-Tailored Physical Activity Intervention for Dutch Adults: A Randomized Controlled Trial. J Med Internet Res 2016 Nov 09;18(11):e278 [FREE Full text] [doi: 10.2196/jmir.6171] [Medline: 27829576]
28. Diez-Canseco F, Zavala-Loayza JA, Beratarrechea A, Kanter R, Ramirez-Zea M, Rubinstein A, et al. Design and Multi-Country Validation of Text Messages for an mHealth Intervention for Primary Prevention of Progression to Hypertension in Latin America. JMIR Mhealth Uhealth 2015 Feb 18;3(1):e19 [FREE Full text] [doi: 10.2196/mhealth.3874] [Medline: 25693595]
29. Mastellos N, Gunn LH, Felix LM, Car J, Majeed A. Transtheoretical model stages of change for dietary and physical exercise modification in weight loss management for overweight and obese adults. Cochrane Database Syst Rev 2014;2:CD008066. [doi: 10.1002/14651858.CD008066.pub3] [Medline: 24500864]
30. Carvalho DMM, Bedeschi LB, Santos LCD, Lopes ACS. Interventions directed at eating habits and physical activity using the Transtheoretical Model: a systematic review. Nutr Hosp 2016 Sep 20;33(5):586 [FREE Full text] [Medline: 27759990]

Abbreviations

GISMAL: Grupo de Investigación en Salud Móvil en América Latina
mHealth: mobile health
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 15.03.18; peer-reviewed by N Darcel, K Bosak; comments to author 10.05.18; revised version received 21.06.18; accepted 16.07.18; published 01.11.18.

Please cite as:
Carrillo-Larco RM, Jiwani SS, Diez-Canseco F, Kanter R, Beratarrechea A, Irazola V, Ramirez-Zea M, Rubinstein A, Martinez H, Miranda JJ, GISMAL Group
Implementation Tells Us More Beyond Pooled Estimates: Secondary Analysis of a Multicountry mHealth Trial to Reduce Blood Pressure
JMIR Mhealth Uhealth 2018;6(11):e10226
URL: http://mhealth.jmir.org/2018/11/e10226/
doi:10.2196/10226
PMID:30389646

Original Paper

Using Mobile Phones to Examine and Enhance Perceptions of Control in Mildly Depressed and Nondepressed Volunteers: Intervention Study

Rachel Msetfi[1], PhD; Donal O'Sullivan[2], MSc; Amy Walsh[3], MA; John Nelson[2], PhD; Pepijn Van de Ven[2], PhD

[1]Department of Psychology, Health Research Institute, University of Limerick, Limerick, Ireland
[2]Department of Electronics and Computer Engineering, University of Limerick, Limerick, Ireland
[3]Department of Psychology, University of Limerick, Limerick, Ireland

Corresponding Author:
Rachel Msetfi, PhD
Department of Psychology
Health Research Institute
University of Limerick
Castletroy
Limerick,
Ireland
Phone: 353 61202709
Email: rachel.msetfi@ul.ie

Abstract

Background: Perceived control is strongly linked to healthy outcomes, mental healthiness, and psychological well-being. This is particularly important when people have little control over things that are happening to them. Perceived control studies have been performed extensively in laboratory settings and show that perceived control can be increased by experimental manipulations. Although these studies suggest that it may be possible to improve people's mental health by increasing their perceived control, there is very little evidence to date to suggest that perceived control can also be influenced in the real world.

Objective: The first aim of this study was to test for evidence of a link between noncontrol situations and psychological well-being in the real world using a mobile phone app. The second and arguably more important aim of the study was to test whether a simple instructional intervention on the nature of alternative causes would enhance people's perceptions of their own control in these noncontrol situations.

Methods: We implemented a behavioral action-outcome contingency judgment task using a mobile phone app. An opportunity sample of 106 healthy volunteers scoring low (n=56, no depression) or high (n=50, mild depression) on a depression scale participated. They were given no control over the occurrence of a low- or high-frequency stimulus that was embedded in everyday phone interactions during a typical day lasting 8 hours. The intervention involved instructions that either described a consistent alternative cause against which to assess their own control, or dynamic alternative causes of the outcome. Throughout the day, participants rated their own control over the stimulus using a quantitative judgment scale.

Results: Participants with no evidence of depression overestimated their control, whereas those who were most depressed were more accurate in their control ratings. Instructions given to all participants about the nature of alternative causes significantly affected the pattern of perceived control ratings. Instructions describing discrete alternative causes enhanced perceived control for all participants, whereas dynamic alternative causes were linked to less perceived control.

Conclusions: Perceptions of external causes are important to perceived control and can be used to enhance people's perceptions. Theoretically motivated interventions can be used to enhance perceived control using mobile phone apps. This is the first study to do so in a real-world setting.

(JMIR Mhealth Uhealth 2018;6(11):e10114) doi:10.2196/10114

KEYWORDS

perception of control; illusory control; well-being; depression; health; intervention; causal learning

Introduction

Background

Perceived control is critical to health outcomes, mental healthiness, and psychological well-being. For example, numerous studies have measured perceived personal control using psychometric questionnaire measures and have shown direct relationships to health outcomes (eg, cancer [1], diabetes [2], heart disease [3], and treatment adherence and effectiveness [4]), with control mediating the negative consequences of adverse conditions [5]. A key idea is that when healthy people have no control over events, they tend towards an "illusory" perception of control (eg, [6]). This is thought of as a protective bias that supports people's sense of control, and therefore well-being, when they cannot control things that are happening to them (eg, [7]). Conversely, people with depression are held to recognize situations in which they have no control all too well [8]. This "depressive realism" phenomenon may represent the absence of a healthy protective mechanism, with illusory control being an ingredient for positive physical and mental health [9,10]. Given the importance of perceived control for health, the aim of the current study was to assess for evidence of this phenomenon outside of the laboratory using a mobile phone app and to test whether a simple theoretically motivated intervention could enhance people's perceptions of control in a healthy manner.

Previous Research

Despite its importance, perceived control research suffers from a lack of studies carried out in real-world or applied settings. So far, laboratory-based research has been the only method of showing whether a person perceives that they have control when there is none. This is because the actual control a person has over a situation needs to be known and adjustable by the experimenter, and an accurate, objective measure of people's experiences is required [11], which is almost never the case in the real world. Some methodologies used in this domain, for example, comparisons between self and observer ratings of a situation [12] or between personal and population risk (eg, of a cancer diagnosis [13]), have provided useful insight but cannot allow a definitive diagnosis of illusory perceived control.

An objective measure of available control is clearly present in laboratory tasks involving "contingency judgments" as participants are exposed to carefully measured contingencies between their actions and outcomes [11]. The contingencies between a person's actions and subsequent outcomes are defined using four event-outcome frequencies (Figure 1):

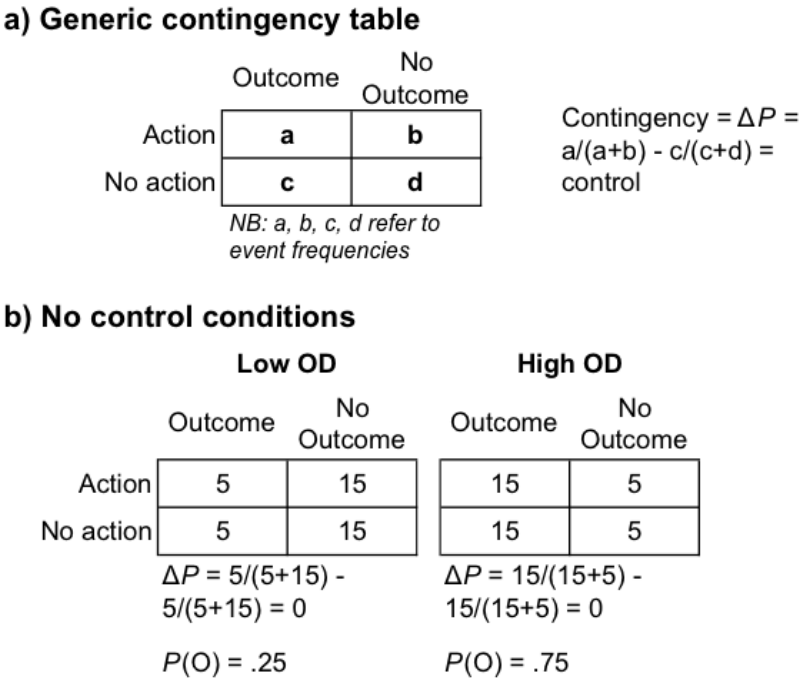
a. a user action is followed by an outcome
b. a user action is not followed by an outcome
c. no action by the user is followed by an outcome
d. no action by the user is not followed by an outcome (ie, no actions and no outcomes)

These events are usually programmed to occur over a short period of time and are quantified using the normative delta *P* (ΔP) metric [14]. Delta *P* is the difference between *P* (O|A), the probability of a user action (A) being followed by an outcome (O) and *P* (O|~A), the probability of the same outcome occurring when the user does not perform the action. Positive and negative ΔP values indicate the user has a certain control over the outcome, though in the case of a negative ΔP the outcome will be more likely to occur when there is no action by the user. A ΔP value equal to zero indicates the user cannot control the outcome through the action. In both of the specific examples given in Figure 1, the person has no control over salient outcomes (ΔP=0) but the frequency or density of outcomes varies from low to high (ie, low outcome density, high outcome density). Therefore, if accurate, people's perceptions of control should not differ between these two conditions. However, numerous studies have shown that healthy people exposed to a high outcome density condition tend to overestimate their control relative to the low outcome density condition (eg, [15]) and relative to people who are depressed (eg, [8,16]). These findings have provided evidence for the link between illusory control and healthy states.

However, the requirement for careful experimental control means that the basic experimental findings have never been tested outside the laboratory. This raises key methodological concerns around external and ecological validity, of generalizability from one very specific control situation to the whole of life [17] and the difference between behavior instructed in the laboratory and that occurring naturally in the real world [18]. Such basic methodological critiques of perceived control research are well acknowledged [19] and have limited the potential for this area of research to result in interventions for applied settings, although laboratory-based interventions have begun to be tested [20].

Where laboratory research has been helpful is in shedding light on our understanding of the psychological processes underpinning perceived control and, theoretically, the factors that will enhance perceptions if used to formulate interventions. So, for example, we know that the perception of alternative causes of outcomes is a key moderator of perceived control [14,21]. Whether a rule-based normative model [14] or a process-based associative model [21] is preferred, one's own control is evaluated against the control exerted by alternative causes. Other potential controlling causes are numerous, both inside and outside the laboratory, and include the environment or context in which events occur. For example, if a person wanted to control the heat level in a room using heating controls (action), an important alternative cause of heat variation would be the room itself and the effectiveness of the central heating system therein. In other words, the context is a key conditionalizer of control experience [22] and has been indicated as a key factor that discriminates healthy and mildly depressed people in their control perceptions [16]. These findings should theoretically [21] lend themselves to interventions that will influence people's perceptions of alternative causes and enhance their feelings of control [20].

Figure 1. Generic (a) and specific (b) contingency relationships between the occurrence of an action and the occurrence of an outcome in conditions (b) in which there is "no control'"over the outcome. NB. The term "outcome density" (OD) refers to the probability of an outcome occurring over all events, P(O).



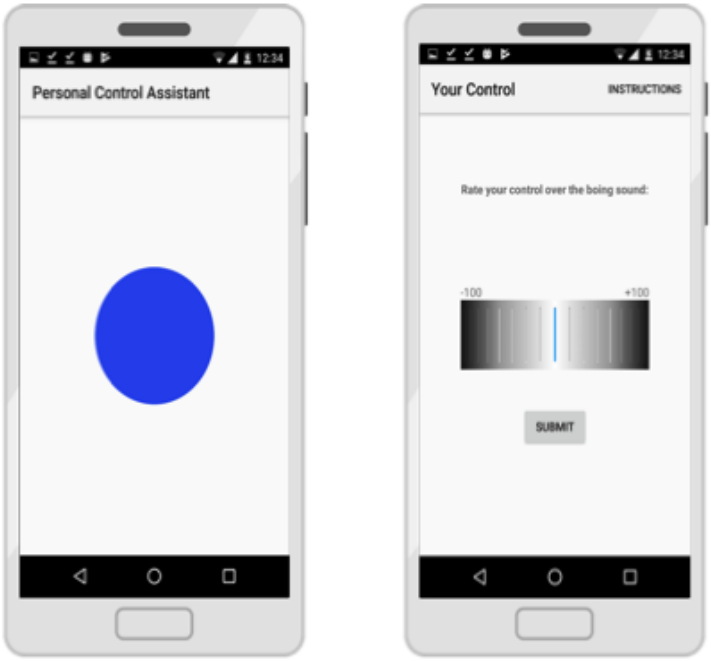
Study Goals

Given the ubiquity of mobile phones in everyday life, and their potential as data gathering and intervention devices, the goals of this study were to investigate perceived control in the participant's normal everyday environment and to test potential interventions that may increase a participant's perceived control, whether or not they are depressed.

To this end, we implemented a contingency judgment control task designed to run on Android mobile phones using the same conditions displayed in Figure 1. In order to maximize ecological validity, we used two strategies. First, experimental trials were embedded into participants' everyday lives through user-phone interactions programmed to take place throughout the activities a participant would experience during a typical day (see Figure 2). Second, the user-phone interactions were modeled on very typical activities, such as when the phone user is alerted to the availability of new information (eg, a message), which is followed up by an action (eg, click to access) and an outcome (eg, message, picture, video). Thus, in our procedure, each trial was prompted by a standard Android alert message and consisted of a user action followed by the occurrence of a brief auditory stimulus or no auditory stimulus at the programmed probability. We also asked participants to rate their control over the auditory outcome at five time points throughout the day. Ratings were performed by scrolling a wheel to a value between -100 and 100, with -100 indicating a perception of complete preventative control and +100 indicating complete generative control. Note that such ratings can be mapped onto the programmed ΔP metric.

In addition, we programmed an intervention by manipulating the instructions given to participants about the nature of alternative causes of the auditory stimulus. We defined the alternative cause as a discrete, static entity constantly present throughout the task (ie, the mobile phone network) or as a dynamic entity that changed throughout the day (ie, the different places they visited during the day). This is important because, theoretically, there is a finite amount of causal control available to any given outcome [14,21]. This means that if one cause is seen as a strong "controller," it is at the expense of all other potential causes, which will be seen as weak "controllers" of the outcome. Any cause that is constantly present (discrete context cause), especially when the outcome is absent, will be seen as a very weak controller, with other causes seeing their control enhanced. Therefore, we predict that participants in the discrete condition will see their actions as strong controllers of the outcome, and the extent of this control will be simply linked to the frequency with which the outcome occurs. The latter will increase the difference in control judgments between the low and high frequency conditions. Conversely, dynamic and therefore multiple alternative causes will have the opposite effect. We predict that people will learn as much about each cause as they do about the action. With causal control shared between so many potential causes, all of them will be perceived as weak, and the illusory control will be reduced (see also [16]).

Figure 2. Contingency judgment task implemented as user-phone interactions, including action and auditory outcome stimulus, with control ratings given after every block of 8 interactions (trials).



To date, mobile devices have been used as data gatherers and to deliver therapy and interventions for numerous conditions (eg, depression [23]). Here, we also study the effects of levels of depressed mood because even minor elevations, so called mild depression [8], dysphoria [16], or even scores above the median on a depression scale [24], can have a significant impact on control perceptions. In this study, in order to test a random sample of the population and examine the effects of depression levels, we use the median as the cut-off score on the depression scale. Importantly, although mobile devices have been used very successfully to support healthy behavior change (eg, exercise [25]) and to promote healthy eating [26], they have never been used to assess or promote a general sense of control. Our study will be the first to do this.

Methods

Recruitment

Participants consisted of 106 university students who volunteered to participate by responding to an email advertisement and fulfilled the inclusion criteria: (1) access to an Android mobile phone and (2) over the age of 17. Volunteers completed the Beck Depression Inventory (BDI) [27] during a visit to the laboratory after which they were supervised in downloading and installing the mobile app.

On the basis of their BDI scores, participants were categorized as members of the low BDI group (BDI ≤5, n=56, female n=27, representing non-depression) or high BDI group (BDI >5, n=50, female n=36, representing mild depression). This cut-off value for group categorization has been used in other similar studies examining the effects of mild depression or dysphoria on contingency learning and represents the median BDI score in most samples [24,28]. Throughout this paper, we refer to high BDI groups or mild depression in order to describe and explain our findings.

Table 1 shows that the BDI groups produced significantly higher scores on other depression relevant scales, including anxiety and stress, specifically the Depression Anxiety Stress Scales (DASS). Given nonrandom assignment to BDI groups, we attempted to match these BDI groups on three characteristics relevant to performance on cognitive tasks and which are affected by levels of depression: age, IQ, and short-term memory capacity (see Measures section). We checked for any statistical differences between groups on these three measures. Thus, the groups were matched on age and estimated IQ scores [29] but not short-term memory capacity, as measured by digit span scores [30] (Table 1).

Design

This study used a mixed 2×2×2×(5) factorial design, in which the between-groups variables were BDI group (2: low BDI, high BDI), outcome density (2: low, high), and alternative cause instructions (discrete, dynamic). The repeated measures variable was judgment block (5: 1-5). Participants made ratings of the control their actions had over the occurrence of the outcome. These were made using a +100 (complete control) through 0 (no control) to -100 (preventative control) scale, presented to the participant as a wheel (Figure 2). In addition, we recorded the number of actions made by each participant and calculated the contingency (actual ΔP) and outcome density experienced (based on the number and types of trials experienced), and the number of trials missed.

Table 1. Participant characteristics for each BDI group.

Measure	Low BDI[a] (n=56)		High BDI (n=50)		Multivariate analysis of variance[b]	
	Mean	SE	Mean	SE	*F*	*P* value
Age	22.98	0.98	25.68	1.03	3.60	.06
Digit span score	7.61	0.18	6.90	0.19	7.47	.01
Estimated IQ	113.00	0.83	112.33	0.88	0.30	.58
BDI	2.30	0.65	12.56	0.68	119.21	<.001
DASS[c]-Anxiety	1.02	0.33	3.62	0.35	29.40	<.001
DASS-Depression	1.05	0.41	4.96	0.43	43.01	<.001
DASS-Stress	2.29	0.49	6.50	0.52	34.95	<.001

[a]BDI: Beck Depression Inventory.

[b]*df*=1, 104.

[c]DASS: Depression Anxiety Stress Scales.

Measures

Participants completed the BDI; the DASS [31]; the digit span test, which provides a measure of short-term memory capacity; an estimated measure of pre-morbid IQ; and a number of other demographic items. Briefly, the digit span test requires participants to retain and repeat a randomized series of digits read to them at a rate of one per second. For the purposes of this study, the test was computerized. IQ was estimated using a formula applied to demographic variables. These measures are described in detail elsewhere [32]. In addition, participants completed a perception of control task, that was administered using a mobile phone.

Perception of Control Task

The task was implemented using an app developed using an Android library for the context-aware delivery of messages to users' mobile devices [33]. The app required wireless or network connectivity with the server initially in order for randomization to groups to take place and the experimental condition settings to be downloaded to the phone. Once this was complete, the app functioned independently and did not require continuous wireless or network connectivity. Incremental data upload to the server was programmed to take place as soon as connectivity was available, both during the task and once the task was finished.

The task was implemented as a discrete trials contingency judgment task, including 40 experimental trials lasting 6 seconds each, divided by intertrial intervals lasting on average 12 minutes (calculated using the Fleshler-Hoffman progression [34]). Each trial was prompted by a standard Android "notification" message (similar to those delivered to alert a user that a text message has arrived and is waiting to be read) that included a brief auditory and visual signal. A time limit of 2 minutes was given for participants to access the alert, after which the trial would be categorized as a "miss." Under these circumstances, the alert would be removed from the screen and the next intertrial interval would commence. This procedure was used to ensure that the procedure lasted for the same duration for all participants. If the participant accessed the alert within the 2-minute time frame, the onscreen button would appear on the screen in 2 seconds. They would then have the opportunity to press a touch-screen button for 3 seconds. Whether or not the button was pressed within 3 seconds, an auditory outcome would sound for 1 second (outcome present), or it would not sound (outcome absent), depending on the programmed probability. Following each of these experimental trials, the intertrial interval would commence. Action-outcome contingencies were programmed as in Figure 1. Participants were either randomized to the low outcome density group (*P* [O|A]=.25, *P* [O|A]=.25) or the high outcome density group (*P* [O|A]=.75, *P* [O|A]=.75). Therefore, for all participants, the programmed contingency was zero, and they had no control over the sound's occurrence. Every 8 trials, the participant would be asked to rate their own control using the previously described wheel (Figure 2). The procedure was programmed to last for approximately 8 hours and function during the participants' typical day, in a similar manner to typical mobile phone interactions.

Procedure

After having been fully briefed on arrival at the lab and having given informed consent, participants were asked for their demographic details, completed a series of questionnaires about their mood, and performed the digit span test. Following this, the experimenter helped participants download the app onto their own mobile phone and guided them through installing and activating the app. Once the app was activated, the participant was prompted to enter a code that would act as a unique identifier to allow matching of lab- and app-generated data. Following this, the instructions (Multimedia Appendix 1) were presented and participants told that they would interact with the phone during the course of the day, which would be opportunities to test if their button pressing controlled the sound occurrence. Following each block of 8 trials and the corresponding control rating, participants received an intervention message to prompt the participant to consider the influence of their context on their control. In the "discrete" condition, participants were told to think about the "control external factors have...this could be factors related to the phone, the mobile network or anything apart from your actions." In the dynamic context condition, participants were asked to think

about "control the place you are located in has...It's important to note that you will change your location throughout the day. The place you are in could affect whether the boing sound occurs, regardless of your actions." When the procedure was complete, the app provided debriefing information and links to support information provided on a webpage.

Power

We conducted a priori power analyses, which indicated that a sample of 152 was required for a power of 0.8 to detect medium-sized between-group effects and a sample of 24 and 48 to detect repeated measures effects and interactions respectively. However, only 106 volunteers kept appointments at the laboratory to participate in the study. Based on the achieved sample size, compromise power to detect the repeated measures effects and interactions, which were the focus of our theoretical predictions, was high (>.99). The power to detect main effects of between-group variables was somewhat lower than we planned (0.70). In spite of this, the size of the key main effects was within the 90% confidence limits (BDI) leading us to conclude that this study provided an adequately powered test of the hypotheses.

Due to participants' completing the task at the same time as their everyday activities, we anticipated missing data, in terms of trials and ratings, as well as issues such as loss of mobile phone battery. In this dataset, 10.6% of judgment values were missing. We therefore carried out multiple imputation, which involves replacing missing data with values generated from a series of multiple regression analyses including standard error and available parameter estimates. The fifth and most conservative iteration was used for the analyses reported in this paper.

Perception of Control Task Validation

As this is the first time a contingency task has been tested outside the laboratory, it was important to track all user-phone interactions and report whether experience was in line with what we had programmed. On average, participants missed 11.4 trials (SE 0.79) of the programmed 40 and, as instructed, pressed the button on around half of the trials they engaged with (press proportion mean 0.58, SE 0.019). The actual contingency (ΔP) experienced was again close to 0 as programmed (mean 0.04, SE 0.02). Participants in the low outcome density condition experienced outcomes on an average of 11.7 trials out of 40 (29.2%, SE 2.7%) whereas participants in the high outcome density condition experienced outcomes on 27.7 trials out of 40 (69.2%, SE 2.5%). Overall, the recorded engagement with experimental trials and contingency experience was as we programmed.

Results

In order to test the hypotheses, that (1) illusory control and depressive realism effects would be present, and (2) that the intervention would enhance ratings of control, a mixed factorial analysis of variance (Multimedia Appendix 2), which included actual ΔP experience as a covariate, was carried out on judgments of control. The alpha level was held at .05 throughout all analyses unless stated otherwise.

Illusory Control and Depression Effects

There are two tests to demonstrate the presence or absence of an illusory control effect. The simplest and first test assumes that judgments that are different to zero represent illusory control. In this test, the higher the absolute number of the rating, the stronger the illusion of control. The second and more rigorous test assumes that people's subjective perception of control scale might differ from the numeric scale on which they are asked to rate it and that positioning of ratings on the judgment scale is rather arbitrary. Therefore, the second test of illusory control is to compare ratings of two conditions (low and high outcome density) that have the same contingency. Here, we carried out both tests simultaneously using the multifactorial analysis of variance (ANOVA) and follow-up tests.

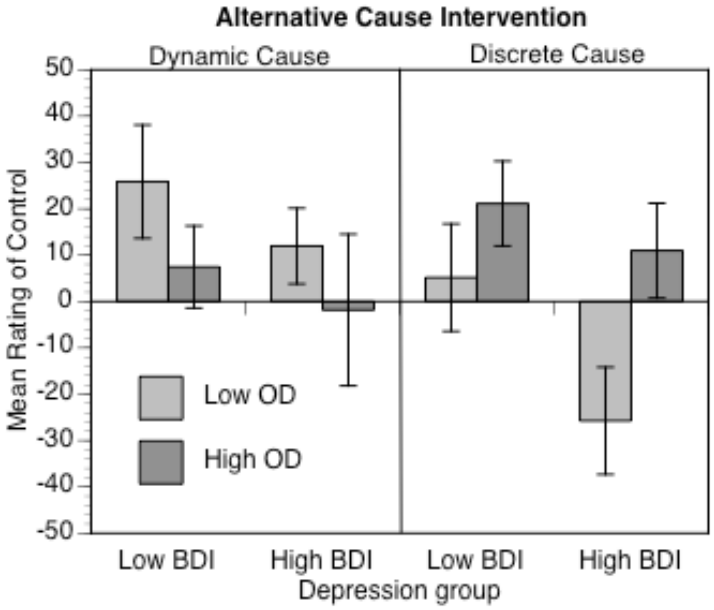
On average, low BDI participants showed the illusion of control. They rated their control as 14.90 (SE 5.37), whereas high BDIs rated their control as nearer to zero (mean 1.16, SE 5.98). This main effect of BDI group was significant and confirmed that low BDI groups produced reliably higher ratings of control than participants with higher levels of depression: $F_{1,91}$=4.05, mean squared error (MSE) 6671.40, P=.047, partial eta square=0.04, 90% CL 0.0004-0.1271. Subsequent single samples t tests comparing ratings to a criterion accuracy value of 0 showed that, for low BDI participants, 4 out of 5 action ratings were significantly higher than 0, all ts>2, all Ps<.05, evidencing illusory control. For more dysphoric, high BDI participants, 5 out of 5 action ratings were not reliably different to 0, all ts<1.16, and all Ps>.25. However, the BDI by outcome density interaction was not reliable ($F_{1,91}$=0.63, MSE 6671.40, P=.43, partial eta square=0.0007), showing that the low BDI trend towards larger absolute ratings was not only evident in high outcome density conditions. This shows that non-depressed participants tended toward perceiving that they had more control over the occurrence of the auditory stimulus than depressed participants did whose ratings represented lower levels of perceived control.

Alternative Cause Intervention Effect

Figure 3 suggests that, as we predicted, instructions on the nature of the alternative cause enhanced perceptions of control. Only when participants were instructed that the alternative cause was a discrete entity did they show evidence of the healthy illusory control.

The analysis described above supported this observation because the interaction between instructions and outcome density was reliable: $F_{1,91}$=7.05, MSE 6671.40, P=.009, partial eta square=.07, 90% CL 0.01-0.1674. Follow-up simple effects analyses confirmed this pattern and showed that the high outcome density conditions received higher ratings than low outcome density conditions only with discrete cause instructions ($F_{1,91}$=6.06, MSE 1334.28, P=.02, partial eta square=.062, 90% CL 0.0064-0.1548) and not with dynamic cause instructions ($F_{1,91}$=1.85, MSE 1334.28, P=.18, partial eta square=.02, 90% CL 0-0.0887).

Figure 3. Mean ratings of control as a function of outcome density, Beck Depression Inventory (BDI) group and alternative cause intervention (error bars correspond to standard errors of the mean).



These findings show that perceived control can be increased by providing people with simple instructions about the nature of alternative causes. Participants who compared their own control to discrete alternative causes produced higher control ratings.

Discussion

Principal Considerations

We used a mobile phone app in order to embed and measure perceived control occurring in participants' everyday lives as naturalistic mobile phone interactions. The app was designed such that participants' actions had no control over the occurrence of an auditory stimulus. Participants without signs of depression overestimated their control, whereas participants showing mild levels of depression rated their control as close to zero. This study provides the first objective demonstration of illusory control and depressive realism in a real-world setting.

Until now, the basic experimental findings of illusory control have never been rigorously tested outside the laboratory. This has raised important methodological and theoretical concerns, in particular around external and ecological validity, of generalizability from one very specific control situation to the whole of life [17] and of the difference between behavior instructed in the laboratory and that occurring naturally in the real world [18]. Such basic methodological critiques of perceived control research are well acknowledged in relation to depression [19] but not really discussed in the general health literature in which the concept of control is so frequently used. This gap has limited how our theoretical understanding of illusory control can be extrapolated to develop interventions, such as that described and tested here.

Notably, the instruction intervention significantly influenced all participants' ratings of control. As we predicted, when the alternative cause was described as a discrete, constantly present entity (the mobile phone network), people judged that they had more control when the auditory stimulus occurred frequently (eg, [15]). Conversely, when people were informed that alternative causes were dynamic (ie, place) and would change throughout the experience, there was no evidence of illusory control effect. This finding suggests that the intervention used in this study, which was based on theoretical accounts of how people use alternative causes to evaluate their own control [14,21] was effective and successful.

This finding is consistent with work that shows that the perceived strength [28], salience [35], or exposure [16] of alternative causes, such as the context or other causes, competes with the cause under consideration for perceived control [15]. However, several things are important to note for those interested in designing perceived control interventions because the direction of the change in perceived control will depend on a number of key factors. First, the degree of control actually present in a situation is critical. When the person does have control, a strong alternative cause will reduce perceived control [28]. Second, the effect of any intervention or experimental manipulation, which changes the salience of or exposure to the alternative cause, will depend on what the person learns about the alternative cause as a result. So, for example, Vadillo et al [35] introduced a "difficult to ignore" alternative cause into their laboratory procedure, which seemed more strongly related to the outcome than participants' actions. Msetfi et al [16] introduced a long delay into the procedure during the time that the alternative cause was present. This allowed participants to learn that the alternative cause was only weakly related to the outcome, which boosted their own perception of control. Both of these findings show how circumstances and information, which are present, can differentially influence people's perceptions of their own control in very similar "no control" situations. This is critical information for those working in the health space, as perception of control is a very important

mediator of health outcomes [4], which can be influenced unintentionally, as well as through intervention.

Strengths and Limitations

It is important to consider whether the experimental procedure was a valid test of the contingencies that we planned for participants to experience. This is because it has been acknowledged that changes in participant behavior can actually affect the contingencies they are exposed to [18,24,36]. Our concern here was missed trials. However, careful scrutiny of the data recorded on each trial showed that even though participants missed some trials, they experienced the contingencies as programmed. This alleviates an important concern.

Another limitation is that our implementation of the contingency judgment task was an extremely crude analogue of real-life user–mobile phone interactions. However, in this first test of illusory control and depressive realism outside the laboratory, it was important to implement a real-life contingency task that was as similar as possible to lab procedures in order to provide the replication required. Our future studies will not only be able to provide a more sophisticated and naturalistic interface but will also collect richer data, including concurrent natural activities and behaviors, as well as mood and well-being ratings over longer periods of time. We also note that we have used similar nonclinical BDI criteria to test for depressive realism as the majority of the key studies in the literature [8,16]. This means that the generalizability of our findings to clinical depression is questionable, and we make no strong claims in this regard other than to state that we have replicated other depressive realism findings as tested elsewhere in the real world.

While we acknowledge the limitations of our study, especially in relation to the relatively controlled nature of the user-phone interactions we designed, a theoretically important aspect of ecological validity has been introduced into the procedure. This is the nature of the context or the alternative cause, which is fundamentally different in the laboratory to the dynamic, constantly changing contexts of real-world control situations. To date, most human learning studies simulate context using discrete cues [37] or places represented by pictures displayed on a computer screen [38] while a few have used an actual place [39]. It has been assumed that all such contexts “work” the same way in learning, and current theoretical models do not differentiate between them (see [40] for an extended discussion on context and learning). Our findings here suggest, to the contrary, that the representational content of context can change patterns of learning. This further emphasizes the importance of translating experimental research into the real world in order to fully use our theoretical knowledge to develop interventions.

Conclusion

The findings of this study show convincingly that when perceptions of control are measured in relation to an objective standard, biased estimates of control do correlate with mental healthiness, with illusory control being the healthiest type of control (eg, [11,19,36]). Importantly, the simple theoretically motivated intervention, which was designed to influence people’s ability to learn about the power of the alternative cause in contrast to their own control, was effective in increasing people’s perceptions of being “in control.” Finally, this study further demonstrates the power of mobile phone technologies for use in experimental and intervention research. This not only provides the opportunity to test psychological theory in novel ways embedded in a person’s everyday environment but also to collect ever richer data about natural behaviors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full instructions.

[PDF File (Adobe PDF File), 38KB - mhealth_v6i11e10114_app1.pdf]

Multimedia Appendix 2

Analysis of variance (ANOVA) table.

[PDF File (Adobe PDF File), 32KB - mhealth_v6i11e10114_app2.pdf]

References

1. Ell K, Mantell J, Hamovitch M, Nishimoto R. Social Support, Sense of Control, and Coping Among Patients with Breast, Lung, or Colorectal Cancer. Journal of Psychosocial Oncology 1989 Nov 07;7(3):63-89. [doi: 10.1300/J077v07n03_05]
2. Macrodimitris SD, Endler NS. Coping, control, and adjustment in Type 2 diabetes. Health Psychol 2001 May;20(3):208-216. [Medline: 11403218]
3. Bosma H, Marmot MG, Hemingway H, Nicholson AC, Brunner E, Stansfeld SA. Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study. BMJ 1997 Feb 22;314(7080):558-565 [FREE Full text] [Medline: 9055714]

4. Geers AL, Rose JP, Fowler SL, Rasinski HM, Brown JA, Helfer SG. Why does choice enhance treatment effectiveness? Using placebo treatments to demonstrate the role of personal control. J Pers Soc Psychol 2013 Oct;105(4):549-566. [doi: 10.1037/a0034005] [Medline: 23915042]
5. Bobak M, Pikhart H, Rose R, Hertzman C, Marmot M. Socioeconomic factors, material inequalities, and perceived control in self-rated health: cross-sectional data from seven post-communist countries. Soc Sci Med 2000 Nov;51(9):1343-1350. [Medline: 11037221]
6. Langer EJ. The illusion of control. Journal of Personality and Social Psychology 1975;32(2):311-328. [doi: 10.1037/0022-3514.32.2.311]
7. Taylor SE, Brown JD. Illusion and well-being: a social psychological perspective on mental health. Psychol Bull 1988 Mar;103(2):193-210. [Medline: 3283814]
8. Alloy LB, Abramson LY. Judgment of contingency in depressed and nondepressed students: sadder but wiser? J Exp Psychol Gen 1979 Dec;108(4):441-485. [Medline: 528910]
9. Alloy L, Albright J, Abramson L, Dykman B. Depressive realism and nondepressive optimistic illusions: The role of the self. In: Ingram RE, editor. Contemporary psychological approaches to depression: Theory, research, and treatment. New York, NY: Plenum Press; 1990:71-86.
10. Alloy L, Abramson L. Depressive realism: Four theoretical perspectives. In: Alloy LB, editor. Cognitive processes in depression. New York, NY: Guilford Press; 1988:223-265.
11. Ackermann R, DeRubeis R. Is depressive realism real? Clinical Psychology Review 1991 Jan;11(5):565-584. [doi: 10.1016/0272-7358(91)90004-E]
12. Ducharme J, Bachelor A. Perception of Social Functioning in Dysphoria. Cognitive Therapy and Research 1993;17(1):53-70.
13. Keller P, Lipkus I, Rimer B. Depressive realism and health risk accuracy: The negative consequences of positive mood. Journal of Consumer Research 2002;29(1):57-69. [doi: 10.1086/339921]
14. Allan L. A note on measurement of contingency between two binary variables in judgment tasks. Bull Psychon Soc 2013 Nov 7;15(3):147-149. [doi: 10.3758/BF03334492]
15. Dickinson A, Shanks D, Evenden J. Judgement of act-outcome contingency: The role of selective attribution. The Quarterly Journal of Experimental Psychology Section A 2007 May 29;36(1):29-50. [doi: 10.1080/14640748408401502]
16. Msetfi RM, Murphy RA, Simpson J, Kornbrot DE. Depressive realism and outcome density bias in contingency judgments: the effect of the context and intertrial interval. J Exp Psychol Gen 2005 Feb;134(1):10-22. [doi: 10.1037/0096-3445.134.1.10] [Medline: 15702960]
17. Skinner EA. A guide to constructs of control. J Pers Soc Psychol 1996 Sep;71(3):549-570. [Medline: 8831161]
18. Matute H. Illusion of Control: Detecting Response-Outcome Independence in Analytic but Not in Naturalistic Conditions. Psychol Sci 2016 May 06;7(5):289-293. [doi: 10.1111/j.1467-9280.1996.tb00376.x]
19. Haaga DA, Beck AT. Perspectives on depressive realism: implications for cognitive theory of depression. Behav Res Ther 1995 Jan;33(1):41-48. [Medline: 7872935]
20. Msetfi RM, Brosnan L, Cavus HA. Enhanced attention to context: An intervention that increases perceived control in mild depression. Q J Exp Psychol (Hove) 2016;69(6):1073-1081. [doi: 10.1080/17470218.2016.1138134] [Medline: 26822778]
21. Rescorla R, Wagner A. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcementnon-reinforcement. In: Black A, Prokasy W, editors. Classical conditioning II: Theory and research. New York, NY: Appleton Century Crofts; 1972:64-99.
22. Msetfi RM, Wade C, Murphy RA. Context and time in causal learning: contingency and mood dependent effects. PLoS One 2013;8(5):e64063 [FREE Full text] [doi: 10.1371/journal.pone.0064063] [Medline: 23691147]
23. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
24. Byrom N, Msetfi R, Murphy R. Two pathways to causal control: use and availability of information in the environment in people with and without signs of depression. Acta psychologica 2015;157:1-12. [doi: 10.1016/j.actpsy.2015.02.0]
25. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
26. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. Health Informatics J 2009 Mar;15(1):17-25 [FREE Full text] [doi: 10.1177/1460458208099865] [Medline: 19218309]
27. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [Medline: 13688369]
28. Chase HW, Crockett MJ, Msetfi RM, Murphy RA, Clark L, Sahakian BJ, et al. 5-HT modulation by acute tryptophan depletion of human instrumental contingency judgements. Psychopharmacology (Berl) 2011 Feb;213(2-3):615-623 [FREE Full text] [doi: 10.1007/s00213-010-1934-4] [Medline: 20631992]
29. Barona A, Reynolds CR, Chastain R. A demographically based index of premorbid intelligence for the WAIS--R. Journal of Consulting and Clinical Psychology 1984;52(5):885-887. [doi: 10.1037/0022-006x.52.5.885]
30. Lezak M. Neuropsychological Assessment. New York, NY: Oxford University Press Inc; 1995.

31. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [Medline: 7726811]
32. Msetfi R, Kornbrot D, Matute H, Murphy R. The relationship between mood state and perceived control in contingency learning: effects of individualist and collectivist values. Front Psychol 2015;6:1430 [FREE Full text] [doi: 10.3389/fpsyg.2015.01430] [Medline: 26483707]
33. van de Ven P, O'Brien H, Henriques R, Klein M, Msetfi R, Nelson J, E-COMPARED Consortium. ULTEMAT: A mobile framework for smart ecological momentary assessments and interventions. Internet Interv 2017 Sep;9:74-81 [FREE Full text] [doi: 10.1016/j.invent.2017.07.001] [Medline: 30135840]
34. Fleshler M, Hoffman HS. A progression for generating variable-interval schedules. J Exp Anal Behav 1962 Oct;5:529-530 [FREE Full text] [doi: 10.1901/jeab.1962.5-529] [Medline: 13945507]
35. Vadillo M, Matute H, Blanco F. Fighting the illusion of control: How to make use of cue competition and alternative explanations. Universitas Psychologica 2013;12(1):261-270 [FREE Full text]
36. Msetfi RM, Murphy RA, Kornbrot DE. Dysphoric Mood States are Related to Sensitivity to Temporal Changes in Contingency. Front Psychol 2012;3:368 [FREE Full text] [doi: 10.3389/fpsyg.2012.00368] [Medline: 23060837]
37. Vallée-Tourangeau F, Murphy R, Drew S, Baker A. Judging the Importance of Constant and Variable Candidate Causes: A Test of the Power PC Theory. The Quarterly Journal of Experimental Psychology Section A 2018 Jan;51(1):65-84. [doi: 10.1080/713755745]
38. Nelson JB, Callejas-Aguilera JE. The role of interference produced by conflicting associations in contextual control. J Exp Psychol Anim Behav Process 2007 Jul;33(3):314-326. [doi: 10.1037/0097-7403.33.3.314] [Medline: 17620029]
39. Gray NS, Williams J, Fernandez M, Ruddle RA, Good MA, Snowden RJ. Context dependent latent inhibition in adult humans. Q J Exp Psychol B 2001 Aug;54(3):233-245. [doi: 10.1080/713932760] [Medline: 11547513]
40. Maren S, Phan KL, Liberzon I. The contextual brain: implications for fear conditioning, extinction and psychopathology. Nat Rev Neurosci 2013 Jun;14(6):417-428 [FREE Full text] [doi: 10.1038/nrn3492] [Medline: 23635870]

Abbreviations

BDI: Beck Depression Inventory
CL: confidence limits
DASS: Depression Anxiety Stress Scales
MSE: mean squared error

Edited by G Eysenbach; submitted 14.02.18; peer-reviewed by H Matute, T Krieger; comments to author 08.07.18; revised version received 05.09.18; accepted 10.09.18; published 09.11.18.

Please cite as:
Msetfi R, O'Sullivan D, Walsh A, Nelson J, Van de Ven P
Using Mobile Phones to Examine and Enhance Perceptions of Control in Mildly Depressed and Nondepressed Volunteers: Intervention Study
JMIR Mhealth Uhealth 2018;6(11):e10114
URL: http://mhealth.jmir.org/2018/11/e10114/
doi:10.2196/10114
PMID:30413398

Original Paper

An mHealth Intervention to Improve Young Gay and Bisexual Men's Sexual, Behavioral, and Mental Health in a Structurally Stigmatizing National Context

Corina Lelu iu-Weinberger[1*], PhD; Monica Manu[2*], MA; Florentina Ionescu[2*], MA; Bogdan Dogaru[3*], BA; Tudor Kovacs[3*], MA; Cristian Dorobăn escu[4*], BA; Mioara Predescu[5*], MD; Anthony Surace[6*], MA; John E Pachankis[7*], PhD

[1]François-Xavier Bagnoud Center, School of Nursing, Rutgers Biomedical and Health Sciences, Newark, NJ, United States

[2]Insight, Bucharest, Romania

[3]Population Services International Romania, Bucharest, Romania

[4]Data Center Solutions, Bucharest, Romania

[5]National Institute of Infectious Diseases "Prof. Dr. Matei Bal ", Bucharest, Romania

[6]Department of Behavioral and Social Sciences, School of Public Health, Brown University, Providence, RI, United States

[7]Social and Behavioral Sciences, Yale School of Public Health, New Haven, CT, United States

[*]all authors contributed equally

Corresponding Author:
Corina Lelu iu-Weinberger, PhD
François-Xavier Bagnoud Center
School of Nursing
Rutgers Biomedical and Health Sciences
65 Bergen Street, 8th Floor
Newark, NJ, 07101
United States
Phone: 1 9739722093
Email: cl1148@sn.rutgers.edu

Abstract

Background: Young gay and bisexual men (YGBM) in some Eastern European countries, such as Romania, face high stigma and discrimination, including in health care. Increasing HIV transmission is a concern given inadequate prevention, travel to high-prevalence countries, and popularity of sexual networking technologies.

Objective: This study aimed to adapt and pilot test, in Romania, a preliminarily efficacious mobile health (mHealth) HIV-prevention intervention, created in the United States, to reduce HIV risk among YGBM.

Methods: After an intervention formative phase, we enrolled 43 YGBM, mean age 23.2 (SD 3.6) years, who reported condomless sex with a male partner and at least 5 days of heavy drinking in the past 3 months. These YGBM completed up to eight 60-minute text-based counseling sessions grounded in motivational interviewing and cognitive behavioral skills training with trained counselors on a private study mobile platform. We conducted one-group pre-post intervention assessments of sexual (eg, HIV-risk behavior), behavioral (eg, alcohol use), and mental health (eg, depression) outcomes to evaluate the intervention impact.

Results: From baseline to follow-up, participants reported significant (1) increases in HIV-related knowledge (mean 4.6 vs mean 4.8; *P*=.001) and recent HIV testing (mean 2.8 vs mean 3.3; *P*=.05); (2) reductions in the number of days of heavy alcohol consumption (mean 12.8 vs mean 6.9; *P*=.005), and (3) increases in the self-efficacy of condom use (mean 3.3 vs mean 4.0; *P*=.01). Participants reported significant reductions in anxiety (mean 1.4 vs mean 1.0; *P*=.02) and depression (mean 1.5 vs mean 1.0; *P*=.003). The intervention yielded high acceptability and feasibility: 86% (38/44) of participants who began the intervention completed the minimum dose of 5 sessions, with an average of 7.1 sessions completed; evaluation interviews indicated that participation was rewarding and an "eye-opener" about HIV risk reduction, healthy identity development, and partner communication.

Conclusions: This first mHealth HIV risk-reduction pilot intervention for YGBM in Eastern Europe indicates preliminary efficacy and strong acceptability and feasibility. This mobile prevention tool lends itself to broad dissemination across various

similar settings pending future efficacy testing in a large trial, especially in contexts where stigma keeps YGBM out of reach of affirmative health interventions.

(JMIR Mhealth Uhealth 2018;6(11):e183) doi:10.2196/mhealth.9283

KEYWORDS

alcohol use; young gay and bisexual men; HIV risk; mental health; mHealth intervention

Introduction

HIV transmission among gay and bisexual men (GBM) is an increasing concern in Central and Eastern European countries, such as Romania, where previously reported low HIV rates are increasing [1,2]. GBM are a relatively hidden [3], yet highly vulnerable emerging HIV risk group in Romania, a country with one of the highest numbers of unrecognized HIV/AIDS cases in that region [4,5]. Currently, HIV transmission is largely driven by male-to-male sexual contact [5-7]. The Joint United Nations Programme on HIV/AIDS reports that HIV cases among GBM in Romania have nearly doubled from 8% in 2009 to 14% in 2011; however, this critical risk group has been suboptimally reached by prevention efforts, with substandard national attention dedicated to GBM health [6]. Furthermore, this group's vulnerability has increased due to frequent travel to high-prevalence countries after Romania's entry in the European Union in 2001 and the advent of sexual networking technologies.

Commensurate preventive resources are needed to counteract the increasing HIV incidence among Romanian young GBM (YGBM) [4,6,8-12]. However, this group faces some of the highest stigma and discrimination in Europe [12,13], including in health care [3], where GBM-specific expertise is limited. Consequently, Romanian YGBM often conceal their sexual orientation [5,14], have low rates of HIV testing [1,2], and often do not disclose their sexual behavior to providers, all of which likely lead to the rising HIV epidemic [15]. Sexual orientation stigma and subsequent discrimination have been shown to be associated with increased vulnerability [10,16], including in relation to HIV transmission, such as condomless sex (in the absence of preexposure prophylaxis), substance and alcohol use, poor mental health, and suboptimal HIV testing [17-21], which combine to form a syndemic (ie, synergistic epidemic) that further exacerbates HIV risk [21-35]. Accurate psychiatric data are not available for Romanian YGBM due to identity concealment motivated by fear and suboptimal medical record tracking [3,36]. However, nonprobability surveys and interviews suggest that Romanian YGBM report high levels of depression and alcohol abuse [37], which has been shown by numerous cross-European studies to be linked to exposure to stigma and discrimination [38-42]. Furthermore, excessive alcohol use, which is normative in Romania [42], has been shown to co-occur with sexual risk behavior [21,28,29]; by addressing mental and behavioral health, sexual risk is concomitantly addressed [26,43-45]. Prevention interventions that address HIV-related syndemics associated with stigma against YGBM are much needed in countries such as Romania.

Culturally sensitive adaptations of HIV-preventive interventions developed in other countries could provide a solution to the largely unaddressed health needs of YGBM in this region. Yet, given the lack of brick-and-mortar venues providing affirming health care for YGBM in Romania, alternative intervention delivery methods are required. Mobile health (mHealth) represents one such tool with capacity for reaching otherwise hidden populations [46]. mHealth possesses particular promise for YGBM, given their preference for Web-based sexual health support [46,47], the pervasiveness of mobile technologies throughout Romania [48-50], and the ability of mHealth technologies to bypass barriers to health care for marginalized populations by bringing prevention "to their pockets" [51].

Motivated by the lack of programming for HIV prevention in Romania, where the HIV epidemic is rapidly increasing and YGBM face difficulties accessing on-the-ground services out of fear of stigmatization, we introduced the first mobile program aimed at reducing the HIV risk among this vulnerable and underserved group. As such, this study was designed to test whether an mHealth HIV-prevention intervention would be feasible, acceptable, and preliminarily efficacious in supporting the unique HIV-prevention and related health needs of YGBM in Romania. Originally developed and successfully pilot-tested in the United States using a one-group pre-post intervention design [52], this intervention was adapted by our US-Romanian team to the local cultural context.

The present intervention is distinct from the original in several ways. First, the original intervention involved an internet (ie, Facebook)-based HIV-prevention intervention with YGBM in New York City, which the majority of participants completed via personal computer, and was not a mobile-based intervention. This study developed and tested a *mobile-based* HIV-prevention intervention in Romanian YGBM, which all participants completed via mobile devices, defined as any device that can be portable such as cellular phones, tablets, or laptops. Second, the original intervention focused on reducing recreational drug use, given the normative nature of drug use and its associations with HIV risk among the gay or bisexual population in New York City [53-57]. The current intervention focused on reducing alcohol abuse, given the normative nature of alcohol abuse and its associations with HIV risk among the gay and bisexual population in Romania [37,58]. Third, the interventions are distinct in that they were delivered in entirely different structural contexts; therefore, the Romanian study addressed structural stigma as a determinant of HIV risk and alcohol abuse. Despite these distinctions, the two interventions share certain features, namely their theoretical framework; both utilized motivational interviewing (MI) and cognitive behavioral skills training (CBST). This paper describes the intervention's adaptation, pilot implementation, and preliminary efficacy results using a sample of Romanian YGBM with the goal of providing an efficient prevention tool that resonates with and is capable of reaching this largely hidden high-risk population.

Methods

Participants and Procedures

Participants were recruited between September 2015 and July 2016 through ads posted on GBM-specific Facebook sites and groups and social sexual networking apps (eg, Grindr and Planet Romeo), word of mouth, and fliers distributed in the main lesbian, gay, bisexual, and transgender (LGBT) nightclub in Bucharest, the capital of Romania. While most recruitment took place in Bucharest, we were able to reach men in other towns of Romania through word of mouth and Facebook. The ads contained a link to a 5-minute Web-based screener assessing sexual behavior, alcohol use, HIV status, and HIV testing patterns. In order to qualify for both formative and intervention phases of the project, men had to (1) be aged between 16 and 29 years; (2) self-report an HIV-negative or unknown status; (3) have had at least one condomless anal sex act with a male partner in the past 3 months; and have either (4) at least 5 heavy drinking days (at least 5 standard drinks on one occasion) or (5) at least one condomless anal sex act under the influence of alcohol in the past 3 months. These criteria are based on established associations between alcohol use and condomless sex [55,59-65]. Those eligible were given the option to enter their contact information in a separate survey for the study coordinator to contact them by phone for enrollment. Each participant received a link to an informed consent document. The study coordinator answered participants' questions and verified their identity and age through an electronic copy of a government-issued identification card sent by the participant. After consent, each participant received a secure link to a 30-45-minute Web-based baseline assessment. Participants completed the same survey approximately 3 months postbaseline, as an immediate postintervention follow-up assessment (given that participants took between 2 and 3 months to complete their 8 weekly intervention sessions). The baseline and follow-up assessments were compensated with the equivalent of US $20 and US $40, respectively, provided to participants in gift cards at the end of their participation. Each completed session was compensated with the equivalent of US $10, allocated as above. Consent, assessments, and counseling sessions were conducted in Romanian, while all translation was overseen by one of the investigators who is a native Romanian speaker fluent in English. As the other investigator did not speak Romanian, whenever participants spoke fluent English (97% of the time), formative and evaluation interviews were conducted in English, given the high English language proficiency in Romania. When participants did not speak English, the bilingual investigator conducted interviews in Romanian, which she also translated into English for the other investigator (who was present for all interviews) and back into Romanian for the participants. The Romanian study team spoke English fluently. All procedures were approved by the Human Research Protections Program at Hunter College of the City University of New York and Yale University, USA.

The Intervention

Adaptation

The intervention, called "Despre Mine. Despre Noi." (DMDN), translated into English as "About Me. About Us.," was adapted for the cultural context of Romania from a US-based Web-based intervention that showed preliminary efficacy for reducing sexual risk and substance use while delivered on Facebook by trained counselors to YGBM at risk for HIV [52]. For the intervention to resonate with the Romanian GBM community, the 2 principal investigators conducted a focus group in December 2013 with YGBM (n=8) as an initial *general* needs assessment. In December 2014, the investigators returned to Romania to conduct 60-90-minute individual interviews with a new set of YGBM (n=22), infectious disease physicians from the National Institute of Infectious Diseases (n=3), and LGBT community advocates (n=3), as a more comprehensive *intervention* needs assessment. The focus group and interviews were audiorecorded and coded for pertinent themes related to the biopsychosocial context of HIV risk and prevention needs for YGBM in Romania and ideal delivery modalities for this unique context. Specifically, the themes assessed the unique sexual (eg, HIV risk, condom use patterns, and partner communication), behavioral (eg, contexts and motivations for excessive alcohol use), and mental (eg, depression, social isolation, relationship strain, coming out) health needs of Romanian YGBM to inform intervention content, structure, and optimal elements for a secure mobile study platform. Once the investigators extracted relevant themes from the focus group and interviews, they modified the original intervention to incorporate these elements and met in May 2015 with 15 of the original interviewees for 60-minute individual interviews, in order to walk them through the modified intervention to gather and incorporate additional feedback. In addition, the participants navigated the mobile platform (a "mobile device" is considered to be a mobile phone, a tablet, or a laptop given their portability) and provided feedback to the technical developer who was present at all the meetings. Final content-related and technical modifications were made before intervention launch based on iterative user testing by the YGBM and study team.

Content

The intervention aimed to support YGBM in reducing their condomless anal sex and alcohol use and improving their mental health. The 8 counseling sessions lasted 60 minutes each and were delivered through the chat feature of our study mobile platform. The intervention was based on the principles and techniques of MI [66,67] utilized to help participants contemplate and prepare for change to reduce HIV-risk behavior. This was supplemented by CBST [68,69] and information for accessing local health-promoting resources. A description of the core intervention elements has been published previously [52]. The DMDN intervention maintained its original 8 60-minute session structure [52]. Focus group and interviews during the adaptation phase yielded the following unique elements to support Romanian YGBM in their highly stigmatizing national context, leading to increased HIV risk behavior yet offering sparse HIV prevention and information resources: (1) basic HIV/AIDS education; (2) raising awareness

of the importance of HIV testing, importance of sexual health communication with partners, and health-promoting resource utilization; (3) support for sexual identity development and coming out, including offering the intervention to YGBM as young as 16 years of age (given their high familial and societal rejection); and (4) specific technical features.

Technical Platform

Given the mobile nature of the intervention, we gave participants the option to use any mobile device they preferred, including mobile phones, laptops, and tablets. DMDN is a mobile intervention, given that it is portable and can be accessed from anywhere by participants. A local software developer constructed and maintained the mobile site used to deliver the intervention and provide HIV-related information or resources to participants. The site had the following four domains: (1) a chat page with interfaces to chat with one's counselor; (2) a resource page containing HIV/AIDS-related information and links to relevant resources (eg, LGBT-affirmative testing clinics, European Centers for Disease Control fact sheets, thebody.com); (3) a behavioral and affective tracking tool where participants entered weekly reports of their number of sex partners and condomless anal sex acts with male partners, alcohol consumption, and negative and positive affect (eg, emotions such as "angry" or "happy"); and (4) account settings. Participants used the behavioral and affective tracking tool as a visual gauge of their behavior over time, which was reviewed in sessions with counselors to motivate behavior change and guide their discussions. The tracking tool was used on a weekly basis for the duration of the intervention.

Once participants completed their baseline assessment, the study coordinator notified their assigned counselor, who scheduled the first session. On the set day and time, the counselor and participant logged into the mHealth study platform and commenced the sessions.

Counselor Training and Supervision

Prior to intervention launch, the 2 US-based investigators held a 2-day training in May 2015 in Bucharest with 3 Master's level Romanian counselors—2 psychologists and 1 LGBT community advocate with experience in MI and HIV-prevention interventions. A training manual, written originally for the US-based intervention and adapted (by including the aforementioned elements) for the Romanian context by the investigators, was used. The training included both didactic and experiential components, a review of principles and techniques of MI and CBST, the unique nature of text-based therapeutic communication, and vignettes and exercises drawn from the investigators' previous interventions. Next, the counselors practiced each session via text-based communication, each taking turns being a mock participant and counselor and receiving remote video supervision from the clinical psychologist investigator. Once the intervention began, the counselors saved the transcripts from each session, translated them into English, and received video-based supervision from the clinical psychologist investigator, guided by the intervention manual, on a weekly basis for the duration of the study.

Measures

Demographics

Demographic items assessed age, gender, sexual orientation, level of education, current employment, income, and relationship status.

Sexual Behavior

To assess *HIV-risk behavior*, we asked participants how often they had anal and vaginal, insertive and receptive, sex with and without condoms with various sex partners (steady male or female, casual male or female, and transgender male or female) over the past 3 months. *Condom use self-efficacy* was assessed using a 13-item scale [70] (alpha=.98). An item example is "How confident are you that you could avoid having sex without a condom when you really need affection?," measured using a 5-point Likert scale ranging from 1 "not at all confident" to 5 "extremely confident."

HIV/AIDS Knowledge and Information Seeking

To assess *HIV/AIDS knowledge*, we used a 13-item scale (alpha=.71) based on the recommendations of the United Nations General Assembly Special Session on HIV/AIDS and the European Centers for Disease Control [71,72] with modifications implemented in previous survey research with European, including Romanian, GBM [5]. An example item is "If someone becomes infected with HIV it may take several weeks before it can be detected in a test." Response options ranged from 1 "I do not believe this" to 5 "I already knew this." Four items (alpha=.71) [5] were used to assess *HIV/AIDS information-seeking*, with an item example being "When was the last time you actively looked for information about HIV or STIs on the internet?" Response options ranged from 1 "within the last 12 months" to 5 "within the last 24 hours." *Recency of HIV testing* was assessed by asking "When was your last HIV test?," with response options ranging from 1 "never" to 5 "3 months or less" [73,74].

Alcohol Use

We asked participants to report the number of days in the past 3 months on which they had had at least 5 standard drinks, considered to be "heavy drinking days" [52,55,75-77]. In addition, participants were asked to report the number of alcoholic drinks consumed on a typical drinking day [78,79]. We also measured *self-efficacy to reduce alcohol use* with a 15-item scale (alpha=.97) [80]. An item example is "I would be able to resist the urge to use alcohol if I were out with friends and they kept suggesting we go somewhere to drink." Response options ranged from 1 "not at all" to 5 "completely."

Psychosocial Outcomes

Depression and anxiety symptoms were assessed with the 12-item Brief Symptom Inventory Scale [81] depression and anxiety subscales (alpha=.92 and alpha=.93, respectively). Participants were asked on a Likert scale ranging from 0 "not at all" to 4 "extremely," how intensely they experienced a variety of symptoms in the previous 7 days, such as "nervousness or shakiness inside" or "feelings of worthlessness."

Training Acceptability

We conducted 15 postintervention interviews with randomly selected participants (15/43, 35%) to gather feedback on their experiences. We randomly selected participants from 2 types of categories, those who completed all 8 intervention sessions (n=12) and those who did not (n=3), in order to obtain input on the intervention experience representative of both groups for future intervention improvements. We only selected 15 participants to interview because based on our interviews for the formative phase of this and other interventions, theme saturation is reached once 12-15 participants are interviewed. We similarly reached theme saturation after interviewing 15 participants in this study. The interview assessed (1) the feasibility of using mobile technologies to receive counseling (eg, "How did it feel for you to communicate via text with your counselor about sexual behavior and everything else you discussed?"); (2) general program perceptions (eg, "What did you like the most/least about this experience and why?"; (3) perceived impact of participation (eg, "How was your health affected by the program?"); (4) the quality of the therapeutic relationship (eg, "Did you feel you could trust your counselor and if so, why? Or why not?"; and (5) evaluation of the mobile platform (eg, "Tell us about what you would keep and/or change about the color scheme, images, text, interface, sections, and tabs." The investigators kept detailed interview notes to measure the intervention's acceptability.

Pre-Post Intervention Analyses

Descriptive statistics were obtained for demographics. The number of sexual acts with and without a condom was calculated by summing across all pertinent items. Prior to computing mean scale scores, necessary items were reverse-coded. We conducted paired-samples *t* tests to detect possible changes from the baseline to the follow-up in our outcomes. The significance level was set at $P<.05$. All analyses were conducted using SPSS 23 (IBM Corp, Armonk, NY, USA) [82].

Results

Participant Enrollment and Intervention Completion Rates

Participants accessed the eligibility screen through apps (635/1411, 45.0%), Facebook (579/1411, 41.0%), and word of mouth or in the local gay nightclub (197/1411, 13.9%). Nearly half (607/1411, 43.0%) of those who screened (n=1411) were from Bucharest. Of the 1411 individuals, 8.6% (121/1411) met the eligibility criteria. Some of the men who screened omitted reporting on sexual behavior (likely due to its stigmatized nature in Romania), which automatically eliminated them from being assessed for potential risk and therefore participation. Due to these missing data, it was not possible to determine whether or not a large proportion of the screening sample did not present risk or simply did not report it. As such, 39.9% (564/1411) of the men who screened did not meet the sex risk criteria and 22.9% (324/1411) did not report on the sex risk criteria; 16.9% (239/1411) did not use alcohol in the past 3 months; 23.9% (338/1411) did not report heavy drinking in the past 3 months; 7.9% (112/1411) were HIV-positive; and 2.9% (42/1411) did not report their status. Given that the eligibility was determined on the basis of a combination of risky sexual behavior, excessive alcohol use, and HIV status, missing data and the need for overlap across these 3 criteria yielded a relatively small final eligible sample. Of the 121 eligible men, 77.6% (94/121) provided contact information, and 52% (49/94) of these enrolled and completed the baseline survey. Of these, 90% (44/49) began the intervention and 86% (38/44) of those who began the intervention completed the minimum dose of 5 sessions. The minimum dose of 5 sessions was established based on the fact that the first 5 sessions convey the core elements of the intervention, while the remaining sessions are dedicated to behavioral skills practice, consolidating treatment gains, and establishing a plan to maintain progress. A high proportion (38/44, 86%) of our participants received the essential intervention elements. The session completion breakdown is as follows: 82% (36/44) completed all 8 sessions, 2% (1/44) completed 7 sessions, 2% (1/44) completed 6 sessions, 4% (2/44) completed 4 sessions, and 9% (4/44) completed 1 session. A total of 88% (43/49) participants completed both the baseline and follow-up questionnaires and constituted the analytic sample. The average number of sessions completed was 7.1; each session lasted 60 minutes. The average length between sessions was 1.5 weeks, with the goal of completing one session per week (inevitable scheduling conflicts arose). Of note, 6 enrolled participants were excluded from the final analyses—1 who seroconverted shortly after the baseline, 3 who were unresponsive to our contact efforts, and 2 who provided false data and were disqualified. Figure 1 provides the Consolidated Standards of Reporting Trials flow diagram of participants from screening to data analyses.

Demographic Distribution

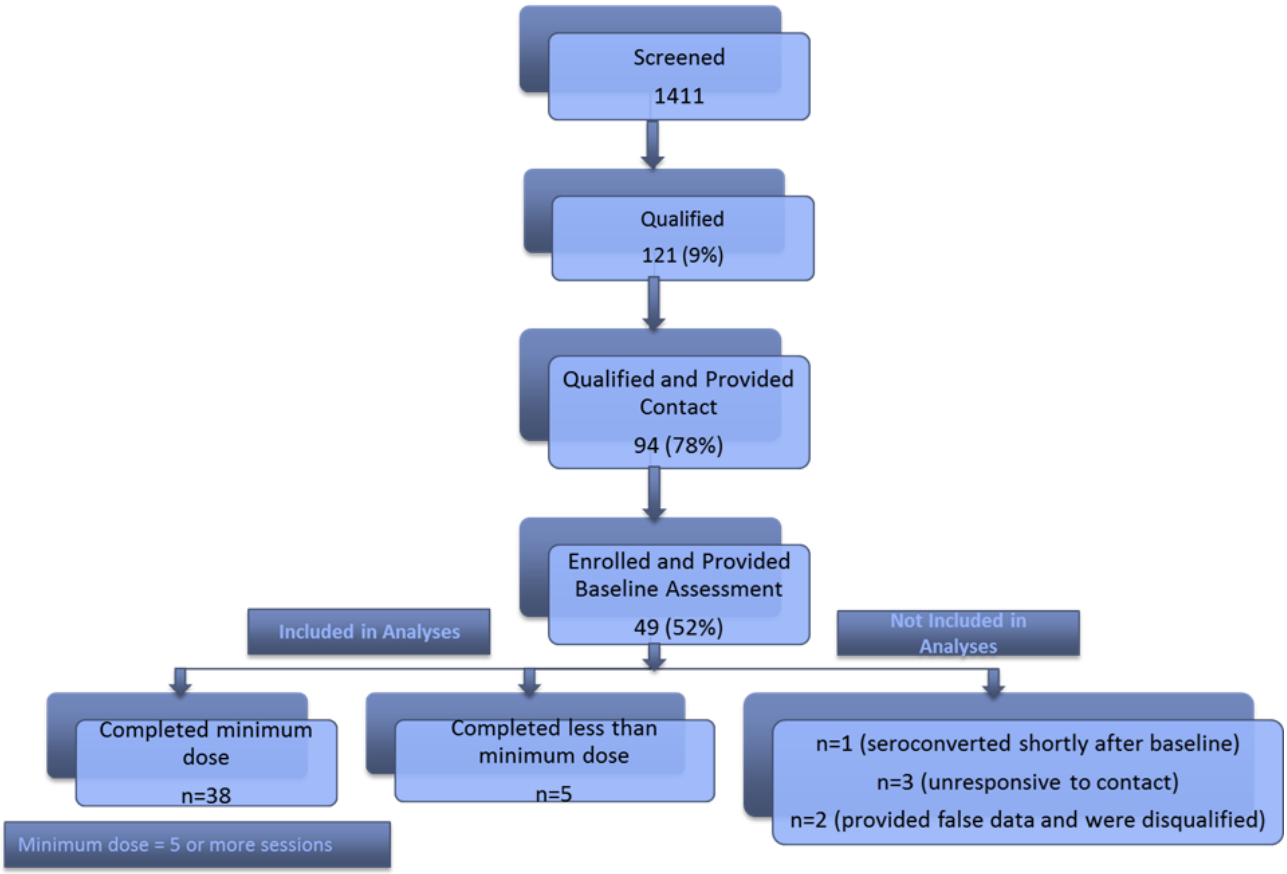
Table 1 describes the baseline sample characteristics. The majority (35/43, 81%) were aged ≤24 years (mean 23.39 years, SD 3.6; range 17-29 years) and gay (30/43, 70%). Nearly half of the participants had at least a college degree (18/43, 42%) and were working full time (20/43, 47%). Nearly one-quarter (10/43, 23%) were students. The majority (26/43, 60%) of the sample indicated being single.

Sexual Behavior and Alcohol Use

Table 2 provides comparisons of participants' sexual behavior, alcohol use, and condom use self-efficacy pre-post intervention. Condom use self-efficacy increased significantly from baseline to follow-up: mean 3.3 (SD 1.0) versus mean 4.0 (SD 1.6), $P=.01$. Participants' number of sex acts without a condom decreased (mean 15.3, SD 13.8 vs mean 14.9, SD 17.4, $P=.90$) and number of sex acts with a condom increased (mean 6.3, SD 7.3 vs mean 8.6, SD 9.8, $P=.12$) from baseline to follow-up, although not significantly.

Participants reported significant reductions in the number of heavy drinking days (mean 12.8, SD 19.3 vs mean 6.9, SD 10.2, $P=.005$) and drinks consumed on a typical drinking day (mean 2.2, SD 0.9 vs mean 1.9, SD 0.8, $P=.01$). In addition, they reported significant increases in self-efficacy to avoid alcohol consumption (mean 3.5, SD 1.7 vs mean 4.2, SD 1.4, $P=.02$).

Figure 1. Consolidated Standards of Reporting Trials flow diagram.



HIV Testing and Knowledge

Table 2 also illustrates participants' changes in HIV testing and knowledge. Participants reported more recent HIV testing at follow-up than at baseline: mean 2.8 (SD 1.5) versus mean 3.3 (SD 1.7), *P*=.05. Specifically, 8 participants reported testing in the previous 3 months at baseline, and 18 participants reported testing in the previous 3 months postintervention, therefore having tested during the intervention period (*P*=.05). Furthermore, participants reported significant increases from baseline to follow-up in their HIV/AIDS-related knowledge (mean 4.6, SD 0.3 vs mean 4.8, SD 0.2, *P*=.001) and information seeking (mean 3.8, SD 1.0 vs mean 4.4, SD 1.1, *P*=.001).

Psychosocial Outcomes

Table 2 presents changes in psychosocial outcomes. Participants reported significant decreases in both depression and anxiety symptoms from baseline to follow-up (*P*=.003 and *P*=.02, respectively).

Intervention Feasibility and Acceptability

First, we examined session completion and reasons for not completing all 8 sessions, as indicators of intervention acceptability. Minimum dose session completion was high at 86% (38/44), with an average of 7.1 completed sessions, indicating high acceptability and feasibility. In addition, we asked participants who did not complete all 8 sessions to provide an explanation for that, and the only stated reason provided was having "too busy" of a schedule, which was unrelated to the intervention.

Second, the investigators distilled the most salient themes reflective of participants' experiences in this pilot intervention. These themes are described below, and interview quotations appear in Textbox 1.

Feasibility of Using Mobile Technologies to Receive Counseling, General Program, and Counseling Relationship Perceptions

The majority completed their sessions from laptops. Communication about all sensitive topics and relevant issues via text was perceived as "natural," and miscommunication was not normative. Several participants mentioned that they preferred communication in writing over face-to-face communication because writing provided them with more opportunities to think thoroughly about what they wanted to express. The counseling relationship was characterized as honest and trustworthy, and this was most participants' first therapeutic encounter, which they characterized as highly positive.

Most Important Lessons Learned

Participants deemed the intervention as instrumental in helping them to achieve personal goals. The most important lessons learned fell under the umbrellas of HIV prevention (sexual risk and alcohol use) and psychological well-being (self-evaluation and partner dynamics).

HIV Prevention: Sexual Risk and Alcohol Use

Participants indicated having gained significant HIV knowledge, including transmission facts, the importance of alcohol reduction, condom use, and testing (Textbox 1).

Psychosocial Well-Being

Participants frequently described improvements in their self-awareness, gay identity, relationships, and well-being as they worked with their counselor to establish personal health goals and learn skills to achieve them.

Intervention Structure and the Mobile Platform

The majority of participants recommended a 10-week program, rather than an 8-week program. One participant subsequently sought counseling outside of the study. Others were less eager to do so because they preferred a texting format over a face-to-face format. In addition, participants suggested flexibility around the 60-minute session to include an additional 10-15 minutes when needed. The platform was found to be easily navigable and intuitive, although a few suggested more vivid color schemes and links to relevant videos.

Table 1. Participant demographic characteristics (n=43).

Characteristics	n (%)
Age (years)	
17-24	35 (81)
25-29	8 (19)
Sexual identity	
Gay	30 (70)
Bisexual	12 (28)
Uncertain	1 (2)
Education	
High school or less	15 (35)
College student	10 (23)
College degree	13 (30)
Graduate degree	5 (12)
Employment	
Full time	20 (47)
Part-time	1 (2)
Self-employed	4 (9)
Unemployed	3 (7)
Other (students or volunteers)	15 (35)
Relationship status	
Partnered	17 (40)
Single	26 (60)
Level of religiosity	
None to low	25 (58)
Moderate to high	18 (42)
High school location	
Small town	22 (51)
Medium to large town	21 (49)
Geographical region of residence	
South	27 (62)
North East	8 (18)
Central	6 (15)
West	2 (5)

Table 2. Reported HIV-related behavior, alcohol use, HIV/AIDS knowledge, and mental health outcomes pre-post intervention.

Outcomes[a]	Baseline, mean (SD)	Follow-up, mean (SD)	Test statistic
Sexual behavior			
Sex acts without a condom	15.3 (13.8)	14.9 (17.4)	Not significant
Sex acts with a condom	6.3 (7.3)	8.6 (9.8)	$t_{19}=-1.6$[b]
Condom use self-efficacy	3.3 (1.0)	4.0 (1.6)	$t_{42}=-2.6$[c]
HIV-related health self-efficacy patterns			
HIV/AIDS knowledge	4.6 (0.3)	4.8 (0.2)	$t_{42}=-3.4$[d]
HIV/AIDS information seeking	3.8 (1.0)	4.4 (1.1)	$t_{42}=-3.5$[d]
HIV testing recency	2.8 (1.5)	3.3 (1.7)	$t_{41}=-2.05$[e]
Alcohol use			
Number of heavy drinking days	12.8 (19.3)	6.9 (10.2)	$t_{34}=3.0$[f]
Number of drinks on a typical drinking day	2.2 (0.9)	1.9 (0.8)	$t_{39}=2.6$[c]
Self-efficacy in reducing drinking	3.5 (1.7)	4.2 (1.4)	$t_{41}=-2.4$[g]
Mental health			
Anxiety	1.4 (1.0)	1.0 (0.9)	$t_{42}=2.45$[g]
Depression	1.5 (1.1)	1.0 (0.9)	$t_{42}=3.26$[h]

[a]Not every respondent provided answers to each question and not every question was applicable to all participants (eg, main or casual partner sexual behavior); therefore, the n varies across certain outcomes.

[b]$P=.12$.

[c]$P=.01$.

[d]$P=.001$.

[e]$P=.05$.

[f]$P=.005$.

[g]$P=.02$.

[h]$P=.003$.

Textbox 1. Intervention feasibility and acceptability (n=15).

Feasibility of Text-Based Counseling

Even talking with someone about things you cannot talk to your friends or family about is a big plus, it helps you for the first time, it's private; she also highlighted the strong parts of me, every time we searched for something, and at the end of the session we had a result. [Age 22 years]

We had wonderful talks, seriously. I started the program very down and she helped. She managed to help me see the glass half full again and not to forget certain things [such as] how not to be clouded in negative emotions. She helped me clarify [the] type of things I [should] change. [Age 27 years]

I thought that I have a friend I was talking to, even though I didn't know her…This was the first time for me when I talked with a stranger very openly. [Age 20 years]

It felt safe in the program because I could be honest. [Age 20 years]

The counselor was very open-minded, nonjudgmental. It was cathartic…We had a great relationship…Excellent communication with her. [Age 27 years]

I felt free to talk, I was home and not in a strange office. It was easy, no misunderstandings. [Age 20 years]

Most Important Lessons Learned

1. HIV Prevention: Sexual Risk and Alcohol Use

[I learned] about how to prevent the transmission of HIV or how responsible I am about it. [Age 20 years]

[We] talked about new partner situations and [when] to ask for condoms. [Age 24 years]

She made me realize that I don't have to seek alcohol when I feel bad, alcohol was not a solution;…even talking about it, it makes you think about it, if you don't, you don't realize you're drinking like that. [Age 22 years]

She helped me explore the cause of my drinking, [for example], self-esteem, why do I drink? [Age 27 years]

I don't use a lot [of alcohol], but through my conversations I realized that sometimes I can exaggerate and this is not a good thing. [She] helped me to become more aware of my alcohol use. [Age 20 years]

[We discussed] alcohol and safe sex, talked about many things that helped and I realized [my] strong points that can help in difficult situations…[I] found my strengths and the program helped me understand myself better. [Age 24 years]

The most important thing that I've learned is about the importance of protected sex. Just talking has made me appreciate how important it is to have a rubber on. [Age 27 years]

Of course it helped to learn more about STIs.…The primary subject [was] how important it is to protect yourself in any situation and about health to be prudent about partners you may have and alcohol use. [Age 20 years]

2. Psychosocial Well-Being

Most important thing I learned was to be more confident and feel better about myself. [Age 24]

I know myself better compared to before;…I can be more calm or I [can] see things from all perspectives. [Age 20 years]

The program boosted my confidence [and] after this [study], I came out to my sister, cousin, two coworkers, friends. [Age 25 years]

We talked about being gay…, also related to finding a job. We sometimes would stray from talking about alcohol and sex to talk about problems at home, having too much sex. [Age 27 years]

It was quite interesting, a pleasant experience. I had more to gain than just talking about HIV and alcohol. I started to realize more things about myself: I don't have much trust in myself, but talking with someone and listening to myself I started to think that I can trust myself… I had a different state of mind after eight sessions, I felt better…to be more prudent and more careful with myself in every way. [Age 20 years]

Discussion

Principal Findings

This study adapts and preliminary tests an HIV-prevention intervention for YGBM [52] in Romania, a country with a high degree of homophobic stigma, increasing HIV prevalence among GBM, and scarce GBM-affirmative resources [3]. After adapting a US-based intervention in close consultation with Romanian key informants, we tested it for preliminary efficacy with 43 at-risk YGBM. Comparison of outcomes before and after participation in the 8 MI-based and CBST-based mobile counseling sessions indicated improvements in sexual (eg, HIV knowledge and testing), behavioral (eg, alcohol use and alcohol and condom use self-efficacy), and mental (eg, depression and anxiety symptoms) health. This intervention shows potential promise for improving the full spectrum of psychosocial health risks that disproportionately affect this vulnerable and underserved group.

Although trends toward reduced condomless sex and increased condom use were not significant in our relatively small pilot sample, participants evidenced markedly increased condom use self-efficacy postintervention. As self-efficacy is a precursor to behavioral change [83], longer follow-ups with larger samples might reveal that this intervention can impact condom use, as mediated by improvements in self-efficacy. Furthermore, we found significant increases in HIV/AIDS-related knowledge, information seeking, and testing, suggesting that DMDN can improve HIV transmission risk prevention.

Both alcohol consumption and self-efficacy for reducing it evidenced marked changes in the intended direction. In addition, participants benefited from markedly improved mental health (eg, reduced depression and anxiety). These gains are salient in a country where alcohol consumption is among the highest in Europe [84] and where LGBT individuals face pervasive stigma deleterious to mental health [3]. Furthermore, alcohol use and mental health difficulties are associated with sexual risk, forming a syndemic (ie, synergistic epidemic) that threatens the health of YGBM [23,55,64,65,85].

The intervention was found to be highly acceptable to the Romanian YGBM community. Postintervention interviews with over one-third of the sample pointed to participants' newfound sense of self-worth and openness toward future counseling. For most participants, this intervention provided a first outlet for exploring their gay or bisexual identity, articulating and setting sexual health goals, and learning to communicate with partners about risk. In addition, participants reported increased awareness of HIV risk and preventive behaviors (eg, testing). Thus, the intervention was embraced as a novel utility to YGBM health.

Awaiting this intervention's efficacy testing via a large randomized controlled trial, it is worth reflecting on its potential for sustainability and dissemination, given that it entails 8 60-minute sessions delivered by trained counselors. The very high-risk nature of that target population for which it is designed—at risk of HIV infection, alcohol abuse, and depression—necessitates a commensurately intense intervention. Whereas a briefer intervention might be warranted for risk reduction in the general population, it is likely unsuitable for syndemically affected sexual minority groups, and 8 sessions fall within the standard range for interventions targeting multiple outcomes in high-risk GBM [43]. Furthermore, structural barriers to implementing a relatively intensive intervention are modifiable to maximize dissemination and sustainability potential. Specifically, in related research [86], we recently trained 180 Romanian mental health providers in GBM-affirmative treatment and have created a mobile platform to provide them with ongoing LGBT-affirmative supervision so that our DMDN intervention, if efficacious pending rigorous testing, will persist after the study concludes. This approach provides a blueprint for sustainability and longer-lasting impact. Furthermore, hybrid individual-structural interventions may be applied concomitantly in similar settings to ameliorate the effects of stigma on individual health, and target it structurally at its source. Finally, similar interventions have included both professional and peer counselors, and regardless of the modality of intervention delivery, it is essential that adequate training be provided.

Limitations and Future Directions

Our findings should be considered alongside several limitations, which point to future research directions. First, although >4000 individuals screened for study eligibility, only 8.6% (121/1411) qualified, of whom 77.7% (94/121) were willing to provide contact information. This is unsurprising given that communication about sexual behavior remains taboo and unsystematically addressed in Romania [87,88], which is further complicated by asking men to communicate about same-sex sexual behavior, another highly stigmatized topic. At this early stage of sexuality-related research among Romanian YGBM, hesitance to provide information to researchers is to be expected. While recruitment was effortless, future recruitment efforts might consider increasing YGBM's comfort with providing accurate personal information, such as further emphasizing confidentiality. Second, this study did not include a control group, which should be considered for a future randomized controlled trial to validate the intervention's efficacy in changing sexual, behavioral, and mental health outcomes. Third, a larger sample will provide sufficient power for determining whether the nonsignificant trends we recorded in sexual risk reduction are statistically and clinically meaningful. Fourth, a longer-term follow-up (eg, up to 24 months) would permit measuring the durability of intervention benefits and the potential impact of booster sessions. Fifth, future studies might consider broader geographical representation of YGBM using social media recruitment outreach campaigns. Expanded outreach beyond major cities and surrounding areas will increase the representation of YGBM who are isolated and outside the perimeter of preventive guidance, and who would likely benefit the most from this intervention. Finally, supplementing the evaluation interviews with a measure of intervention acceptability as a rating of its usefulness, ease of use, likelihood of future use, or recommendation to peers in the future would have bolstered our evaluation of this intervention. The inclusion of such scales in future intervention evaluations is warranted.

Conclusions

This first step in addressing Romanian YGBM's sexual, behavioral, and mental health through mHealth interventions lays the groundwork for larger trials capable of establishing the intervention's efficacy before broad implementation. Awaiting proof of efficacy, this intervention could eventually be adopted by public health care entities and LGBT-affirmative practitioners looking to reach broader segments of the population than possible through face-to-face means. Furthermore, the DMDN intervention may lend itself to wide dissemination to GBM without access to conventional sources of health support.

Acknowledgments

We gratefully acknowledge the contributions of Ms Patricia Mihaescu and Ms Luiza Rotaru, staff at Population Services International, Bucharest, our study site and pioneers in HIV prevention among GBM communities nationally; Mr Florin Buhuceanu, executive director of ACCEPT Association, Romania's largest LGBT rights association; Dr Mardarescu and Dr Streinu-Cercel of the National Institute of Infectious Diseases "Prof. Dr. Matei Bal "; and Dr Maria Georgescu (executive director) and Ms Nicoleta Dascalu, from the Romanian Association against AIDS, a comprehensive health care organization for marginalized

groups. We also thank our participants who gave their time to this study, and Dr Laura Povlich, our project official at the Fogarty International Center, for her support.

This project was funded by R21TW009925 from the Fogarty International Center of the National Institutes of Health (CLW and JEP, principal investigators).

Conflicts of Interest

None declared.

References

1. Folch C, Ferrer L, Montoliu A, Munoz R, Stanekova D, Smolen M, et al. SIALON II. 2016 Jul 01. SIALON II City Profiles: Recommendations on HIV prevention gaps and opportunities for health promotion among men who have sex with men URL: http://www.sialon.eu/data2/file/133_Sialon [accessed 2018-09-13] [WebCite Cache ID 72PYKdXYS]
2. Ferrer L, Furegato M, Foschia J, Folch C, González V, Ramarli D, et al. Undiagnosed HIV infection in a population of MSM from six European cities: results from the Sialon project. Eur J Public Health 2015 Jun;25(3):494-500. [doi: 10.1093/eurpub/cku139] [Medline: 25161202]
3. Moraru A. Social perception of homosexuality in Romania. Procedia-Social and Behavioral Sciences 2010;2(2):45-49. [doi: 10.1016/j.sbspro.2010.07.048]
4. CM. AIDS in Romania: When ignorance is lethal. The Economist; 2012. URL: https://www.economist.com/blogs/easternapproaches/2012/04/aids-romania [WebCite Cache ID 6uVGVyzw7]
5. EMIS Network. EMIS 2010: The European men-who-have-sex-with-men internet survey: Findings from 38 countries. European Centre for Disease Prevention and Control; 2013. URL: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/EMIS-2010-european-men-who-have-sex-with-men-survey.pdf [WebCite Cache ID 72Pb49Smy]
6. UNAIDS. 2012. Romania: Country progress report on AIDS reporting period January 2010 - December 2011 URL: http://www.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2014countries [WebCite Cache ID 72PoBZKLG]
7. Likatavicius G, Klavs I, Devaux I, Alix J, Nardone A. An increase in newly diagnosed HIV cases reported among men who have sex with men in Europe, 2000-6: implications for a European public health strategy. Sex Transm Infect 2008 Nov;84(6):499-505. [doi: 10.1136/sti.2008.031419] [Medline: 19028955]
8. European Union Agency for Fundamental Rights. Homophobia, transphobia and discrimination on grounds of sexual orientation and gender identity in the EU Member States. Summary of findings, trends, challenges and promising practices. In: European Union Agency for Fundamental Rights. Wein, Austria: European Union Agency for Fundamental Rights; Jun 01, 2011.
9. Beyrer C, Sullivan PS, Sanchez J, Dowdy D, Altman D, Trapence G, et al. A call to action for comprehensive HIV services for men who have sex with men. Lancet 2012 Jul 28;380(9839):424-438 [FREE Full text] [doi: 10.1016/S0140-6736(12)61022-8] [Medline: 22819663]
10. Beyrer C, Baral SD. The case of gay, bisexual and other men who have sex with men (MSM), working paper for the Third Meeting of the Technical Advisory Group of the Global Commission on HIV and the Law. Global Commission on HIV and the Law; 2011 Jul 07. URL: https://hivlawcommission.org/wp-content/uploads/2017/06/MSM-HIV-and-the-Law-The-Case-of-Gay-Bisexual-and-other-men-who-have-sex-with-men.pdf [WebCite Cache ID 6uVHLqiN5]
11. Amon JJ. The political epidemiology of HIV. J Int AIDS Soc 2014;17:19327 [FREE Full text] [Medline: 25059653]
12. Quinn S. Accessing health: the context and the challenges for LGBT people in Central and Eastern Europe. ILGA Europe; 2006. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.507.6786&rep=rep1&type=pdf [WebCite Cache ID 6uVHXSJFM]
13. International Lesbian, Gay, Bisexual, Trans, Intersex Association Europe. 2015. Rainbow Europe Map and Country Index URL: https://www.ilga-europe.org/sites/default/files/Attachments/side_a_rainbow_europe_map_2015_a3_no_crops.pdf [WebCite Cache ID 72VLTmVJy]
14. Gios L, Mirandola M, Toskin I, Marcus U, Dudareva-Vizule S, Sherriff N, et al. Bio-behavioural HIV and STI surveillance among men who have sex with men in Europe: the Sialon II protocols. BMC Public Health 2016;BMC Public Health(16):212 [FREE Full text]
15. Pachankis JE, Hatzenbuehler ML, Hickson F, Weatherburn P, Berg RC, Marcus U, et al. Hidden from health: structural stigma, sexual orientation concealment, and HIV across 38 countries in the European MSM Internet Survey. AIDS 2015 Jun 19;29(10):1239-1246. [doi: 10.1097/QAD.0000000000000724] [Medline: 26035323]
16. Mahajan AP, Sayles JN, Patel VA, Remien RH, Sawires SR, Ortiz DJ, et al. Stigma in the HIV/AIDS epidemic: a review of the literature and recommendations for the way forward. AIDS 2008 Aug;22 Suppl 2:S67-S79 [FREE Full text] [doi: 10.1097/01.aids.0000327438.13291.62] [Medline: 18641472]
17. Millett GA, Peterson JL, Wolitski RJ, Stall R. Greater risk for HIV infection of black men who have sex with men: a critical literature review. Am J Public Health 2006 Jun;96(6):1007-1019. [doi: 10.2105/AJPH.2005.066720] [Medline: 16670223]

18. Millett GA, Peterson JL, Flores SA, Hart TA, Jeffries WL, Wilson PA, et al. Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis. Lancet 2012 Jul 28;380(9839):341-348. [doi: 10.1016/S0140-6736(12)60899-X] [Medline: 22819656]
19. Jeffries WL, Marks G, Lauby J, Murrill CS, Millett GA. Homophobia is associated with sexual behavior that increases risk of acquiring and transmitting HIV infection among black men who have sex with men. AIDS Behav 2013 May;17(4):1442-1453. [doi: 10.1007/s10461-012-0189-y] [Medline: 22569942]
20. Oster AM, Wiegand RE, Sionean C, Miles IJ, Thomas PE, Melendez-Morales L, et al. Understanding disparities in HIV infection between black and white MSM in the United States. AIDS 2011 May 15;25(8):1103-1112. [doi: 10.1097/QAD.0b013e3283471efa] [Medline: 21505305]
21. Lelutiu-Weinberger C, Pachankis JE, Golub SA, Walker JJ, Bamonte AJ, Parsons JT. Age cohort differences in the effects of gay-related stigma, anxiety and identification with the gay community on sexual risk and substance use. AIDS Behav 2013 Jan;17(1):340-349 [FREE Full text] [doi: 10.1007/s10461-011-0070-4] [Medline: 22038078]
22. Stall R, Mills TC, Williamson J, Hart T, Greenwood G, Paul J, et al. Association of co-occurring psychosocial health problems and increased vulnerability to HIV/AIDS among urban men who have sex with men. Am J Public Health 2003 Jun;93(6):939-942. [Medline: 12773359]
23. Stall R, Paul JP, Greenwood G, Pollack LM, Bein E, Crosby GM, et al. Alcohol use, drug use and alcohol-related problems among men who have sex with men: the Urban Men's Health Study. Addiction 2001 Nov;96(11):1589-1601. [doi: 10.1080/09652140120080723] [Medline: 11784456]
24. Mustanski BS, Newcomb ME, Du Bois SN, Garcia SC, Grov C. HIV in young men who have sex with men: a review of epidemiology, risk and protective factors, and interventions. J Sex Res 2011 Mar;48(2-3):218-253 [FREE Full text] [doi: 10.1080/00224499.2011.558645] [Medline: 21409715]
25. Mustanski B, Garofalo R, Herrick A, Donenberg G. Psychosocial health problems increase risk for HIV among urban young men who have sex with men: preliminary evidence of a syndemic in need of attention. Ann Behav Med 2007 Aug;34(1):37-45 [FREE Full text] [doi: 10.1080/08836610701495268] [Medline: 17688395]
26. Pachankis JE. A transdiagnostic minority stress treatment approach for gay and bisexual men's syndemic health conditions. Arch Sex Behav 2015 Oct;44(7):1843-1860 [FREE Full text] [doi: 10.1007/s10508-015-0480-x] [Medline: 26123065]
27. King M, McKeown E, Warner J, Ramsay A, Johnson K, Cort C, et al. Mental health and quality of life of gay men and lesbians in England and Wales: controlled, cross-sectional study. Br J Psychiatry 2003 Dec;183:552-558 [FREE Full text] [Medline: 14645028]
28. Hatzenbuehler ML, Pachankis JE, Wolff J. Religious climate and health risk behaviors in sexual minority youths: a population-based study. Am J Public Health 2012 Apr;102(4):657-663. [doi: 10.2105/AJPH.2011.300517] [Medline: 22397347]
29. Hatzenbuehler ML, Nolen-Hoeksema S, Erickson SJ. Minority stress predictors of HIV risk behavior, substance use, and depressive symptoms: results from a prospective study of bereaved gay men. Health Psychol 2008 Jul;27(4):455-462. [doi: 10.1037/0278-6133.27.4.455] [Medline: 18643003]
30. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
31. Meyer IH, Schwartz S, Frost DM. Social patterning of stress and coping: does disadvantaged social statuses confer more stress and fewer coping resources? Soc Sci Med 2008 Aug;67(3):368-379 [FREE Full text] [doi: 10.1016/j.socscimed.2008.03.012] [Medline: 18433961]
32. Hatzenbuehler ML, Wieringa NF, Keyes KM. Community-level determinants of tobacco use disparities in lesbian, gay, and bisexual youth: results from a population-based study. Arch Pediatr Adolesc Med 2011 Jun;165(6):527-532. [doi: 10.1001/archpediatrics.2011.64] [Medline: 21646585]
33. Hatzenbuehler ML, Corbin WR, Fromme K. Trajectories and determinants of alcohol use among LGB young adults and their heterosexual peers: results from a prospective study. Dev Psychol 2008 Jan;44(1):81-90 [FREE Full text] [doi: 10.1037/0012-1649.44.1.81] [Medline: 18194007]
34. Cochran SD, Mays VM, Sullivan JG. Prevalence of mental disorders, psychological distress, and mental health services use among lesbian, gay, and bisexual adults in the United States. J Consult Clin Psychol 2003 Feb;71(1):53-61 [FREE Full text] [Medline: 12602425]
35. Bränström R, Pachankis JE. Validating the syndemic threat surrounding sexual minority men?s health in a population-based study with national registry linkage and a heterosexual comparison. JAIDS 2018;78(4):376-382. [doi: 10.1097/QAI.0000000000001697]
36. Nachescu V. Hierarchies of difference: national identity, gay and lesbian rights, and the church in postcommunist Romania. In: Coleman E, Sandfort T, editors. Sexuality and gender in postcommunist Eastern Europe and Russia. New York, NY: Haworth Press; 2005:57-78.
37. Popova S, Rehm J, Patra J, Zatonski W. Comparing alcohol consumption in central and eastern Europe to other European countries. Alcohol Alcohol 2007;42(5):465-473. [doi: 10.1093/alcalc/agl124] [Medline: 17287207]

38. Bränström R, Hatzenbuehler ML, Pachankis JE, Link BG. Sexual Orientation Disparities in Preventable Disease: A Fundamental Cause Perspective. Am J Public Health 2016 Jun;106(6):1109-1115. [doi: 10.2105/AJPH.2016.303051] [Medline: 26985608]
39. Bränström R, Pachankis J, Hatzenhuehler M. Sexual orientation based differences in mental health morbidity: Age effects in a population-based longitudinal study in Sweden. Eur J Pub Health 2015;25(supplement 3):220. [doi: 10.1093/eurpub/ckv173.014]
40. Bränström R, van der Star A. All inclusive Public Health--what about LGBT populations? Eur J Public Health 2013 Jun;23(3):353-354. [doi: 10.1093/eurpub/ckt054] [Medline: 23704057]
41. Bränström R, Hatzenhuehler M, Pachankis J. Sexual orientation disparities in mental health: The moderating and mediating role of sociodemographic and minority stress factors. Eur J Pub Health 2016 Nov 02;26(suppl 1). [doi: 10.1093/eurpub/ckw170.047]
42. Hatzenbuehler ML, Bränström R, Pachankis JE. Societal-level explanations for reductions in sexual orientation mental health disparities: Results from a ten-year, population-based study in Sweden. Stigma and Health 2018 Feb;3(1):16-26. [doi: 10.1037/sah0000066]
43. Lelutiu-Weinberger C, Rendina H, Mirandola M, Gios L, Folch C, Rafila A, et al. The role of gay-related stigma in HIV-risk behaviour among sexual minority men in Europe. AIDS Behav 2018. [doi: 10.1007/s10461-018-2306-z] [Medline: 30302655]
44. Pachankis JE, Hatzenbuehler ML, Rendina HJ, Safren SA, Parsons JT. LGB-affirmative cognitive-behavioral therapy for young adult gay and bisexual men: A randomized controlled trial of a transdiagnostic minority stress approach. J Consult Clin Psychol 2015 Oct;83(5):875-889 [FREE Full text] [doi: 10.1037/ccp0000037] [Medline: 26147563]
45. Bozicevic I, Voncina L, Zigrovic L, Munz M, Lazarus JV. HIV epidemics among men who have sex with men in central and eastern Europe. Sex Transm Infect 2009 Sep;85(5):336-342. [doi: 10.1136/sti.2008.035436] [Medline: 19773456]
46. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011 Feb;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]
47. Ybarra ML, Bull SS. Current trends in Internet- and cell phone-based HIV prevention and intervention programs. Curr HIV/AIDS Rep 2007 Dec;4(4):201-207. [Medline: 18366952]
48. Pachankis JE, Lelutiu-Weinberger C, Golub SA, Parsons JT. Developing an online health intervention for young gay and bisexual men. AIDS Behav 2013 Nov;17(9):2986-2998 [FREE Full text] [doi: 10.1007/s10461-013-0499-8] [Medline: 23673791]
49. Ipsos MediaCT. Our mobile planet: Romania. The Media, Content and Technology Research Specialists; 2013 May 20. URL: http://services.google.com/fh/files/misc/omp-2013-ro-en.pdf [WebCite Cache ID 72UYoIaOH]
50. Virtopeanu D. Mobile Romania Overview. In: Breeze Mobile. 2014 Oct 27 Presented at: Breeze Mobile Reporting; October 27, 2014; Bucharest, Romania p. 1-15 URL: http://storage0.dms.mpinteractiv.ro/media/401/1881/22647/13479113/205/21-mobilerooverview-virtopeanu.pdf
51. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
52. Lelutiu-Weinberger C, Pachankis JE, Gamarel KE, Surace A, Golub SA, Parsons JT. Feasibility, Acceptability, and Preliminary Efficacy of a Live-Chat Social Media Intervention to Reduce HIV Risk Among Young Men Who Have Sex With Men. AIDS Behav 2015 Jul;19(7):1214-1227 [FREE Full text] [doi: 10.1007/s10461-014-0911-z] [Medline: 25256808]
53. Nanín JE, Parsons JT. Club Drug Use and Risky Sex Among Gay and Bisexual Men in New York City. J Gay Lesbian Psycho 2006 Aug 14;10(3-4):111-122. [doi: 10.1300/J236v10n03_10]
54. Rotheram-Borus MJ, Rosario M, Meyer-Bahlburg HFL, Koopman C, Dopkins SC, Davies M. Sexual and substance use acts of gay and bisexual male adolescents in New York City. J Sex Res 1994;31(1):47-57. [doi: 10.1080/00224499409551729]
55. Wells BE, Kelly BC, Golub SA, Grov C, Parsons JT. Patterns of alcohol consumption and sexual behavior among young adults in nightclubs. Am J Drug Alcohol Abuse 2010 Jan;36(1):39-45 [FREE Full text] [doi: 10.3109/00952990903544836] [Medline: 20141395]
56. Kelly BC, Parsons JT, Wells BE. Prevalence and predictors of club drug use among club-going young adults in New York city. J Urban Health 2006 Sep;83(5):884-895 [FREE Full text] [doi: 10.1007/s11524-006-9057-2] [Medline: 16937088]
57. Parsons JT, Halkitis PN, Bimbi DS. Club Drug Use Among Young Adults Frequenting Dance Clubs and Other Social Venues in New York City. Journal of Child & Adolescent Substance Abuse 2006 Feb 27;15(3):1-14. [doi: 10.1300/J029v15n03_01]
58. Longfield K, Astatke H, Smith R, McPeak G, Ayers J. Men who have sex with men in Southeastern Europe: Underground and at increased risk for HIV/STIs. Cult Health Sex 2007;9(5):473-487. [doi: 10.1080/13691050601161864] [Medline: 17687673]
59. Peterson JL, Coates TJ, Catania JA, Middleton L, Hilliard B, Hearst N. High-risk sexual behavior and condom use among gay and bisexual African-American men. Am J Public Health 1992 Nov;82(11):1490-1494. [Medline: 1443298]
60. Parsons JT, Lelutiu-Weinberger C, Botsko M, Golub SA. Predictors of day-level sexual risk for young gay and bisexual men. AIDS Behav 2013 May;17(4):1465-1477 [FREE Full text] [doi: 10.1007/s10461-012-0206-1] [Medline: 22614745]

61. Lyles CM, Kay LS, Crepaz N, Herbst JH, Passin WF, Kim AS, et al. Best-evidence interventions: findings from a systematic review of HIV behavioral interventions for US populations at high risk, 2000-2004. Am J Public Health 2007 Jan;97(1):133-143. [doi: 10.2105/AJPH.2005.076182] [Medline: 17138920]
62. Herbst JH, Sherba RT, Crepaz N, Deluca JB, Zohrabyan L, Stall RD, HIV/AIDS Prevention Research Synthesis Team. A meta-analytic review of HIV behavioral interventions for reducing sexual risk behavior of men who have sex with men. J Acquir Immune Defic Syndr 2005 Jun 01;39(2):228-241. [Medline: 15905741]
63. Chesney MA, Barrett DC, Stall R. Histories of substance use and risk behavior: precursors to HIV seroconversion in homosexual men. Am J Public Health 1998 Jan;88(1):113-116. [Medline: 9584015]
64. Cooper ML, Orcutt HK. Alcohol use, condom use and partner type among heterosexual adolescents and young adults. J Stud Alcohol 2000 May;61(3):413-419. [Medline: 10807212]
65. Vanable PA, McKirnan DJ, Buchbinder SP, Bartholow BN, Douglas JM, Judson FN, et al. Alcohol use and high-risk sexual behavior among men who have sex with men: the effects of consumption level and partner type. Health Psychol 2004 Sep;23(5):525-532. [doi: 10.1037/0278-6133.23.5.525] [Medline: 15367072]
66. Miller W, Rollnick S. Motivational Interviewing: Preparing People for Change. New York, NY: Guilford Press; 2002.
67. Rollnick S, Miller W. What is motivational interviewing? Behav Cog Psychother 1995;23:325-334 [FREE Full text]
68. Hoffman SG. An Introduction to Modern CBT: Psychological Solutions to Mental Health Problems. Malden, MA: Wiley-Blackwell; 2011.
69. Goldfried, MR, Davison, GC. Clinical Behavior Therapy, Expanded. Oxford, England: John Wiley & Sons, Inc; 1994.
70. Parsons JT, Halkitis PN, Bimbi D, Borkowski T. Perceptions of the benefits and costs associated with condom use and unprotected sex among late adolescent college students. J Adolesc 2000 Aug;23(4):377-391. [doi: 10.1006/jado.2000.0326] [Medline: 10936012]
71. Joint United Nations Programme on HIV/AIDS. 2002. Monitoring the declaration of commitment on HIV/AIDS: guidelines on construction of core indicators URL: http://www.who.int/hiv/strategic/me/en/isbn9291732389.pdf [WebCite Cache ID 72VOUCOKH]
72. European Center for Disease Prevention Control. 2009. Mapping of HIV/STI behavioural surveillance in Europe URL: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0909_TER_Mapping_of_HIV_STI_Behavioural_Surveillance_in_Europe.pdf [WebCite Cache ID 72VOFvEhJ]
73. Centers for Disease Control Prevention. 2014. Gay and bisexual men's health URL: https://www.cdc.gov/msmhealth/HIV.htm [WebCite Cache ID 6uVR8PScR]
74. European Centers for Disease Control Prevention. 2010. HIV testing: Increasing uptake and effectiveness in the European Union URL: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/101129_GUI_HIV_testing.pdf [WebCite Cache ID 6uVRImRbA]
75. Parsons JT, Kutnick AH, Halkitis PN, Punzalan JC, Carbonari JP. Sexual risk behaviors and substance use among alcohol abusing HIV-positive men who have sex with men. J Psychoactive Drugs 2005 Mar;37(1):27-36 [FREE Full text] [doi: 10.1080/02791072.2005.10399746] [Medline: 15916249]
76. Parsons JT, Lelutiu-Weinberger C, Botsko M, Golub SA. Efficacy of a motivational interviewing intervention to reduce sexual risk and substance use across time in young men who have sex with men. 2012 Jun 18 Presented at: Paper presented at: International Conference on Motivational Interviewing; 2012; Venice, IT.
77. Parsons JT, Lelutiu-Weinberger C, Botsko M, Golub SA. A randomized controlled trial utilizing motivational interviewing to reduce HIV risk and drug use in young gay and bisexual men. J Consult Clin Psychol 2014 Feb;82(1):9-18 [FREE Full text] [doi: 10.1037/a0035311] [Medline: 24364800]
78. Bohn MJ, Babor TF, Kranzler HR. The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol 1995 Jul;56(4):423-432. [Medline: 7674678]
79. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
80. Sklar SM, Annis HM, Turner NE. Development and validation of the drug-taking confidence questionnaire: a measure of coping self-efficacy. Addict Behav 1997;22(5):655-670. [Medline: 9347068]
81. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. Psychol Med 1983 Aug;13(3):595-605. [Medline: 6622612]
82. George D, Mallery P. SPSS 23. In: IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference. New York, NY: Routledge; 2016.
83. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
84. Rehm J, Zatonksi W, Taylor B, Anderson P. Epidemiology and alcohol policy in Europe. Addiction 2011 Mar;106 Suppl 1:11-19. [doi: 10.1111/j.1360-0443.2010.03326.x] [Medline: 21324017]
85. Conway KP, Compton W, Stinson FS, Grant BF. Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 2006 Feb;67(2):247-257. [Medline: 16566620]

86. Lelutiu-Weinberger C, Pachankis JE. Acceptability and Preliminary Efficacy of a Lesbian, Gay, Bisexual, and Transgender-Affirmative Mental Health Practice Training in a Highly Stigmatizing National Context. LGBT Health 2017 Oct;4(5):360-370 [FREE Full text] [doi: 10.1089/lgbt.2016.0194] [Medline: 28891750]
87. Rada C. Sexual behaviour and sexual and reproductive health education: a cross-sectional study in Romania. Reprod Health 2014 Jun 23;11:48 [FREE Full text] [doi: 10.1186/1742-4755-11-48] [Medline: 24957900]
88. Jordan M, Isaev O. Transitions Online. 2010. Silence makes babies URL: http://www.tol.org/client/article/21599-silence-makes-babies.html [WebCite Cache ID 6uVRXpPu9]

Abbreviations

CBST: cognitive behavioral skills training
DMDN: Despre Mine. Despre Noi.
GBM: gay and bisexual men
LGBT: lesbian, gay, bisexual, and transgender
MI: motivational interviewing
mHealth: mobile health
YGBM: young gay and bisexual men

Edited by G Eysenbach; submitted 26.10.17; peer-reviewed by BT Tulbure, L Hightow-Weidman, R Leeman; comments to author 15.03.18; revised version received 30.04.18; accepted 25.06.18; published 14.11.18.

Please cite as:
Lelu iu-Weinberger C, Manu M, Ionescu F, Dogaru B, Kovacs T, Dorobăn escu C, Predescu M, Surace A, Pachankis JE
An mHealth Intervention to Improve Young Gay and Bisexual Men's Sexual, Behavioral, and Mental Health in a Structurally Stigmatizing National Context
JMIR Mhealth Uhealth 2018;6(11):e183
URL: https://mhealth.jmir.org/2018/11/e183/
doi:10.2196/mhealth.9283
PMID:30429117

Original Paper

mHealth Supportive Care Intervention for Parents of Children With Acute Lymphoblastic Leukemia: Quasi-Experimental Pre- and Postdesign Study

Jingting Wang[1], PhD; Doris Howell[2,3], PhD; Nanping Shen[4], PhD; Zhaohui Geng[1], PhD; Fulei Wu[1], PhD; Min Shen[5], BA; Xiaoyan Zhang[6], BA; Anwei Xie[7], MSc; Lin Wang[8], PhD; Changrong Yuan[9], PhD, FAAN

[1]School of Nursing, Second Military Medical University, Shanghai, China

[2]Department of Supportive Care, Princess Margaret Cancer Center, Toronto, ON, Canada

[3]Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[4]Department of Nursing, Shanghai Children's Medical Center, Shanghai, China

[5]Department of Hematology, Children's Hospital of Soochow University, Suzhou, China

[6]Department of Hematology and Oncology, Shanghai Children's Medical Center, Shanghai, China

[7]Department of Infectious Disease, Children's Hospital of Soochow University, Suzhou, China

[8]United States Census Bureau, Washington, DC, United States

[9]School of Nursing, Fudan University, Shanghai, China

Corresponding Author:
Changrong Yuan, PhD, FAAN
School of Nursing
Fudan University
Room 403, Building 8
305 Fenglin Road
Shanghai, 200032
China
Phone: 86 02164431803
Fax: 86 02164431803
Email: yuancr@fudan.edu.cn

Abstract

Background: Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy. Caring for children with ALL is challenging for parents. A mobile health (mHealth) supportive care intervention was developed to meet parents' needs.

Objective: This study aims to evaluate the potential effectiveness of this mHealth supportive care intervention on emotional distress, social support, care burden, uncertainty in illness, quality of life, and knowledge.

Methods: We conducted a quasi-experimental pre- and postdesign study from June 2015 to January 2016. In total, 101 parents were enrolled in the study, with 50 in the observation group and 51 in the intervention group. Parents in the observation group received the standard health education and were observed for 3 months. Parents in the intervention group received the mHealth supportive care intervention, in addition to the standard health education. The intervention consisted of 2 parts—an Android smartphone app "Care Assistant (CA)" and a WeChat Official Account. The CA with 8 modules (Personal Information, Treatment Tracking, Family Care, Financial and Social Assistance, Knowledge Center, Self- Assessment Questionnaires, Interactive Platform, and Reminders) was the main intervention tool, whereas the WeChat Official Account was supplementary to update information and realize interaction between parents and health care providers. Data of parents' social support, anxiety, depression, care burden, uncertainty in illness, quality of life, their existing knowledge of ALL and care, and knowledge need were collected before and after the 3-month study period in both groups. For the intervention group, parents' experience of receiving the intervention was also collected through individual interviews.

Results: Overall, 43 parents in the observation group and 49 in the intervention group completed the study. Results found that the intervention reduced parents' anxiety ($D_{int(Post-Pre)}$=−7.0 [SD 13.1], $D_{obs(Post-Pre)}$=−0.4 [SD 15.8], t_{90}=−2.200, *P*=.03) and uncertainty in illness ($D_{int(Post-Pre)}$=−25.0 [SD 8.2], $D_{obs(Post-Pre)}$=−19.8 [SD 10.1], t_{90}=−2.761, *P*=.01), improved parents' social function ($D_{int(Post-Pre)}$=9.0 [SD 32.8], $D_{obs(Post-Pre)}$=−7.5 [SD 30.3], t_{90}=2.494, *P*=.01), increased parents' knowledge of ALL and

care ($D_{int(Post-Pre)}$=28.4 [SD 12.4], $D_{obs(Post-Pre)}$=17.2 [SD 11.9], t90=4.407, *P*<.001), and decreased their need for knowledge ($D_{int(Post-Pre)}$=−9.9 [SD 11.6], $D_{obs(Post-Pre)}$=−1.9 [SD 6.4], t_{90}=−4.112, *P*<.001). Qualitative results showed that parents were satisfied with the intervention and their role in the caregiving process.

Conclusions: The mHealth intervention in supporting parents of children with ALL is effective. This study is informative for other future studies on providing mHealth supportive care for parents of children with cancer.

(JMIR Mhealth Uhealth 2018;6(11):e195) doi:10.2196/mhealth.9981

KEYWORDS

acute lymphoblastic leukemia; effectiveness; mHealth; mobile phone; parent; supportive care

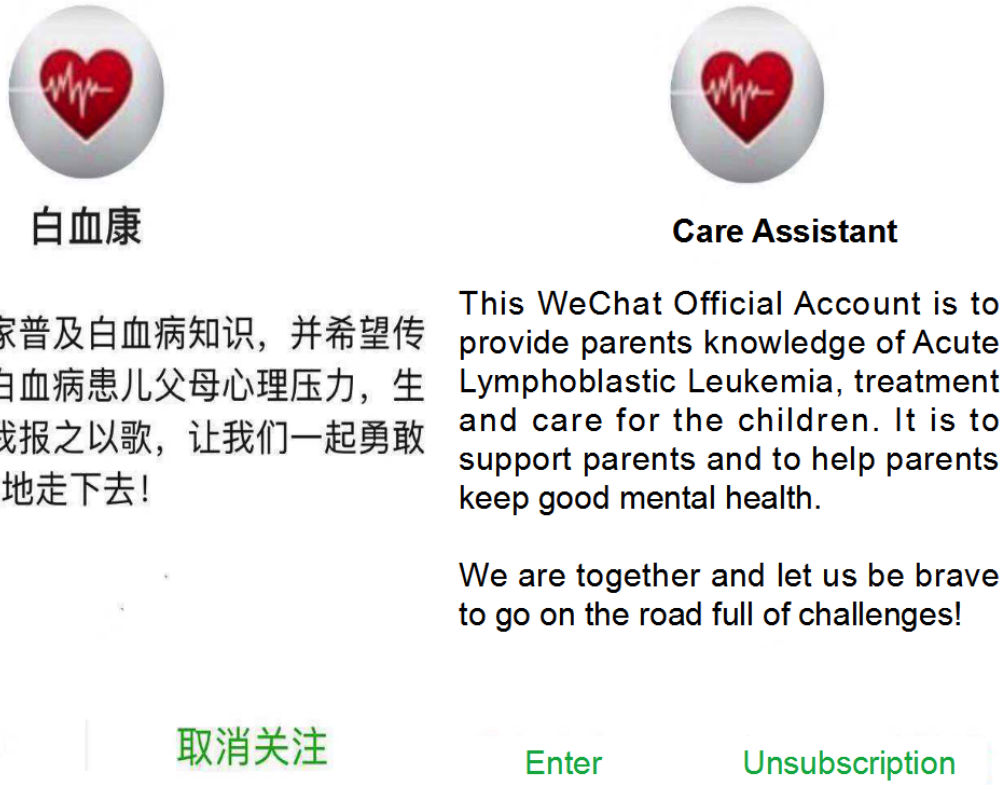
Introduction

Cancer is the second leading cause of pediatric death in developed countries [1]. Acute lymphoblastic leukemia (ALL) accounts for 26.8% of all pediatric cancer cases and is the most common childhood malignancy among children younger than 15 years [2]. The peak incidence of childhood ALL occurs in children aged 2-4 years [3]. The 5-year survival rate in childhood ALL is around 85% in developed countries [3] and slightly lower at 80% in China [4]. Regardless of a relatively good prognosis, the diagnosis of childhood ALL is still a life-altering event for parents. Parents are often challenged with taking on the tasks of caring for their children with symptom and emotion management when inadequately equipped with knowledge of the disease and caregiving. In the meantime, parents themselves are in need of supportive care in health, emotion, finance, and social support.

Supportive care was first described in 1994 as "the provision of necessary services as defined by those living with or affected by cancer to meet their physical, psychosocial, informational and spiritual needs during the prediagnostic, diagnostic, treatment, and follow-up phases, while encompassing issues of survivorship, palliation, and bereavement" [5]. In 2000, Cancer Care Ontario in Canada developed the Supportive Care Needs Framework and outlined 7 domains of need, including practical, spiritual, social, psychological, informational, emotional, and physical needs [6]. The Supportive Care Needs Framework was developed to help guide health professionals to meet the supportive care needs of patients with cancer along the disease continuum [6]. A recent study found that parents of children with cancer have the needs for these aspects of supportive care—communication with physician, timely provision of information, and suitable and accessible psychosocial care [7].

Traditionally, support interventions for parents were delivered through face-to-face education and telephone follow-up. Two studies on traditional supportive care for Chinese parents showed effects in reducing parents' emotional distress and in improving parents' knowledge of care, satisfaction with care, and relationship with health care providers [8,9].

Figure 1. WeChat Official Account welcome page.

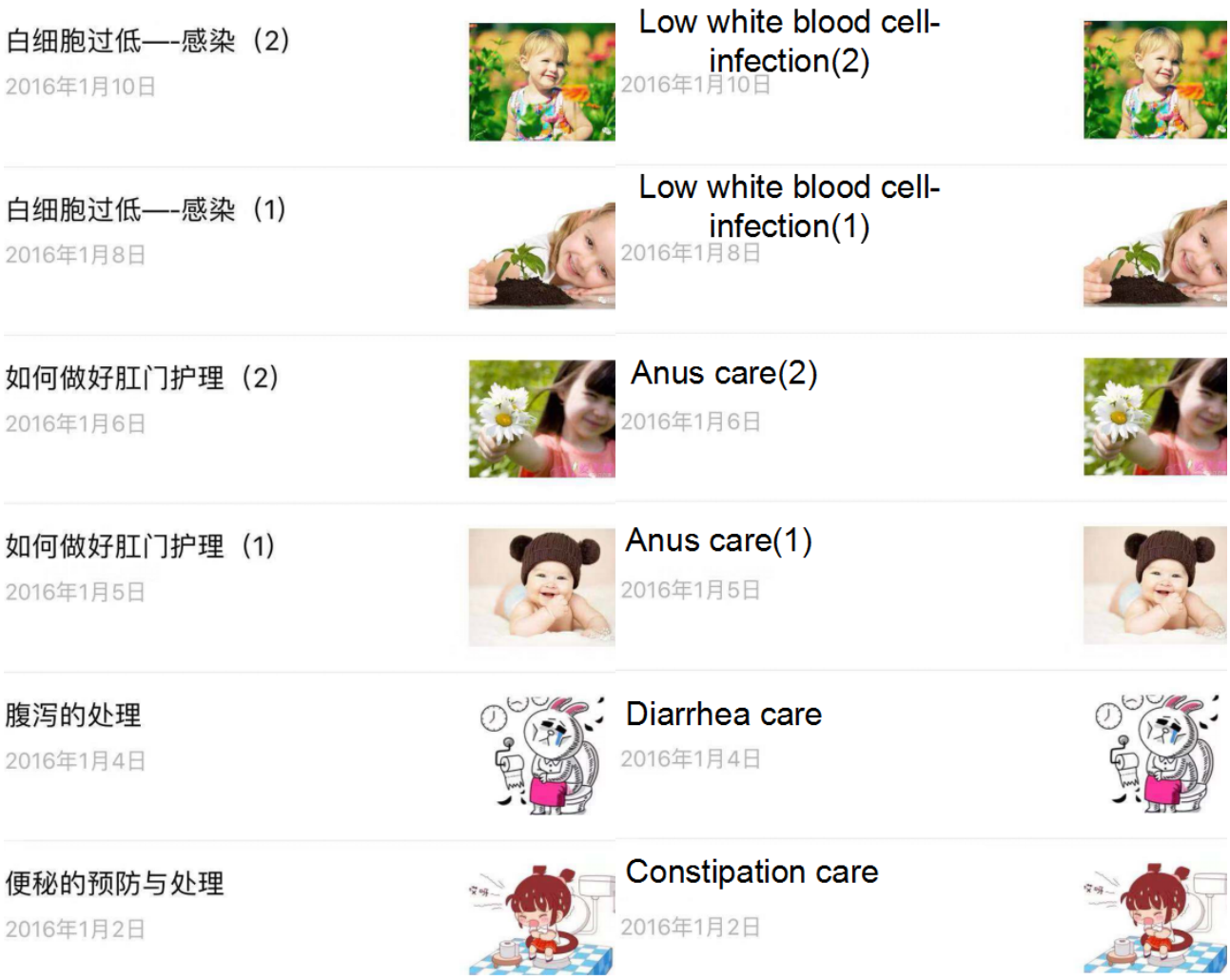


In addition, systematic reviews indicated that supportive care interventions improved cancer caregivers' psychological status, emotion, well-being, disease-related knowledge, quality of life, and coping strategy [10]. However, the traditional modes of delivering interventions have a major disadvantage of poor accessibility. For example, some parents may be less motivated to attend intervention training courses or unable to find substitute caregiver for their children to allow them to attend the training [10].

Mobile health (mHealth) is an emerging technology to make supportive care more accessible for parents of children with ALL. mHealth interventions can be used to manage side effects, improve drug adherence, provide cancer care information, plan and follow-up care, and detect and diagnose cancer [11,12]. Smartphone apps are important mHealth tools and tend to be easy to be acquired. Smartphone apps allow patients to access health information and health care services anytime and anywhere [12]. Studies have documented the feasibility and effectiveness of mobile apps in supporting adult patients with cancer and survivors in health care and clinical practice [13-16]. However, to the best of our knowledge, there has not been much use of apps in supporting parents of pediatric patients with cancer.

To meet the supportive care needs of parents of children with ALL, we designed and implemented a mHealth supportive care intervention that consisted of 2 parts—a smartphone app called "Care Assistant (CA)" that runs on the Android system with good usability [17,18] and a WeChat Official Account. WeChat is a multipurpose social media mobile app that is the most popular in China. The WeChat Official Account is a module of WeChat, which can publish papers to users who subscribe to it. In this study, the CA was the main intervention tool, whereas WeChat was supplementary. The CA has 8 modules as follows: Personal Information, Treatment Tracking, Family Care, Financial and Social Assistance, Knowledge Center, Self-Assessment Questionnaires, Interactive Platform, and Reminders [17]. The selected screenshots to visualize the app can be found in former published papers [17,18]. The WeChat Official Account was added to this intervention because of parents' needs for information across the children's illness trajectory. It is a cost-saving strategy to update information and realize interaction between parents through the addition of a WeChat Official Account. Figures 1 and 2 are selected screenshots provided to visualize the WeChat Official Account.

Figure 2. WeChat Official Account health education article example.



This study aims to evaluate the potential effectiveness of the mHealth supportive care intervention, with respect to supporting parents of children with ALL on emotional distress, social support, care burden, uncertainty, quality of life, and knowledge. In addition, this study examined parents' experience of receiving this mHealth supportive care intervention.

Methods

Participants

We recruited 101 parents of children with ALL from Shanghai Children's Medical Center and Soochow University Affiliated Children's Hospital to participate in this study. The first 50 parents were assigned to the observation group, followed by 51 parents to the intervention group 3 months later. The rationale for the assignment method is provided in the "Study Design" section.

As the CA app has its target users, the eligibility criteria for participants included the following: parents were the main caregiver of a child diagnosed with ALL at the age <15 years; the child had been diagnosed ALL within the past 30 days at the time when the parent was enrolled in the study; parents had the education of high school or higher; parents were able to communicate fluently in Mandarin Chinese; patient's average monthly family income was ≥2000 Chinese yuan (equivalent to US $300); parents were willing to initiate and continue ALL treatment for the child; and parents had an Android smartphone and a frequently used WeChat account. The exclusion criteria included the following: the child had other severe diseases or chronic diseases and the child is in a critical condition.

The study received ethical review approval from the Second Military Medical University, Shanghai Children's Medical Center, and Soochow University, China. Participants attended an information session and were informed of the aim of this study and participant's rights. All participants provided written informed consent prior to their participation in this study.

Study Design

A quasi-experiment (nonrandomized experiment) pre- and postdesign was used to evaluate the potential effectiveness of the mHealth supportive care intervention. The study was conducted from June 2015 to January 2016. To prevent parents in the observation group from getting the CA app and WeChat Official Account through interacting with parents in the intervention group and consequently resulting in contamination in data collection, a 3-month data collection of the observation group was carried out first, followed by the intervention trial. Parents in the observation group received the standard health education in the hospital and were observed for 3 months. The standard health education in hospital included regular health education sessions, which were available to all caregivers of children, health education fliers or booklets, and one-on-one education by nurses. The standard health education neither included any information about the CA app, WeChat Official Account, nor the mHealth intervention. Parents in the intervention group received the mHealth supportive care intervention, in addition to the standard health education. The intervention period lasted 3 months.

Intervention of the Mobile Health Supportive Care

Study researchers helped parents in the intervention group install the CA app and trained parents using the app. In addition, parents subscribed the WeChat Official Account used for this study. Parents used the app freely during the intervention period and also received health education papers sent through the WeChat Official Account. Furthermore, parents could communicate with our care management team with questions about caring for their children.

The care management team included 4 members—1 software engineer, 1 clinical nurse, and 2 nursing researchers. The team performed following functions: maintaining the CA app to ensure that the app functioned well without software failure; selecting and distributing weekly supportive care health education papers in 3-5 topics, based on parents' needs and clinical judgments, to parents through the WeChat Official Account, with one paper being an example of successful ALL treatment; and answering parents' questions through the WeChat Official Account.

Outcome Measures

We applied 7 measures to assess the effectiveness of the mHealth supportive care. The 7 measures are defined below.

Anxiety

Zung's Self-Rating Anxiety Scale (SAS) was used to measure anxiety. The SAS is a 20-item, 4-point scale ranging from never (1) to often (4) [19]. The total score of 1.25 is the final standard score. A higher score indicates a severer anxiety symptom. The Chinese version of the SAS is widely used in China, and its reliability has been established (Cronbach alpha=.85) [19,20].

Depressive Disorder

Zung's Self-Rating Depression Scale (SDS) was used to measure depression. The SDS is a 20-item, 4-point scale ranging from never (1) to often (4) [21]. The total score of 1.25 is the final standard score. A higher score indicates severer depressive symptoms. The Chinese version of SDS is with good reliability and widely used among Chinese populations (Cronbach alpha=.86) [19].

Perceived Social Support

The Perceived Social Support Scale (PSSS) was used to measure perceived social support [22]. The PSSS is a 12-item, 7-point scale ranging from never (1) to every day (7). A lower score indicates poorer social support. The Chinese version of the PSSS has demonstrated good reliability in prior studies (Cronbach alpha=.88) [23].

Burden of Care

The Zarit Burden Inventory (ZBI) was used to measure parents' burden of care [24]. The ZBI is a 22-item, 5-point scale ranging from never (0) to always (4). A higher score indicates higher care burden. The Chinese version of the ZBI has demonstrated good reliability (Cronbach alpha=.87) [24], and was used in caregivers of Chinese adult patients with ALL [25].

Parents' Perception of Uncertainty in Illness

The Parents' Perception of Uncertainty Scale (PPUS) was used to measure parents' uncertainty in illness [26]. The PPUS is a 28-item, 5-point scale ranging from strongly disagree (1) to strongly agree (5). A higher score indicates a stronger perception of uncertainty. The Chinese version of the PPUS is with good reliability (Cronbach alpha=.84) [26].

Quality of Life

The Medical Outcomes Study 36-item Short Form (SF-36) was used to measure the parents' quality of life [27]. The SF-36 comprises 36 items summarized into 8 dimensions—physical functioning, social functioning, pain, mental health, vitality, general health, role limitation owing to physical problems, role limitation owing to emotional problems, and a single-item subscale on health transition. The Chinese version of the SF-36 is widely used with Cronbach alpha coefficients ranging from.72 to.88, except.39 for the social functioning scale and.66 for the vitality scale [27]. In this study, we used the 8 validated subscales of the SF-36.

Parents' Existing Knowledge and Knowledge Needs

Parents' existing knowledge of ALL and care and their need for related knowledge were measured by the Knowledge Questionnaire. The questionnaire was designed by the authors and had 2 parts—existing knowledge and knowledge needs. The existing knowledge and knowledge needs subscales have the same 50 items. For the existing knowledge part, it is a 5-point scale ranging from "do not know at all" (1) to "know very well" (5). A lower score indicates less existing knowledge. For the knowledge needs part, it is a 5-point scale ranging from "do not want to know at all" (1) to "want to know very much" (5). A higher score indicates higher knowledge needs.

Data Collection

Observation Group

Data were collected at 2 time-points—at the beginning (preobservation data collection) and the end (postobservation data collection) of the 3-month study period. In the preobservation data collection, 2 sets of the data were collected as follows: (1) parents' and their child's sociodemographic characteristics pertinent to the study, collected using an in-house developed questionnaire and (2) outcome data as defined in the preceding "Outcome Measures" section. In the postobservation data collection, only the outcome data were collected in the same way as in the preobservation data collection. In both data collections, participants answered the questionnaires according to their experience in the previous week. No exit interviews were conducted among participants in the observation group because interviews were held to collect parents' experience of receiving the intervention, and parents in the observation group did not receive any intervention.

Intervention Group

Similar to the observation group, the following additional data collection, semistructured one-on-one and face-to-face interviews with active participants of the intervention group, were conducted in the postintervention data collection. The purpose of the interviews was to understand parents' experience of and attitude toward receiving this intervention and their willingness of receiving this intervention in the future. Moreover, parents were asked to talk about their support needs and if they had unmet needs still to be addressed. The interviews were conducted one at a time until data saturation was achieved [28]. In total, 11 parents were interviewed. This was a triangulation mixed-methods design. The qualitative data we chose to collect included important information that could help to demonstrate the effectiveness of this study and explain the quantitative results. Researchers obtained parents' true feelings of the intervention, why parents felt they benefited from the intervention, and more suggestion to improve further intervention after this potential effectiveness study.

Data Analysis

Quantitative Data

Data collected from 43 participants in the observation group and 49 participants in the intervention group were analyzed. All data were entered into a centralized database and analyzed with IBM SPSS 21.0 statistical software. Descriptive data were presented as mean (SD) for continuous variables, and as absolute and relative frequencies for categorical variables. The comparison of the baseline sociodemographic characteristics in the 2 groups was made using the independent- sample *t* test or one-way analysis of variance. The difference in outcomes between the 2 groups was examined using the independent-sample *t* test or paired-sample *t* test. The level of significance was set at alpha=.05.

Qualitative Data

The audiorecording of the qualitative data was transcribed and analyzed by 2 researchers independently within 24 hours after the interview. A thematic framework of content analysis was applied to analyze the information [28]. In addition, a stepwise approach was adopted for the content analysis. First, the transcribed data were read several times by the researchers to find the theme of the whole. Second, the segmentation of information was done to organize the segments and subsegments of information. Third, significant information related to research questions was extracted. Finally, data were coded and grouped into categories and abstracted into subthemes and the main theme.

Results

Overview of Participants at Each Stage of the Evaluation Progression

A total of 50 parents in the observation group and 51 parents in the intervention group participated in this study. In the observation group, 86% (43/50) parents completed the study. Among 7 parents in the observation group who did not complete the study, 3 parents rejected to complete the data collection in the postobservation phase, 2 children transferred to other hospitals, 1 child's parents gave up their child's treatment, and 1 parent's data were invalid for the postobservation phase questionnaire was not finished by the only parent who was being surveyed and observed. In the intervention group, 96% (49/51) parents completed the study. Two parents did not complete the

study as 1 parent stopped using Android smartphone on which the CA app only works, and the other parent's child died during the study period. Figure 3 displays the progression of participants at each stage of the evaluation.

Sociodemographic Characteristics

Multimedia Appendix 1 summarizes participants' and their children's baseline sociodemographic characteristics. No significant differences were observed between the intervention and observation groups (P>.05).

Baseline Outcome Questionnaire Results

Table 1 shows the baseline outcome questionnaire results. The data were collected during the preintervention data collection. There were no significant differences in questionnaire results at the baseline between the intervention and observation groups (P>.05).

Figure 3. The progression of participants at each stage of the evaluation. CNY: Chinese Yuan.

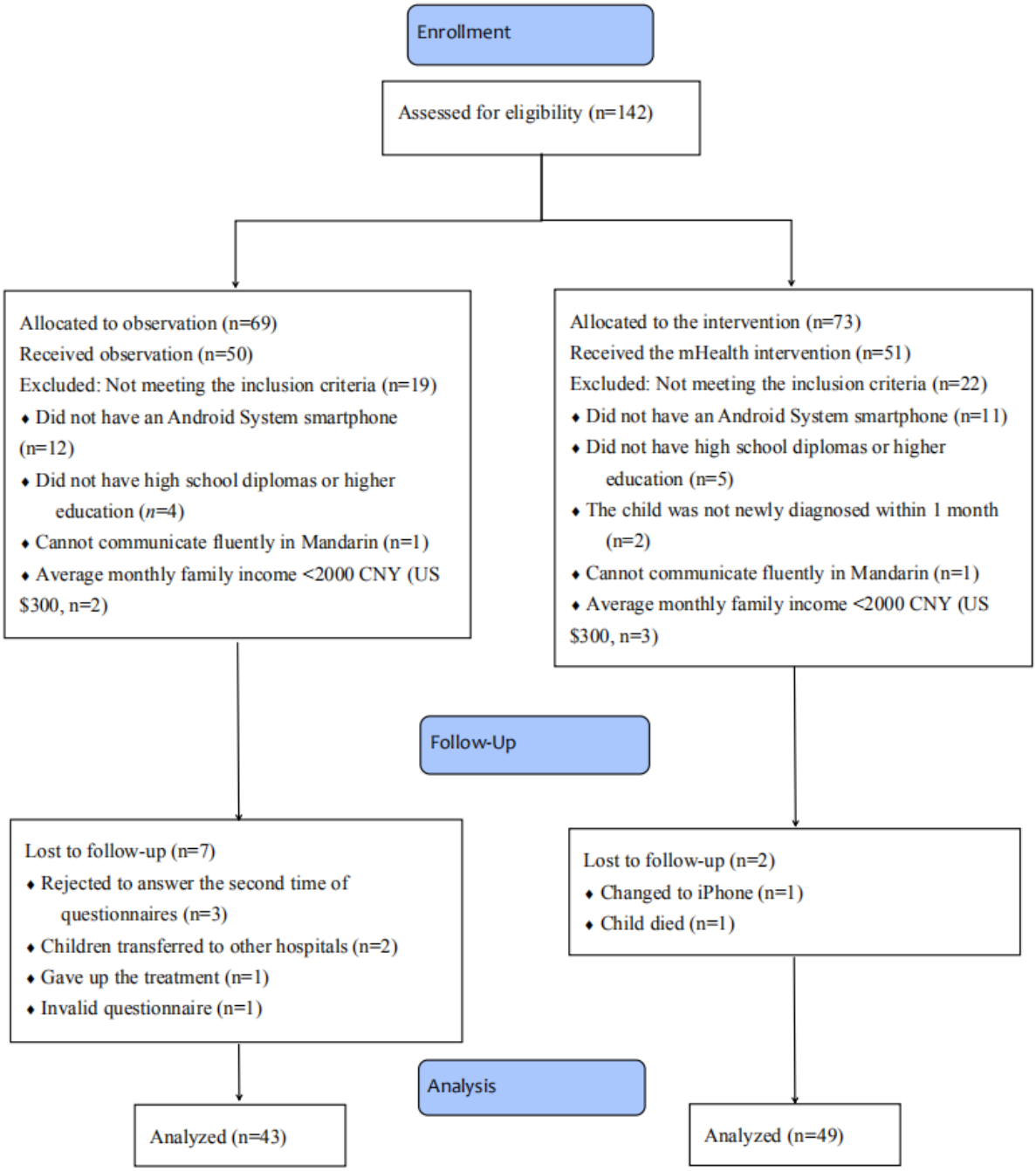


Table 1. The comparison of participants' baseline questionnaire results.

Name of the scales	Intervention group (n=49), mean (SD)	Observation group (n=43), mean (SD)	*t* value	*P* value
Perceived Social Support Scale	66.7 (10.0)	65.2 (11.5)	0.644	.52
Self-Rating Anxiety Scale	49.8 (10.7)	49.5 (10.8)	0.155	.88
Self-Rating Depression Scale	56.9 (12.1)	55.9 (13.7)	0.377	.71
Zarit Burden Inventory	37.6 (10.4)	37.2 (11.1)	0.180	.86
Parents' Perception of Uncertainty Scale	114.4 (7.1)	113.7 (6.8)	0.507	.61
Medical Outcomes Study 36-item Short Form				
Physical functioning	85.6 (18.2)	86.2 (19.6)	−0.140	.89
Social functioning	69.6 (23.4)	70.6 (23.0)	−0.206	.84
Pain	77.6 (17.8)	77.5 (20.5)	0.020	.98
Mental health	58.1 (18.3)	58.4 (22.9)	−0.069	.95
Vitality	58.5 (19.5)	58.4 (20.1)	0.024	.98
General health	64.2 (20.6)	66.0 (19.3)	−0.428	.67
Role limitation owing to physical problems	39.8 (46.5)	43.6 (47.3)	−0.389	.70
Role limitation owing to emotional problems	31.9 (42.4)	29.4 (41.9)	0.285	.78
Knowledge questionnaire				
Existing knowledge	23.3 (5.0)	23.6 (7.3)	−0.259	.80
Knowledge needs	98.9 (4.1)	99.5 (3.1)	−0.807	.42

The Effectiveness of the Mobile Health Supportive Care Intervention

Comparison of Outcome Questionnaires Results' D Value (Postintervention-Preintervention) Between the Intervention and Observation Groups

Multimedia Appendix 2 shows participants' outcome questionnaires results' *D* values (postintervention-preintervention) in the intervention and observation groups. Results showed that the mHealth supportive care intervention significantly reduced parents' anxiety (*P*=.03), reduced uncertainty in illness (*P*=.01), improved parents' social function (*P*=.01), increased parents' knowledge of ALL and care (*P*<.001), and decreased parents' need of knowledge (*P*<.001).

Comparison of Preintervention and Postintervention Questionnaires Results in the Intervention Group

Multimedia Appendix 2 presents preintervention and outcome questionnaire results in the intervention group. The findings showed that after the 3-month intervention, parents' anxiety (*P*<.001), depression (*P*=.01), uncertainty in illness (*P*<.001), and knowledge needs significantly decreased(*P*<.001), and the knowledge of ALL and care parents had significantly increased (*P*<.001).

Comparison of Preintervention and Postintervention Questionnaires Results in the Observation Group

Multimedia Appendix 2 shows preintervention and postintervention questionnaire results in the observation group. The results showed that over the course of 3-month observation, parents' care burden significantly increased (*P*=.01), and the knowledge of ALL and care significantly increased (*P*<.001), whereas uncertainty in illness significantly decreased (*P*<.001).

Qualitative Results

Eleven parents in the intervention group were interviewed regarding their experience of, attitude toward, and willingness to participate in the intervention. While 4 parents were fathers, 7 were mothers of children with ALL. Parents aged 22-35 years with the education of high school or above. Their family income ranged from 2000 to 8000 Chinese yuan a month. Their children aged 1-7 years, and 6 of them were boys. Multimedia Appendix 3 shows the detailed qualitative data by category. The qualitative results indicated that the overall experience of receiving this mHealth supportive care intervention was positive. Parents agreed that this intervention provided them useful information of ALL and care for their children. They were benefited from receiving this intervention but still had unmet supportive care.

Discussion

Principal Findings

This study indicated the potential effectiveness of the mHealth intervention in supporting parents of children with ALL. Results found that the intervention reduced parents' anxiety and uncertainty in illness, improved parents' social function, increased parents' knowledge of ALL and care, and decreased their need for knowledge. Parents were satisfied with the intervention and their role in the caregiving process. Moreover, this study is informative for other future studies on providing mHealth supportive care for parents of children with cancer.

Parents of children with ALL are engaged with a variety of care activities because of the special characteristics of ALL and children's limited cognitive ability. Studies have indicated parents' supportive care needs and poor health status [7]. To provide parents with accessible and effective supportive care is an urgent need. Interventions provided through smartphone apps have been proved to be feasible and effective in the health management of patients with chronic diseases [12]. However, few studies have focused on providing supportive care intervention to caregivers of patients with cancer, particularly children with cancer. Thus, this study aimed at providing supportive care to parents of children with ALL using mHealth. This is an innovative approach to meet parents' needs. Our research team took the lead in developing a mHealth smartphone app to support parents when the methodology and clinical application of mHealth interventions were at the very beginning stage in China.

The intervention was developed with rigorous methodology. A qualitative study with parents of children with ALL was initially conducted to gain an in-depth understanding of the characteristics and supportive care needs of parents [29]. The CA app was then developed according to a user-centered development process with a multidisciplinary group including software engineers and researchers [17]. A scoping review of the usability evaluation of mobile apps for cancer was conducted, which guided the app's usability evaluation. Parents, physicians, and nurses participated in the usability evaluation. Their use experience data were collected and analyzed [18]. Guided by the supportive care framework and based on parents' needs, preference, and the usability evaluation results, this mHealth supportive care intervention was developed. Then, an optimization discussion with professional health care providers was conducted to refine the intervention. This systematic process promoted the effectiveness of the intervention.

We prototyped a multicomponent mHealth intervention program to support parents. Based on the suggestions of health care providers, we had the CA app as the main intervention tool. "WeChat," the most popular chat tool in China, was added to the intervention to enable communication between parents and the care management team and provide the information that is not available in the app. Two intervention components work collaborated to provide parents with accessible and convenient supportive care and meet parents' constantly changing needs. The multidisciplinary care management team, including nurses, researchers, and software engineers, facilitates communication between parents and professional health care providers with high efficiency.

In the outcome measurement, we included both quantitative and qualitative outcomes. The quantitative results demonstrated the potential effectiveness of this mHealth supportive care intervention in improving parents' psychological health status, quality of life, care satisfaction and confidence, knowledge of ALL, and care for children. The effectiveness is in accordance with traditional interventions' results [8-10]. In this study, parents accepted and were willing to go on using this app and the WeChat Official Account. In the interviews, parents mentioned several times about the ease of access to scientific knowledge about the disease and care for their children, which improved their care confidence and communication efficiency with physicians and nurses. The easy access of information increased parents' existing knowledge and decreased their knowledge needs, which matched with the quantitative data.

In addition, accessible information and response from the care management team helped to reduce parents' anxiety and uncertainty of illness. Parents mentioned that there were professional health care providers to answer their personalized questions, which made them feel safe and less anxious. Parents felt they were more confident in handling some problems they encounter during the caring process. However, depression is a more serious and complicated psychological symptom. The intervention group parents' depression symptom decreased significantly (P=.08), whereas there was no significant decrease in the observation group nor the D value between the 2 groups (P=.08, P=.83). These results indicated that more professional psychological intervention need to be added in our mHealth intervention and further psychological effectiveness of the intervention need to be evaluated.

With the progress of children's disease and treatment, parents need to take on more care responsibilities and also have the heavier economic burden, which would increase parents' overall care burden. The observation group parents' care burden increased significantly (P=.09), whereas there was no significant increase in the intervention group and no significant difference between the D value in the 2 groups (P=.62, P=.08). Parents in the intervention group got more guide and direction from the CA app, WeChat Official Account, and professional health care providers, which might lead to better care for children, fewer complications, less treatment fee, or less increased care burden. The reason for a change in the care burden need to be explored in further study.

This intervention did not provide effective social support to parents, as the change of parents' perceived social support did not increase significantly (P=.94). However, the intervention group parents' social function improved compared with the observation group (P=.01). However, the intervention was still far from improving parents' quality of life because the other dimensions of quality of life did not change significantly: physical functioning (P=.56), pain (P=.66), mental health (P=.95), vitality (P=.47), general health (P=.86), role limitation owing to physical problems (P=.64), and role limitation owing to emotional problems (P=.32). These results indicated that family and friends' support needs to be improved in further study and have more family members participating in different aspects of caring for the child may be helpful for parents' physical and psychological status. Furthermore, the quality of life is a long-term indicator, which may not change immediately with the intervention, so further effectiveness evaluation is necessary.

Moreover, with this intervention, parents still expressed needs that were to be met, such as the need of an app to run on iOS phones, long-term support, and the worry about their children's disease progression and relapse. To a certain degree, these qualitative results explain why parents need more disease-related knowledge, though the need decreased in comparison with the preintervention period. These qualitative results help explain

causes for the quantitative results and provide us with information about the plausible relationship between the variables. This mixed-methods design provides us with the advantage of comprehensively examining the effectiveness of the intervention.

This intervention met parents' needs and improved the efficiency of communication between health care providers and parents. Information on patient care provided in this intervention has the potential effectiveness in helping reduce health care providers' educational burden and provided an approach to standardizing health education. Basic disease-related information decreased parents' frequency of asking health care providers common questions. However, parents always hope there could be someone reliable that would respond to them when they occurred with some new personalized problem. We got this need from our former study [17,29], and this is why we set the care management team with clinical nurse and nursing researchers. Further evaluation of the intervention's influence on professional health care providers will be conducted. Moreover, effective and efficient communication between professional health care providers and parents is essential to build a harmonious patient-provider relationship, which highlights the promotional value of further clinical application of this mHealth supportive care intervention.

Limitations and Future Directions

The CA app currently runs only on an Android smartphone. Many parents use iOS phones. A CA app running on iOS will be developed in the near future. This study was a nonrandomized controlled trail. In the future, we will address the issue of cross-group contamination and conduct a randomized controlled study to validate the intervention effectiveness. The CA app had its target users, which led to this mHealth supportive intervention being not suitable for every parent; only parents with basic education background can fully understand the text content in the CA app and WeChat Official Account. This study was conducted for 3 months with a sample size of about 50 for each group. The effect of intervention over a longer course of care continuum on all caregivers is yet to be assessed. In future studies, we will increase the sample size by recruiting main caregivers, including parents and other people, and extend the observation period to the entire treatment course.

Conclusions

This mHealth supportive care intervention showed its potential effectiveness in reducing parents' anxiety and uncertainty in illness, improving parents' social function, increasing parents' knowledge of ALL and care for children, and decreasing parents' need of information. Parents are satisfied with this intervention and willing to continue receiving the intervention. This study presents informative methodologies of assessing the effectiveness of providing supportive care to parents of children with cancer through a mHealth intervention.

Acknowledgments

This study was supported by the International Science and Technology Cooperation Program (Number 18410711700) and Shanghai Sailing Program (Number 18YF1429400) from Science and Technology Commission of Shanghai Municipality, China.

Authors' Contributions

CY was responsible for the entire study design. JW, NS, ZG, FW, XZ, MS, and AX conducted the study. JW drafted the manuscript. DH performed critical revisions of the manuscript. LW performed additional revisions of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Participants' baseline sociodemographic characteristics stratified by the study group.

[PDF File (Adobe PDF File), 134KB - mhealth_v6i11e195_app1.pdf]

Multimedia Appendix 2

Comparison of outcomes' changes in the intervention group (n=49), in the observation group (n=43), and D value between the 2 groups.

[PDF File (Adobe PDF File), 93KB - mhealth_v6i11e195_app2.pdf]

Multimedia Appendix 3

The qualitative results from the interviews in the intervention group (n=11).

[PDF File (Adobe PDF File), 97KB - mhealth_v6i11e195_app3.pdf]

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin 2013 Jan;63(1):11-30. [doi: 10.3322/caac.21166] [Medline: 23335087]
2. Kaatsch P. Epidemiology of childhood cancer. Cancer Treat Rev 2010 Jun;36(4):277-285. [doi: 10.1016/j.ctrv.2010.02.003] [Medline: 20231056]
3. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. CA Cancer J Clin 2014 Apr;64(2):83-103. [doi: 10.3322/caac.21219] [Medline: 24488779]
4. Subspecialty Group of Hematology, the Society of Pediatrics, Chinese Medical Association, Editorial Board, Chinese Journal of Pediatrics, Subspecialty Group of Hematology the Society of Pediatrics Chinese Medical Association, Editorial Board Chinese Journal of Pediatrics. [Guidelines for the diagnosis and treatment of childhood acute lymphoblastic leukemia]. Zhonghua Er Ke Za Zhi 2014 Sep;52(9):641-644. [Medline: 25476423]
5. Fitch M. Providing supportive care for individuals living with cancer (Task Force Report). Toronto: Ontario Cancer Treatment and Research Foundation 1994.
6. Fitch MI. Supportive care framework. Can Oncol Nurs J 2008;18(1):6-24. [Medline: 18512565]
7. Tenniglo LJA, Loeffen EAH, Kremer LCM, Font-Gonzalez A, Mulder RL, Postma A, et al. Patients' and parents' views regarding supportive care in childhood cancer. Support Care Cancer 2017 Dec;25(10):3151-3160 [FREE Full text] [doi: 10.1007/s00520-017-3723-7] [Medline: 28456909]
8. Wu L, Liu L, He J, Jin T, Liu Y. Effect evaluation on establishment of support system of parents of children with leukemia. Chin Nur Res 2014;3:964-965.
9. Xian W, Jin H. Social club in improving of adherence for parents of children with ALL. Nur Prac Res 2008;5(7):82-83.
10. Applebaum AJ, Breitbart W. Care for the cancer caregiver: a systematic review. Palliat Support Care 2013 Jun;11(3):231-252. [doi: 10.1017/S1478951512000594] [Medline: 23046977]
11. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Optimizing cancer care through mobile health. Support Care Cancer 2015 Jul;23(7):2183-2188. [doi: 10.1007/s00520-015-2627-7] [Medline: 25649121]
12. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
13. Klasnja P, Hartzler A, Powell C, Pratt W. Supporting cancer patients' unanchored health information management with mobile technology. AMIA Annu Symp Proc 2011;2011:732-741 [FREE Full text] [Medline: 22195130]
14. Agboola S, Flanagan C, Searl M, Elfiky A, Kvedar J, Jethwani K. Improving outcomes in cancer patients on oral anti-cancer medications using a novel mobile phone-based intervention: study design of a randomized controlled trial. JMIR Res Protoc 2014 Dec 23;3(4):e79 [FREE Full text] [doi: 10.2196/resprot.4041] [Medline: 25537463]
15. Harder H, Holroyd P, Burkinshaw L, Watten P, Zammit C, Harris PR, et al. A user-centred approach to developing bWell, a mobile app for arm and shoulder exercises after breast cancer treatment. J Cancer Surviv 2017 Dec;11(6):732-742 [FREE Full text] [doi: 10.1007/s11764-017-0630-3] [Medline: 28741202]
16. Sun Y, Jiang F, Gu JJ, Wang YK, Hua H, Li J, et al. Development and Testing of an Intelligent Pain Management System (IPMS) on Mobile Phones Through a Randomized Trial Among Chinese Cancer Patients: A New Approach in Cancer Pain Management. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e108 [FREE Full text] [doi: 10.2196/mhealth.7178] [Medline: 28743681]
17. Wang J, Yao N, Wang Y, Zhou F, Liu Y, Geng Z, et al. Developing 'Care Assistant': A smartphone application to support caregivers of children with acute lymphoblastic leukaemia. J Telemed Telecare 2016 Apr;22(3):163-171. [doi: 10.1177/1357633X15594753] [Medline: 26271029]
18. Wang J, Yao N, Shen M, Zhang X, Wang Y, Liu Y, et al. Supporting Caregivers of Children With Acute Lymphoblastic Leukemia via a Smartphone App: A Pilot Study of Usability and Effectiveness. Comput Inform Nurs 2016 Nov;34(11):520-527. [doi: 10.1097/CIN.0000000000000265] [Medline: 27385540]
19. Wu WY. Self-Rating Anxiety Scale. Zhang ZJ. editor. Behavioral Medicine Inventory Manual. Beijing: The Chinese Medicine Electronic Audio and Video Publishing House, 2005:213-214.
20. Wang Z. Chinese version of Zung's self-rating anxiety scale. J Shanghai Psychiatry 1984;2:73-74.
21. Wu WY. Self-Rating Depression Scale. Zhang ZJ. editor. Behavioral Medicine Inventory Manual. Beijing: The Chinese Medicine Electronic Audio and Video Publishing House, pp 2005:215-217.
22. Huang L, Jiang QJ, Ren WH. Coping style, social support,psychosomatic symptoms in patients with cancer. Chin Ment Health J 1996;10:160-161.
23. Huang L, Jiang QJ, Ren WH. The correlational study on coping modes, social support, and psychosomatic symptoms among cancer patients. Chin Ment Health J 1999;8(3):199-201.
24. Lu L, Wang L, Yang X, Feng Q. Zarit Caregiver Burden Interview: development, reliability and validity of the Chinese version. Psychiatry Clin Neurosci 2009 Dec;63(6):730-734 [FREE Full text] [doi: 10.1111/j.1440-1819.2009.02019.x] [Medline: 19781014]
25. Yang X, Hao Y, Yang Y, Zhu K, Liu Z, Wang L. Study on influencing factors associated with nursing burden of caregivers of leukemia patients. Pract Prev Med 2012;9(2):176-178.

26. Mai J, Xie W, Ma C, Deng Y, Dai Y. Initial revision of Chinese Version of parents? perception of uncertainty scale. Chin J Prac Nurs 2013;29(28):46-50.
27. Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. J Epidemiol Community Health 2003 Apr;57(4):259-263 [FREE Full text] [Medline: 12646540]
28. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
29. Wang Y, Wang J, Yu S, Zhou F, Yuan C. Parents' needs of early diagnosed children with acute lymphoblastic leukemia: a qualitative study. Nurs J Chin PLA 2016;33(4):6-10.

Abbreviations

ALL: acute lymphoblastic leukemia
CA: Care Assistant
mHealth: mobile health
PPUS: Parents' Perception of Uncertainty Scale
PSSS: Perceived Social Support Scale
SAS: Self-Rating Anxiety Scale
SDS: Self-Rating Depression Scale
SF-36: Medical Outcomes Study 36-item Short Form
ZBI: Zarit Burden Inventory

Edited by G Eysenbach; submitted 06.02.18; peer-reviewed by S Choi, M Schoen; comments to author 30.07.18; revised version received 18.08.18; accepted 22.08.18; published 19.11.18.

Please cite as:
Wang J, Howell D, Shen N, Geng Z, Wu F, Shen M, Zhang X, Xie A, Wang L, Yuan C
mHealth Supportive Care Intervention for Parents of Children With Acute Lymphoblastic Leukemia: Quasi-Experimental Pre- and Postdesign Study
JMIR Mhealth Uhealth 2018;6(11):e195
URL: http://mhealth.jmir.org/2018/11/e195/
doi:10.2196/mhealth.9981
PMID:30455166

Review

Mobile Ecological Momentary Diet Assessment Methods for Behavioral Research: Systematic Review

Susan M Schembre[1,2], RD, PhD; Yue Liao[1], MPH, PhD; Sydney G O'Connor[3], BA; Melanie D Hingle[4], MPH, RDN, PhD; Shu-En Shen[5]; Katarina G Hamoy[6]; Jimi Huh[3], PhD; Genevieve F Dunton[3], MPH, PhD; Rick Weiss[7], MS; Cynthia A Thomson[8], RD, PhD; Carol J Boushey[9], RD, MPH, PhD

[1]Department of Behavioral Science, Division of Cancer Control and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[2]Department of Family and Community Medicine, College of Medicine-Tucson, University of Arizona, Tucson, AZ, United States

[3]Institute for Health Promotion & Disease Prevention, Department of Preventive Medicine, University of Southern California, Los Angeles, CA, United States

[4]Department of Nutritional Sciences, College of Agriculture & Life Sciences, University of Arizona, Tucson, AZ, United States

[5]Department of Kinesiology, Wiess School of Natural Sciences, Rice University, Houston, TX, United States

[6]Department of Health and Human Performance, College of Liberal Arts and Social Sciences, University of Houston, Houston, TX, United States

[7]Viocare, Inc, Princeton, NJ, United States

[8]Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

[9]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, United States

Corresponding Author:
Susan M Schembre, RD, PhD
Department of Family and Community Medicine
College of Medicine-Tucson
University of Arizona
Abrams Building Room 3345E
3950 South Country Club Road
Tucson, AZ, 85714
United States
Phone: 1 520 626 7735
Email: sschembre@email.arizona.edu

Abstract

Background: New methods for assessing diet in research are being developed to address the limitations of traditional dietary assessment methods. Mobile device–assisted ecological momentary diet assessment (mEMDA) is a new dietary assessment method that has not yet been optimized and has the potential to minimize recall biases and participant burden while maximizing ecological validity. There have been limited efforts to characterize the use of mEMDA in behavioral research settings.

Objective: The aims of this study were to summarize mEMDA protocols used in research to date, to characterize key aspects of these assessment approaches, and to discuss the advantages and disadvantages of mEMDA compared with the traditional dietary assessment methods as well as implications for future mEMDA research.

Methods: Studies that used mobile devices and described mEMDA protocols to assess dietary intake were included. Data were extracted according to Preferred Reporting of Systematic Reviews and Meta-Analyses and Cochrane guidelines and then synthesized narratively.

Results: The review included 20 studies with unique mEMDA protocols. Of these, 50% (10/20) used participant-initiated reports of intake at eating events (event-contingent mEMDA), and 50% (10/20) used researcher-initiated prompts requesting that participants report recent dietary intake (signal-contingent mEMDA). A majority of the study protocols (60%, 12/20) enabled participants to use mobile phones to report dietary data. Event-contingent mEMDA protocols most commonly assessed diet in real time, used dietary records for data collection (60%, 6/10), and provided estimates of energy and nutrient intake (60%, 6/10). All signal-contingent mEMDA protocols used a near real-time recall approach with unannounced (ie, random) abbreviated diet surveys. Most signal-contingent protocols (70%, 7/10) assessed the frequency with which (targeted) foods or food groups were consumed. Relatively few (30%, 6/20) studies compared mEMDA with the traditional dietary assessment methods.

Conclusions: This review demonstrates that mEMDA has the potential to reduce participant burden and recall bias, thus advancing the field beyond current dietary assessment methods while maximizing ecological validity.

(JMIR Mhealth Uhealth 2018;6(11):e11170) doi:10.2196/11170

KEYWORDS

diet surveys; diet records; mobile phone; mobile apps; ecological momentary assessment

Introduction

Diet plays significant direct and indirect roles in the etiology and prevention of chronic diseases including type 2 diabetes, coronary heart disease, cancer, and obesity [1]. It has been estimated that an unhealthy diet was the leading cause of premature death in the United States, contributing to more than 500,000 deaths in 2016 [1]. Despite these statistics, we lack a clear understanding of how patterns of dietary intake affect health through the life span because of dietary measurement limitations. Current dietary assessment methods including 24-hour dietary recalls, food frequency questionnaires, and dietary records have advantages and disadvantages in research settings [2]. The 24-hour dietary recalls are considered the gold standard in dietary assessment because they produce the highest quality data, but they rely heavily on participant memory and can require the greatest amount of time to acquire when administered by the interviewer. Food frequency questionnaires are more easily administered and provide good estimates of patterns of intake; however, they are also subject to participant memory and are most likely to underestimate energy intake because of limitations including a lack of cultural tailoring or limited food lists. Diet records or diet diaries minimize reliance on memory when foods and beverages are recorded when consumed; however, participants require a high level of motivation as well as training to improve recording accuracy. Furthermore, research shows that people pay little attention to when and what they eat [3], and factors such as age, sex, and weight status can influence the accurate recall of food and estimation of portion size [2,4-8]. Finally, methods that do not use unannounced recalls (eg, diet records) may be prone to biases such as social desirability or reactivity bias, which may lead to participants underreporting or omitting of foods or beverages consumed or changing their usual dietary behaviors because of the awareness that they are being observed [9]. Such errors in reporting are known to create conflicting evidence linking diet to health outcomes that could be addressed if we could more robustly measure diet [10,11]. Furthering our understanding of the connection between diet and disease will require improvements in the dietary assessment methodologies. For this reason, the research community has recognized the need for new dietary assessment methods that can reduce misreporting and recall biases [12,13].

Recent advancements in digital technology and computational sciences have laid the foundation for emerging dietary assessment solutions. These advancements have catalyzed the development of new methods aimed at automating the assessment of dietary intake, thereby limiting or eliminating the need for self-report. In particular, 2 such new dietary assessment methods have been developed: image-based dietary assessment [14-17] and the detection of food intake by biomechanical sensors [18-29]. Single image-based assessment methods use photos of foods and beverages along with fiduciary markers before and after consumption. The time-stamped images are either reviewed and coded into nutrition software by trained research staff (image-assisted assessment) or are analyzed by software designed to identify the type and volume of foods in the image (image-based assessment). Alternatively, approaches using gyroscopes, microphones, and mechanical or electrical impedance sensors have been integrated into wearable devices such as watches and headsets or are designed to be mounted on teeth to detect wrist or hand motion or patterns of chewing or swallowing indicative of food intake (eg, number of bites). However, the automation of dietary assessment using these mobile-based approaches remains in the proof-of-concept stage. There is a lack of large-scale validation studies demonstrating their utility to assess dietary intake in community-dwelling populations. For instance, the mean detection accuracy of image detection and wearable devices has been acceptable in controlled, laboratory settings (range 73%-99%) [19,21,23,24,26-35], but limited testing has been done in natural settings. The use of mechanical sensors in research is further hindered by poor battery life, having to remember to wear or use the device, needing to turn the device on or off to avoid detection errors, and the conspicuousness or general discomfort of having to wear collars, wires, or harness accessories. Substantial work will be needed before these methods can accurately quantify energy or nutrient intake for research purposes. Therefore, novel dietary assessment methods addressing limitations of the traditional dietary assessment methods and methods that bridge the gap between traditional and newer methods of dietary assessment are needed [12,13].

Another less developed dietary assessment method with the potential to improve the validity and reliability of dietary assessment is the mobile device–assisted ecological momentary assessment (mEMA). mEMA is based on the foundation of ecological momentary assessment (EMA) described by Shiffman et al in 2008 [36]. EMA involves the repeated sampling of a person's current behaviors and experiences in real time, in their natural environments. Currently, there are 2 mobile device–assisted ecological momentary diet assessment (mEMDA) approaches: event-contingent mEMDA and signal-contingent mEMDA. Event-contingent mEMDA most often occurs in real time at the time of eating (or drinking). The frequency of sampling is determined by the number of times a participant reports eating. Here, the act of initiating a meal or snack triggers either the real-time recording of dietary intake (eg, dietary records or image-assisted dietary records). The advantage of real-time diet records is that they are intended to capture all foods and beverages consumed without having to recall the events at a later time. Although this is an advantage,

the key limitation of event-contingent mEMDA is that there are no unannounced sampling events. The self-monitoring of dietary intake can be influenced by psychosocial and behavioral factors that introduce reactivity and measurement bias including eating behaviors (dietary restraint or disinhibition), social desirability, body image, or depression and anxiety [5]. Furthermore, with image-assisted diet records, there is the potential for data entry bias by research staff viewing images or for intentional or unintentional reporting errors (eg, inaccurate report of portion sizes) by participants, particularly if foods are omitted from the images or the images are not taken at multiple points of a meal (eg, before and after a meal).

Signal-contingent mEMDA relies on researcher-initiated, signaled prompts to participants that trigger the recall of current or recent dietary intake. Although study participants are often prompted multiple times per day, signal-contingent mEMDA does not always allow for the real-time assessment of dietary intake. Rather, assessment surveys often include questions referencing dietary intake occurring within the most recent interval of time (eg, past 30 min). Moreover, study participants are most often asked to report the consumption of specific foods or foods from specified food groups (eg, fruits and vegetables) by means of a brief survey. The frequency of sampling using signal-contingent mEMDA is determined by the researcher and can occur randomly at fixed or semifixed times or randomly within fixed or semifixed time intervals. As with image-assisted or image-based dietary records, these momentary dietary assessments can be time-stamped and are either stored or transmitted for later database integration. Although this method benefits from random (unannounced) sampling, short recall intervals, and reduced participant burden, the commonly used sampling schemes, limited study durations, or limited food lists can hinder the quantification of energy or nutrient intakes.

Due to recent advancements made in mobile device hardware and software and the pervasive use of mobile devices, EMA-based methods leverage the capabilities of mobile technology, offering researchers an opportunity to assess the dietary intake of study participants as they are occurring in natural settings. Both event-contingent and signal-contingent mEMDA seek to reduce recall bias and thereby improve the accuracy of dietary assessment by eliminating or shortening the recall interval and reducing participant burden while maximizing ecological validity as compared with the traditional dietary approaches. However, neither approach has been well characterized nor adequately compared against objective biomarkers or other methods of energy intake assessment (eg, 24-hour dietary recalls). Focused efforts are needed to develop mEMDA methods for their consistent and replicable application in research settings. Therefore, the goals of this systematic review were to summarize the event-contingent and signal-contingent mEMDA protocols that have been used in research to date, to characterize key aspects of these assessment approaches (eg, design, data collection, data processing and dietary analysis, and dietary outcomes), and to discuss the advantages and disadvantages of each as well as implications for future mEMDA research. The focus of the review was on studies summarizing unique dietary assessment protocols using mobile devices to facilitate event-contingent or signal-contingent EMA to assess dietary intake.

Methods

Literature Search

A systematic strategy was devised by 6 authors (SMS, YL, MDH, JH, GFD, CAT, and CJB) to search the MEDLINE, EMBASE, PubMed, PsycINFO, and IEEE explore databases for all relevant literature published through February 2018. The search was limited to articles written in the English language and conducted with humans. The database search included the use of controlled vocabulary and keywords to identify studies addressing dietary assessment, mobile devices, and ecological momentary assessment. Keywords such as "nutrition assessment," "diet surveys," "diet records," "energy intake," "meals," or "eating" combined with "text messaging," "mobile phone," "mobile applications," "micro-electrical-mechanical systems," or "wearable electronic devices" were included as MeSH search terms. In addition, non-MeSH search terms were included to be complete: "caloric intake," "food diary," "diet monitoring," "food tracking," "diet tracking," "diet assessment," or "calorie tracking" and "text messages," "cell phone," "smartphone," "tablet computer," "mobile health," "eHealth," "mHealth," "digital health," "mobile technology," or "experience sampling." Search terms synonymous with the terms above that did not produce additional references (eg, "food record") were omitted from the final search conducted by the author (SMS). References cited in all included studies and studies citing included studies (referred to as "other sources" in the Preferred Reporting of Systematic Reviews and Meta-Analyses [PRISMA] diagram) were also reviewed to identify any additional studies to be screened for inclusion.

Study Inclusion and Exclusion Criteria

Eligible studies had protocols using mobile devices and event-contingent or signal-contingent EMA approaches to assess dietary intake in research settings. These included diet assessment studies as well as behavioral trials where dietary intake was assessed. Dietary intake was defined as the quantification of energy intake, macro- or micronutrients, discrete foods, servings, or food groups. Literature returned by the search was screened first by article type then by title, abstract, and the described methods by 2 authors (SMS and YL). Only original research articles were included. Abstracts, review papers, editorials, etc, were excluded. Additionally, studies were excluded if the title, abstract, or methods indicated (1) the study was not diet-related (eg, nondiet-related papers, proof-of-concept, or technology design papers); (2) the studies were interventions with non-EMA dietary assessment methods (24-hour dietary recalls and food frequency questionnaires); (3) did not assess dietary intake (eg, binge eating lapses, availability of snack foods, and food craving); (4) used self-monitoring approaches without dietary analysis; (5) were described in an earlier study or were considered a secondary analysis; or (6) were not peer-reviewed journal articles (eg, abstracts, editorials, discussions, evaluations, reviews, reports, news, notes, surveys, or content analysis). Additional papers referencing the included

studies were used to obtain methodological details not otherwise provided in the included studies.

Data Extraction and Analysis

Data were extracted into a structured coding form according to PRISMA guidelines [37] and the Cochrane Handbook for Systematic Reviews of Interventions [38]. A data extraction form developed for this review was used by 2 authors (SES and KGH) to independently extract and review characteristics from all studies. Extracted data represented details on mEMDA protocols and included but were not limited to (1) mobile platform, (2) sampling duration, (3) prompt approach (signal-contingent only), (4) prompt frequency (signal-contingent only), (5) data collection method, (6) data processing and nutrient analysis, (7) diet data outcomes, and (8) protocol adherence. A comprehensive list of data extracted from included studies is provided as Multimedia Appendix 1. Discrepancies in the extracted data were resolved by a discussion between 2 expert reviewers (SMS and YL) to complete the dataset. In several cases, studies closely related to the included studies were reviewed for additional information to resolve issues of missing or unclear data. Extracted data were descriptive in nature. The data were synthesized narratively and tabularized with the intent of summarizing available protocols for assessing diet using mobile EMA methods.

Results

Literature Search

The literature search yielded 1462 studies, of which 173 were duplicates, leaving 1289 articles to be screened for eligibility. A total of 463 articles were excluded based on an initial screening indicating these were not journal articles. Thus, 826 articles were screened by title, abstract, and methods for eligibility. After 806 articles that did not meet the inclusion criteria were excluded, 20 studies were included in the review (see PRISMA diagram, Figure 1). Among the 20 studies included in the review, 10 used event-contingent mEMDA protocols and 10 used signal-contingent mEMDA protocols to assess dietary intake. An additional 19 related journal articles were used to obtain methodological details not provided in the included studies. Of the 20 included studies, 7 were behavioral trials where dietary intake was assessed. The remaining 13 were diet assessment studies.

Summary of Event-Contingent Studies

The protocols in studies using event-contingent mEMDA are summarized in Table 1. A total of 10 studies used event-contingent mEMDA protocols in nutrition-related research [39-48]. Additional details about protocols used in the included studies that were not described fully were extracted from related journal articles that used the same mEMDA protocol [30,49-60].

Figure 1. Preferred Reporting of Systematic Reviews and Meta-Analyses diagram. EMA: ecological momentary assessment.

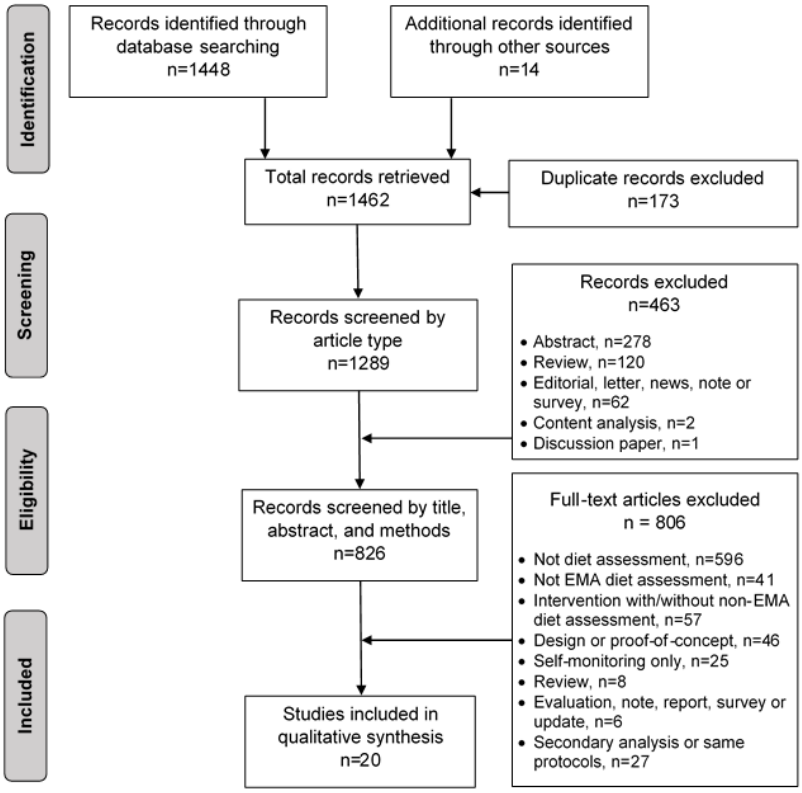


Table 1. Event-contingent, mobile ecological momentary dietary assessment.

First author, year	Mobile platform and device	Sample period	Data collection[a]	Data processing and nutrient analysis	Diet data outcomes
Ashman et al, 2017 [39]	Internet-based, mobile phone app	3 days	Image-assisted dietary record: images taken before and after meals with fiducial marker	Dietitians analyzed food images with FoodWorks software (The Nutrition Company)	Energy, protein, dietary fat, carbohydrates, and select micronutrients
Boushey et al, 2017 [40]	Mobile phone app	7.5 days	Image-assisted dietary record: images taken before and after meals with fiducial marker	Trained analysts analyzed food images with Food and Nutrient Database for Dietary Studies (United States Department of Agriculture)	Energy intake
Della-Torre et al, 2017 [41]	Internet-based, mobile phone app	4 days	Dietary record: food and beverages chosen from 900 options	Automated app output (study-specific food composition database)	Energy, protein, dietary fat, carbohydrate, fruit and vegetables, and dairy
Grenard et al, 2013 [42]	PDA[b] device and software	7 days	Dietary record: food and beverages chosen from 3 groups	Data downloaded from PDA by researchers (no nutrient database used)	Number of sweetened drinks, sweet snacks, salty snacks, and sweet or salty snacks
Hingle et al, 2013 [43]	Social media (mobile phone app; Twitter)	3 days	Dietary record: food and beverages chosen from 24 groups	Web-based data capture app (ViBE) used to automatically calculate output (no nutrient database used)	Number of times each food category was reported
Martin et al, 2012 [44]	Mobile phone app	6 days	Image-assisted dietary record: images taken before meals with fiducial marker	Image analysis by 2-step process: human raters and computer automation with Food and Nutrient Database for Dietary Studies (United States Department of Agriculture)	Energy, protein, dietary fat, carbohydrates, and select micronutrients
Schuz et al, 2015 [45]	Mobile phone app	10 days	Dietary record: items labeled as breakfast, lunch, dinner, snacks, and drinks	Data downloaded from app by researchers (no nutrient database used)	Frequency of meals, snacks, nonalcoholic drinks, or alcoholic drinks
Seto et al, 2016 [46]	Mobile phone	6 days	Voice-annotated video with time stamp	Dietitians analyzed the videos and coded the portion size and food groups (no nutrient database used)	Portions of total meal, dairy, protein, grains, vegetables, and fruits
Thomas et al, 2011 [47]	PDA device and software	6 days	Dietary record: food and beverages chosen from 8 groups with manual entry of food type and portion size	Data downloaded from PDA by researchers (no nutrient database used)	Food group servings
Waki et al, 2014 [48]	Mobile phone app	3 months	Image-assisted dietary record: images taken before meals	Automatic photo processing by study-specific software and Dietary Reference Intakes	Energy, protein, dietary fat, carbohydrate, dietary fiber, and sodium

[a]All food and beverage recorded unless otherwise noted.

[b]PDA: personal digital assistant.

Design

Of the 10 studies, 5 were mobile phone–based [40,44-46,48]: 3 used mobile phone apps [40,44,48] and 2 used the mobile phone camera function [44,46]. The remaining 5 studies used a PDA device with customized software [42,47], an internet-based app [39,41], or social media (Twitter) [43]. All studies assessed dietary intake continuously throughout each day. The sampling duration ranged from 3 days to 3 months, with 6 days being the most common.

Data Collection Methods

A total of 6 studies collected data by dietary records [39,41-43,45,47]. A note-taking app with image capture was used in 1 study [39]. Others had study participants record the consumption of predefined food types or food groups [41-43,45,47]. Participants were asked to take photos of all food and beverage consumed without additional note taking in 3 studies [40,44,48], and 1 study collected dietary data by voice-annotated video taken with a mobile phone [46]. In 1 study, all food and beverages were provided to participants to

take home during the study period and encouraged them to supplement with usual foods and beverages not provided [40]. All other studies collected dietary data based on a participant's usual eating behaviors. Hingle et al [43] collected 1756 food-related hashtags via Twitter across all participants over 3 days. In Seto et al's [46] 6-day study, 72 food items were reported via video per participant. Participants on average reported 7 food entries via dietary record per day in Ashman et al's 3-day study [39]. Alternative methods to capture missed meals (ie, pen and pencil or voice recording) were used in 2 studies. EMA prompts were used at standard or usual breakfast, lunch, dinner, and snack times as a reminder to log eating events in 1 study [44].

Dietary Analysis and Outcomes

Trained dietitians or research staff were involved in 5 studies to analyze the data based on a food composition database or similar software [39-41,44,46]; 3 studies downloaded data from the mobile device to perform further data analysis without the use of a nutrient database or software [42,45,47]; these studies assessed frequency or servings of food intake. Output automatically generated by a non-nutrient-related software or app was used in 2 studies [41,43]; 1 study estimated intake of nutrients, energy, and food groups [41], the other assessed frequency of food category [43]. The nutrient analysis was automated within the study app in 1 study [48]. Participants received feedback about nutritional balance and energy balance of the meal as well as advice on dietary modification within the app. With respect to the primary outcomes assessed, 5 studies estimated energy intake [39,40,44,45,48]. Macro- or micronutrients were estimated in 4 studies [39,40,44,48]. Portions or servings consumed from designated food groups were estimated in 2 studies [46,47]. Dietary data at within-day level (ie, for each meal) were provided in 4 studies [43,45,46,48]; the remaining studies provided dietary data summary at the day level.

Protocol Adherence

Data regarding adherence to the dietary data collection protocols described in the studies (eg, reporting all eating events) were not provided in any of the event-contingent studies. This was most likely because of not having objective knowledge of when eating actually occurred. However, 3 studies reported the number of eating events captured or frequency with which foods were consumed [39,43,46].

Comparison Testing

There were 4 studies comparing their mEMDA approach against a traditional dietary assessment method [39-41,44]. In 2 studies, estimated energy intake was compared with doubly labeled water [41,44]. Martin et al [44] found no significant difference in energy intake between the estimation from their mEMDA approach (Remote Food Photography Method) versus the doubly labeled water in a sample of overweight and obese adults (−152 ± 694 kcal/day, P=.16). However, in another comparison test related to the mEMDA method used by Martin et al, Nicklas et al [50] found the Remote Food Photography Method underestimated energy intake when compared with doubly labeled water by an average of 222 kcal/day (−15.6%, P<.001) in a sample of minority (Hispanic and African American) preschool children (data reported by their caregivers). Boushey's study [40] aimed to test the accuracy of the estimated energy intake from the mobile Food Record (mFR) against energy expenditure assessed by doubly labeled water in a community sample of 45 adults aged 21-65 years. On the basis of the comparison, the mean percent of underreporting on the mFR was 12% (SD 11) for men and 10% (SD 10) for women. Estimated intake to 24-hour dietary recalls was compared in 2 studies [39,41]. Della-Torre et al [41] developed and evaluated an electronic mobile-based food record, electronic carnet alimentaire (e-CA) for a research setting. They evaluated e-CA's accuracy in terms of energy, macronutrient, and food group intake in a convenience sample of 21 adults and found the primary diet data had more than 85% agreement with the 24-hour dietary recall. Ashman et al [39] evaluated relative validity of the DietBytes image-based dietary assessment method for assessing energy and nutrient intakes in 25 pregnant women and found the macronutrient and energy intake had more than 90% agreement with the 24-hour dietary recall.

Summary of Signal-Contingent Studies

The protocols in articles that only used signal-contingent mEMDA approaches are summarized in Table 2. A total of 10 studies described signal-contingent mEMDA used in nutrition-related research [61-70]. Additional details were extracted from multiple related journal articles [66,71-76].

Design

Of the 10 studies, prompts were delivered via mobile phone in 7 studies [61-63,66,67,69,70]. Of these 7 studies, 5 used mobile apps [62,63,66,67,69], 1 used short message service text messaging [61], and 1 used a Web-based survey [70]. A wrist-worn electronic diary device was used in 2 other studies [64,65], and another study used an iPod Touch [68].

Of the 10 studies, 5 studies used *random* intervals for prompting [62-64,69,70] with frequencies ranging from 3 to 10 prompts per day. Of these 5 studies, 3 studies assessed dietary intake "since the last prompt" at varied time intervals [64,69,70], 1 study assessed dietary intake in the past 2 hours [63], and 1 study assessed dietary intake in real time [62].

The other 5 studies prompted surveys at *fixed* intervals [61,65-68], and frequencies ranged from 4 to 14 prompts per day. Of these 5 studies, 2 studies assessed dietary intake "since the last prompt" at varied time intervals [66,68], 2 studies assessed dietary intake in the past 1-3.5 hours [65,67], and 1 study assessed dietary intake in real time [61]. The sampling duration for the 10 studies ranged from 4 days to 6 weeks, with the most common duration being 7 days (n=4).

Table 2. Signal-contingent, mobile ecological momentary dietary assessment.

First author, year	Mobile platform	Sample period	Prompt approach	Prompt frequency (recall interval)	Diet data collection (format, source)	Diet data output outcomes (units)
Berkman et al, 2014 [61]	Mobile phone SMS[a] text messages	14 days	Individualized fixed time	4 prompts: 3 real time, 1 retrospective (since last prompt)	1 survey item (open-ended, preselected snack food)	Frequency of snack intake
Bruening et al, 2016 [62]	Mobile phone app	4 days	Random interval	8 prompts: 7 real time, 1 retrospective prompt (past 3 hours)	2 survey items (multiple choice, 8 food groups, and 8 beverage groups)	Bread or grains, entrée, fruit and vegetables, salty foods, and sweets intake (number and percent of prompts)
Dunton et al, 2015 [63]	Mobile phone app	8 days	Random interval	Mother: 4 or 8 retrospective prompts (past 2 hours); Child: 3 or 7 retrospective prompts (past 2 hours)	1 survey item (multiple choice, 5 food groups)	Healthy and unhealthy eating (frequency of prompts)
Miller et al, 2016 [64]	Wrist-worn electronic diary	6 weeks	Random interval	3 retrospective prompts (since last prompt)	1 survey item (open-ended)	Low glycemic index foods (servings)
Powell et al, 2017 [65]	Wrist-worn electronic diary	7 days	Fixed time (±10 min)	14 retrospective prompts (past hour)	8 survey items (8 food groups, yes or no)	Snack and fruit and vegetable intake (ranked portion sizes)
Richard et al, 2017 [66]	Mobile phone app	7 days	Fixed time	5 retrospective prompts (since last prompt)	1 survey item (open-ended)	Snack intake density (kcal/100 g)
Spook et al, 2013 [67]	Mobile phone app	7 days	Fixed time (±30 min)	5 retrospective prompts (past 3.5 hours)	3 survey items (multiple choice and visual analog scales, 3 food groups)	Number and frequency of snack, fruit and vegetable, and soda intake
Strahler and Nater, 2018 [68]	iPod Touch app	4 days	Fixed time	5 retrospective prompts (since last prompt)	3 survey items (multiple choice recoded to yes or no)	Frequency of meal type, main component, and drink consumption
Wouters et al, 2016 [69]	Mobile phone app	4 days	Quasi-random interval (average 90 min)	10 retrospective prompts (since last prompt)	Digital food log of snacks (open ended)	Energy intake carbohydrate, fat, and protein
Zenk et al, 2014 [70]	Mobile phone Web-based survey	7 days	Random interval	5 retrospective prompts (since last prompt)	9 Web-based survey items (9 food groups, yes or no)	Number of snacks consumed (0 or more than 1)

[a]SMS: short message service.

Data Collection Method

All studies used an abbreviated survey format to collect dietary data. The number of diet-related survey items ranged from 1 to 9 items. There were 3 studies choosing their dietary variables from intake patterns specific to the targeted population [62,67,70]. A search function linked with a national food composition database within the study app was provided in 1 study [69]. The number of overall diet-related survey items ranged from 1 to 16 items. Five studies used multiple choice options for participants to answer the survey [62,63,67,68,70]; 1 study used yes/no choice [65]; 1 study asked participants to enter number of servings [64]; 1 study asked participants to enter specific study codes that indicate servings, craving, and hunger [61]; 1 study used free text to record snacks [66]; and 1 study asked participants to record intake by searching the food database within the app [69].

Dietary Analysis and Outcomes

Response data were downloaded from the respective mobile platform by researchers to perform analysis without the use of a nutrient database for all studies. There were 7 signal-contingent studies reporting on the occurrence or frequency of (targeted) food or food group intakes at the day-level that were most relevant to the research [61-63,65,67,68,70]. Three studies focused on snack intake only [65,66,70]. Only 2 studies estimated energy intake [66,69]. Another study estimated servings of low glycemic index foods [64].

Protocol Adherence

EMA prompt response rate was reported in 8 of the 10 studies [61,63,64,66-70]. Response rates ranged from 23%-63% per day to 98% across the study period. The mean response rate across all studies was 79% with a median of 74%.

Comparison Testing

There were 2 studies comparing their mEMDA protocols against 24-hour dietary recalls [62,63]. In Bruening's study [62], participants completed 3 days of dietary recalls (2 weekdays and 1 weekend day). Each food item reported in the dietary recalls was coded to match the food groups used in the mEMDA protocol. The concordance rate at the day level ranged from 79% for entrees to 94% for fruit and vegetables in a sample of college students. In Dunton's study [63], children completed 2 days of dietary recalls. Each food item reported in the dietary recalls was coded to match the food groups used in the mEMDA protocol. Furthermore, time of food intake from the dietary recalls was matched with the 2-hour recall time windows that were used in the mEMDA protocol. The 2-hour concordance rate ranged from 65% for fruits/vegetables to 90% for soda/energy drinks in a sample of children (mean age=10 years) [74].

Discussion

Summary of Key Findings

This systematic review summarized the existing protocols for measuring dietary intake using 2 mEMDA approaches: event-contingent and signal-contingent mEMDA. Studies describing 20 unique mEMDA protocols were included in the review. Half of the studies used event-contingent mEMDA protocols and half used signal-contingent mEMDA protocols. Most studies used mobile phones to collect dietary data. Studies that used event-contingent mEMDA most commonly assessed diet in real time, used dietary records to collect data, and provided estimates of energy and nutrient intake for data collection purposes. All signal-contingent mEMDA studies used near real–time recalls and unannounced abbreviated diet surveys and assessed the frequency of consumption of foods or food groups most relevant to the research. Only 6 (30%) mEMDA studies directly compared mEMDA and dietary outcomes measured by the traditional dietary assessment methods (eg, 24-hour dietary recalls). As such, the evolving body of literature identified in this review supports the application of mEMDA as the next step for advancement in the field of dietary assessment, bridging the gap between traditional methods and newer, more technologically advanced methods (ie, biomechanical sensing and image-based food detection), which are currently under development.

Key Strengths and Limitations of Mobile Device–Assisted Ecological Momentary Diet Assessment

The mEMDA approaches described in this review have strengths and limitations potentially impacting the quality of estimated dietary intake. Event-contingent mEMDA protocols have several strengths based primarily on the fact that food (and beverage) consumption generally occurs as a discrete event; it serves as a cue to record intake [77]. These studies most commonly assessed diet in real time (eg, at the time of consumption) or used image-assisted or image-based dietary assessment methods to reduce recall bias (eg, dependence on memory) when intake was reported at a later time. Moreover, when integrated with food composition databases or similar software, event-contingent mEMDA protocols most often facilitate the calculation of or automatically calculated estimates of energy and nutrient intake. Finally, the mobile platform of event-contingent mEMDA enables eating events to be time-stamped. This removes the time-keeping burden from participants, allows for better specificity for eating occasions, and enhances the ability to examine the distribution and frequency of eating events across days or weeks [77]. There are also several limitations inherent to event-contingent mEMDA that are equally problematic as the similar traditional methods of dietary assessment (eg, written food records). First, event-contingent mEMDA requires participants to initiate the self-reporting of each eating event over the study period, which may be perceived as burdensome and may lead to omitted data when participants forget or decline to report [36]. Second, the routine anticipation of self-report of dietary intake (vs unannounced recalls) consistent with event-contingent mEMDA is more likely to be biased by psychological reactance or social desirability, whereby people change their usual eating behaviors or intentionally misreport intake so as to not be judged for making diet-related decisions perceived by the individual to be less healthful [5]. Finally, compared with traditional dietary assessment methods, event-contingent mEMDA may be limited by existing on a mobile platform, which might be less acceptable to some populations including people who are less comfortable with technology (eg, older adults) [78] or have low electronic health literacy [79]. For these individuals, written records might be preferred. In addition, compared with signal-contingent mEMDA protocols, which typically employ an abbreviated survey, event-contingent mEMDA more often involves more high-resource data processing to process raw video, photos, or sound files into analyzable food group or nutrient data at meal level or day level requiring software that might not be readily accessible to researchers. Overall, event-contingent mEMDA appears useful for capturing individuals' intake as it occurs, by eliciting a time-stamped log of all eating events and their contents; however, limitations include greater participant burden (ie, recording all food and beverage consumed and remembering to do so), the increased likelihood of psychological reactance or social desirability biases, and lower acceptance levels in some populations.

The remaining studies (n=10) used signal-contingent mEMDA sampling. All the signal-contingent mEMDA studies used a near real–time recall approach with abbreviated diet surveys, whereas some also incorporated real-time prompting. Most signal-contingent studies assessed the frequency with which foods or food groups most relevant to the research were consumed. Strengths of signal-contingent mEMDA include lower participant burden related to survey brevity and unannounced sampling, which provides a random sampling of eating events throughout the day. Here, participants receive prompts to report their recent intake throughout the day, providing a representative sample of overall daily intake without

having to proactively input details about each eating occasion as they occur. Additionally, the majority of existing signal-contingent studies used simplified reporting methods, asking participants whether or not they consumed certain target foods within a recent interval of time (eg, past 30 min or past 2 hours). Though signal-contingent mEMDA has several strengths, there are also limitations to note. First, it may be subject to incomplete data, particularly when sampling windows do not cover the entire day. As a result, some eating events may be omitted. For example, when a meal occurs outside the daily sampling window, such as very early or late in the day, participants may not be prompted to report it. Second, dietary intake captured by signal-contingent mEMDA versus 24-hour recalls is subject to lower specificity of timing, particularly when the recall window is longer (eg, >2 hours), as participants are typically asked to report whether or not food (or beverage) consumption has occurred, without elaboration about the specific time it occurred. Finally, although existing signal-contingent methods have typically used lower resource processing methods (eg, not requiring advanced training or specialized software), the resulting data may be limited to quantifying frequencies of intake as opposed to estimates of energy intake or nutrients due, in part, to the brevity of the surveys, a key aspect of signal-contingent mEMDA. Although survey brevity (and frequency of administration) is important for maintaining high response rates [80], it poses limitations to data quality with regard to how the data are summarized and the extent to which they can be used to describe dietary intake. Future research will be needed to determine how best to balance response rates with the collection of high-quality diet data. To summarize, signal-contingent mEMDA is able to capture a representative (ie, randomly selected) sample of individuals' daily intake while minimizing the participant burden associated with participant-initiated reporting; however, it may not be suited for time-stamping eating events or quantifying energy or nutrient intakes.

Another limitation of mEMDA methods is that few studies have compared mEMDA protocols with the traditional dietary assessment methods. However, the few that have conducted comparison studies have generally found acceptable agreement in estimated energy intakes and reported concordance. For instance, 2 studies that compared mEMDA reports of estimated energy intake against doubly labeled water [40,44] found that mEMDA underestimated energy intake by an average of 222 kcal/day (−15.6%) [50] and approximately 11% [40], respectively. In 4 studies event-contingent [39,41] and signal-contingent recalls [62,63] mEMDA estimated intake was compared with 24-hour dietary recalls. The 2 event-contingent studies demonstrated 85% agreement for primary diet data [41] and 90% agreement for macronutrient and energy intake [39] when compared with 24-hour dietary recalls. The 2 signal-contingent studies found day-level concordance between foods and beverages reported by mEMDA and next day 24-hour dietary recalls ranged from 79% (entrees) to 94% (fruit and vegetables) in a sample of college students [62] and 2-hour concordance ranged from 65% (fruits and vegetables) to 90% (soda and energy drinks) for children [63]. Similar magnitudes of difference have been reported when comparing traditional methods of dietary assessment (eg, Web- and computer-based 24-hour dietary recalls) with a more objective reference method (eg, direct dietary observation, doubly labeled water, and controlled-feeding studies) [81]; however, it is difficult to determine the validity of any method of dietary assessment because of the inherent measurement errors that obscure differences between the observed and the true reports. Moreover, it is important to note that these studies did not report compliance rates for mEMDA versus the comparison method. Despite these limitations, the studies highlighted in this review demonstrate the feasibility and preliminary evidence of accuracy of mEMDA in research settings. It will be important for developing mEMDA methods to conduct similar comparison studies against more objective reference methods and to compare their protocol adherence rates with the traditional methods of dietary assessment.

Strengths and Limitations of This Review

Our systematic review is the first to summarize the existing literature of mEMDA for the measurement of diet in research studies and to discuss corresponding strengths and weaknesses. This study was based on a comprehensive search across multiple domains (images, biomechanical approaches, EMA, etc) using current guidance for a robust systematic review process. This study was limited in breadth as a result of many burgeoning methods not yet being applied in research settings and were, therefore, excluded from this review. Furthermore, senior reviewers made final decisions about study inclusion rather than an independent expert, which may have introduced some selection bias. In addition, because of the wide divergence of study measures and reporting of relevant items (eg, protocol adherence), we were only able to narratively describe data collection methods. Furthermore, this review was limited in its ability to describe how accurately mEMDA methods capture diet compared with traditional methods of dietary assessment because of limited reports of such comparison statistics in the existing studies.

Implications for Future Research

Currently, there is a need within the field of momentary diet assessment to maximize data quality while minimizing participant burden. Ecologically valid and reliable data on individuals' dietary intake are essential to understand the role of diet on human health through the life span. Due to the known limitations of the existing dietary assessment methods, the research community is motivated to develop new solutions aimed at (semi)automating the assessment of dietary intake. Although the automated methods of real-time image-based detection and real-time detection of food intake by biomechanical sensors or hand-held devices have seen significant progress in terms of identifying foods and estimating portion sizes (detecting wrist or hand motion or patterns of chewing or swallowing indicative of food intake), the design and proof-of-concept data suggest the automation of dietary assessment remains out of reach for the time being. Given these limitations, mEMDA, characterized by the repeated assessment of an individual's behaviors and experiences in real time or near-real time in their natural settings, represents a novel dietary assessment method. By reviewing the literature and identifying key patterns, strengths, and weaknesses of the existing

momentary diet assessment methods, a unique topic of high relevance in the dietary assessment community and EMA community, researchers may better understand and move forward with improving and incorporating mEMDA into their own research. Compared with the traditional methods, mEMDA may reduce participant burden and recall biases while maximizing ecological validity. Therefore, mEMDA has the potential to bridge the gap between currently available methods (eg, 24-hour dietary recall) and newer methods (eg, biomechanical sensors), which are currently under development.

The strengths of event-contingent mEMDA (eg, ability to capture the full day of dietary intake and estimate energy and nutrient intakes) and of signal-contingent mEMDA (eg, lower participant burden and unannounced prompting schemes) could potentially be leveraged to design novel mEMDA methods that reduce their individual limitations. On the basis of these early studies, efforts now need to be focused on standardizing mEMDA methods with the goal of maximizing dietary data quality and ecological validity while minimizing participant burden. With improved standardization, it is likely that validated mEMDA tools will become more widely available to researchers in the future as most methods have been study-specific. Existing studies illustrate the wide range of dietary outcomes assessed through mEMDA methods, and although the ability to develop and tailor assessment items based on a particular study's needs is an advantage, the divergence of outcome measures and lack of validation remains a major challenge.

Acknowledgments

The preparation of this manuscript was supported in part by federal grants NSF1654213, R21CA215415, R01HL119255, R01CA186700, F31HL137346, R25ECA056452, and P30CA016672; the Chandler Cox Foundation; the Duncan Family Institute for Cancer Prevention and Risk Assessment; The UT MD Anderson Bionutrition Research Core; and the Center for Energy Balance in Cancer Prevention and Survivorship.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Extracted data list.

[PDF File (Adobe PDF File), 23KB - mhealth_v6i11e11170_app1.pdf]

References

1. US Burden of Disease Collaborators, Mokdad AH, Ballestros K, Echko M, Glenn S, Olsen HE, et al. The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States. J Am Med Assoc 2018 Dec 10;319(14):1444-1472 [FREE Full text] [doi: 10.1001/jama.2018.0158] [Medline: 29634829]
2. Thompson F, Subar A. Dietary assessment methodology. Amsterdam, Netherlands: Elsevier Inc; 2017. URL: https://epi.grants.cancer.gov/dietary-assessment/Chapter%201_Coulston.pdf
3. Wansink B. Mindless eating: Why we eat more than we think. New York: Bantam Books; 2006.
4. Johnson RK, Soultanakis RP, Matthews DE. Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall: a doubly labeled water study. J Am Diet Assoc 1998 Oct;98(10):1136-1140. [doi: 10.1016/S0002-8223(98)00263-6] [Medline: 9787719]
5. Maurer J, Taren DL, Teixeira PJ, Thomson CA, Lohman TG, Going SB, et al. The psychosocial and behavioral characteristics related to energy misreporting. Nutr Rev 2006 Feb;64(2 Pt 1):53-66. [doi: 10.1111/j.1753-4887.2006.tb00188.x] [Medline: 16536182]
6. Mendez MA, Popkin BM, Buckland G, Schroder H, Amiano P, Barricarte A, et al. Alternative methods of accounting for underreporting and overreporting when measuring dietary intake-obesity relations. Am J Epidemiol 2011 Feb 15;173(4):448-458 [FREE Full text] [doi: 10.1093/aje/kwq380] [Medline: 21242302]
7. Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't Veer P. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. Br J Nutr 2009 Jul;101(Suppl 2):S73-S85. [doi: 10.1017/S0007114509990602] [Medline: 19594967]
8. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol 2003 Jul 01;158(1):1-13. [Medline: 12835280]
9. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
10. Rhee JJ, Sampson L, Cho E, Hughes MD, Hu FB, Willett WC. Comparison of methods to account for implausible reporting of energy intake in epidemiologic studies. Am J Epidemiol 2015 Feb 15;181(4):225-233 [FREE Full text] [doi: 10.1093/aje/kwu308] [Medline: 25656533]

11. Mendez MA. Invited commentary: dietary misreporting as a potential source of bias in diet-disease associations: future directions in nutritional epidemiology research. Am J Epidemiol 2015 Feb 15;181(4):234-236. [doi: 10.1093/aje/kwu306] [Medline: 25656531]
12. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010 Jan;110(1):48-51 [FREE Full text] [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
13. McCabe-Sellers B. Advancing the art and science of dietary assessment through technology. J Am Diet Assoc 2010 Jan;110(1):52-54. [doi: 10.1016/j.jada.2009.10.014] [Medline: 20102827]
14. Zhu F, Bosch M, Boushey CJ, Delp EJ. An image analysis system for dietary assessment and evaluation. Proc Int Conf Image Proc 2010:1853-1856 [FREE Full text] [doi: 10.1109/ICIP.2010.5650848] [Medline: 22025261]
15. Probst Y, Nguyen DT, Tran MK, Li W. Dietary assessment on a mobile phone using image processing and pattern recognition techniques: algorithm design and system prototyping. Nutrients 2015 Jul 27;7(8):6128-6138 [FREE Full text] [doi: 10.3390/nu7085274] [Medline: 26225994]
16. Jia W, Yue Y, Fernstrom JD, Yao N, Sclabassi RJ, Fernstrom MH, et al. Imaged based estimation of food volume using circular referents in dietary assessment. J Food Eng 2012 Mar;109(1):76-86 [FREE Full text] [doi: 10.1016/j.jfoodeng.2011.09.031] [Medline: 22523440]
17. Nelson M, Atkinson M, Darbyshire S. Food photography II: use of food photographs for estimating portion size and the nutrient content of meals. Br J Nutr 1996 Jul;76(1):31-49. [Medline: 8774215]
18. Desendorf J, Bassett DR, Raynor HA, Coe DP. Validity of the Bite Counter device in a controlled laboratory setting. Eat Behav 2014 Aug;15(3):502-504. [doi: 10.1016/j.eatbeh.2014.06.013] [Medline: 25064306]
19. Dong Y, Hoover A, Scisco J, Muth E. A new method for measuring meal intake in humans via automated wrist motion tracking. Appl Psychophysiol Biofeedback 2012 Sep;37(3):205-215 [FREE Full text] [doi: 10.1007/s10484-012-9194-1] [Medline: 22488204]
20. Scisco JL, Muth ER, Hoover AW. Examining the utility of a bite-count-based measure of eating activity in free-living human beings. J Acad Nutr Diet 2014 Mar;114(3):464-469. [doi: 10.1016/j.jand.2013.09.017] [Medline: 24231364]
21. Kalantarian H, Sarrafzadeh M. Audio-based detection and evaluation of eating behavior using the smartwatch platform. Comput Biol Med 2015 Oct 01;65:1-9. [doi: 10.1016/j.compbiomed.2015.07.013] [Medline: 26241487]
22. Amft O, Tröster G. Recognition of dietary activity events using on-body sensors. Artif Intell Med 2008 Feb;42(2):121-136. [doi: 10.1016/j.artmed.2007.11.007] [Medline: 18242066]
23. Fontana JM, Sazonov ES. Evaluation of chewing and swallowing sensors for monitoring ingestive behavior. Sens Lett 2013 Mar;11(3):560-565 [FREE Full text] [Medline: 25484630]
24. Päßler S, Wolff M, Fischer WJ. Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food. Physiol Meas 2012 Jun;33(6):1073-1093. [doi: 10.1088/0967-3334/33/6/1073] [Medline: 22621915]
25. Sazonov E, Schuckers S, Lopez-Meyer P, Makeyev O, Sazonova N, Melanson EL, et al. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. Physiol Meas 2008 May;29(5):525-541 [FREE Full text] [doi: 10.1088/0967-3334/29/5/001] [Medline: 18427161]
26. Kalantarian H, Alshurafa N, Le T, Sarrafzadeh M. Monitoring eating habits using a piezoelectric sensor-based necklace. Comput Biol Med 2015 Mar;58:46-55. [doi: 10.1016/j.compbiomed.2015.01.005] [Medline: 25616023]
27. Dong B, Biswas S. Wearable diet monitoring through breathing signal analysis. Conf Proc IEEE Eng Med Biol Soc 2013;2013:1186-1189. [doi: 10.1109/EMBC.2013.6609718] [Medline: 24109905]
28. Farooq M, Fontana JM, Sazonov E. A novel approach for food intake detection using electroglottography. Physiol Meas 2014 May;35(5):739-751 [FREE Full text] [doi: 10.1088/0967-3334/35/5/739] [Medline: 24671094]
29. Farooq M, Sazonov E. A novel wearable device for food intake and physical activity recognition. Sensors (Basel) 2016 Jul 11;16(7) [FREE Full text] [doi: 10.3390/s16071067] [Medline: 27409622]
30. Zhu F, Bosch M, Boushey CJ, Delp EJ. An image analysis system for dietary assessment and evaluation. Proc Int Conf Image Proc 2010:1853-1856 [FREE Full text] [doi: 10.1109/ICIP.2010.5650848] [Medline: 22025261]
31. Chen N, Lee YY, Rabb M, Schatz B. Toward dietary assessment via mobile phone video cameras. AMIA Annu Symp Proc 2010 Nov 13;2010:106-110 [FREE Full text] [Medline: 21346950]
32. Daugherty BL, Schap TE, Ettienne-Gittens R, Zhu FM, Bosch M, Delp EJ, et al. Novel technologies for assessing dietary intake: evaluating the usability of a mobile telephone food record among adults and adolescents. J Med Internet Res 2012;14(2):e58 [FREE Full text] [doi: 10.2196/jmir.1967] [Medline: 22504018]
33. Kawano Y, Yanai K. Real-time mobile food recognition system. In: IEEE Comput Soc Conf. 2013 Presented at: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops; June 23-28, 2013; Potland, OR, US p. 1-7. [doi: 10.1109/CVPRW.2013.5]
34. Kong FY, He HS, Raynor HA, Tan JD. DietCam: multi-view regular shape food recognition with a camera phone. Pervasive Mob Comput 2015 May;19:108-121. [doi: 10.1016/j.pmcj.2014.05.012]
35. Rahman M, Pickering M, Kerr D, Boushey C, Delp E. A new texture feature for improved food recognition accuracy in a mobile phone based dietary assessment system. 2012 Presented at: IEEE Int Conf Multi Expo Workshops; July 9-13, 2012; Melbourne, Australia p. 418-423. [doi: 10.1109/Icmew.2012.79]

36. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
37. 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 health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
38. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: Wiley, John & Sons, Incorporated; 2008.
39. Ashman AM, Collins CE, Brown LJ, Rae KM, Rollo ME. Validation of a smartphone image-based dietary assessment method for pregnant women. Nutrients 2017 Jan 18;9(1) [FREE Full text] [doi: 10.3390/nu9010073] [Medline: 28106758]
40. Boushey CJ, Spoden M, Delp EJ, Zhu F, Bosch M, Ahmad Z, et al. Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. Nutrients 2017 Mar 22;9(3) [FREE Full text] [doi: 10.3390/nu9030312] [Medline: 28327502]
41. Bucher Della Torre S, Carrard I, Farina E, Danuser B, Kruseman M. Development and evaluation of e-CA, an electronic mobile-based food record. Nutrients 2017 Jan 18;9(1) [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
42. Grenard JL, Stacy AW, Shiffman S, Baraldi AN, MacKinnon DP, Lockhart G, et al. Sweetened drink and snacking cues in adolescents: a study using ecological momentary assessment. Appetite 2013 Aug;67:61-73 [FREE Full text] [doi: 10.1016/j.appet.2013.03.016] [Medline: 23583312]
43. Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, et al. Collection and visualization of dietary behavior and reasons for eating using Twitter. J Med Internet Res 2013;15(6):e125 [FREE Full text] [doi: 10.2196/jmir.2613] [Medline: 23796439]
44. Martin CK, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, et al. Validity of the remote food photography method (rfpm) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899 [FREE Full text] [doi: 10.1038/oby.2011.344] [Medline: 22134199]
45. Schüz B, Bower J, Ferguson SG. Stimulus control and affect in dietary behaviours. An intensive longitudinal study. Appetite 2015 Apr;87:310-317. [doi: 10.1016/j.appet.2015.01.002] [Medline: 25579222]
46. Seto E, Hua J, Wu L, Shia V, Eom S, Wang M, et al. Models of individual dietary behavior based on smartphone data: The influence of routine, physical activity, emotion, and food environment. PLoS One 2016;11(4):e0153085 [FREE Full text] [doi: 10.1371/journal.pone.0153085] [Medline: 27049852]
47. Thomas JG, Bond DS, Ryder BA, Leahey TM, Vithiananthan S, Roye GD, et al. Ecological momentary assessment of recommended postoperative eating and activity behaviors. Surg Obes Relat Dis 2011;7(2):206-212. [doi: 10.1016/j.soard.2010.10.007] [Medline: 21130703]
48. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. Dialbetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
49. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr 2009 Feb;101(3):446-456 [FREE Full text] [doi: 10.1017/S0007114508027438] [Medline: 18616837]
50. Nicklas T, Saab R, Islam NG, Wong W, Butte N, Schulin R, et al. Validity of the remote food photography method against doubly labeled water among minority preschoolers. Obesity (Silver Spring) 2017 Dec;25(9):1633-1638 [FREE Full text] [doi: 10.1002/oby.21931] [Medline: 28758370]
51. Kato S, Waki K, Nakamura S, Osada S, Kobayashi H, Fujita H, et al. Validating the use of photos to measure dietary intake: the method used by DialBetics, a smartphone-based self-management system for diabetes patients. Diabetol Int 2015 Oct 14;7(3):244-251. [doi: 10.1007/s13340-015-0240-0]
52. Seto E, Hua J, Wu L, Bestick A, Shia V, Eom S, et al. The Kunming CalFit study: modeling dietary behavioral patterns using smartphone data. Conf Proc IEEE Eng Med Biol Soc 2014;2014:6884-6887. [doi: 10.1109/EMBC.2014.6945210] [Medline: 25571578]
53. Aflague TF, Boushey CJ, Guerrero RTL, Ahmad Z, Kerr DA, Delp EJ. Feasibility and use of the mobile food record for capturing eating occasions among children ages 3-10 years in Guam. Nutrients 2015 Jun;7(6):4403-4415 [FREE Full text] [doi: 10.3390/nu7064403] [Medline: 26043037]
54. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, et al. Evidence-based development of a mobile telephone food record. J Am Diet Assoc 2010 Jan;110(1):74-79 [FREE Full text] [doi: 10.1016/j.jada.2009.10.010] [Medline: 20102830]
55. Bathgate KE, Sherriff JL, Leonard H, Dhaliwal SS, Delp EJ, Boushey CJ, et al. Feasibility of assessing diet with a mobile food record for adolescents and young adults with down syndrome. Nutrients 2017 Mar 13;9(3) [FREE Full text] [doi: 10.3390/nu9030273] [Medline: 28335382]
56. Kerr DA, Dhaliwal SS, Pollard CM, Norman R, Wright JL, Harray AJ, et al. BMI is associated with the willingness to record diet with a mobile food record among adults participating in dietary interventions. Nutrients 2017 Mar 07;9(3) [FREE Full text] [doi: 10.3390/nu9030244] [Medline: 28272343]
57. Kerr DA, Harray AJ, Pollard CM, Dhaliwal SS, Delp EJ, Howat PA, et al. The connecting health and technology study: a 6-month randomized controlled trial to improve nutrition behaviours using a mobile food record and text messaging support

in young adults. Int J Behav Nutr Phys Act 2016 Apr 21;13:52 [FREE Full text] [doi: 10.1186/s12966-016-0376-8] [Medline: 27098449]
58. Kerr DA, Pollard CM, Howat P, Delp EJ, Pickering M, Kerr KR, et al. Connecting health and technology (chat): Protocol of a randomized controlled trial to improve nutrition behaviours using mobile devices and tailored text messaging in young adults. BMC Public Health 2012;12:477 [FREE Full text] [doi: 10.1186/1471-2458-12-477] [Medline: 22726532]
59. Elliston KG, Ferguson SG, Schüz N, Schüz B. Situational cues and momentary food environment predict everyday eating behavior in adults with overweight and obesity. Health Psychol 2017 Dec;36(4):337-345. [doi: 10.1037/hea0000439] [Medline: 27669177]
60. Waki K, Aizawa K, Kato S, Fujita H, Lee H, Kobayashi H, et al. Dialbetics with a multimedia food recording tool, foodlog: smartphone-based self-management for type 2 diabetes. J Diabetes Sci Technol 2015 May;9(3):534-540 [FREE Full text] [doi: 10.1177/1932296815579690] [Medline: 25883164]
61. Berkman ET, Giuliani NR, Pruitt AK. Comparison of text messaging and paper-and-pencil for ecological momentary assessment of food craving and intake. Appetite 2014 Oct;81:131-137. [doi: 10.1016/j.appet.2014.06.010] [Medline: 24930596]
62. Bruening M, Ohri-Vachaspati P, Brewis A, Laska M, Todd M, Hruschka D, et al. Longitudinal social networks impacts on weight and weight-related behaviors assessed using mobile-based ecological momentary assessments: study protocols for the SPARC study. BMC Public Health 2016 Dec 30;16:901 [FREE Full text] [doi: 10.1186/s12889-016-3536-5] [Medline: 27576358]
63. Dunton GF, Liao Y, Dzubur E, Leventhal AM, Huh J, Gruenewald T, et al. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: protocol for the MATCH study. Contemp Clin Trials 2015 Jul;43:142-154 [FREE Full text] [doi: 10.1016/j.cct.2015.05.007] [Medline: 25987483]
64. Miller CK, Weinhold KR, Mitchell DC. Using ecological momentary assessment to track goal progress toward the adoption of a low glycemic index diet among adults with type 2 diabetes: a pilot study. Top Clin Nutr 2016;31(4):323-334. [doi: 10.1097/TIN.0000000000000083]
65. Powell DJ, McMinn D, Allan JL. Does real time variability in inhibitory control drive snacking behavior? An intensive longitudinal study. Health Psychol 2017 Dec;36(4):356-364. [doi: 10.1037/hea0000471] [Medline: 28192005]
66. Richard A, Meule A, Reichenberger J, Blechert J. Food cravings in everyday life: an EMA study on snack-related thoughts, cravings, and consumption. Appetite 2017 Dec 01;113:215-223. [doi: 10.1016/j.appet.2017.02.037] [Medline: 28249745]
67. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
68. Strahler J, Nater UM. Differential effects of eating and drinking on wellbeing-an ecological ambulatory assessment study. Biol Psychol 2018 Dec;131:72-88. [doi: 10.1016/j.biopsycho.2017.01.008] [Medline: 28119068]
69. Wouters S, Thewissen V, Duif M, Lechner L, Jacobs N. Assessing energy intake in daily lifeignal-contingent smartphone application versus event-contingent paper and pencil estimated diet diary. Psychologica Belgica 2016;56(4):357-369. [doi: 10.5334/pb.339] [Medline: 120684512]
70. Zenk SN, Horoi I, McDonald A, Corte C, Riley B, Odoms-Young AM. Ecological momentary assessment of environmental and personal factors and snack food intake in African American women. Appetite 2014 Dec;83:333-341. [doi: 10.1016/j.appet.2014.09.008] [Medline: 25239402]
71. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A mobile ecological momentary assessment tool (devilsparc) for nutrition and physical activity behaviors in college students: a validation study. J Med Internet Res 2016 Jul 27;18(7):e209 [FREE Full text] [doi: 10.2196/jmir.5969] [Medline: 27465701]
72. Liao Y, Schembre SM, O'Connor SG, Belcher BR, Maher JP, Dzubur E, et al. An electronic ecological momentary assessment study to examine the consumption of high-fat/high-sugar foods, fruits/vegetables, and affective states among women. J Nutr Educ Behav 2018 Mar 21;50(6):626-631. [doi: 10.1016/j.jneb.2018.02.003] [Medline: 29573964]
73. O'Connor SG, Koprowski C, Dzubur E, Leventhal AM, Huh J, Dunton GF. Differences in mothers' and children's dietary intake during physical and sedentary activities: an ecological momentary assessment study. J Acad Nutr Diet 2017 Aug;117(8):1265-1271. [doi: 10.1016/j.jand.2017.02.012] [Medline: 28392348]
74. O'Connor SG, Ke W, Dzubur E, Schembre S, Dunton GF. Concordance and predictors of concordance of children's dietary intake as reported via ecological momentary assessment and 24 h recall. Public Health Nutr 2018 Apr;21(6):1019-1027. [doi: 10.1017/S1368980017003780] [Medline: 29352820]
75. Wouters S, Jacobs N, Duif M, Lechner L, Thewissen V. Affect and between-meal snacking in daily life: the moderating role of gender and age. Psychol Health 2018 Apr;33(4):555-572. [doi: 10.1080/08870446.2017.1380813] [Medline: 28934860]
76. McMinn D, Allan JL. The SNAPSHOT study protocol: SNAcking, Physical activity, Self-regulation, and Heart rate Over Time. BMC Public Health 2014 Sep 26;14:1006 [FREE Full text] [doi: 10.1186/1471-2458-14-1006] [Medline: 25261200]
77. Shiffman S. Designing protocols for ecological momentary assessment. In: Stone AA, Shiffman S, Atienza AA, Nebling L, editors. The Science of Real-Time Data Capture: Self-Reports in Health Research. New York, NY: Oxford University Press; 2007:27-53.

78. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
79. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
80. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [Medline: 12173681]
81. Timon CM, van der Berg R, Blain RJ, Kehoe L, Evans K, Walton J, et al. A review of the design and validation of web- and computer-based 24-h dietary recall tools. Nutr Res Rev 2016 Dec;29(2):268-280. [doi: 10.1017/S0954422416000172] [Medline: 27955721]

Abbreviations

e-CA: electronic carnet alimentaire
EMA: ecological momentary assessment
mEMA: mobile device–assisted ecological momentary assessment
mEMDA: mobile device–assisted ecological momentary diet assessment
mFR: mobile Food Record
PDA: personal digital assistant
PRISMA: Preferred Reporting of Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 31.05.18; peer-reviewed by D Steinberg, Y Probst; comments to author 09.07.18; revised version received 03.10.18; accepted 10.10.18; published 20.11.18.

Please cite as:
Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen SE, Hamoy KG, Huh J, Dunton GF, Weiss R, Thomson CA, Boushey CJ
Mobile Ecological Momentary Diet Assessment Methods for Behavioral Research: Systematic Review
JMIR Mhealth Uhealth 2018;6(11):e11170
URL: http://mhealth.jmir.org/2018/11/e11170/
doi:10.2196/11170
PMID:30459148

Original Paper

Providing a Smart Healthy Diet for the Low-Income Population: Qualitative Study on the Usage and Perception of a Designed Cooking App

Faustine Régnier[1], PhD; Manon Dugré[2], MA; Nicolas Darcel[2], PhD; Camille Adamiec[1], PhD

[1]Institut National de la Recherche Agronomique, Alimentation et Sciences Sociales Unité de Recherche 1303, University of Paris Saclay, Ivry-sur-Seine cedex, France

[2]AgroParisTech, UMR 914 Physiology of Nutrition and Ingestive Behavior, Paris, France

Corresponding Author:
Faustine Régnier, PhD
Institut National de la Recherche Agronomique
Alimentation et Sciences Sociales Unité de Recherche 1303, University of Paris Saclay
65 boulevard de Brandebourg
Ivry-sur-Seine cedex, 94205
France
Phone: 33 49596914
Fax: 33 49596990
Email: faustine.regnier@inra.fr

Abstract

Background: Health behaviors among low-income groups have become a major issue in the context of increasing social inequalities. The low-income population is less likely to be receptive to nutritional recommendations, but providing cooking advice could be more effective. In this domain, taking advantage of digital devices can be a bonus with its own challenges.

Objective: The aim of this study was to develop and deploy NutCracker, a social network–based cooking app for low-income population, including cooking tips and nutritional advices, aiming at creating small online communities. We further determined the usefulness, perceptions, barriers, and motivators to use NutCracker.

Methods: The smartphone app, designed jointly with beneficiaries of the social emergency services, was implemented in a disadvantaged neighborhood of Magny, (Paris region, France). Once the app became available, 28 subjects, living in the neighborhood, tested the app for a 6-month period. Logs to the app and usages were collected by the software. In total, 12 in-depth, semistructured interviews were conducted among the users and the social workers to analyze their uses and perceptions of the app relative to their interest in cooking, cooking skills, socioeconomic constraints, and social integration. These interviews were compared with 21 supplementary interviews conducted among low-income individuals in the general population.

Results: NutCracker was developed as a social network–based app, and it includes cooking tips, nutritional advice, and Web-based quizzes. We identified barriers to uses (especially technical barriers, lack of knowledge in the field of new technologies and written comprehension, and search for real contacts) and motivators (in particular, good social integration, previous use of social networks, and help of children as intermediaries). Cooking skills were both a barrier and a lever.

Conclusions: Targeting the low-income groups through a cooking app to promote healthier behaviors offers many advantages but has not been fully explored. However, the barriers in low-income milieu remain high, especially among the less socially integrated strata. Lessons from this intervention allow us to identify barriers and possible levers to improve nutrition promotion and awareness in deprived areas, especially in the time of social crisis.

(JMIR Mhealth Uhealth 2018;6(11):e11176) doi:10.2196/11176

KEYWORDS

health; cooking; mobile phone; low-income population; social networks; utilization

Introduction

Background

Current health promotion campaigns that aim at motivating individuals toward healthier eating habits are inefficient among low-income populations. This situation is particularly concerning in the current context where social inequalities in health are very marked and are rising in western countries [1]. Health behaviors among the lower strata appear, more than ever, as a major issue. In the context of a deep economic crisis, leading to increasing health and food consumption inequalities [2], low-income populations appear to face more food insecurity than before [3]. They eat less recommended food products, especially fruits and vegetables [4].

Moreover, public health campaigns, developed by the French National Nutrition and Health Program, focusing on the effect of a proper diet on health, were not able to reach people in low-income categories. Although the disadvantaged populations are well aware of the guidelines for healthy eating [2,5], they do not implement them because of their social lifestyle, tastes, and preferences.

Consequently, we observe an unequal prevalence of obesity [2] and diabetes [6], much more frequently in low-income population, leading to the need to focus on these disadvantaged social categories.

There is need of more efficient strategies to enhance nutritional awareness and promote healthy behaviors in disadvantaged populations through novel interventions specifically targeting them. Hence, the use of interventions that include a cooking component seem to be an effective way to enhance the adoption of better eating behaviors [7], especially among low-income individuals [8].

Objectives

Digital technologies could offer a good platform to help build and disseminate appropriate individualized recommendations targeting special social groups. As described by the European Commission's Green Paper on mobile health (mHealth) [9], mHealth solutions appear as a solid basis for people's empowerment. However, not many culinary apps with health promotion prospects have been developed for mobile devices. NutCracker—continued in the FacilEat4All project—is an interventional research project based on the development of a cooking app to promote healthy nutrition in low-income groups and on its evaluation. This app was jointly designed with the final beneficiaries.

The project took place in disadvantaged neighborhoods of Magny (Yvelines, Paris Region, France). This city had a total population of 32,639 inhabitants in 2015 [10], mainly of a low social level, with 24.6% of its population living below the poverty threshold (national average 14.1%) [11]. The rate of unemployment is high (19.4% vs national average 10.4% [11]). Participants in this study were volunteers recruited via social care services and local nongovernmental organizations (NGOs).

In this study, we aimed to (1) present the coconstruction of NutCracker, specifically designed for low-income people, promoting a healthy diet; (2) describe the results of its usage and determine the barriers and levers of this process; and (3) discuss the main lessons that can be drawn from this study.

Methods

App Design: Recruitment and Coconstruction

Our study was composed of 2 parts—an app design and an app evaluation—and was based on 2 subsamples.

We worked with the disadvantaged households of Magny, Paris region, France to design a mobile app for healthy cooking and nutrition. The app was jointly designed with its future beneficiaries in a 2-step process. First, questionnaires were distributed among 46 participants with the help of the local social workers. The questionnaire focused on the socioeconomic status of the respondents, the foods commonly consumed, food supply, expectations and general perception toward food, importance of eating, importance granted to nutritional issues, level of culinary skills, and the availability of cooking equipment. Second, a collective discussion with 8 participants was organized to discuss about a new mobile phone app providing healthy and affordable cooking advices and recipes to determine the participants' expectations toward such a device. The objective of this discussion was to launch a collective dynamic as well as to reach a consensus on the overall design of the mobile app and provide general recommendations on its features. We used the *persona* method [12] to remove social inhibition stemming from individuals own situations [13]. During the collective discussion, which was audio-recorded and transcribed, we followed a discussion guide in which a series of 4 fictional characters (established on the categories derived from the previous questionnaires), representing an individual that might use the mobile app, was introduced. Participants were then asked to give their opinion on the features of the app that would best suit this *persona*'s needs and expectations. Following a suggestion by 1 of the participants, others were asked to add their thoughts and voice their opinion about this suggestion and to go into as much detail as possible. The outcomes of these sessions were used as the outline for the app development (see Multimedia Appendix 1).

When this app was released, it was tested by 28 volunteers (8 participants from the collective discussion plus 20 supplementary volunteers) from the same neighborhood and recruited in collaboration with beneficiaries of 3 social emergency services in Magny.

App Evaluation: Recruitment and Semistructured Interviews

Data on individual app usage (logs and visited features) were collected, and a qualitative study was performed to assess this app.

We conducted 33 semistructured interviews using the 32-item Consolidated Criteria for Reporting Qualitative Research checklist [14]. To evaluate the usage and perceptions of the participants on the NutCracker app, a field study, including semistructured, face-to-face interviews, was conducted on individual project participants (12 in total, comprising 8 users

and 4 social workers who were involved in the project; Figure 1). Our interview guide contained a series of open-ended questions about their use of the app, their favorite features, the barriers encountered and, more generally, their use of digital devices and the Web with regard to cooking and in other areas, that is, social networking, gaming, Web-based purchases, and performing administrative tasks. All participants were from the municipality of Magny.

Along with this sample, and to compare with digital practices in low-income population, 21 interviews were conducted in the Paris region, among individuals of the lower strata (Figure 1). We investigated these individuals' exiting attitudes toward and acceptance of the digital world and their use of new technologies in all their forms. This additional sample allowed us to determine whether the field results for NutCracker revealed specificities linked to the app or if they reflected a larger, shared trends related to digital cooking tools in deprived areas.

These interviews, conducted by the 2 project sociologists, lasted on average of 1 hour, were face-to-face interviews, were recorded (except for 1 participant who refused), transcribed in their entirety by a team of transcription consultants, and anonymized. Analyses of the data were double-checked by the 2 sociologists, and then discussed and validated with the team involved in the project (ie, the 4 researchers who were directly involved and the scientific committee of 5 experts). We developed a content analysis, and the 3 main themes investigated using interview guide were as follows: (1) uses; (2) barriers and levers in using NutCracker; and (3) digital devices in the field of cooking, diet, and other related areas. Among the themes, subthemes were identified in line with the themes in the interview guide, and new themes derived from the data collected were also included.

Samples Included in the Study

Our study was based on 2 subsamples (Figure 1).

In the NutCracker sample (Table 1), participants were all women, which further shows that in low-income families, women are often responsible for domestic tasks related to food [15]. These were women from underprivileged backgrounds, who were foreign housewives. Their living conditions were unstable and highly dependent on social services. Due to unemployment, the social integration of these women was based entirely on the social group they belong to and on family integration [16]. In comparison, the general population sample contained individuals from low-income categories who were better off socially, either because of having a job (they were employed either as blue collar or employee) or because they lived in more socially diverse municipalities than Magny. Participants from both samples were aged between 28 and 58 years, with a median age of 46 years.

Ethical Consideration

For the collective discussion, participants were informed about the purpose of the discussion before the interview. For the semistructured interviews, the goals of this study were explained to the interviewees, and individual consent was obtained for the recording. Personal data from the interviews, including the name of the place the project took place in, were anonymized.

Figure 1. Subsamples of the study.

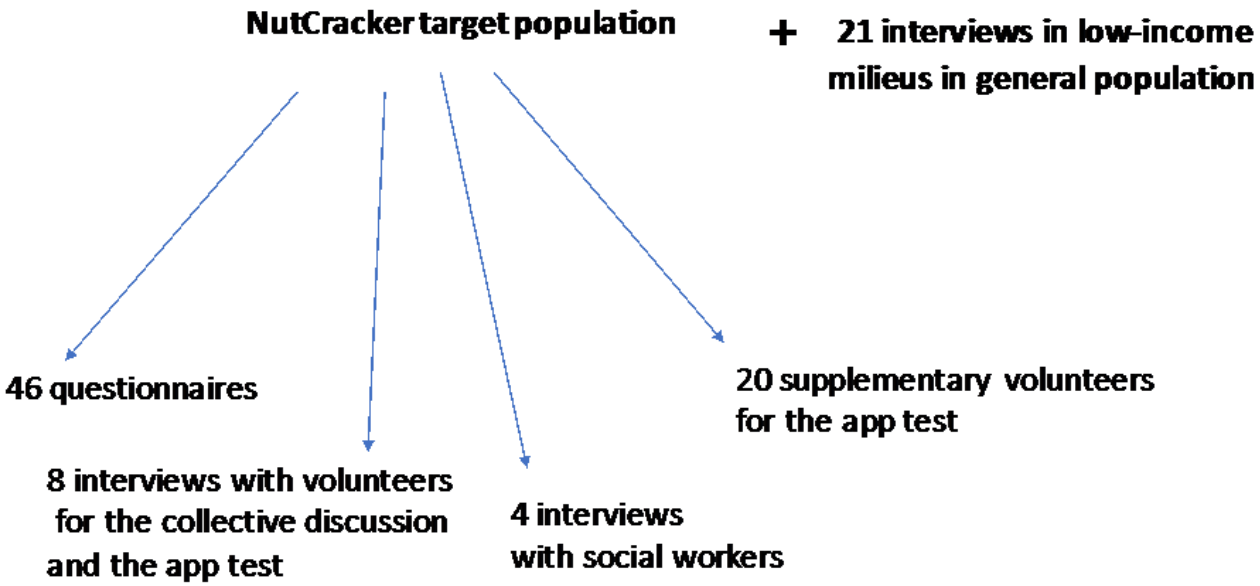


Table 1. Sociodemographic characteristics of the participants.

Sociodemographic variables	NutCracker target population (n=8), n (%)	NutCracker social workers (n=4), n (%)	Low-income milieus in the general population (n=21), n (%)	Total
Sex				
Male	0 (0)	2 (50)	3 (14)	5
Female	8 (100)	2 (50)	18 (86)	28
Age (in years)				
19-29	0 (0)	1 (25)	7 (33)	8
30-39	2 (25)	2 (50)	5 (24)	9
40-49	5 (63)	1 (25)	5 (24)	11
50+	1 (13)	0 (0)	4 (19)	5
Occupation				
Intermediate professions	0 (0)	4 (100)	0 (0)	4
Employee	2 (25)	N/A[a]	12 (57)	14
Manual workers	0 (0)	N/A	4 (19)	4
Unemployed	6 (75)	N/A	2 (10)	8
Students	0 (0)	N/A	3 (14)	3
Total	8 (100)	4 (100)	21 (100)	33

[a]N/A: not applicable.

Results

Joint Design of the App

The results from analyzing the 46 questionnaires emphasized the importance of food abundance and variety, the desire to please children with food, and the value put on national brands as markers of food quality. Although generic nutrition recommendations are very well known, the importance of nutritional issues, such as the criteria for food choices, appeared to be less important for the low-income population than for people with higher social ranking. Respondents declared the frequent use of mobile technologies, with 46 owning a smartphone (ie, 72%, whereas the national average was 58% in 2015 [17]). We noted a substantial use of social networks (ie, Snapchat, WhatsApp, Facebook, Twitter, and Instagram). When questioned about their expectations from a *nutrition program*, respondents pointed to the need for tips on cheap and healthy foods. Indeed, in the questionnaire, 18 out of 46 responded that they had insufficient fund to eat well, and 26 responded that they would eat better quality foods if they had more money (maybe as an echo of the several public health campaigns launched by the French National Nutrition and Health Program). Using leftovers appeared in the top 5 responses to the question on what they would like to find in a mobile nutrition app. Respondents also indicated the need for cooking tips (1 of the top 5 responses to the question on their expectations from a mobile nutrition app). The keys for better understanding of nutritional information and food labeling were also part of the expectations, with 9 of the 46 respondents claiming they had insufficient knowledge of nutrition. During the collective discussion, most of the participants (6/8) hinted on the importance of providing the price or low budget labeling choices for the suggested recipes as well as tips to cook with leftover food. They also asked for practical solutions to meet the dietetic advices: theoretical knowledge in health and nutrition had to be, for the respondents, translated into cooking tips and daily assistance. The mobile app created, based on this groundwork, consisted of 4 features: (1) a collaborative Web-based recipe book, which included recipes with inexpensive and good nutritional quality foods; (2) a series of simple nutritional information and cooking tips; (3) a collection of nutritional quizzes; and (4) a social network component allowing users to share, like, and comment on recipes, or share their achievements in the quizzes. The app was also designed to allow the collection of logs and usages from individual users. Data on the number and dates of logs to the app per user; number of viewed recipes per user; quizzes taken and percentage of correct responses; and recipes entered, shared, and liked were collected.

Once the app was created, it was released and made available for download to all study participants. A total of 28 households who met the selection criteria mentioned above were selected to use the mobile app over a period of 6 months (from May 2016 to October 2016).

Limited Uses of NutCracker and New Technologies

The postintervention field research highlighted a very limited use of NutCracker among its intended target population, namely, the women (Table 2). Of the 28 women who were presented with the app, only 7 downloaded it, 4 used it, and only in 2 was, thus, a regular occurrence. Only 1 participant used NutCracker actively, published recipes (9 recipes shared), and answered quizzes (she answered more than 200 questions).

By contrast, the app was much more successful among the general population, where it was shared through a network of

interconnections. Between May 2016 and October 2016, 378 connections were made, 600 recipes were viewed, 18 recipes were shared via the app, and responses to 1200 quiz question were supplied by the users.

Generally, participants from Magny had more limited access to the internet and to new technologies than participants from the low-income milieus in the general population (Table 2). Participants in the general population used the internet to access cooking websites and to perform Web-based trade, games, or administrative tasks more frequently. Diet or physical self-tracking devices were not used by the target population.

Barriers to the Use of NutCracker and New Technologies

On the basis of the inductive thematic analysis methodology described in the study by Peng et al [18], several types of reasons for not using NutCracker were identified (Table 3).

Technical Barriers

The most frequent barrier cited during the interviews was the technical instability of the prototype NutCracker leading to recurrent interruptions during the sessions. Another factor was the necessity to provide user name and password at every session as well as not being able to quickly reset the password.

Table 2. Uses of NutCracker and new technologies by the NutCracker sample and the low-income milieu population.

Uses	Verbatim	NutCracker target population (n=8), n (%)	Low-income milieus in the general population (n=21), n (%)
NutCracker app	"[NutCracker], I don't have it on my phone anymore, I don't know why. I had downloaded it and it was really interesting."	4 (50)	Not applicable
Cooking apps or websites	"Marmiton website...it's complete, and it works really well."	5 (63)	14 (67)
Web-based trade	"I sell things on eBay."	0 (0)	5 (24)
Web-based games	"I play poker on the Internet."	0 (0)	3 (14)
Administrative Web-based tasks	"My bank accounts, I have everything, the RATP, I have WhatsApp, etc. and I even have Carrefour, Wish, I have my accounts, I have Mappy to get around, I have music stuff."	4 (50)	11 (52)
Diet or physical activity self-tracking	"I use the app MyFitnessPal."	0 (0)	1 (5)

Table 3. Barriers to using NutCracker and the new technologies by the Nutcracker sample and the low-income milieu population.

Barriers	Verbatim	NutCracker target population (n=8), n (%)	Low-income milieus in the general population (n=21), n (%)
Technical barriers	"I don't think my telephone is good enough...I will change it, but I don't know when [...] It's another expense."	3 (75)[a]	10 (48)
Unfamiliarity with new technologies	"Administration services computerize, but people are lost. Not everyone has an email address."	4 (50)	5 (24)
Comprehension and written barriers	"We can speak French well, we know how to write, but there are things that we have trouble deciphering."	4 (50)	2 (10)
Time constraints	"As I work mornings and evenings, I don't have a lot of time. [...] I don't have the time."	3 (38)	3 (14)
Competition from television	"I watch (the Samira channel) all the time - as soon as my children leave, I watch it."	3 (38)	2 (10)
Search for real contacts	"Cooking workshops give you the opportunity to get out from of home and meet other people."	8 (100)	1 (5)
Live social network	"So, do you like it (what I published)? I'm the only one who publish, it's a shame."	1 (13)	Not applicable
Fear of being stalked on the internet	"There are always dangers. And I'm scared of...[...] I don't like it, it's scary. And...well [...] it can cause damage."	4 (50)	4 (19)
Interest for cooking and cooking skills	"I have everything in my head."	6 (75)	13 (62)

[a]We only considered the 4 participants who used the app.

These technical barriers, although unique to NutCracker, reflect an exacerbated trend of limited use or underperformance, if it was not a defective equipment in the low-income milieus. A total of 9 individuals mentioned these issues, which were essentially caused by the insufficient availability of memory or an overtaxing app, because most individuals preferred to keep what space they had for personal photos or videos:

> *I don't have many apps on my telephone, I have to buy a card, I don't have enough memory. All this stuff takes space!* [General population, 43 years, secretary, 2 children]

Lack of Knowledge: New Technologies, French Language, and Written Comprehension

A lack of knowledge on new technologies was also an obstacle. In the least well-integrated fringes of low-income categories (unemployed women and education level lower than a bachelor degree), everyday internet tasks are neglected, unless they attend specific training programs in social centers where they learn how to create and use an email account and a password on a computer before moving to the mobile phone usage.

All women in the NutCracker intervention group were foreigners, and some of them only had a basic command of French or were not comfortable with the written French (4 of 8). These 2 barriers explained both the low rate of usage of NutCracker and the overall use of digital technologies:

> *I am taking French courses [...] because there are some words that I don't understand. I can read, but I don't understand. [...] I have some difficulty.* [NutCracker, 44 years, housewife, 4 children]

This lack of knowledge regarding the new technologies did not appear as a real barrier to our sample in the general population.

Time Constraints

The project participants who did not download NutCracker were characterized by a few financial constraints as they were either unemployed or earned a very low salary. They also had time constraints. For the unemployed women, domestic tasks, most notably those with large families, were cited as barriers to using NutCracker. Eveline and Hawa were employed, but they had long commutes to work, and their workdays as housekeepers were split in 2 (early mornings and evenings).

Competition From Television

In Magny, 3 of the women interviewed expressed a preference for passive media (television), which allowed them to have their hands free, unlike a smartphone, tablet, or computer. It gave them the opportunity to do household chores, with the television playing in the background.

Regarding cooking more specifically, and for the most underprivileged participants in our study, television was the main source of culinary expertise: television channels, such as the Algerian channel Samira, which is entirely devoted to cooking and gastronomy, were greatly valued.

Among the participants from the general population, the television rarely prevented them from using digital devices.

Search for Social Contact in Real Life

The NutCracker app was aimed to connect women from underprivileged areas to Web-based micronetwork; however, the interviews with these women showed that they preferred to socialize *in the real world.* Given their extreme social isolation, these women sought social contact outside of the domestic realm [17]. Workshops in social centers where they were introduced to the app represented a place of freedom for them. This attachment to *real* encounters was associated with low use of digital communities.

This refusal of being part of virtual socializing was observed in only 1 participant from the low-income milieus sample; she was better positioned socially, but her real-life socializing was affected by a break-up. Following her divorce, Elga was seeking stronger social ties:

> *There are already so many things that we can buy on the internet, if we add food to that, we will never leave the house. I need to get out of the house.* [General population, 47 years, employed, 2 children]

Large Number of Members of a Living Social Network

The small number of members of a living social network prevented the creation of an active and thriving social network. Individuals, such as Martine (NutCracker, 39 years, housewife, 3 children), who were the most enthusiastic, found themselves to be quite alone in their investment in NutCracker, which led them to stop using the app. For example, in her interview, Martine said that her enthusiasm waned once she realized that she was one of the few active people in the network and that no one was responding to her posts.

Fear of the Internet and of Participation in Social Networks

The NutCracker app depended on women publishing their recipes; however, women expressed reluctance to publishing recipes and to sharing their own experience. Furthermore, putting one's own recipes on the Web implies exposing oneself, as food is closely connected to identity, whether individual or social [19]. It is quite likely that for many of these women, publishing a recipe was seen as exposing an intimate realm, that of their home cooking or their cultural origins, which they perceived as being devalued in a migratory context.

This also explains why publishing recipes as part of a group in a workshop was much easier than doing so individually. Instead of involving risky personal exposure, the collective posting was perceived as developing a collective identity:

> *...putting up their recipes, they were happy, it was gratifying.* [NutCracker, social worker]

In the general population, participants more rarely mentioned the fear of a digital tracking (4 of 21), and in these cases, individuals expressed a fear of being overwhelmed by the internet and becoming addicted to digital tools. Furthermore, participants from a better social position presented limiting the use of digital tools as a choice; thus, the refusal of digital technologies was chosen, and not endured.

Levers for Using NutCracker and New Technologies

The results below show some levers for using NutCracker and new technologies (Table 4).

Social Status and Familiarity With New Technologies

As expected, the participants who were more familiar with new technologies were the ones who most easily took to the NutCracker app, which was related to the social position of the individuals. They had a slightly higher social status (most notably through their husbands' professions in the case of the 2 most active users); a higher level of education (high school qualification), which in itself requires a certain familiarity with digital and information technology devices; a functional equipment (a powerful smartphone or a home personal computer); and a keen interest in cooking. This outcome was confirmed by our general population sample, where participants had higher social positions (in terms of occupation or level of education), and almost all participants used the internet and new technologies.

Former Uses of Social Networks and Leadership

The participants who were most active in NutCracker were also those who were already familiar with social network apps most used in low-income areas (Table 4). This Web-based sociability was bolstered by already well-developed real-life sociability. Martine represents a case study in this area. Martine's social insertion proved to be successful, as it was strongly linked to local social bonds, notably within the NGOs, alongside a high level of Web-based sociability (on Facebook, Twitter, Snapchat, Pinterest, and YouTube).

Furthermore, Martine characterized herself through her position as an intermediary, which drove her to invest herself in the digital culinary realm, whether on NutCracker or, previously, other sites or social networks (especially Marmiton or her personal Facebook page where she published recipes). Originally from Senegal, Martine arrived in France a few days after she was born:

> *I'm familiar with French culture, as well as my homeland's culture, so I quite enjoy a kind of mixing of both cultures and all, yes, that's me.*

This distinguished her from the other participants who had arrived in France much later in life, and often as adults. This sense of double belonging, associated with pride in her 2 cultures, led her to see cooking as a way of exchanging and discovering new cultures. This positive self-image also helped her feel more comfortable publishing her personal recipes on the Web.

Children as Intermediaries

The interview study highlighted the role of children as essential intermediaries for digital technologies as they were regarded as the experts in the field. All the more, as there is a tendency for the underprivileged to please their kids with the symbols of consumer society [5], which now includes mobile devices, quite often, the women's mobile phone was not capable of connecting to the internet, whereas the children have smartphones. Kelthoum is proud of her children's expertise:

> *I know how to use it, but there are a lot of things we don't know [...] but young people, they know everything.* [NutCracker, 44 years, housewife, 2 children]

Eveline's case was interesting as well:

> *It's my daughter who knows, because she's the one who installed my [Viber] app (...) She knows how to do everything on the internet.*

Among the low-income milieus in general, children appear as mediators, and they may help in case of technical difficulties, but their role was not as essential as in the deprived households, and they were rarely mentioned in the discourses we collected (once only).

Table 4. Levers for using NutCracker and new technologies by the NutCracker sample and the low-income milieu population.

Levers	Verbatim	NutCracker target population (n=8), n (%)	Low-income milieus in the general population (n=21), n (%)
Social integration	"I'm an outgoing person...I'm too outgoing [...] I'm organizing a pastry competition (at school) for the mums."	2 (25)	17 (81)
Knowledge of new technologies	"Compared to before, we had no Internet...we just had a blank shit [...] Now, everything is	4 (50)	18 (86)
Familiarity with other social networks	"I already had a Facebook page precisely on cooking, I was used to publish recipes every week."	6 (75)	10 (48)
Children acting as intermediaries	"It's the kids who know about that."	2 (25)	2 (10)
Digital and modernity	"I'm a geek of my smartphone."	1 (13)	5 (24)
Videos on internet	"But there's everything on YouTube [...] You just type "chicken" or "Tajine" and it gives you [...] there are tons of different videos."	3 (38)	7 (33)
Interest for cooking and cooking skills	"Maybe I'll find some ideas (on the app) in order to make my children eat vegetables."	6 (75)	15 (71)

Added Value of Videos on the Internet

Those participants who were the least comfortable with the French language preferred using visual material on the internet or in apps to increase their culinary expertise (3 of 8 in the NutCracker group); watching cooking videos on YouTube was preferred over using websites or apps that were predominantly text-based. Thus, Fatima (Magny), who had difficulty with the French language, used YouTube videos—the links were sent to her by her friends via WhatsApp:

> *I start the video, I put it on full screen (...) it's faster, it's easier.* [NutCracker, 44 years, housewife, 4 children]

In addition, when she looked something up on the internet, she used the *microphone* function on Google so that she did not have to write anything.

This approach was also a trend we observed in low-income milieus in the general population (eg, Rabia, 38 years, employee: "I watch videos (on YouTube) for recipes") that was exacerbated in Magny.

Cooking Skills: Both a Barrier and a Lever

Cooking skills appear as both a barrier and a lever. The 2 women who actually used NutCracker had a keen interest in cooking—for Martine, the appeal was more on the *cooking* side and, for Kelthoum, who was mostly looking for fast and easy ways to get her children to eat vegetables, it was more on the *health* side. She explained why she really got into the project:

> *Maybe I'll find a way to get my children to eat my food, vegetables included. [...] There are a lot of moms who have trouble getting their kids to eat vegetables.* [NutCracker, 43 years, housewife, 3 children]

On the other end of the spectrum, great culinary traditions based on oral transmission rendered the use of the internet unnecessary and unattractive when it came to cooking. Eveline echoed this when explaining why she did not need the internet to nourish her culinary inspiration:

> *It's all in my head.* [NutCracker, 44 years, employed]

Exactly the same words ("it's all in my head") were expressed in the general population by Savina (45 years, old manual worker).

Furthermore, in this specific social milieu, cooking was not necessarily connected to the digital domain, as emphasized by a social worker in Magny:

> *...making the connection with the app is very, very difficult.*

Cooking is the realm of the tangible and the emotional, which is disconnected from the digital world. It is difficult to determine, at this stage of our investigation, whether this separation is specific to people from underprivileged groups or not.

Finally, and somewhat unexpectedly, those who were completely uninterested in cooking were the ones who demonstrated the most interest in the app; they figured that the digital tool could provide them with ideas to turn what they saw as a tedious task into something more fun. For Clarisse, cooking was like:

> *a chore; when we get home from work, we're tired.*

Therefore, regarding an app:

> *Yes, why not? It could provide some ideas.* [General population, 54 years, employed, 2 children)]

Discussion

Principal Findings

NutCracker, developed as a social network-based app, includes cooking tips, nutritional advices, and Web-based quizzes. We identified barriers to its uses (including technical barriers, lack of knowledge in the field of new technologies and written comprehension, and search for real contacts) and motivators (in particular, good social integration, previous use of social networks, and help of children as intermediaries). Cooking skills were both a barrier and a lever. Although the designers of the project had great expectations and foresaw a high membership rate, the real results may appear disappointing. Several explanations can shed light on this discrepancy. To begin with, a social desirability bias was linked to the group's situation in the collective discussion, where the participants found it difficult to express reluctance toward a project concerning a pleasant and well-liked area such as cooking and the symbols of modernity, that is, digital tools. Furthermore, the participation of these women in this project revealed a desire to move upwards socially in very underprivileged areas by gaining access to information deemed good and desirable (eating healthy, at a low cost) as these recommendations by the National Program for Nutrition and Health have been widely promoted since 2001 [2]. For these women, culinary innovation and having access to affordable recipes containing ingredients promoted by public health campaigns (fruits and vegetables) to prepare for their children are ways to project themselves as being *good cooks* and *good mothers* and to conform to the values and practices of those who are more well off. Finally, the prospect of using a socially valued tool gives these women the feeling of participating in, and benefitting from, the consumer society in the same way as individuals from more privileged backgrounds.

The reality of the field study itself also revealed some barriers. First, those related to the slowness or capacity issues of the devices, especially because smartphones were often the only internet access for the households in our study. Our work corroborates previous studies, which noted that the speed the apps use was critical to the satisfaction of users [20]. Studies also show that although digital tools are relatively common among the French population, the distribution of smartphones is much more unequal than that of mobile phones [21], which our study corroborates.

We have also added some new elements to the digital divide such as the difficulties experienced by our participants, which reflect the working class's unfamiliarity with new technologies. Our results demonstrated that the lowest percentile of the low-income population, which is the least socially integrated, are at greater risk of digital inequalities, which is in line with previous reports [17,22]. Conversely, a good social integration

promotes the use of cooking apps and the knowledge of social norms affecting food choices [23] and uses of new technologies [22].

Our study confirms a gap in the use of digital self-tracking tools. Although individuals from well-off backgrounds choose to use these tools [19], individuals from underprivileged backgrounds are compelled to use them. The low usage rate of NutCracker by its target population was mainly because of constraints in time and equipment, whereas individuals, who were more socially integrated, including those from underprivileged areas, exercised their choice and decision-making power in using the digital world.

This research also confirms the ability of this app to act as a culinary lever to promote healthier eating habits [7,24] through the interest of people in using digital tools and not through providing health information. Cooking apps are among the tools often used in underprivileged areas, whereas self-tracking devices, based on a dietary and quantified approach, are not familiar.

However, the connection between cooking and the digital world is very weak and should be explored further in future. Our results highlight the widespread reliance of the target population on videos, which allow them to overcome the written-language barrier and to rematerialize cooking through direct access to tips and tricks, emphasizing the significance of visual material. The use of videos to develop cooking skills [25] is currently being tested in the FacilEat4All project, which is expanding on the NutCracker project.

Finally, we have demonstrated the difficulty experienced by women of very underprivileged backgrounds to integrate and participate in online social networks, barring women who were already well-integrated in their real world and were active on social media. Social inequalities were increased by gender inequalities. Our results corroborate studies performed on digital technologies that highlight the complementarity and overlapping between real-world and Web-based sociability [26,27]. To compensate for the difficulty in creating an autonomous ex nihilo social network, our team suggests the use of existing social networks through a private Facebook group to foster a sense of community spirit more quickly. Recent work has shown the circulation of a prevention program based on social networks [28]. Finally, the NutCracker experience shows the relevance of individuals-as-intermediaries; their investment in digital tools makes them opinion leaders, and their importance has recently been emphasized in the field of electronic health [29,30]. We consider the position of cultural intermediary to be crucial to this investment, which made Martine a *champion* of the app [31].

Limitations

The app was disseminated among a small sample of participants: this prevented the creation of an active social network. Our current research project FacilEat4All—a continuation of NutCracker—is based on a broader community (approximately 100 volunteers). Another limitation of the NutCracker project is that the choice of the participants, which was limited to individuals from underprivileged backgrounds, led to the selection of people with heavy combination of constraints (budget, social integration, and language). The FacilEat4All project now includes people from modest categories with less difficulties (white or blue collars, in socially more diverse areas).

Finally, the method of face-to-face interview was particularly interesting for the postintervention study to have access to individual opinions on the app. On the contrary, collective discussions appeared as useful tool to launch a collective dynamic for the project, albeit the group's discussion led to a general consensus regarding the expectations concerning the app itself.

In the most underprivileged milieus, it was particularly difficult to motivate individuals to test the app over a long term (3 months) and to participate in the social network. One conclusion of this limitation is that in those underprivileged milieus, digital devices are more easily accepted if used as collective tools, in workshops for example.

Conclusions

The NutCracker and FacilEat4all projects bring new elements to a theme that had not been studied closely until now. They promote healthy eating through culinary levers and digital tools. The NutCracker app and the study of the use and perception of digital *diet* tools (nutrition or cooking) by people with underprivileged backgrounds have highlighted the numerous barriers in using cooking apps for people from modest backgrounds. In addition to technical barriers, a lack of skills related to new technologies, a reluctance toward written material, and a combination of time and financial constraints also restricted the use of the app. What stopped the participants from inserting themselves into the online social micronetwork were difficulties regarding self-expression on the Web and a need to integrate socially in real life.

Our study also shed light on the levers we rely on, such as prior experience and use of social networks, which led certain participants to become leaders. We also observed the importance of children as intermediaries of new technologies. Finally, our research has demonstrated the importance of culinary levers in the development of digital tools for people in the low-income categories. Further interventions should assess the advantages of a cooking-based communication platform to promote healthier behaviors.

Acknowledgments

The authors are very grateful to the *Point Services aux Personnes* of Saint Quentin en Yvelines and *Fondation Agir Contre l'Exclusion* for their involvement and assistance throughout the research. The authors wish to thank Jacques Bibai and Julien Ollivier (NutCracker) for their work in producing the app. This study was sponsored by the regional council of Paris Île-de-France

and Ferrero within a research contract (grant numbers 11001167, 11000623, and FacilEat4all 11001083 – p00000946). Neither of the sponsors has been involved in analyzing the results nor in the review or approval of the manuscript.

Authors' Contributions

FR developed the protocol for the field study, performed a part of the interviews, analyzed the interviews, and wrote the manuscript. MD conceived the pedagogical program, prepared, and distributed the questionnaires, prepared and conducted the collective discussion, prepared the app-design. ND developed the protocol for the questionnaires and collective discussion, and wrote the manuscript. CA performed a part of the interviews, analyzed the interviews, and participated in writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of a website; SPSS files containing original data.

[PDF File (Adobe PDF File), 941KB - mhealth_v6i11e11176_app1.pdf]

References

1. Marmot M. European Commission. 2013 Dec. Health inequalities in the EU: Final report of a consortium URL: https://ec.europa.eu/health//sites/health/files/social_determinants/docs/healthinequalitiesineu_2013_en.pdf [accessed 2018-11-09] [WebCite Cache ID 73nlgWz3x]
2. Saint-Pol T. Déterminants sociaux et culturels de l'alimentation. In: Institut national de la santé et de la recherche médicale (Inserm) , editor. Inégalités sociales de santé en lien avec l'alimentation et l'activité physique. Paris, France: Pôle expertise collective ITMO Santé Publique, Aviesan; Jun 2015:217-235.
3. Martin-Fernandez J, Grillo F, Parizot I, Caillavet F, Chauvin P. BMC Public Health. 2013 Dec. Prevalence and socioeconomic and geographical inequalities of household food insecurity in the Paris region, France, 2010 URL: https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-13-486 [accessed 2018-11-09] [WebCite Cache ID 73nmC8hxd]
4. Martin P, Consalès J, Scheromm P, Marchand P, Ghestem F, Darmon N. Community gardening in poor neighborhoods in France: a way to re-think food practices? Appetite 2017 Sep 01;116:589-598 [FREE Full text] [doi: 10.1016/j.appet.2017.05.023] [Medline: 28527952]
5. Régnier F, Masullo A. Obésité, goûts et consommation. Intégration des normes d'alimentation et appartenance sociale. Revue Française Sociologie 2009;50(4):747-773. [doi: 10.3917/rfs.504.0747]
6. Mandereau-Bruno L, Fosse-Edorh S. INVS Santé publique France. 2017 Aug 22. Prevalence du diabete traite pharmacologiquement (tous types) en France en 2015. Disparites territoriales et socio-economiques URL: http://invs.santepubliquefrance.fr/beh/2017/27-28/pdf/2017_27-28_3.pdf [accessed 2018-11-09] [WebCite Cache ID 73o7vXB1t]
7. Reicks M, Kocher M, Reeder J. Impact of cooking and home food preparation interventions among adults? A systematic review. J Nutr Educ Behav 2018 Feb;50(2):148-172.e1. [doi: 10.1016/j.jneb.2017.08.004] [Medline: 28958671]
8. Davies J, Damani P, Margetts B. Intervening to change the diets of low-income women. 2009 May Presented at: Proceedings of the Nutrition Society; July 4, 2008; Nottingham University p. 210 URL: http://doi.org/10.1017/S0029665109001128 [doi: 10.1017/S0029665109001128]
9. European Commission. European Commission. 2014 Apr 10. Green Paper on mobile health ("mHealth") URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2018-11-10] [WebCite Cache ID 73pVacGFm]
10. INSEE. Institut national de la statistique et des études économiques. 2017 Dec 27. Recensement de la population 2015 en géographie au 01/01/2017 URL: https://www.insee.fr/fr/statistiques/3293086?geo=COM-78621 [accessed 2018-11-12] [WebCite Cache ID 73rq3ZDb6]
11. INSEE. Institut national de la statistique et des études économiques.: INSEE; 2017 Sep 20. Fichier localisé social et fiscal en géographie au 01/01/2015 - Revenu et pauvreté des ménages URL: https://www.insee.fr/fr/statistiques/3126710?geo=COM-78621 [accessed 2018-11-12] [WebCite Cache ID 73rmXj6LB]
12. Blomkvist M. Department of Information Technology, Uppsala University. 2006 Nov. The User as a Personality: A Reflection on the Theoretical and Practical Use of Personas in HCI Design (Department of Information Technology, Uppsala University), November 2006 URL: https://www.it.uu.se/research/publications/reports/2006-049/2006-049-nc.pdf [accessed 2018-11-12] [WebCite Cache ID 73rmnAcZZ]
13. Cooper A. The Innmates Are Running the Asylum. USA: SAMS; 2004.
14. 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;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

15. Champagne C, Pailhé A, Solaz A. Le temps domestique et parental des hommes et des femmes: quels facteurs d'évolutions en 25 ans? Economie et Statistique 2015;478(1):209-242 [FREE Full text] [doi: 10.3406/estat.2015.10563]
16. Régnier F. Vous consultez « Goût de liberté » et self-quantification Perceptions et appropriations des technologies de self-tracking dans les milieux modestes. Réseaux 2018 Feb;208-209:95-120. [doi: 10.3917/res.208.0095]
17. CREDOC. Baromètre du Numérique Edition 2015. Paris: CREDOC; 2015.
18. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
19. Fischler C. Food, self and identity. Soc Sci Inf 1998 Oct;27:275-292 [FREE Full text] [doi: 10.1177/053901888027002005]
20. Jospe MR, Fairbairn KA, Green P, Perry TL. Diet app use by sports dietitians: a survey in five countries. JMIR Mhealth Uhealth 2015 Jan 22;3(1):e7 [FREE Full text] [doi: 10.2196/mhealth.3345] [Medline: 25616274]
21. CREDOC. La diffusion des technologies de l'information et de la communication dans la Société. Paris: CREDOC; 2013.
22. Régnier F, Chauvel L. Digital inequalities in the use of self-tracking diet and fitness apps: interview study on the influence of social, economic, and cultural factors. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e101 [FREE Full text] [doi: 10.2196/mhealth.9189] [Medline: 29678807]
23. Higgs S. Social norms and their influence on eating behaviours. Appetite 2015 Mar;86:38-44. [doi: 10.1016/j.appet.2014.10.021] [Medline: 25451578]
24. Lavelle F, Hollywood L, Caraher M, McGowan L, Spence M, Surgenor D, et al. Increasing intention to cook from basic ingredients: a randomised controlled study. Appetite 2017 Sep 01;116:502-510. [doi: 10.1016/j.appet.2017.05.024] [Medline: 28526477]
25. Surgenor D, Hollywood L, Furey S, Lavelle F, McGowan L, Spence M, et al. The impact of video technology on learning: a cooking skills experiment. Appetite 2017 Jul 01;114:306-312. [doi: 10.1016/j.appet.2017.03.037] [Medline: 28363814]
26. Beuscart J, Dagiral E, Parasie S. Sociologie d'internet. Paris: Armand Colin; 2016.
27. Bouchillon BC, Gotlieb MR. Making them count. Soc Sci Comput Rev 2016 Feb 07;35(3):299-318. [doi: 10.1177/0894439315626422]
28. Looyestyn J, Kernot J, Boshoff K, Maher C. A web-based, social networking beginners' running intervention for adults aged 18 to 50 years delivered via a facebook group: randomized controlled trial. J Med Internet Res 2018 Feb 26;20(2):e67 [FREE Full text] [doi: 10.2196/jmir.7862] [Medline: 29483065]
29. Gentina E, Kilic D, Dancoine P. Distinctive role of opinion leaders in the social networks of school adolescents: an investigation of e-cigarette use. Public Health 2017 Mar;144:109-116. [doi: 10.1016/j.puhe.2016.11.020] [Medline: 28274370]
30. Shi J, Salmon CT. Identifying opinion leaders to promote organ donation on social media: network study. J Med Internet Res 2018 Jan 09;20(1):e7 [FREE Full text] [doi: 10.2196/jmir.7643] [Medline: 29317384]
31. Partridge SR, Gallagher P, Freeman B, Gallagher R. Facebook groups for the management of chronic diseases. J Med Internet Res 2018 Jan 17;20(1):e21 [FREE Full text] [doi: 10.2196/jmir.7558] [Medline: 29343460]

Abbreviations

mHealth: mobile health
NGO: nongovernmental organization

Edited by G Eysenbach; submitted 30.05.18; peer-reviewed by M Boaz, J Shi; comments to author 09.07.18; revised version received 23.09.18; accepted 25.10.18; published 23.11.18.

Please cite as:
Régnier F, Dugré M, Darcel N, Adamiec C
Providing a Smart Healthy Diet for the Low-Income Population: Qualitative Study on the Usage and Perception of a Designed Cooking App
JMIR Mhealth Uhealth 2018;6(11):e11176
URL: http://mhealth.jmir.org/2018/11/e11176/
doi:10.2196/11176
PMID:30470677

Original Paper

An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study

Becky Inkster[1], DPhil; Shubhankar Sarda[2], BSc (Physics); Vinod Subramanian[3], MBA

[1]School of Clinical Medicine, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

[2]Wysa, London, United Kingdom

[3]Wysa, Bangalore, India

Corresponding Author:
Becky Inkster, DPhil
School of Clinical Medicine
Department of Psychiatry
University of Cambridge
Box 189
Cambridge Biomedical Campus
Cambridge, CB2 2QQ
United Kingdom
Phone: 44 773 847 8045
Email: becky.inkster@gmail.com

Abstract

Background: A World Health Organization 2017 report stated that major depression affects almost 5% of the human population. Major depression is associated with impaired psychosocial functioning and reduced quality of life. Challenges such as shortage of mental health personnel, long waiting times, perceived stigma, and lower government spends pose barriers to the alleviation of mental health problems. Face-to-face psychotherapy alone provides only point-in-time support and cannot scale quickly enough to address this growing global public health challenge. Artificial intelligence (AI)-enabled, empathetic, and evidence-driven conversational mobile app technologies could play an active role in filling this gap by increasing adoption and enabling reach. Although such a technology can help manage these barriers, they should never replace time with a health care professional for more severe mental health problems. However, app technologies could act as a supplementary or intermediate support system. Mobile mental well-being apps need to uphold privacy and foster both short- and long-term positive outcomes.

Objective: This study aimed to present a preliminary real-world data evaluation of the effectiveness and engagement levels of an AI-enabled, empathetic, text-based conversational mobile mental well-being app, Wysa, on users with self-reported symptoms of depression.

Methods: In the study, a group of anonymous global users were observed who voluntarily installed the Wysa app, engaged in text-based messaging, and self-reported symptoms of depression using the Patient Health Questionnaire-9. On the basis of the extent of app usage on and between 2 consecutive screening time points, 2 distinct groups of users (*high users* and *low users*) emerged. The study used mixed-methods approach to evaluate the impact and engagement levels among these users. The quantitative analysis measured the app impact by comparing the average improvement in symptoms of depression between high and low users. The qualitative analysis measured the app engagement and experience by analyzing in-app user feedback and evaluated the performance of a machine learning classifier to detect user objections during conversations.

Results: The average mood improvement (ie, difference in pre- and post-self-reported depression scores) between the groups (ie, high vs low users; n=108 and n=21, respectively) revealed that the high users group had significantly higher average improvement (mean 5.84 [SD 6.66]) compared with the low users group (mean 3.52 [SD 6.15]); Mann-Whitney *P*=.03 and with a moderate effect size of 0.63. Moreover, 67.7% of user-provided feedback responses found the app experience helpful and encouraging.

Conclusions: The real-world data evaluation findings on the effectiveness and engagement levels of Wysa app on users with self-reported symptoms of depression show promise. However, further work is required to validate these initial findings in much larger samples and across longer periods.

(JMIR Mhealth Uhealth 2018;6(11):e12106) doi:10.2196/12106

KEYWORDS

mental health; conversational agents; artificial intelligence; chatbots; coping skills; resilience, psychological; depression; mHealth; emotions; empathy

Introduction

Background

Major depression is a disabling disorder with symptoms such as feelings of sadness, worthlessness, and losing interest in activities. Depression is the single largest contributor to global disability with an estimated 300 million or approximately 4.4% of the world's population (2015) affected by it [1]. Severe depression can lead to suicide, which was the second leading cause of death among people aged 15 to 29 years globally in 2015 [1]. Major depression has been found to impair quality of life [2] and psychosocial functioning [3,4], which is a person's ability to perform daily activities and to maintain interpersonal relationships.

The economic burden of depression is rising. The cost of major depression in the United States was estimated at US $210.5 billion per year in 2010, an increase of 21.5% from 2005 [5]. For every dollar spent treating major depression in 2010, US $4.70 was spent on direct cost of related illnesses, and an additional US $1.90 was spent on reduced workplace productivity and costs associated with suicide linked to depression [5]. According to the Centre for Mental Health policy paper (2010), the total cost of mental ill health in England was estimated at £105.2 billion a year from 2009 to 2010, an increase of 36% from 2002 to 2003 [6]. The Farmer-Stevenson review that was launched by the UK Parliament in 2017 on mental health in the workplace placed the cost to employers due to poor mental health at £33 to £42 billion a year, with over half of it coming from presenteeism [7]. According to the World Health Organization (WHO) Mental Health Atlas 2017, government spend globally on mental health in 2015 was less than 2% of the global median of government's health expenditures overall, which has only exacerbated the situation [8].

Mood disorders can be treated by pharmacotherapy or psychotherapy [9]; however, significant treatment barriers remain, such as major shortage of mental health professionals, long waiting lists for treatment, and stigma. The WHO Mental Health Atlas 2017 reported that there is a global median of 9 mental health workers including approximately 1 psychiatrist per 100,000 people [8]. In India, there are approximately 10 mental health professionals for 100,000 people affected by mental health problems [10]. According to the Impact Assessment report from the UK Department of Health (October 2014), access to services for people with mental health problems is more restricted, and waiting times are longer than for other health care services [11]. A 2018 British Medical Association research briefing stated that two-thirds of the National Health Service (NHS) mental health trusts in the United Kingdom had year-long waiting periods before therapy started, and in some locations, waiting periods were close to 2 years [12]. Perceived public stigma, a known barrier, is the degree to which the general public holds negative views and discriminates against a specific group. Young adults who reported higher scores on the Patient Health Questionnaire-2 (PHQ-2) showed greater associations with perceived public stigma than personal stigma [13]. The WHO World Mental Health Surveys show that apart from perceived stigma, structural barriers such as finance and lack of service availability were the most reported barriers to treatment among those with severe disorders [14].

Prior Work

Face-to-face therapy and guided self-help techniques such as cognitive behavioral therapy (CBT) and behavioral activation are known to be effective in treating depression [15,16]. Face-to-face therapy only provides point-in-time support and cannot scale quickly to address growing mental health challenges. Innovative delivery methods are required to supplement care. Studies have shown that certain user groups are opening up to technology about their mental health problems. A recent study showed that participants reported more posttraumatic stress disorder symptoms when asked by a virtual human interviewer compared with a gold standard assessment [17]. Guided internet-based self-help interventions have been observed to have positive effects on patients with symptoms of depression and to reduce risk of symptom deterioration [18-22]. Mobile app–administered therapy either stand-alone or in blended mode has been found to show positive effects on patients with depression across severity levels in randomized controlled trial (RCT) studies [23-28]. However, there are studies with mixed findings about the benefits of smartphone or online-administered interventions. A recent RCT study that examined the effects of an online mindfulness meditation app compared with an active sham meditation control app found that mindfulness improved across university student participants in both groups, and there seemed no added benefit from offering progressive and varied mindfulness tools [29].

Text-based messaging (internet or smartphone) either with a human coach or with a machine (chatbots) has found increasing adoption in recent years. Artificial intelligence (AI) text-based conversational agents have the ability to offer contextual and always-available support. Studies using internet-based, one-to-one text-based chat interventions for psychological support have shown feasibility and positive improvement in mental health outcomes when compared with wait-list conditions [30]. Two recent studies measured the efficacy of a fully automated mobile conversational agent in the delivery of mental well-being [31,32]. Our study aims to add to the research and evidence base on the effectiveness and engagement levels of AI-enabled, text-based, conversational mobile mental well-being apps.

Wysa, a Smartphone-Based Empathetic Artificial Intelligence Chatbot App for Mental Well-Being

Wysa, developed by Touchkin, is an AI-based *emotionally intelligent* mobile chatbot app aimed at building mental

resilience and promoting mental well-being using a text-based conversational interface. The Wysa app assists users to develop positive self-expression by using AI to create an external and responsive self-reflection environment. Engaging with the app is free and available 24×7, but accessing a human coach via the app is a paid service. We used an early in-the-market app version (see Multimedia Appendix 1) that included only the free always-available chatbot service (not the paid coach service). The app responds to emotions that a user expresses over written conversations and, in its conversation, uses evidence-based self-help practices such as CBT, dialectical behavior therapy, motivational interviewing, positive behavior support, behavioral reinforcement, mindfulness, and guided microactions and tools to encourage users to build emotional resilience skills. The Wysa scientific advisory board approves all content and tools. The conversation-based tools and techniques encourage users to manage their anxiety, energy, focus, sleep, relaxation, loss, worries, conflicts, and other situations.

The app can be downloaded from the Google Play Store and from the Apple App Store. There is no user registration to sign in and no personal identifiable information is asked at any time during app use. Wysa was described as “friendly” and “easy to use” in a youth user study conducted by Wellcome Trust, United Kingdom, Neuroscience, Ethics, and Society Young People’s Advisory Group at the University of Oxford, and BBC Tomorrow's World [33]. The app was adapted and implemented at Columbia University’s SAFE Lab as a tool to provide support to at-risk communities in inner cities (Brooklyn and Chicago), many of whom are gang-involved youth. Although, Wysa is not a medical device, when used as a health and well-being support tool, it can support clinical services as seen from its use at the NHS North East London Foundation Trust [34].

Study Objective

The primary study objective was to determine the effectiveness of delivering positive psychology and mental well-being techniques in a text-based conversational mode using the Wysa app on users with self-reported symptoms of depression. Users were presented with the validated Patient Health Questionnaire (PHQ-9) during their conversations and screened for selection based on their 2-item (PHQ-2) score. The average improvement in self-reported symptoms of depression (Pre-PHQ-9 minus Post-PHQ-9) was compared between 2 comparison groups: (1) more engaged app users (“high users” group) and (2) less engaged app users (“low users” group).

Our secondary study objective was to understand users’ in-app experiences during app use. A qualitative thematic analysis, as proposed by Braun and Clarke, 2006 [35,36], on in-app feedback responses was performed.

Methods

Ethics

The study involved a remotely screened, anonymous nonclinical global population (ie, real-world *in-the-wild* data) and was, therefore, exempt from registration in a public trials registry. The users downloaded the app after having agreed to the Wysa app Terms of Service and Privacy Policy, which included consent to use anonymized data for research purposes. Minimal deidentified data required for the study were used. For details on app specific ethical practices, see Multimedia Appendix 2.

Study Design

The Wysa app was downloaded from the Google Play Store voluntarily by geographically dispersed users. The users were filtered for eligibility from a pool of anonymous Wysa app users based on the inclusion criteria (see Figure 1). For the study, we solely looked at user-provided data that were collected by the app during active use. Given the anonymity and nonavailability of user profiles, qualitative and quantitative data were collected concurrently during the study period on and between July 11, 2017, and Sept 5, 2017. These data consisted of user responses to the app’s inbuilt assessment questionnaire and responses to the app-designed text-based conversations and questions. No additional research-framed questionnaires or user feedback questions were designed or issued for repeated interval data collection.

On the basis of the extent of app usage on and between 2 consecutive PHQ-9 screenings, 2 comparison groups emerged (“high users” and “low users”). The users in both groups voluntarily reported 2 valid time point PHQ-9 scores: one at onboarding (first assessment, “Pre-PHQ-9”) and the other on or after 2 weeks (second assessment, “Post-PHQ-9”). The 2 screening time points were considered valid if during the study period only 2 surveys were responded to within a gap of 14 or more days. The “high users” consisted of users who engaged with the app on the 2 screening days as well as at least once between those days. The “low users” consisted of users who only engaged on the 2 screening days but never between those days.

The authors decided to implement a quasi-experimental (simple pre-post) mixed-methods approach given our study objective and the nature of the data being collected. For details on the mixed-methods design and approach, see Multimedia Appendices 2 and 3. See the study recruitment flow diagram in Multimedia Appendix 4.

Figure 1. The study inclusion criteria. PHQ: Patient Health Questionnaire.

Inclusion criteria	Exclusion criteria
1. Use a smartphone with android operating system 2. Install app voluntarily from Google Play Store 3. Agree to app Terms of Service and Privacy Policy 4. Use app on and between July 11, 2017 to September 5, 2017 5. Self-report only 2 valid PHQ-9 screenings 6. Self-report PHQ-2 score of 6 (pre screenings)	1. Report more than 2 valid PHQ-9 screenings 2. Pre screening PHQ-2 score of less than equal to 5

Quantitative Measurement and Screening

The inbuilt app-administered assessment questionnaire (PHQ-9) required users to recollect problems over the last 2 weeks; notably, this form of data collection is neither momentary nor *real-time* capture. For details about PHQ-9, see Multimedia Appendix 2. The PHQ-2 score was generated from responses to the first 2 items of the PHQ-9 (ie, range: 0-6). The PHQ-2 is intended for use as an initial screening of depression symptoms, whereas the PHQ-9 score is then used for monitoring depression symptoms [37]. As the app engaged with anonymous users, there was no information available about clinical history and diagnosis. Remote digital screening for depressive symptoms in anonymous populations is very challenging in the absence of face-to-face clinical interviews; therefore, we selected the most stringent threshold based on recommendations in the scientific literature [37], which required a PHQ-2 score of 6.

Data Collection and Analysis

The app takes the user through conversational pathways based on a user's interaction. This path varies for every user, based on their messages and context. At various points in a user's conversational journey, a user is presented with app-designed open- and closed-ended questions that check the helpfulness of these sessions and seeks user feedback (in-app feedback; eg, at the end of every wellness session or at end of every mindfulness or physical activity tool-based session). This voluntary feedback provided by the users was not scheduled repeatedly nor was it used to measure changes in behavior or emotions of an individual over time. Instead, the objective was to understand the users' experiences and engagement with the app. For the in-app feedback questions, see Multimedia Appendix 5. All transmissions to and from the app were encrypted using recognized security standards and were securely stored in a private cloud server. All user-generated conversations and screening responses were checked for compromise (eg, malicious bots) and deidentified for app identifiers. At onboarding, the following user context information was collected:

1. Major event or recent changes: The response to the question, "What has been the major event or change in your life recently?" was collected by the app in free-text before a Pre-PHQ-9 screening.
2. Ability to cope with daily tasks: Immediately after the Pre-PHQ-9 screening, based on the score, users were asked about their ability to cope with daily tasks. For high severity PHQ-9 scores, users were asked "Is it getting hard for you to cope with your daily tasks?," whereas for none to mild severity, they were asked "Are you happy with how life is going at the moment?" The user could respond either by clicking preformatted options or by free-text.

For a typical user app engagement, see Multimedia Appendix 6. Microsoft Excel software was used for data wrangling and analysis. Open-source python software on Jupyter Notebook was used for machine learning (ML) modeling.

Quantitative Analysis Method

Impact (Pre-Post) Analysis

To quantify the app impact, the average improvement (pre-PHQ-9 minus post-PHQ-9) was compared between the 2 user groups. A Mann-Whitney *U* test was carried out to test the hypothesis that high users would have greater average improvement than low users. The effect size was measured using the nonparametric common language effect size (CL), calculated as $[1-(U/n_{h*} n_l)]$, where U was the Mann-Whitney *U* and n_h and n_l are the numbers of high users and low users, respectively [38]. The CL gives the probability that a user picked at random from the high users group will have a higher average improvement than a user picked at random from the low users group [38,39].

Context/Descriptive Analysis

To maintain user anonymity, the app did not capture personal identifiable information or sociodemographic information (except time zone). To capture useful context about users, an analysis of the qualitative responses to key app-based questions was performed, including days of active use, recent major event or changes, ability to cope with daily tasks, and completion of wellness tools.

Qualitative Analysis Method

Engagement Effectiveness

An analysis of users' in-app feedback responses was performed using thematic analysis [35,36] to measure engagement effectiveness. Main themes and subthemes, derived from the analysis, helped understand users' app experience and engagement. Prevalence of a theme was measured based on count of response instances and number of responding users. Further insights were identified by intersecting derived user context with the main themes. For details on thematic analysis approach, see Multimedia Appendix 2.

Engagement Efficiency

To measure the app's engagement efficiency, an analysis of objections raised by users was performed. It is important for a real-world conversational app to understand users' written messages with high accuracy, precision, and recall to provide

empathetic listening and to correctly interpret and respond to a user every single time. This is critical to provide seamless user engagement and experience, which in turn leads to higher app usage and retention. All the conversation messages (instances) the users had with the app were manually tagged for "objection" or "no objection." Objections took 2 forms: refusals (ie, when the user objects to a bot's understanding of what was said; for eg, "I don't want to do this") and complaints (ie, when the user raises a complaint to a bot's response; for eg, "That's not what I said"). See Multimedia Appendix 7 for examples on objections. The proportion of objections raised by a user was measured for prevalence. The tagged dataset was also used to evaluate the performance of an existing supervised ML classifier algorithm deployed to automatically detect objections in real-world use. For details about this analysis, see Multimedia Appendix 2.

Results

Analysis Size

The mixed-methods analysis was performed on 129 users (high users, n_h=108; low users, n_l=21) who had met the inclusion criteria.

Quantitative Analysis

Impact (Pre-Post) Analysis

The study first screened for users who self-reported a Pre-PHQ-2 score equal to 6. We initially checked that users' PHQ-9 scores had improved (ie, reduced going from pre- to post), on average, between time points. Both comparison groups showed a significant reduction in PHQ-9 score (within groups) as measured by a Wilcoxon signed-rank test (Table 1). The authors expected that regression to the mean (whereby values that are initially measured as extreme are more likely to be moderate on subsequent measurement) might play a role in this apparent large improvement [40].

Therefore, a between-groups comparison of the average improvement (Pre-PHQ-9 minus Post-PHQ-9) was performed using a Mann-Whitney *U* test (Table 2). We found that the high users group showed significantly higher average improvement compared with the low users group (*P*=.03). The effect size was found to be approximately 0.63. For the purposes of post hoc comparisons, other studies have found that a CL of 0.63 is roughly equivalent to a Cohen *d* of 0.47 [39]. For quality control purposes, as discussed in the paper by Zimmerman [41], an unpaired *t* test with outliers removed was then conducted. This also produced a significant result (*P*=.028).

As a post hoc analysis, the PHQ-2 screening cutoff score was reduced so that additional Wysa users could be added to the sample. With a PHQ-2 cutoff score of 5, the high users group still showed higher average improvement compared with the low users group, but the effect was less significant (*P*=.06). With a PHQ-2 cutoff score of 4, the same effect was observed but at an even lower significance (*P*=.09).

Context/Descriptive Analysis

In total, 83.3% (90/108) of high users actively used the app for more than 4 days on and between 2 consecutive PHQ-9 screenings (see Multimedia Appendix 8). Given the natural app-use environment, each user in both groups had different pre- and postscreening days that were spaced at least 2 weeks apart within the study period.

Table 1. Within-group analysis.

Users with self-reported PHQ[a]-2=6	Number of users (N)	Mean (scores)	Median (scores)	W-value (*P* value[b])
High users				
Pre-PHQ-9	108	18.92	19.50	478.5 (*P*<.001)
Post-PHQ-9	108	13.07	12.00	—
Low users				
Pre-PHQ-9	21	19.86	21.00	32.5 (*P*=.01)
Post-PHQ-9	21	16.33	17.00	—

[a]PHQ: Patient Health Questionnaire.

[b]95% significance.

Table 2. Between-group analysis.

Users with self-reported PHQ-2[a]=6	Number of users (N)	Mean improvement (SD)	Median improvement	Mann-Whitney *U* (*P* value[c])	Effect size (CL[b])
High users (n_h)	108	5.84 (6.66)	6.00	835.5 (*P*=.03)	0.632
Low users (n_l)	21	3.52 (6.15)	2.00	—	—

[a]PHQ-2: Patient Health Questionnaire-2.

[b]CL: common language effect size.

[c]95% significance.

In addition, 80.6% (104/129) of users gave a postscreening within 18 days of a prescreening (see Multimedia Appendix 9). The users came from diverse time zones (see Multimedia Appendix 10); 48.1% (62/129) of users came from America, followed by 26.4% (34/129) from Europe and 18.6% (24/129) from Asia. A total of 89.9% (116/129) users reported a recent major event or change in their life (see Multimedia Appendix 11). A total of 26.7% (31/116) cited "relationship issues/changes" as a recent major event. Among relationship issue/change, "break-up" was the top cited issue (11 of the 31), followed by "concerns and challenges with close family member" (8 of the 31). Other relationship issues or changes included issues with friends (3 of the 31), issues with other relations (3 of the 31), conflicts in marriage (3 of the 31), and getting into a new relation (3 of the 31). A total of 12.9% users (15/116) reported "mental well-being changes" as a recent event. Moreover, 5 of the 15 acknowledged they had multiple well-being issues, and 4 of the 15 acknowledged going through depression. In addition, 10.3% (12/116) mentioned "change of location" and 9.5% (11/116) mentioned facing a "personal loss or bereavement." Furthermore, 90.7% (117/129) of users reported "hard to cope" or "slightly hard to cope" (see Multimedia Appendix 12), signifying a high percentage of users giving themselves a negative self-rating on their current ability to cope with daily tasks. A total of 59.7% (77/129) of users assessed and completed at least 1 wellness tool provided by the app (see Multimedia Appendix 13). Among those who completed, 72 were high users and 5 were low users. The remaining 40.3% (52/129) who did not complete a wellness tool only conversed with the app and likely assessed a wellness tool but not complete it. For details on most frequently reported major events or changes by 2 or more users, see Multimedia Appendix 14. The authors recognize that there would be overlap among the defined major event categories, which was a challenge to address given the anonymity of the users.

Qualitative Analysis

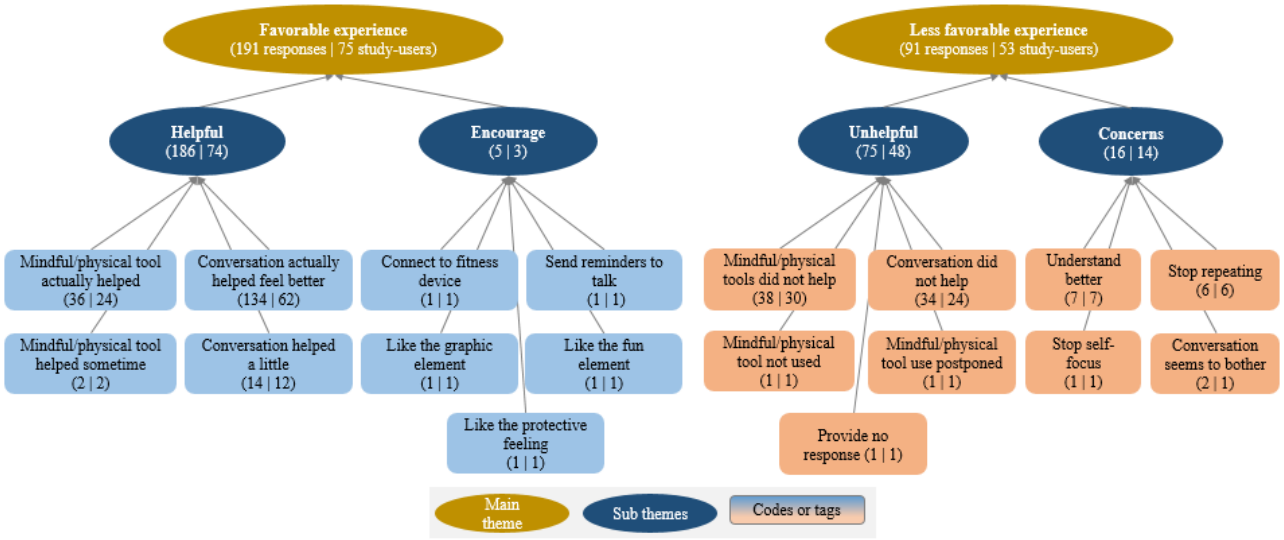
Engagement Effectiveness

In all, 73.6% (95/129) of users provided at least one response to the in-app feedback questions. Of those who responded, 86 were from the high users group and 9 were from the low users group. A total of 282 feedback responses were received from these 95 users. In total, 60.9% (172/282) responses were received for the in-app question "Have I been able to help you feel better yet?" that was asked at the end of each user session. A total of 90.8% (256/282) semistructured responses were received by choosing app-provided preformatted options. The remaining 9.2% (26/282) responses were by way of free-text and were provided by 17 of the 129 users.

Thematic analysis was carried out on the 282 responses received from the users. Two main themes emerged, one "Favorable Experience" with the subthemes Helpful and Encourage and the other "Less Favorable Experience" with the subthemes Unhelpful and Concerns. The thematic map with prevalence can be seen in Figure 2. A total of 67.7% (191/282) responses provided by 75 users found the app experience favorable. Of those favorable, 97.4% (186/191) responses found the conversation with the app and the tools helpful. A total of 32% (91/282) responses provided by 53 users found the app experience less favorable. Of those less favorable, 82% (75/91) responses found the conversation and tools either not helpful or did not use the tools; 13 responses (14%, 13/91) pointed to the app as not understanding or repeating, and a small fraction of 3 responses (3%, 3/91) mentioned that the app was self-focused and conversations seemed to bother the user.

Only 17 of the 129 users provided free-text feedback responses that provided additional insight into users' in-app experience. The free-text responses were analyzed keeping in perspective the user context as identified in the Context/Descriptive Analysis subsection within the Quantitative Analysis Results section. For a detailed analysis of the free-text in-app feedback responses, see Multimedia Appendix 15.

Figure 2. Thematic map with prevalence.



Favorable experience was the dominant theme from the user responses. Almost all of the favorable experiences were attributed to the helpfulness of the app in users actually feeling better after their conversation sessions and also after their use of app-provided mindfulness and physical activity techniques. Users mostly chose the preformatted response option of "Yes, Actually" in response to the feedback question "Have I been able to help you feel better yet?" that acknowledged that the app conversation and mindfulness or physical activity techniques were actually helping them feel better. If users found app-based conversations or mindfulness and physical activity techniques not helpful or expressed any concern, it was classified as a less favorable experience. Among those who provided a less favorable experience, 2 users postponed use or did not use the techniques or tools during the study period. These were also considered as a less favorable experience given that the users were not motivated enough to try out the techniques or tools. Users mostly chose the preformatted response option of "Not, Really" or "Not yet" in response to the feedback question "Have I been able to help you feel better yet?" that acknowledged that the app conversation and mindfulness or physical activity techniques did not help the user feel better. Some users chose the preformatted option of "Understand me better" or "Too repetitive" in response to in-app feedback question "Anything specific you'd like to improve?"

Of the 95 users who provided the 282 responses, those who reported hard to cope with daily tasks reported a higher proportion of favorable experience responses compared with less favorable experience responses (Figure 3).

Among those who reported hard to cope, those who reported relationship issues or changes as a major event expressed a significantly higher proportion of favorable experience responses compared with less favorable experience responses (Figure 4). Those who did not face coping challenges were mostly found to be mixed about their experience with the app.

Figure 3. Coping experience–based feedback response distribution.

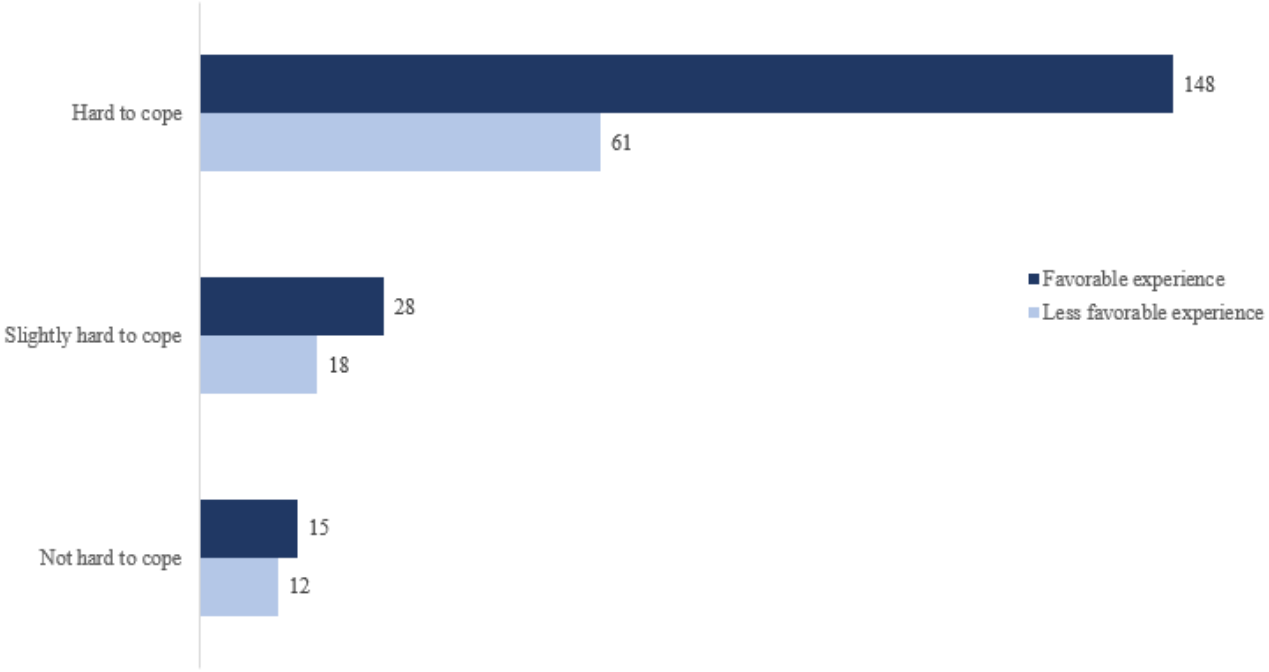
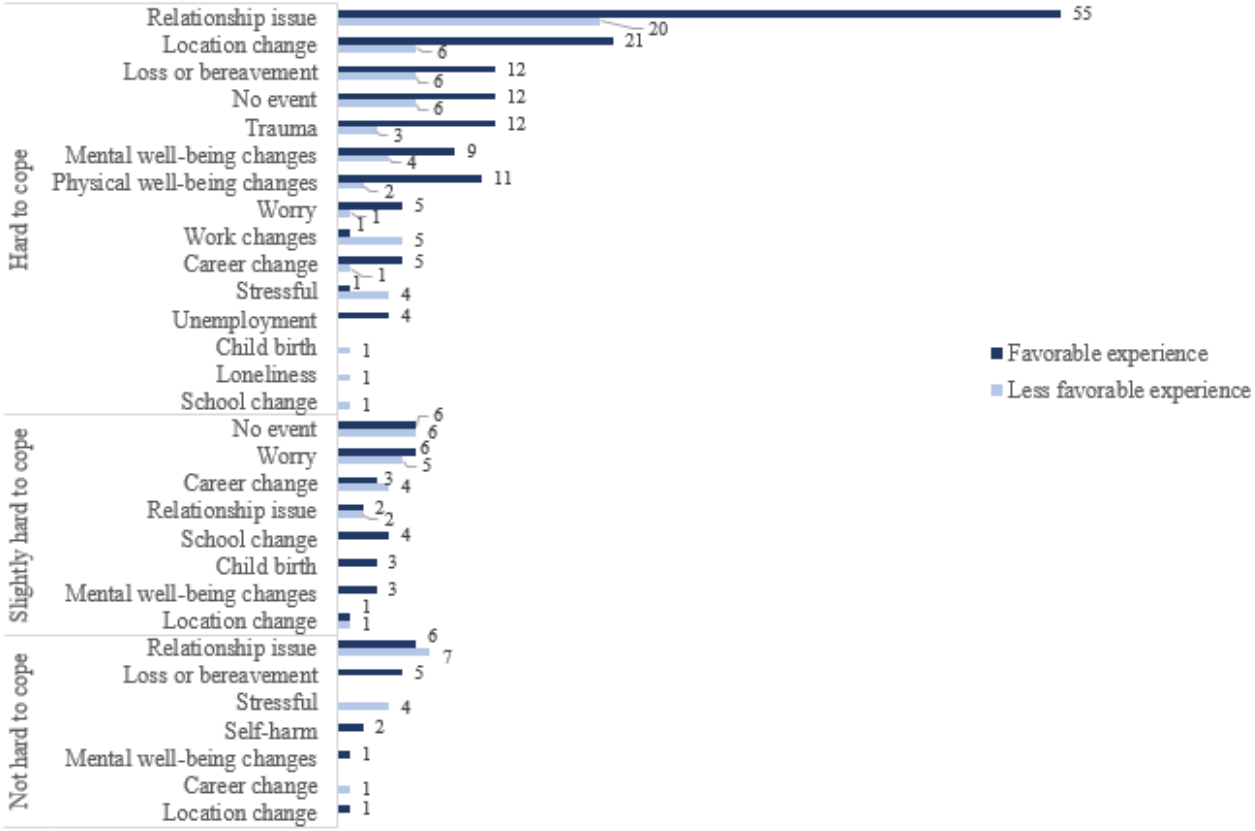


Figure 4. Coping major event-app experience-based feedback response distribution.



Engagement Efficiency

A total of 8075 anonymized conversational instances were obtained from 129 users during the study period. A relatively small proportion, 1.58% (128/8075) instances, of objections were observed in the conversation with the app.

The existing supervised classification-based ML algorithm that was deployed to classify objections in real time was tested on these 6611 instances. The remaining 18.13% (1464/8075) instances were ignored by the algorithm as the messages contained emoticons, texts in multiple lines, and special characters. The classifier model provided the following performance:

- Accuracy: 99.2% of objections and no objections that was detected was actually correct
- Specificity: 99.7% of no objections that was detected was actually correct
- Precision: 74.7% of objections detected (classified) was actually correct
- Recall: 62.1% of actual objections was detected (classified) correctly

See Figure 5 for the confusion matrix.

Figure 5. Confusion matrix of the objection handling machine learning model.

Number of conversational instances		Actual		
		Objection (True)	No objection (False)	
Predicted	Objection (True)	59 (True positive)	20 (False positive)	79
	No objection (False)	36 (False negative)	6496 (True negative)	6532
		95	6516	6611

Discussion

Principal Findings

The study revealed that the high users group had a significantly higher average improvement score in self-reported symptoms of depression compared with the low users group at a stringent PHQ-2 cutoff.

We found a significant reduction in PHQ-9 scores in high users and low users groups. We attribute the latter to the regression to the mean, suspecting that regression to the mean also plays a role in the high users group. Although the comparison group of "low users" does not fully constitute a control group, it provided an attempt to account for regression to the mean, as the reduction in PHQ-9 score seen in the high users group was significantly greater than that of the low users group. Users in both groups used the app during the full study period; therefore, they had comparable expectations that possibly reduced some biases.

A less significant effect was observed when the stringent cutoff PHQ-2 score was reduced. One explanation is that the app is most effective for people who show more severe symptoms of depression. As this is an in-the-wild study with no face-to-face screening, it is likely that lowering the PHQ-2 threshold score increased the number of people in the sample who were not mentally unwell and thus introduced additional unaccounted-for variability. Future work should deploy repeated measure questionnaires such as Resilience Scale RS-14 [42], which may be more sensitive to changes in resilience in the general population.

Relationship issues, mental well-being issues, location change, loss or bereavement, and career change formed the top major events or changes reported by users. Breakups and challenges with family members were the most common relationship issues. A recent study [43] has found that good mental health is not only the absence of symptoms but also what the user rates about his or her current ability to cope. Individuals who rated their current mental health as good had 30% lower probability of having a mental health problem at follow-up. Given the high proportion of negative self-rating on ability to cope in this study, the average improvement in self-reported symptoms of depression among high app users in a relatively short time period appears promising.

A high percentage of our study users (74%) provided in-app feedback. Most preferred to respond by clicking preformatted options presented by the app rather than free-text. A higher proportion of feedback found the app helpful and encouraging. There was an almost equal proportion of users who found the mindfulness and physical activity tools and techniques both helpful and not helpful, suggesting mixed experiences. Some suggested improvements included wanting the app to understand them better and wanting to avoid repetitions. Users who expressed hard to cope with daily tasks and who reported facing relationship issues in the recent past found the app helpful and gave a higher favorable experience feedback.

User objections (refusals or complaints) formed a relatively small proportion (1.58%). The existing objection detection ML model gave higher values for accuracy but lower for recall and precision, suggesting a need for further tuning of the model to reduce false positives and false negatives. A high performing ML model would become a necessity when conversation volumes increase to ensure high user engagement and retention. Continuous measurement of the objection rate can help provide an internal benchmark for chatbot apps to improve upon their engagement efficiency.

Comparison With Prior Work

Our study results were compared with other RCT studies [31,32] using an automated text-based conversational agent intervention to study impact on participants' mental well-being. One feasibility study ("first study") compared reduction in symptoms of depression from 2-week use of a CBT-oriented instant messenger-based conversational agent against an information control group in a nonclinical college population (n=70) [31]. The other pilot study ("second study") compared increased levels of psychological well-being from 2-week use of a positive psychology-oriented smartphone-based conversational agent against a wait-list control group in a nonclinical population (n=28) [32]. Both the studies reported between-group effect sizes based on the parametric Cohen *d*. The first study used PHQ-9 reporting a medium effect size of d=0.44 (from intent-to-treat analysis). The second study used the Flourishing Scale, Perceived Stress Scale, and Satisfaction with Life Scale and reported an effect size range of d=0.01 to 0.91 (from intent-to-treat analysis). The equivalent Cohen *d* of 0.47 (for CL of 0.63) from our study was comparable with that reported from the first study.

Both studies processed qualitative data gathered from responses to open-ended questions at postmeasurement using thematic analysis (Braun and Clarke, 2006). Although the approach taken differed from our study, there were similarities in observed experiences. The proportion of favorable responses (58 of 89 participants; 65%) to less favorable responses (31 of 89 participants; 35%) in the first study was similar to our study (68%:32%), suggesting users in both the studies reported a similar experience with a chatbot app. This observation will need validation in future studies. Users in our study and the first study highlighted the helpfulness of the conversation and the encouragement received, along with the feedback that chatbot provided an element of fun. Among the less favorable experiences, users (our study and first study) pointed to the repetitiveness of the conversation and a need for the app to understand the user better.

We also compared between-group effect sizes from 2 other RCTs that compared a Web-based human therapy intervention for depression with a waiting list [44,45]. We observed that our study effect size fell within the range of effect sizes reported (0.18-0.81) in those studies and closer to the larger effect size at follow-up. Our study effect size was also compared with the effect sizes reported in a 2018 meta-analysis [22] of RCT studies published before September 2016. The effect size from the 32 studies on major depressive disorders was found to range between 0.51 and 0.81 (Hedges *g*). Our study effect size was close to this effect range.

There are no known published metrics to compare how the "Objection Rate" fares among chatbot users with self-reported symptoms of depression. The observed objection rate of 1.58% when compared with the overall objection rate of 0.83% when all app users during the study period were considered (including excluded users); it is seen that the objection rate of users with self-reported symptoms of major depression (PHQ-2=6) was higher. This might indicate that users with high self-reported symptoms tend to object more during their conversation with a well-being app. Extensive research is needed in this area, especially given the ethical issues that may arise.

Value of the Study

The study design allows for scalability to conduct large longitudinal studies and, therefore, a relatively easier and early assessment of a chatbot's real-world effectiveness and engagement. The in-app based feedback approach allowed for real-time insights into the users' experience using a personalized intervention, without the danger of losing vital feedback and insights due to delays in collection. The study outlines a way to use existing conversational inputs to gather additional context about the user when no personally identifiable information or demographic information is collected. This is an approach that will aid in personalizing the user experience when conversing with a chatbot app. There exists tremendous value and potential for the app to enable Ecological Momentary Assessment (EMA) or Experience Sampling Method (ESM). Our study team supports the adoption of EMA or ESM as a research method for future studies where the objectives involve a more intensive, repeated, and momentary capture to assess changes in behavior, emotions, and mood of users. In future longitudinal studies, it will also add value to report on important app engagement measures such as user retention to complement the study findings. In a real-world context as conversations scale, the study recommends a need to evaluate and build high-performing ML models, including evaluation of unsupervised learning approaches, to detect objections in real-time while ensuring better control and interpretability of the model results. This allows for early handling of user objections to help make the chatbot more empathetic, enhance user engagement and retention, and strive for high ethical standards.

Limitations of the Study

A study of this nature has a number of limitations. A lack of a randomized controlled environment would lead to nonhandling of biases. No prior health information exists about the users, particularly their past or ongoing clinical history, diagnosis or treatment, or presence of comorbidities that could impact the effect. Both PHQ-2 and PHQ-9 have good acceptability for screening but do not confirm clinical diagnosis of depression (ie, participants with high PHQ-9 scores need not necessarily have depression and vice versa). This study design is a form of quasi-experimental design and is slightly lower in design quality compared with interrupted time-series designs (multiple pretest and posttest observations spaced at equal intervals of time). Statistical limitations include small and unbalanced comparison group sizes and not being able to account for variables such as age, gender, or socioeconomic status. A lack of detailed feedback responses on users' app experience limits the data available to gain insights using a qualitative analysis.

Bias may also exist in the form of increased exposure to certain features in the app for the high users group, which may partly contribute to influencing users in unknown ways. There is a need to insulate the app's design (such as color themes, font types, text alignments, icons, and emoticons) from contributing to the effects observed. The study sample size was too small to examine how people reacted to the app design elements and how that impacts their symptoms of depression. The authors intend to further delineate these issues in future research with larger samples.

Handling of these limitations would be a subject for future studies including the conduct of more elaborate comparison studies.

Conclusions

Our study identified a significantly higher average improvement in symptoms of major depression and a higher proportion of positive in-app experiences among high Wysa users compared with low Wysa users. These findings are encouraging and will help in designing future studies with larger samples and more longitudinal data points.

Acknowledgments

The authors would like to thank Wysa for providing access to their mobile app and the anonymized data during the study period for research purposes. Wysa funded the publication fees for the paper.

Authors' Contributions

BI and VS designed and performed research, analyzed data, and wrote the manuscript; SS and VS performed data wrangling. The manuscript was reviewed by all the authors.

Conflicts of Interest

BI is a scientific advisor to Wysa with no fiduciary associations. VS is an independent research consultant at Wysa and draws a consulting fee. SS is a technical lead and a paid employee at Wysa.

Multimedia Appendix 1

Wysa app study version.

[PNG File, 313KB - mhealth_v6i11e12106_app1.png]

Multimedia Appendix 2

Supplementary methods.

[PDF File (Adobe PDF File), 36KB - mhealth_v6i11e12106_app2.pdf]

Multimedia Appendix 3

Mixed methods approach followed for the study.

[PNG File, 16KB - mhealth_v6i11e12106_app3.png]

Multimedia Appendix 4

Study recruitment chart. "Enrollment" depicts inclusion of users who provided only 2 valid PHQ-9 assessment (pre and post over 14 days apart). “Allocation” splits users into 2 comparison groups based on their app usage on and between the two screening time-points. “Analysis” includes users who scored a total of “6” in the first 2 items of their PHQ-9.

[PNG File, 39KB - mhealth_v6i11e12106_app4.png]

Multimedia Appendix 5

In-app feedback question and responses.

[PDF File (Adobe PDF File), 34KB - mhealth_v6i11e12106_app5.pdf]

Multimedia Appendix 6

A typical user engagement with the Wysa app. Time period 1 denotes the start of app use by a user. Time period n denotes the end of the study period.

[PNG File, 50KB - mhealth_v6i11e12106_app6.png]

Multimedia Appendix 7

Sample objection quotes from users.

[PDF File (Adobe PDF File), 24KB - mhealth_v6i11e12106_app7.pdf]

Multimedia Appendix 8

Distribution of users based on Total Active Days on and between screening days.

[PNG File, 11KB - mhealth_v6i11e12106_app8.png]

Multimedia Appendix 9

Distribution of Wysa app users based on number of days between the screening days.

[PNG File, 10KB - mhealth_v6i11e12106_app9.png]

Multimedia Appendix 10

Region and time-zones of all included users.

[PNG File, 36KB - mhealth_v6i11e12106_app10.png]

Multimedia Appendix 11

Distribution of self-reported major events and changes of all included users.

[PNG File, 17KB - mhealth_v6i11e12106_app11.png]

Multimedia Appendix 12

Distribution of self-reported ability to cope among all included users.

[PNG File, 15KB - mhealth_v6i11e12106_app12.png]

Multimedia Appendix 13

Distribution of users who completed wellness tools.

[PNG File, 8KB - mhealth_v6i11e12106_app13.png]

Multimedia Appendix 14

Major events or changes reported by users.

[PDF File (Adobe PDF File), 27KB - mhealth_v6i11e12106_app14.pdf]

Multimedia Appendix 15

Analysis of free-text in-app feedback responses from study users.

[PDF File (Adobe PDF File), 38KB - mhealth_v6i11e12106_app15.pdf]

References

1. World Health Organization. Geneva: World Health Organization; 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL: http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/ [WebCite Cache ID 71tDp00UM]
2. Ishak WW, Balayan K, Bresee C, Greenberg JM, Fakhry H, Christensen S, et al. A descriptive analysis of quality of life using patient-reported measures in major depressive disorder in a naturalistic outpatient setting. Qual Life Res 2013 Apr;22(3):585-596. [doi: 10.1007/s11136-012-0187-6] [Medline: 22544416]
3. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One 2014;9(2):e90311 [FREE Full text] [doi: 10.1371/journal.pone.0090311] [Medline: 24587318]
4. Pulcu E, Elliott R. Neural origins of psychosocial functioning impairments in major depression. Lancet Psychiatry 2015 Sep;2(9):835-843. [doi: 10.1016/S2215-0366(15)00237-0] [Medline: 26360902]
5. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Feb;76(2):155-162 [FREE Full text] [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
6. Centre for Mental Health. 2010. The economic and social costs of mental health problems in 2009/10 URL: https://www.centreformentalhealth.org.uk/economic-and-social-costs-of-mental-health-problems [accessed 2018-08-23] [WebCite Cache ID 71tDUFW8c]
7. Farmer P, Stevenson D. GOV.UK. 2017 Oct 26. Thriving at Work: a review of mental health and employers URL: https://www.gov.uk/government/publications/thriving-at-work-a-review-of-mental-health-and-employers [WebCite Cache ID 71tDyIsy2]
8. World Health Organization. Geneva: World Health Organization; 2018. Mental Health Atlas 2017 URL: http://www.who.int/mental_health/evidence/atlas/mental_health_atlas_2017/en/ [WebCite Cache ID 71tEB6iJK]
9. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 2013 Jun;12(2):137-148 [FREE Full text] [doi: 10.1002/wps.20038] [Medline: 23737423]
10. Patel V. SUNDAR: mental health for all by all. BJPsych Int 2015 Feb;12(1):21-23 [FREE Full text] [doi: 10.1192/S2056474000000118] [Medline: 29093840]
11. Department of Health, UK. British Library. 2014 Sep 25. Access and waiting times standards for 2015-16 in mental health services: Impact Assessment URL: https://www.bl.uk/britishlibrary/~/media/bl/global/social-welfare/pdfs/non-secure/a/c/c/access-and-waiting-time-standards-for-201516-in-mental-health-services-impact-assessment.pdf [accessed 2018-08-23] [WebCite Cache ID 71tEOfm0q]
12. British Medical Association. 2018 Feb 26. New BMA research unveils blindspot in mental healthcare URL: https://www.bma.org.uk/news/media-centre/press-releases/2018/february/new-bma-research-unveils-blindspot-in-mental-healthcare [WebCite Cache ID 71tETlIw7]
13. Pedersen ER, Paves AP. Comparing perceived public stigma and personal stigma of mental health treatment seeking in a young adult sample. Psychiatry Res 2014 Sep 30;219(1):143-150 [FREE Full text] [doi: 10.1016/j.psychres.2014.05.017] [Medline: 24889842]
14. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]

15. Tindall L, Mikocka-Walus A, McMillan D, Wright B, Hewitt C, Gascoyne S. Is behavioural activation effective in the treatment of depression in young people? A systematic review and meta-analysis. Psychol Psychother 2017 Dec;90(4):770-796 [FREE Full text] [doi: 10.1111/papt.12121] [Medline: 28299896]
16. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
17. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front Robot AI 2017 Oct 12;4. [doi: 10.3389/frobt.2017.00051]
18. Callan JA, Wright J, Siegle GJ, Howland RH, Kepler BB. Use of computer and mobile technologies in the treatment of depression. Arch Psychiatr Nurs 2017 Dec;31(3):311-318. [doi: 10.1016/j.apnu.2016.10.002] [Medline: 28499574]
19. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
20. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med 2016 Oct;46(13):2679-2693. [doi: 10.1017/S0033291716001562] [Medline: 27649340]
21. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. J Am Med Assoc Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
22. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
23. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
24. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
25. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
26. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. Br Med J Open 2014;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
27. Ly KH, Janni E, Wrede R, Sedem M, Donker T, Carlbring P, et al. Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. Internet Interv 2015 Mar;2(1):60-68. [doi: 10.1016/j.invent.2014.12.002]
28. Ben-Zeev D, Brian RM, Jonathan G, Razzano L, Pashka N, Carpenter-Song E, et al. Mobile health (mHealth) versus clinic-based group intervention for people with serious mental illness: a randomized controlled trial. Psychiatr Serv 2018 May 25:appips201800063. [doi: 10.1176/appi.ps.201800063] [Medline: 29793397]
29. Noone C, Hogan MJ. A randomised active-controlled trial to examine the effects of an online mindfulness intervention on executive control, critical thinking and key thinking dispositions in a university student sample. BMC Psychol 2018 Apr 05;6(1):13 [FREE Full text] [doi: 10.1186/s40359-018-0226-3] [Medline: 29622047]
30. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
31. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
32. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46. [doi: 10.1016/j.invent.2017.10.002]
33. BBC Tomorrow's world. BBC. 2017. Would you trust a chatbot therapist? URL: http://www.bbc.co.uk/guides/zt8h2nb [accessed 2018-09-04] [WebCite Cache ID 72B9r5lgZ]
34. Emma S. Wysa. Wysa Case Studies; 2018 Apr 18. NHS Children Services: How the North East London NHS Foundation Trust uses Wysa for children's mental health URL: https://www.wysa.io/blog/nhs-children-services [accessed 2018-09-04] [WebCite Cache ID 72BAtBWuZ]
35. Braun A, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Braun V, Clarke V. What can “thematic analysis” offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]

37. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. BJPsych Open 2016 Mar;2(2):127-138 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001685] [Medline: 27703765]
38. Ruscio J. A probability-based measure of effect size: robustness to base rates and other factors. Psychol Methods 2008 Mar;13(1):19-30. [doi: 10.1037/1082-989X.13.1.19] [Medline: 18331151]
39. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law Hum Behav 2005 Oct;29(5):615-620. [doi: 10.1007/s10979-005-6832-7] [Medline: 16254746]
40. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005 Feb;34(1):215-220. [doi: 10.1093/ije/dyh299] [Medline: 15333621]
41. Zimmerman DW. Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. J Exp Educ 1998 Jan;67(1):55-68. [doi: 10.1080/00220979809598344]
42. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. J Nurs Meas 1993;1(2):165-178. [Medline: 7850498]
43. McAlpine DD, McCreedy E, Alang S. The meaning and predictive value of self-rated mental health among persons with a mental health problem. J Health Soc Behav 2018 Jun;59(2):200-214. [doi: 10.1177/0022146518755485] [Medline: 29406825]
44. Kramer J, Conijn B, Oijevaar P, Riper H. Effectiveness of a web-based solution-focused brief chat treatment for depressed adolescents and young adults: randomized controlled trial. J Med Internet Res 2014;16(5):e141 [FREE Full text] [doi: 10.2196/jmir.3261] [Medline: 24874006]
45. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. Lancet 2009 Aug 22;374(9690):628-634. [doi: 10.1016/S0140-6736(09)61257-5] [Medline: 19700005]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
CL: common language effect size
EMA: Ecological Momentary Assessment
ESM: Experience Sampling Method
ML: machine learning
NHS: National Health Service
PHQ-2: 2-item Patient Health Questionnaire
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.09.18; peer-reviewed by M Mulvenna, KL Ong; comments to author 26.09.18; revised version received 09.10.18; accepted 09.10.18; published 23.11.18.

Please cite as:
Inkster B, Sarda S, Subramanian V
An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study
JMIR Mhealth Uhealth 2018;6(11):e12106
URL: http://mhealth.jmir.org/2018/11/e12106/
doi:10.2196/12106
PMID:30470676

Original Paper

A Mobile Technology Intervention With Ultraviolet Radiation Dosimeters and Smartphone Apps for Skin Cancer Prevention in Young Adults: Randomized Controlled Trial

Elke Hacker[1], PhD; Caitlin Horsham[1], MSc; Dimitrios Vagenas[1], PhD; Lee Jones[1], BSc (Hons); John Lowe[2], PhD; Monika Janda[1,3], PhD

[1]Institute of Health and Biomedical Innovation, School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia
[2]School of Health and Sport Sciences, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore, Australia
[3]Centre of Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

Corresponding Author:
Elke Hacker, PhD
Institute of Health and Biomedical Innovation
School of Public Health and Social Work
Queensland University of Technology
60 Musk Avenue
Kelvin Grove
Brisbane, 4059
Australia
Phone: 61 07 3138 9674
Email: elke.hacker@qut.edu.au

Abstract

Background: Skin cancer is the most prevalent and most preventable cancer in Australia. Despite Australia's long-running public health campaigns, young Australian adults continue to report high levels of ultraviolet radiation (UVR) exposure and frequent sunburns. Young people are now increasingly turning away from traditional media, such as newspapers and TV, favoring Web-based streaming, which is challenging the health care sector to develop new ways to reach this group with targeted, personalized health promotion messages. Advances in technology have enabled delivery of time- and context-relevant health interventions.

Objective: The primary aim of this randomized controlled trial was to test the effect of UVR feedback from a smartphone app or a UVR dosimeter feedback device on sun protection habits, sun exposure behaviors, sunburn, and physical activity levels in young adults.

Methods: Young adults aged 18-35 years (n=124) were recruited from Queensland, Australia, between September 2015 and April 2016, via social or traditional media campaigns and outreach activities in the local community. Participants were randomized into 3 groups for a 4-week intervention: (1) no intervention control group; (2) UVR monitor group, who were asked to wear a UVR dosimeter feedback device set to their skin type; and (3) a SunSmart app group, who were asked to download and use the SunSmart phone app. Data were self-assessed through Web-based surveys at baseline and 1 week and 3 months postintervention.

Results: Complete data were available for 86.2% (107/124) of participants (control group, n=36; UVR monitor group, n=36; and SunSmart app group, n=35). Intervention uptake in the UVR monitor group was high, with 94% (34/36) of participants using the device all or some of the time when outdoors. All SunSmart app group participants downloaded the app on their smartphone. There was no significant difference in the change in the sun protection habits (SPH) index (main outcome measure) across the 3 groups. However, compared with the control group, a significantly greater proportion of the participants in the UVR monitor group reduced their time unprotected and exposed to UVR on weekends during the intervention compared with the baseline (odds ratio [OR]: 2.706, 95% CI 1.047-6.992, *P*=.04). This significant effect was sustained with greater reductions observed up to 3 months postintervention (OR: 3.130, 95% CI 1.196-8.190, *P*=.02). There were no significant differences between the groups in weekday sun exposure, sunscreen use, sunburn, suntan, or physical activity.

Conclusions: Using technology such as apps and personal UVR monitoring devices may improve some sun exposure behaviors among young adults, but as the SPH index did not increase in this study, further research is required to achieve consistent uptake of sun protection in young people.

Trial Registration: The Australian and New Zealand Clinical Trials register ACTRN12615001296527; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=368458 (Archived by WebCite at http://www.webcitation.org/731somROx)

(JMIR Mhealth Uhealth 2018;6(11):e199) doi:10.2196/mhealth.9854

KEYWORDS

health promotion; melanoma; preventive medicine; public health; skin neoplasms; mobile phone; sunburn; sunlight; technologies; Web apps

Introduction

Ultraviolet radiation (UVR) or sunlight exposure is the main environmental risk factor for melanoma and keratinocyte skin cancers (including basal cell carcinoma and squamous cell carcinoma). It is predicted that in the United States, new cases of melanoma will rise from around 70,000 in 2007-2011 to 116,000 in 2026-2031 [1]. Melanoma is the most common cancer in those aged 15-39 years in Australia [2]. Consistently across the United States, several European countries, and Australia, young adults are reporting higher levels of sunburn compared with older adults, despite having good knowledge and sun-protective intentions [3-5]. In Australia, people aged 18-24 years were seven times more likely to report sunburn on the previous weekend than those aged >65 years [6]. Young people, men, and those from a lower socioeconomic class or education level are all less likely to engage in preventive activities [7].

Over the past 30 years, Australia has successfully implemented world-class skin cancer prevention campaigns such as Slip! Slop! Slap!, SunSmart, and "Protect yourself in five ways from skin cancer" delivered mostly using traditional public health and media channels such as posters, brochures, television, radio, and newspaper advertising [8,9]. These programs have raised public awareness and improved preventive behaviors among Australians and have thought to have led to a slight reduction in melanoma incidence in younger generations [10]. Despite this success, the achievable impact of traditional media is waning due to the increased use of personalized internet-delivered multimedia content, especially among young people [11-13].

While the increasing use of mobile technology offers many opportunities for providing and collecting information and delivering time-, person-, or context-sensitive health interventions, very few studies to date have tested sun protection interventions with personalized messaging [14]. Buller et al provided personalized time until sunburn information using a mobile phone app to >600 US residents, aged above 18 years, which led to a marked increase in sun protection behavious [15]. A greater proportion of intervention group participants reported they kept time in the sun to a minimum (60% for app users vs 49% for nonusers; *P*=.04) and used more sun protection (39% vs 34%; *P*=.04). Our previous study recruited 574 participants aged 18-42 years from Queensland, Australia, sending 21 personalized motivational sun protection short message service text messages [16]. At 12 months postrandomization, sun protection group participants (mean change, 0.12) had significantly greater improvement in their sun protection habits (SPH) index than the physical activity attention control group participants (mean change, 0.02; *P*=.03) [16]. A change in the SPH index of 0.2 translates to the additional consistent use of at least one sun protection behavior. Furthermore, Djaja et al [17] have shown using the item response theory that a change to "using a hat consistently" moves a person from above-average to below-average skin cancer risk [17].

In addition, personalized feedback information could be received using UVR-detecting dosimeters. The personal UV dosimeter provides feedback by sounding an alarm at a defined UVR threshold, alerting users the need for sun protection to reduce the risk of sunburn. Commercial interest has seen a large number of UVR-detecting devices being marketed directly toward the public. The devices can be worn as watches (attached to a strap) or pinned to clothing such as hats or shirts. There is a lack of evidence whether they aid consumers' sun protection behaviors. To build the evidence for their efficacy for sun protection behavior change, the objective of this intervention trial was to evaluate one mobile phone app (SunSmart app, Cancer Council Victoria) and one personal UVR dosimeter monitor (Healthtronics SunSafe Pty Ltd), which has been shown in pretesting to provide accurate readings, and to assess the impact these have on young adults' sun exposure and sun protection habits compared with a no intervention control group.

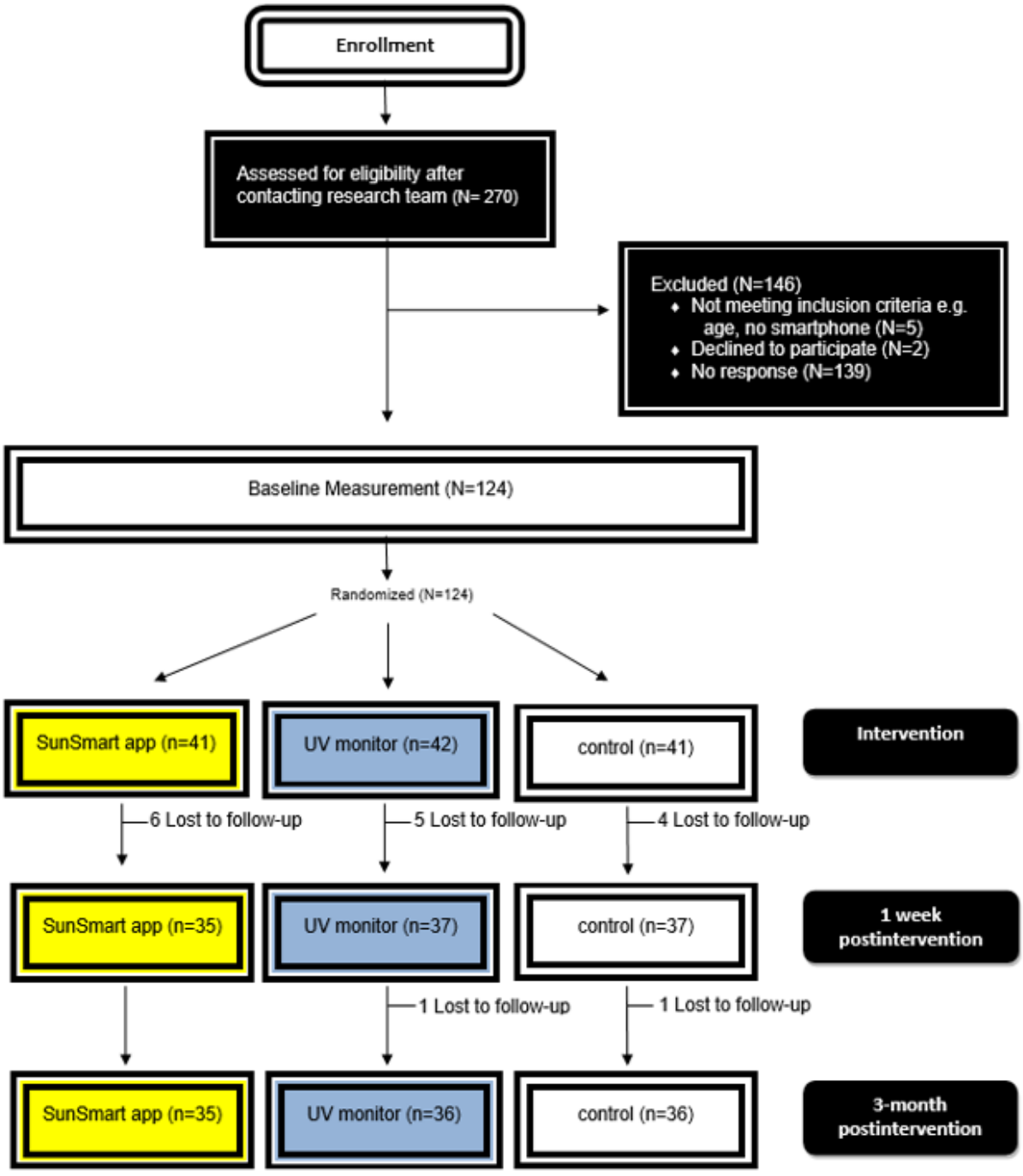
Methods

Study Design and Participants

The SknTec trial, conducted in Queensland, Australia, used a randomized controlled design with 2 intervention groups (SunSmart app or UVR monitor) and a measurement-only control group. Study approval was obtained from the Queensland University of Technology's (QUT) Human Research Ethics Committee, and the study was conducted in accordance with the Declaration of Helsinki with written informed consent from all participants (approval number QUT 1400000302). This trial adheres to the Consolidated Standards of Reporting Trials EHEALTH for randomized controlled trials checklist (Multimedia Appendix 1). Eligibility criteria included young adults who were aged 18-35 years, had never been diagnosed with melanoma, had Fitzpatrick skin type 1-3, owned a smartphone, and were not a regular user of the SunSmart app or a personal UVR dosimeter. Participants (n=124) were recruited through Web via emails at the university or social media. Traditional media such as posters at sporting centers and in the local community were also used for recruitment (Figure 1). Prospective participants completed a screening telephone call or in-person visit at the university. The project was outlined,

eligibility was determined, and written informed was consent obtained.

Figure 1. Flowchart of study recruitment. UV: ultraviolet.



The data collection was performed using a staggered recruitment process during September-December 2015, with data collection running over Spring, Summer, and Autumn in Brisbane, Australia, when the UV index is consistently >6 and can reach 14+, requiring sun protection every day (Multimedia Appendix 2). UVR data recorded during the study were captured using the Australian Radiation Protection and Nuclear Safety Agency UV-Biometer model 501 detector (Solar Light Co, Philadelphia, PA, USA) located in Brisbane (latitude 27°S, 153°E). During the baseline 2-week period, participants completed a Web-based questionnaire and recorded their daily sun exposure, as well as physical activity, using a Web-based sun diary described previously [18]. After the baseline period, participants were block randomized (permuted blocks of 6), stratified by gender, using a computer-generated random number list created by the study software engineer independent from other study procedures. Participants were randomized at baseline when their allocation was created in the database. The research nurse and participants were blinded to the allocation of the intervention arms during the baseline phase, which was only revealed by the research nurse to participants at the commencement of the intervention phase.

For the 4-week intervention phase, participants were separated into 3 groups: (1) no intervention, measurement-only control group; (2) UVR monitor group, where participants were asked

to wear a UVR dosimeter feedback device while outdoors (Healthtronics SunSafe Pty Ltd; Figure 2, image on left); or (3) the SunSmart app group, where participants were asked to download and use the free SunSmart phone app on their personal mobile phone (Cancer Council Victoria, Australia; Figure, image on right). During the 4-week intervention phase, participants were asked to complete daily sun diaries. Participants were emailed by the research team if the daily sun diaries had not been completed for >3 days in a row. At the completion of the intervention phase, participants in the UVR monitor group returned their monitors via mail. Participants in the SunSmart app group were emailed instructions to remove the app from their phone and asked to confirm it was uninstalled via return email.

Follow-up posttest measurements were taken at 1 week and 3 months postintervention, with participants completing a Web-based questionnaire and recording their sun exposure daily for 2 weeks using the Web-based sun diary. Of note, participants were reimbursed for their time at the end of the study with an Aus $70 gift card.

Intervention Devices

Ultraviolet Radiation Monitor

The Healthtronics SunSafe UV dosimeter device (Figure 2) pins to an individual's clothing and can be personalized for skin type alarming when UVR thresholds are met. This device computes a daily maximum UV dose for each particular skin type on a scale of 1-5 based on the Fitzpatrick skin types (1=very fair skin type, sunburns always, never suntans; 2=fair skin type, sunburns easily, minimal suntan; 3=light-medium skin type, some sunburn, gradual suntan; 4=medium skin type, minimally sunburns, always suntans; 5=medium-dark skin type, rarely sunburns, always suntan). Its UV detectors are housed on a sloping panel to adjust for the different positional orientation areas of the body exposed to the sun when standing up or lying down. The UV-B-SAFE 1 model is water resistant and solar powered with a charge time of 30 seconds in the sun. The UVR monitor plays a short tune when it is turned on to advise it has sufficient power. When the alarm sounds continuously, it means the user has reached his or her maximum UVR threshold for the day. The UV monitor is capable of providing real-time dose levels only and does not record personal UV exposure data to a Web-based database. All participants assigned to the UVR monitor device confirmed receiving the device, which contained the manufacturer's printed instructions (Multimedia Appendix 3).

Figure 2. Intervention devices; left: personal ultraviolet radiation dosimeter monitor (Healthtronics SunSafe Pty Ltd, Australia); right: SunSmart app (Cancer Council Victoria, Australia).

SunSmart App

The SunSmart app (Figure 2) displays the daily UV index, the weather for a range of Australian locations, and a daily time period when sun protection is required based on the most sun-sensitive skin type-1. During download, the app notification function was enabled to send daily reminders of the time periods sun protection is required. Furthermore, the SunSmart app gives recommendations for how people can best protect themselves from the sun, how the UV index works, and a vitamin D tracker and sunscreen calculator tool [19].

Outcome Measures

Outcomes were assessed at baseline and 1 week and 3 months postintervention. Sociodemographic data were collected at baseline, including skin cancer risk factors (hair color, eye color, tendency to burn, ability to tan, personal or family history of skin excisions, or skin cancer) and sun protection attitudes and intentions using questionnaires previously developed [20]. An evaluation questionnaire was conducted at the end of the study and satisfaction with the intervention devices and the intervention delivery was assessed. Participants were asked to rate on a 10-point Likert scale how satisfied they were with the intervention device (1=not at all satisfied, 5=moderately satisfied, 10=extremely satisfied). Participants were asked to self-report their use of the intervention devices.

Sun Protection Habits Index

The primary outcome measure was the SPH index developed by Glanz et al [21] measured at baseline and evaluation time points. It queries the frequency of 6 sun-protective methods that are used when outdoors using a 4-point Likert scale (1=never or rarely to 4=always), which are averaged to derive the score, including wearing a shirt with sleeves, wearing a hat, wearing sunglasses, using sunscreen, staying in the shade, and limiting time in the sun during midday hours. SPH index test-retest reliability has good internal consistency (0.76) and test-retest reliability (0.78), and estimates of its validity have been previously reported [22].

Ultraviolet Radiation Exposure, Sunburn, and Physical Activity

Data on frequency of sunburn and suntan (number of times), time spent in the sun unprotected on weekdays and weekends (minutes and hours, body areas exposed unprotected), sunscreen use (yes or no), and physical activity (minutes and hours) were collected using the Web-based sun diary [18].

Statistical Analysis

Generalized estimating equations (GEE) models were used for analyzing changes in the mean combined SPH index over time and for individual items, including wearing a shirt with sleeves, sunglasses, staying in the shade, sunscreen use, limiting time in the sun, and wearing a hat. The model contained the group, gender, and skin type. In addition, we fitted the interaction of time with (1) gender and (2) group. The sun diary variables (UVR unprotected exposure, UVR unprotected torso exposure, sunscreen use, and physical activity) were dichotomized into (1) yes (improvement of ≥5 minutes from every individual's baseline) or (2) no (no improvement or improvement of <5 minutes). Number of participants in each category and percentage are presented in Table 1. Furthermore, logistic binary regression analyses were used to detect the odds of improvement in each intervention group compared with the control group.

Regarding sample size calculations, to detect an effect size of 0.4, which allows a change in the mean SPH index of 0.2 (from a mean score of 2.3-2.5 at follow-up), given a common SD of 0.5 among 3 groups where 2 of the groups are compared with one control group (based on the Dunnett multiple comparison test), an optimal sample size of 300 was determined. However, only 55% of the requested funding was received, allowing us to recruit a maximum of 200 participants; of them, 124 completed the study, allowing the ability to detect an effect size of 0.6, or a 0.3 difference in SPH among groups [23].

Table 1. Participant characteristics.

Characteristics	Total (n=124)	Control (n=41)	SunSmart App (n=41)	Ultraviolet radiation monitor (n=42)
Mean age in years	25.8	25.4	26.5	25.4
Gender, n (%)				
Female	85 (68.5)	26 (63.4)	32 (78.0)	27 (64.3)
Male	39 (31.5)	15 (36.6)	9 (22.0)	15 (35.7)
Highest completed education, n (%)				
Completed high school	21 (16.9)	8 (19.5)	7 (17.1)	6 (14.3)
Trade or technical certificate or diploma	12 (9.7)	6 (14.6)	3 (7.3)	3 (7.1)
University or college degree	91 (73.4)	27 (65.9)	31 (75.6)	33 (78.6)
Current work situation, n (%)				
Employed full-time	24 (19.4)	9 (22.0)	9 (22.0)	6 (14.3)
Part-time or casual	20 (16.1)	6 (14.6)	8 (19.5)	6 (14.3)
Student	80 (64.5)	26 (63.4)	24 (58.5)	30 (71.4)
Is your main job now…, n (%)				
Mainly indoors	109 (87.9)	38 (92.7)	35 (85.4)	36 (85.7)
Mainly outdoors	2 (1.6)	N/A[a]	1 (2.4)	1 (2.4)
About equal amounts indoors and outdoors	13 (10.5)	3 (7.3)	5 (12.2)	5 (11.9)
Born in Australia, n (%)				
Yes	58 (46.8)	20 (48.8)	20 (48.8)	18 (42.9)
No	66 (53.2)	21 (51.2)	21 (51.2)	24 (57.1)
Eye color, n (%)				
Blue or gray	28 (22.6)	9 (22.0)	13 (31.7)	6 (14.2)
Green	16 (12.9)	5 (12.2)	4 (9.8)	7 (16.7)
Brown	65 (52.4)	24 (58.5)	19 (46.3)	22 (52.4)
Other[b]	15 (12.1)	3 (7.3)	5 (12.2)	7 (16.7)
Natural hair color at the age of 21 years (or now if younger), n (%)				
Red (including auburn)	7 (5.6)	3 (7.3)	2 (4.9)	2 (4.8)
Fair or blonde (including white)	8 (6.4)	1 (2.5)	4 (9.7)	3 (7.1)
Light brown	25 (20.2)	8 (19.5)	10 (24.4)	7 (16.7)
Dark brown	42 (33.9)	16 (39.0)	10 (24.4)	16 (38.1)
Black	42 (33.9)	13 (31.7)	15 (36.6)	14 (33.3)
Skin color, n (%)				
Fair	72 (58.1)	24 (58.6)	21 (51.2)	27 (64.3)
Medium	38 (30.6)	13 (31.7)	15 (36.6)	10 (23.8)
Olive or dark	13 (10.5)	3 (7.3)	5 (12.2)	5 (11.9)
Black	1 (0.8)	1 (2.4)	N/A	N/A
Skin reaction in strong summer sun for 30 minutes without protection, n (%)				
My skin would not burn at all	10 (8.1)	5 (12.2)	1 (2.5)	4 (9.5)
My skin would burn lightly	44 (35.5)	13 (31.7)	13 (31.7)	18 (42.9)
My skin would burn moderately	47 (37.9)	15 (36.6)	16 (39.0)	16 (38.1)
My skin would burn severely	23 (18.5)	8 (19.5)	11 (26.8)	4 (9.5)

Characteristics	Total (n=124)	Control (n=41)	SunSmart App (n=41)	Ultraviolet radiation monitor (n=42)
Skin reaction if you spend several weeks at the beach and you are often in the strong sun, without any protection, n (%)				
My skin would not tan	10 (8.0)	4 (9.8)	2 (4.9)	4 (9.5)
My skin would tan lightly	25 (20.2)	8 (19.5)	8 (19.5)	9 (21.5)
My skin would tan moderately	60 (48.4)	21 (51.2)	18 (43.9)	21 (50.0)
My skin would tan deeply	29 (23.4)	8 (19.5)	13 (31.7)	8 (19.0)
Previous skin cancer, mole, or other spot(s) removed or treated, n (%)				
Yes	15 (12.1)	5 (12.2)	7 (17.1)	3 (7.1)
No	108 (87.1)	36 (87.8)	34 (82.9)	38 (90.5)
Unsure or do not know	1 (0.8)	N/A	N/A	1 (2.4)

[a]N/A: not applicable.

[b]Other: mixed or undefined eye color.

Results

Participant Characteristics

The mean age of all participants was 25.8 (range 18-35) years and 68.5% (85/124) were females; 73.3% (91/124) participants had a university degree, and 87.9% (109/124) of participants worked mainly indoors (Table 1). More than half of the participants (72/124, 58.1%) had fair skin that was sun sensitive, including skin that moderately or severely burns after 30 minutes of sun exposure in summer without protection. The characteristics were quite evenly distributed among the groups.

Intervention and Data Collection Completeness

In the UVR monitor group, 94% (34/36) of participants self-reported using the UVR monitor all or some of the time when outside. All participants in the UVR monitor group confirmed receiving the device and returned the UVR monitor postintervention. All SunSmart app group participants downloaded the app on their smartphone, and 97% (34/35) of participants reported they received the daily UV index sun protection pop-up notifications. All participants completed the baseline questionnaire (n=124); 87.9% (109/124) completed the 1-week and 86.3% (107/124) completed the 3-month postintervention questionnaire. The Web-based sun diary was completed by 95.2% (118/124) participants at baseline, 88.7% (110/124) participants during the intervention and 1 week postintervention, and 86.3% (107/124) participants at 3 months postintervention.

Sun Protection Habits Index

At baseline, the mean SPH index value was 2.42 (SE 0.08) for the control group, 2.36 (SE 0.08) for the UV monitor group, and 2.47 (SE 0.07) for the SunSmart app group (Multimedia Appendix 4). At the 3-month time point, the SPH index had improved by +0.13, +0.14, and +0.06 in the UV monitor, SunSmart app, and control groups, respectively (P=.001). This increase did not differ significantly by group, resulting in a nonsignificant group by time interaction (P=.35, GEE).

Ultraviolet Radiation Exposure, Sunscreen Use, and Physical Activity

Compared with the control group, a significantly greater proportion of the UVR monitor group participants improved their sun protection on weekends during the intervention phase (OR 2.706, 95% CI 1.047-6.992, P=.04; Table 2). This reduction in weekend unprotected exposure in the UVR monitor group was on average 58.78 (SE 13.41) minutes each day, a 58.5% (SE 5.18) reduction from baseline exposure in those who improved (n=23). The UVR monitor group continued demonstrating an improvement in sun protection on weekends 3 months postintervention (OR 3.130, 95% CI 1.196-8.190, P=.02; Table 2). This reduction in weekend unprotected exposure was on average 61.05 (SE 12.51) minutes each day, a 75.3% (SE 5.66) reduction from baseline exposure in those who improved (n=23). Weekend UVR exposure did not differ significantly between the control and the SunSmart app groups at any time point. Total and weekday UVR exposure did not differ significantly at any time point between the intervention group and the control group. Sunscreen use and physical activity levels remained largely unchanged across the study period in all 3 groups (Multimedia Appendix 4). The number of weekdays that sunscreen was used was observed in the SunSmart app group 1 week postintervention, but these results were not statistically significant (OR 2.808, 95% CI 0.854-9.238, P=.09; Multimedia Appendix 4, Supplementary Table 2), which was an increase in sunscreen use from 3 out of 10 days at baseline compared with 4 out of 10 days at 1 week postintervention.

Unprotected Ultraviolet Radiation Torso Exposure Incidence

We observed that 52.3% (56/107) of participants who completed the study reported unprotected UVR torso exposure at one or more time points during the study. Unprotected UVR torso exposure did not differ by gender, with 55% (41/75) of females and 47% (15/32) of males in the study. Unprotected torso exposure incidence did not differ significantly between the control and intervention groups at any time point (Multimedia Appendix 4).

Table 2. Participants who reduced their time in the sun unprotected.

Characteristics	Yes[a], n (%)	No, n (%)	OR	95% CI	*P* value
During intervention[b]					
Weekday and weekend ultraviolet radiation (UVR) unprotected exposure					
Control	14 (38.9)	22 (61.1)	1.00	N/A[c]	Ref[d]
SunSmart app	16 (44.4)	20 (55.6)	1.379	0.525-3.627	.51
UVR monitor	18 (47.4)	20 (52.6)	1.407	0.557-3.549	.47
Weekday UVR unprotected exposure					
Control	15 (41.7)	21 (58.3)	1.00	N/A	Ref
SunSmart app	11 (30.6)	25 (69.4)	0.622	0.231-1.670	.35
UVR monitor	18 (47.4)	20 (52.6)	1.280	0.508-3.225	.60
WeekEnd UVR unprotected exposure					
Control	13 (36.1)	23 (63.9)	1.00	N/A	Ref
SunSmart app	15 (41.7)	21 (58.3)	1.533	0.563-4.176	.40
UVR monitor	23 (60.5)	15 (39.5)	2.706	1.047-6.992	.04
1 week after the intervention[e]					
Weekday and weekend UVR unprotected exposure					
Control	18 (48.6)	19 (51.4)	1.00	N/A	Ref
SunSmart app	21 (58.3)	15 (41.7)	1.582	0.611-4.095	.35
UVR monitor	20 (54.1)	17 (45.9)	1.242	0.498-3.095	.64
Weekday UVR unprotected exposure					
Control	19 (51.4)	18 (48.6)	1.00	N/A	Ref
SunSmart app	17 (47.2)	19 (52.8)	0.972	0.375-2.514	.95
UVR monitor	14 (37.8)	23 (62.2)	0.576	0.228-1.455	.24
Weekend UVR unprotected exposure					
Control	14 (37.8)	23 (62.2)	1.00	N/A	Ref
SunSmart app	19 (52.8)	17 (47.2)	2.238	0.831-6.023	.11
UVR monitor	21 (56.8)	16 (43.2)	2.173	0.853-5.535	.10
3 months after the intervention[f]					
Weekday and weekend UVR unprotected exposure					
Control	16 (44.4)	20 (55.6)	1.00	N/A	Ref
SunSmart app	20 (57.1)	15 (42.9)	1.748	0.666-4.587	.26
UVR monitor	18 (50.0)	18 (50.0)	1.250	0.495-3.162	.64
Weekday UVR unprotected exposure					
Control	14 (38.9)	22 (61.1)	1.00	N/A	Ref
SunSmart app	15 (42.9)	20 (57.1)	1.205	0.456-3.182	.71
UVR monitor	14 (38.9)	22 (61.1)	1.000	0.385-2.595	.99
Weekend UVR unprotected exposure					
Control	13 (36.1)	23 (63.9)	1.00	N/A	Ref
SunSmart app	18 (51.4)	17 (48.6)	1.898	0.717-5.026	.20
UVR monitor	23 (63.9)	13 (36.1)	3.130	1.196-8.190	.02

[a]Yes: reduced average daily minutes of time unprotected in the sun compared with baseline based on the self-reported diary entry.

[b]n=110 for the intervention measurement period.

[c]N/A: not applicable.

[d]Ref: Reference *P* value.

[e]n=109 for the 1 week after intervention measurement period.

[f]n=107 for the 3 months after intervention measurement period.

Sunburn and Suntan Incidence

Sunburn rates were high with 58.0% (62/107) of participants reporting one or more (range 1-11) sunburns during the study. In this study, 63% (46/75) of all females and 50% (16/32) of all males reported one or more sunburns. No differences were observed between groups in the incidence of sunburn and suntan (Multimedia Appendix 4). Deliberate suntanning behavior during the study was reported by 29.9% (32/107) of participants (range 1-7), which was observed in 37% (28/75) of females and 13% (4/32) of males. Sunburn and suntan incidence did not differ significantly between the control and intervention groups at any time point (Multimedia Appendix 4).

Satisfaction With Intervention Devices

Two-thirds of participants (UVR monitor group, 23/36, 64%; SunSmart app group, 23/35, 66%) found their intervention helpful to guide their sun-protective behavior (Multimedia Appendix 4). About half of the participants (UVR monitor group, 17/36, 47%; SunSmart app group, 19/35, 54%) self-reported that they changed or modified their behavior in response to the output from the device. In the UVR monitor group, 47% (17/36) participants found the device to be encouraging to engage in sun protection; however, only 19% (7/36) would purchase one. In the SunSmart app group, 63% (22/35) of participants found the app encouraging to engage in sun protection, and 40% (14/35) would download it in the future. In addition, 36% (13/36) of participants in the UVR monitor group and 60% (21/35) in the SunSmart app group reported the device repeated what they already knew. UVR monitor participants' mean response for intervention device satisfaction out of a scale from 1 to 10 was 5.19 (SE 0.47), and the SunSmart app participants' mean response was 5.66 (SE 0.36). Qualitative feedback from the open-ended responses was grouped by themes, showing that young adults found the SunSmart app likable and easy to navigate. However, feedback for the content was that "it never changes," "it's always 8 am to 4 pm use sun protection," and "gets boring." Participants liked the personalized feedback by the UVR monitor; however, they were unlikely to carry a separate device for UVR detection. Participants wanted more control over the feedback method that the UVR monitor provides and would find it more appealing if they could tailor the alert or alarm for their specific preferences (eg, if it allowed the users to select their song as an alarm or a more subtle vibration alert).

Discussion

Principal Findings

This randomized controlled trial (RCT) examined the impact of using digital or electronic technologies to improve young peoples' sun protection or sun exposure behaviors. We found no consistent benefit of providing participants with either a mobile phone app or electronic dosimeter for their sun protection habits compared with a no intervention control group. Despite all 3 groups reporting a similar improvement in the main outcome measure, the SPH index, our young participants continued to experience high sunburn rates throughout the study period.

While we did not observe a significant effect on the SPH index, some differences in specific measures of sun exposure were noted among the groups. The reduction in weekend unprotected sun exposure for the UVR monitor group was encouraging as weekend sun exposure is common, with 21% of adolescents reporting being sunburnt on an average summer weekend in Australia [3]. Previous studies have suggested a suboptimal understanding of the UV index and peak UVR times among young adults as reasons for this [24,25]. Our findings suggest using a personal UV monitor that produces an auditory alarm may be helpful to educate young adults on what is a safe level of UVR exposure for their skin type. However, this was not enough to also change their sun protection habitual behavior as measured by the SPH index over and above the change achieved in the control group. Moreover, the change was not strong enough to reduce sunburn rates. Participants commented that they would have liked to further personalize the alarm sounds, which could be tested in future studies. Furthermore, carrying a separate device for UVR detection was mentioned as being burdensome by some participants, and future work could explore the potential to utilize movement and light sensors in smart devices already carried by people as a way of capturing UVR exposure.

Previous studies testing UVR monitors have reported varying results. Carli et al [26] (n=91) found longer sun exposure (*P*=.003) and more frequent sunburns (*P*=.004) in the UV monitor group compared with the control group. These unfavorable adverse outcomes may have been due to limitations of the UV monitor's detector (SunCast UV monitor), which when not ideally positioned toward the sun may have underreported UVR exposure. The UVR monitor (UV-B-Safe model) tested in our study had a sloping panel of detectors to better accommodate positional body orientations and alerted users with an auditory alarm in contrast to the SunCast UV monitor, which displays the UVR measurement on a UV index scale and did not have an alarm function. A Swedish study tested a UVR intensity indicator (Teraco, Inc, USA), which changed color if the UVR levels were moderate, high, or extreme [27]. These authors reported no statistically significant differences in the frequency of sunbathing, sunburn, or attitudes toward being in the sun between groups receiving either the UVR monitor or written information about sun protection in those aged 18-37 years. However, participants' use of the UVR intensity indicator device was low, with only 42% using it compared with 94% (34/36) of participants in this study.

In a randomized clinical trial evaluating the Solar Cell app, which provided personalized, real-time sun protection advice and alerts, it was found that a greater proportion of app users

reduced the use of sunscreen (29% vs 35%) in the control group (*P*=.048), with a greater proportion of app users increasing the use of shade instead (41% vs 34% control; *P*=.03) [15]. Findings from our study and a study by Buller et al [15] showed no difference in the number of sunburns between app users and control group participants. Of note, 57.9% (62/107) of participants reported one or more sunburns during this study, which is slightly higher than the 47% rate observed among young adults in the United States by Buller et al [28]. The prevalence of sunburn among young men was not significantly different from the prevalence among young women, consistent with previous reports from the United States [28]. High-risk sun exposure behaviors such as torso exposure were commonly reported, and suntanning behaviors were similar to those reported in previous studies, with a higher proportion of females reporting suntanning compared with males [29].

Participants provided a satisfaction rating score of 5 out of 10, indicating they were moderately satisfied with the intervention devices, and the qualitative feedback received demonstrated that personalized, tailored engaging feedback is preferred by young adults. They recommended combining elements from both intervention devices, which may be advantageous to reduce sunburn rates in this population. In addition, Buller et al [15] reported the beneficial impact of tailoring information to each user "in the moment," promoting a sense of volition, choice, and control. Likewise, our previous work provides evidence for the benefit of personalized approaches. The Healthy Text study recruited 574 participants (age, 18-42 years) who received 21 motivational messages on sun protection compared with physical activity attention control messages [20]. At 12 months postrandomization, the sun protection group had significantly greater improvement in its SPH index than the control group [16], and building behavioral capacity (eg, obtaining information and receiving reminders) was the most valued aspect of the messages [30]. Heckman et al [31] recently reported significant decreases in UV exposure and increases in SPH index 3 and 12 weeks after baseline for participants who received a tailored multimedia internet intervention program (UV4.me). These studies have illustrated that further improvements to the technology platforms are needed to reduce the sunburn prevalence in young adult populations.

The observation of improvement in the measurement-only control group may be due to participants completing multiple surveys about sun protection, which may have increased attention to their own sun-protective behavior. Campbell and Stanley [32] described the impact of exposure to a pretest or intervening assessment influences performance on a posttest, and a more recent work by Koster et al [33] reported that simply keeping a UV exposure diary increased attention toward the behavior examined. The change in the control group could also be explained by the Hawthorne effect, which is when a person's behavior changes because he or she is knowingly under observation. Future study design could incorporate an attention control group that receives equivalent information about an alternative activity, for example, physical activity. Other reasons the control group improved could be exposure to other sun protection programs implemented over the study period (not measured) or seasonal variation with baseline data collected in Spring and follow-up data collected at the end of Summer. Seasonal variations may be partly responsible for some of the changes observed; for example, if young people were more likely to use sunscreen later in the summer or get sunburnt in early summer, or be employed seasonally with changing work schedules.

The strengths of this study include the RCT design, which created equivalent groups, high participant retention, and testing relatively low-cost and readily available intervention components. Limitations of this study include self-reported outcome measures, which are subject to recall and social desirability biases. The proper use of sun protection methods (eg, the adequate thickness in the application of sunscreen) was not objectively assessed. A further limitation is that UV monitor compliance was self-reported and personal UV exposure data were not captured in a Web-based study database. Our sample size was relatively small and may have led to low statistical power, contributing to the nonsignificant findings. We used convenience sampling in a university setting, and participation also involved time-intensive activities including completing a screening questionnaire and daily sun diaries, which may be perceived as too burdensome, leading to a unique sample. Furthermore, participants were mainly highly educated (university or college degree) and worked indoors, and results may not be generalizable to other subgroups of the population.

Conclusions

We aimed to provide evidence for the effectiveness of digital and mobile technologies to improve sun protection behaviors among young adults. Self-monitoring devices for maintaining wellness are becoming more widespread. Tracking one's health may enable consumers to improve health outcomes, but we found a relatively limited impact on important sun protection behaviors in this young population. Hence, an even more personalized approach to public health efforts may be needed to facilitate UVR protection and avoid increases in skin cancer cases.

Acknowledgments

The authors would like to thank all participants in the SknTec Study for their time and the project staff Linda Finch and Sam Vaartjes for their contributions to the study. UVR measurements were collected by the Australian Radiation Protection and Nuclear Safety Agency UV network and data provided to the research team. The study was funded by a research grant from Cancer Council Queensland (#1081370). MJ is funded by the National Health and Medical Research Council Trip Fellowship (#1151021). The sponsors of the study (Cancer Council Queensland) had no role in the study design, collection, analysis, and interpretation of data; in the writing of this manuscript; and in the decision to submit the paper for publication. The corresponding author had full access to all data in the study and final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 490KB - mhealth_v6i11e199_app1.pdf]

Multimedia Appendix 2

Supplementary Figure 1.

[PDF File (Adobe PDF File), 260KB - mhealth_v6i11e199_app2.pdf]

Multimedia Appendix 3

Supplementary Figure 2.

[PDF File (Adobe PDF File), 743KB - mhealth_v6i11e199_app3.pdf]

Multimedia Appendix 4

Supplementary Tables 1-7.

[PDF File (Adobe PDF File), 45KB - mhealth_v6i11e199_app4.pdf]

References

1. Whiteman DC, Green AC, Olsen CM. The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031. J Invest Dermatol 2016 Jun;136(6):1161-1171 [FREE Full text] [doi: 10.1016/j.jid.2016.01.035] [Medline: 26902923]
2. Australian Institute of Health and Welfare. Cancer in Australia: an overview. 2014. URL: https://www.aihw.gov.au/reports/cancer/cancer-in-australia-an-overview-2014/contents/table-of-contents [accessed 2018-10-09] [WebCite Cache ID 732CIUjtU]
3. Makin JK, Warne CD, Dobbinson SJ, Wakefield MA, Hill DJ. Population and age-group trends in weekend sun protection and sunburn over two decades of the SunSmart programme in Melbourne, Australia. Br J Dermatol 2013 Jan;168(1):154-161. [doi: 10.1111/bjd.12082] [Medline: 23039760]
4. Centers for Disease Control and Prevention (CDC). Sunburn and sun protective behaviors among adults aged 18-29 years--United States, 2000-2010. MMWR Morb Mortal Wkly Rep 2012 May 11;61(18):317-322 [FREE Full text] [Medline: 22572977]
5. Diffey BL, Norridge Z. Reported sun exposure, attitudes to sun protection and perceptions of skin cancer risk: a survey of visitors to Cancer Research UK's SunSmart campaign website. Br J Dermatol 2009 Jun;160(6):1292-1298. [doi: 10.1111/j.1365-2133.2009.09149.x] [Medline: 19416235]
6. Green AC, Marquart L, Clemens SL, Harper CM, O'Rourke PK. Frequency of sunburn in Queensland adults: still a burning issue. Med J Aust 2013 May 06;198(8):431-434. [Medline: 23641994]
7. Seité S, Del Marmol V, Moyal D, Friedman AJ. Public primary and secondary skin cancer prevention, perceptions and knowledge: an international cross-sectional survey. J Eur Acad Dermatol Venereol 2017 May;31(5):815-820. [doi: 10.1111/jdv.14104] [Medline: 28045207]
8. Dobbinson S, Wakefield M, Hill D, Girgis A, Aitken JF, Beckmann K, et al. Prevalence and determinants of Australian adolescents' and adults' weekend sun protection and sunburn, summer 2003-2004. J Am Acad Dermatol 2008 Oct;59(4):602-614. [doi: 10.1016/j.jaad.2008.06.011] [Medline: 18691790]
9. Volkov A, Dobbinson S, Wakefield M, Slevin T. Seven-year trends in sun protection and sunburn among Australian adolescents and adults. Aust N Z J Public Health 2013 Feb;37(1):63-69. [doi: 10.1111/1753-6405.12012] [Medline: 23379808]
10. Iannacone MR, Youlden DR, Baade PD, Aitken JF, Green AC. Melanoma incidence trends and survival in adolescents and young adults in Queensland, Australia. Int J Cancer 2015 Feb 01;136(3):603-609 [FREE Full text] [doi: 10.1002/ijc.28956] [Medline: 24806428]
11. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
12. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]

13. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
14. Finch L, Janda M, Loescher LJ, Hacker E. Can skin cancer prevention be improved through mobile technology interventions? A systematic review. Prev Med 2016 Sep;90:121-132. [doi: 10.1016/j.ypmed.2016.06.037] [Medline: 27374946]
15. Buller DB, Berwick M, Lantz K, Buller MK, Shane J, Kane I, et al. Smartphone mobile application delivering personalized, real-time sun protection advice: a randomized clinical trial. JAMA Dermatol 2015 May;151(5):497-504 [FREE Full text] [doi: 10.1001/jamadermatol.2014.3889] [Medline: 25629710]
16. Youl PH, Soyer HP, Baade PD, Marshall AL, Finch L, Janda M. Can skin cancer prevention and early detection be improved via mobile phone text messaging? A randomised, attention control trial. Prev Med 2015 Feb;71:50-56. [doi: 10.1016/j.ypmed.2014.12.009] [Medline: 25524612]
17. Djaja N, Janda M, Lucas RM, Harrison SL, van der Mei I, Ebeling PR, et al. Self-Reported Changes in Sun-Protection Behaviors at Different Latitudes in Australia. Photochem Photobiol 2016 Dec;92(3):495-502. [doi: 10.1111/php.12582] [Medline: 26914695]
18. Hacker E, Horsham C, Allen M, Nathan A, Lowe J, Janda M. Capturing Ultraviolet Radiation Exposure and Physical Activity: Feasibility Study and Comparison Between Self-Reports, Mobile Apps, Dosimeters, and Accelerometers. JMIR Res Protoc 2018 Apr 17;7(4):e102 [FREE Full text] [doi: 10.2196/resprot.9695] [Medline: 29666044]
19. Cancer Council Victoria. 2018. Free SunSmart app URL: http://www.sunsmart.com.au/tools/interactive-tools/free-sunsmart-app [accessed 2018-02-20] [WebCite Cache ID 6xMdjas6e]
20. Janda M, Youl P, Marshall AL, Soyer HP, Baade P. The HealthyTexts study: a randomized controlled trial to improve skin cancer prevention behaviors among young people. Contemp Clin Trials 2013 May;35(1):159-167. [doi: 10.1016/j.cct.2013.03.009] [Medline: 23557730]
21. Glanz K, Schoenfeld ER, Steffen A. A randomized trial of tailored skin cancer prevention messages for adults: Project SCAPE. Am J Public Health 2010 Apr;100(4):735-741 [FREE Full text] [doi: 10.2105/AJPH.2008.155705] [Medline: 20167900]
22. O'Riordan DL, Nehl E, Gies P, Bundy L, Burgess K, Davis E, et al. Validity of covering-up sun-protection habits: Association of observations and self-report. J Am Acad Dermatol 2009 May;60(5):739-744 [FREE Full text] [doi: 10.1016/j.jaad.2008.12.015] [Medline: 19278750]
23. Georgiev GZ. Sample Size Calculator. URL: https://www.gigacalculator.com/calculators/power-sample-size-calculator.php [WebCite Cache ID 732CVmjue]
24. Wong CC, Liu W, Gies P, Nixon R. Think UV, not heat!. Australas J Dermatol 2015 Nov;56(4):275-278. [doi: 10.1111/ajd.12272] [Medline: 25496328]
25. Carter OBJ, Donovan RJ. Public (Mis)understanding of the UV Index. J Health Commun 2007;12(1):41-52. [doi: 10.1080/10810730601093371] [Medline: 17365348]
26. Carli P, Crocetti E, Chiarugi A, Salvini C, Nardini P, Zipoli G, et al. The use of commercially available personal UV-meters does cause less safe tanning habits: a randomized-controlled trial. Photochem Photobiol 2008;84(3):758-763. [doi: 10.1111/j.1751-1097.2007.00256.x] [Medline: 18086243]
27. Bränström R, Ullén H, Brandberg Y. A randomised population-based intervention to examine the effects of the ultraviolet index on tanning behaviour. Eur J Cancer 2003 May;39(7):968-974. [Medline: 12706366]
28. Buller DB, Cokkinides V, Hall HI, Hartman AM, Saraiya M, Miller E, et al. Prevalence of sunburn, sun protection, and indoor tanning behaviors among Americans: review from national surveys and case studies of 3 states. J Am Acad Dermatol 2011 Nov;65(5 Suppl 1):S114-S123. [doi: 10.1016/j.jaad.2011.05.033] [Medline: 22018060]
29. Kasparian NA, McLoone JK, Meiser B. Skin cancer-related prevention and screening behaviors: a review of the literature. J Behav Med 2009 Oct;32(5):406-428. [doi: 10.1007/s10865-009-9219-2] [Medline: 19521760]
30. Finch L, Youl P, Marshall AL, Soyer HP, Baade P, Janda M. User preferences for text message-delivered skin cancer prevention and early detection. J Telemed Telecare 2015 Jun;21(4):227-234. [doi: 10.1177/1357633X15571652] [Medline: 25697490]
31. Heckman CJ, Darlow SD, Ritterband LM, Handorf EA, Manne SL. Efficacy of an Intervention to Alter Skin Cancer Risk Behaviors in Young Adults. Am J Prev Med 2016 Jul;51(1):1-11 [FREE Full text] [doi: 10.1016/j.amepre.2015.11.008] [Medline: 26810358]
32. Campbell D, Stanley J. Experimental and Quasi-Experimental Designs for Research. Boston, USA: Houghton Mifflin Company; 1963.
33. Køster B, Søndergaard J, Nielsen JB, Allen M, Bjerregaard M, Olsen A, et al. Effects of smartphone diaries and personal dosimeters on behavior in a randomized study of methods to document sunlight exposure. Prev Med Rep 2016 Jun;3:367-372 [FREE Full text] [doi: 10.1016/j.pmedr.2016.04.002] [Medline: 27419038]

Abbreviations

GEE: generalized estimating equations

OR: odds ratio
QUT: Queensland University of Technology
RCT: randomized controlled trial
SPH: Sun protection habits index
UVR: ultraviolet radiation

Edited by G Eysenbach; submitted 21.02.18; peer-reviewed by C Heckman, S Clemens; comments to author 10.07.18; revised version received 02.08.18; accepted 14.08.18; published 28.11.18.

Please cite as:
Hacker E, Horsham C, Vagenas D, Jones L, Lowe J, Janda M
A Mobile Technology Intervention With Ultraviolet Radiation Dosimeters and Smartphone Apps for Skin Cancer Prevention in Young Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2018;6(11):e199
URL: https://mhealth.jmir.org/2018/11/e199/
doi:10.2196/mhealth.9854
PMID:30487115

Original Paper

The Association Between the Use of Antenatal Care Smartphone Apps in Pregnant Women and Antenatal Depression: Cross-Sectional Study

Yushi Mo[1], BS; Wenjie Gong[1], PhD; Joyce Wang[1], BA; Xiaoqi Sheng[2], MS; Dong R Xu[3], PhD, MPP

[1]XiangYa School of Public Health, Central South University, Changsha, China

[2]Hunan Provincial Maternal and Child Health Hospital, Changsha, China

[3]Sun Yat-sen Global Health Institute, School of Public Health and Institute of State Governance, Sun Yat-sen University, Guangzhou, China

Corresponding Author:
Wenjie Gong, PhD
XiangYa School of Public Health
Central South University
238 Shangmayuanling Lane
Xiangya Road, KaiFu District
Changsha,
China
Phone: 86 13607445252
Email: gongwenjie@csu.edu.cn

Abstract

Background: Antenatal care smartphone apps are increasingly used by pregnant women, but studies on their use and impact are scarce.

Objective: This study investigates the use of antenatal care apps in pregnant women and explores the association between the use of these apps and antenatal depression.

Methods: This study used a convenient sample of pregnant women recruited from Hunan Provincial Maternal and Child Health Hospital in November 2015. The participants were surveyed for their demographic characteristics, use of antenatal care apps, and antenatal depression. Factors that influenced antenatal pregnancy were analyzed using logistic regression.

Results: Of the 1304 pregnant women, 71.31% (930/1304) used antenatal care apps. Higher usage of apps was associated with urban residency, nonmigrant status, first pregnancy, planned pregnancy, having no previous children, and opportunity to communicate with peer pregnant women. The cutoff score of the Edinburgh Postnatal Depression Scale was 10, and 46.11% (601/1304) of the pregnant women had depression. Logistic regression showed that depression was associated with the availability of disease-screening functions in the apps (odds ratio (OR) 1.78, 95% CI 1.03-3.06) and spending 30 minutes or more using the app (OR 2.05, 95% CI 1.19-3.52). Using apps with social media features was a protective factor for antenatal depression (OR 0.33, 95% CI 0.12-0.89).

Conclusions: The prevalence of the use of prenatal care apps in pregnant women is high. The functions and time spent on these apps are associated with the incidence of antenatal depression.

(JMIR Mhealth Uhealth 2018;6(11):e11508) doi:10.2196/11508

KEYWORDS

antenatal care; antenatal depression; app; mobile phone

Introduction

With the advent of the information age, the use of smartphones and apps is becoming more common. It is predicted that the world's smartphone ownership rate will reach 66% in 2018, and China will have 1.3 billion smartphone users [1]. Meanwhile, the number of global app downloads has exceeded to 175 billion in 2017, with China becoming the world's largest smartphone app market. In the 4th quarter of 2017 alone, the usage time of Chinese users of iOS, Google Play, and third-party Android apps reached 200 billion hours [2]. China's huge smartphone user base and high accessibility of Wi-Fi or 4G

have made people's frequent use of mobile apps a reality, and the use of these apps in health care is worthy of attention. As of December 2016, the number of people using the Internet for health purposes in China reached 195 million, accounting for 26.6% of all Internet users. Among them, the medical information inquiry use rate was the highest, accounting for 10.8% [3]. Consultation in obstetrics and gynecology was also the most used service in a leading Web-based medical consultation app in China [4]. Due to high demand and adherence, pregnant women have become a key target group for app developers.

Antenatal care apps (acAPPs) are a type of smartphone apps that provide prenatal care services and information targeting prepregnant and pregnant women. It is a new way of channeling information and interpersonal interaction in the current era of mobile technology development, which has been widely accepted [5,6]. A study in the United States showed that apps for women's health and pregnancy accounted for 7% of all kinds of apps [7]. We found 110 maternal health-related apps on the Chinese app market. Functions of these apps were diverse although the provision of maternity-related information to pregnant women was the major one. This form of health guidance is superior to traditional methods and can even improve the quality of pregnancy care in areas with scarce medical resources [5,8-10]. Many apps also have social functions, professional counseling functions, and special tools, such as calculation of fetal movement, calculation of due date, and measurement of changes in body weight. These self-monitored health status functions can be used to track pregnant women's health [5,8,11,12].

Although acAPPs are widely used, research on the relationship of such use with user health and especially mental health is scarce. At present, existing studies mainly focus on the use of apps for monitoring or intervening mental disorders. Sensors and apps of smartphones have been used to predict daily mood [13,14], detect depression [15-18], intervene depression [19], relieve pressure [20], and treat depression [18,19,21]. Mental health and app-related researches are even rarer, and their use and effect evaluation among pregnant women have not been reported. Pregnancy is a period of high incidence of mental disorders, which increases not only the risk of postpartum depression but also its severity [22-25].

There may be intergenerational negative effects from both physical and environmental aspects. Therefore, gestational depression has received much attention in recent years and was included in the category of perinatal depression by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, in 2013. The use period of acAPPs overlaps with the occurrence of gestational depression. Whether there is a certain relationship between the two and what that relationship may be are the factors not only worthy of researchers' attention in the field of maternal and child health but also important information for acAPP developers.

This study investigates the use of acAPPs in pregnant women at a specialized hospital for obstetrics and gynecology, examines antenatal depression, and explores the correlation between the use of acAPPs and mental health during pregnancy. These results can provide direction and theoretical basis for the development of future acAPPs.

Methods

Sample

Participants (N=1304) were pregnant women who visited the maternity department of the Maternal and Child Health Hospital of Hunan Province from November 16, 2015, to November 21, 2015.

Procedure

A convenience sampling approach was employed. During the investigation period, we explained the research purpose, content, and possible risks and benefits to all pregnant women who visited the Maternity Department of the Hunan Provincial Maternal and Child Health Hospital and invited them to participate in the survey. Questionnaires were issued to pregnant women aged 18 years and over who gave verbal consent. Women who were unable to understand the contents of the questionnaire were excluded from this study. We issued 1800 questionnaires, recollected 1520 questionnaires, and ultimately determined that 1304 questionnaires were valid, corresponding to an effective recovery rate of 72.44% (1304/1800). Flow of participants is presented in Figure 1. The study has been approved by the institutional review board of the Institute of Nursing and Behavioral Medicine Research, School of Nursing, Central South University (# 2015062).

Measures

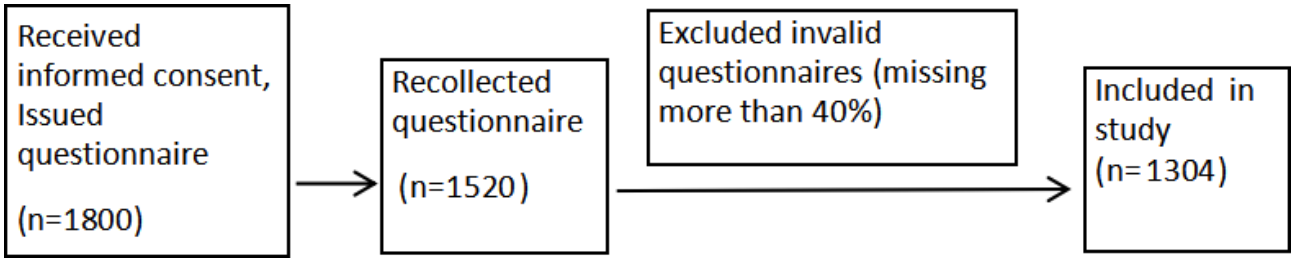
A self-developed questionnaire, optimized after preinvestigation, was used to collect information on demographics (eg, birth year, ethnicity, type of birthplace and residence, migrant population status, family income, and educational levels) and acAPP usage (eg, acAPP download channel, start usage time, duration of each use, frequency of use, and common functions).

The Edinburgh Postnatal Depression Scale (EPDS) is commonly used to screen for perinatal depression. It is a 10-item self-rated questionnaire, with each item scored from 0 to 3, giving a score ranging from 0 to 30. EPDS used to screen for antenatal depression in this study was translated by Wang Yuqiong [26], and several studies [27-30] have validated that it can be used for perinatal depression screening, including antenatal depression. The critical value was 9.5.

Statistical Methods

Different usages of acAPPs among pregnant women were analyzed using the chi-square test. The association between acAPP usage and antenatal depression was analyzed using binary logistic regression.

Figure 1. Flow of participants.



Results

Demographic Information of Pregnant Women

The demographic characteristics of the sample and univariate analysis are shown in Table 1. The average age of the pregnant women was 28.66 (SD 3.964) years. They mainly lived in a city, were nonmigrants, had a family monthly income of 5000-10,000 yuan (US $785-1570), and had completed undergraduate or college education. More than half of the pregnant women were pregnant for the first time and about two-thirds had no previous children.

Use of Antenatal Care Apps

In this study, 71.31% (930/1304) of the pregnant women used acAPPs. Higher usage of acAPPs was associated with urban residency, nonmigrant status, first pregnancy, planned pregnancy, having no previous children, and opportunity to communicate with peer pregnant women. There were differences in the utilization rate of acAPPs among pregnant women of different ages, family incomes, and education levels (P=.02, P=.001, P<.001, respectively). Usage of acAPPs among pregnant women aged 25-29 years was higher than that among women aged ≥35 years. App usage among those with family incomes in the 5000-10,000 renminbi (RMB)/month and 10,000-15,000 RMB/month brackets was higher than that among those with family incomes less than or equal to 5000 RMB/month. Usage among women with education levels of junior high school and below was lower than that among women with higher education levels (P<.001; see Table 2).

When choosing the kind of acAPP, pregnant women paid the most attention to the user rating and information content of acAPPs, and app stores and official websites were the main channels of acAPP downloads (Table 3). Apps commonly used by pregnant women are shown in Table 4. Baobaoshu, accounting for 45.8% of the total usage, was the most commonly used acAPP among pregnant women, followed by Meiyou and Huaiyunguanjia (14.9% and 7.4%, respectively). Other apps included Haoyunma, Yunqiguanjia, Yunqitixing, Qinbaobao, Lamabang, and more than 10 other kinds of acAPPs.

Among the 10 possible functions of the health care acAPP during pregnancy, antenatal care tips, health information, changes in pregnancy records, and dietary recommendations were the most commonly used functions by pregnant women (Table 5). Moreover, 67.1% (619/930) of the pregnant women started using the acAPP mainly 12 weeks after pregnancy. Almost half (455/930, 49.1%) used acAPPs 1-2 times per day, the average duration of each use was mainly 15 minutes, and 70.86% (659/930) of the pregnant women had been using acAPPs for 3-12 months (Table 6). Comprehensive, informative, user-friendly, and reliable content were features most users said appealed to them about acAPPs. Increasing knowledge was the greatest benefit pregnant women reported receiving from using acAPPs (Figure 2).

Antenatal Depression and Its Association With Using Antenatal Care Apps

Among pregnant women, 46.16% (602/1304) screened positive for depression by EPDS. Controlling for demographic characteristics and pregnancy situations, logistic regression showed that functions of and time spent on these apps were associated with the incidence of antenatal depression (Table 7). Moreover, acAPPs containing disease-screening functions and using apps for more than 30 minutes at a time were positively related to the occurrence of depression. AcAPPs containing social functions were negatively correlated with depression.

Table 1. Demographic information of pregnant women.

Item	Value, n (%)
Age, years	
<25	146 (11.1)
25-29	685 (52.1)
30-34	322 (24.5)
≥35	126 (9.6)
Missing	35 (2.7)
Residence	
City	858 (65.3)
Rural area	310 (23.6)
Missing	146 (11.1)
Migrant population	
Yes	313 (23.8)
No	933 (71.0)
Missing	68 (5.2)
Family income, renminbi/month (US $/month)	
≤5000 yuan (785)	252 (19.2)
5000-10,000 yuan (785-1570)	674 (51.3)
10,000-15,000 yuan (1570-2355)	192 (14.6)
>15,000 yuan (2355)	110 (8.4)
Missing	86 (6.5)
Education level	
Junior high school and below	80 (6.1)
High school or secondary school	236 (18.0)
Undergraduate or college	729 (60.3)
Master’s and above	101 (7.7)
Missing	105 (8.0)
First pregnancy	
Yes	713 (56.4)
No	552 (42.0)
Missing	49 (3.7)
Planned pregnancy	
Yes	841 (64.0)
No	431 (32.8)
Missing	42 (3.2)
Previous children	
Yes	427 (32.5)
No	865 (65.8)
Missing	22 (1.7)
Opportunity to communicate with peer pregnant women	
Have	1054 (80.2)
Do not have	245 (18.6)
Missing	15 (1)

Table 2. Different usages of antenatal care apps among pregnant women.

Item	Using antenatal care app, n (%)	Not using antenatal care app, n (%)	χ^2	*df* [a]	*P* value
Age, years			9.6	3	0.02
<25	100 (11.0)	46 (12.5)			
25-29	500 (54.8)	185 (50.4)			
30-34	236 (25.9)	86 (23.4)			
≥35	76 (8.3)	50 (13.6)			
Residence			11.7	1	0.001
City	636 (76.3)	222 (66.5)			
Rural area	198 (23.7)	112 (33.5)			
Migrant population			11.2	1	0.001
Yes	203 (22.6)	110 (31.7)			
No	696 (77.4)	237 (68.3)			
Family income, renminbi/month (US $/month)			17.3	3	0.001
≤5000 yuan (785)	155 (17.6)	97 (28.1)			
5000-10,000 yuan (785-1570)	506 (57.3)	168 (48.7)			
10,000-15,000 yuan(1570-2355)	142 (16.1)	50 (14.5)			
>15,000 yuan (2355)	80 (9.1)	30 (8.7)			
Education level			56.5	3	<0.001
Junior high school and below	30 (3.5)	50 (14.2)			
High school or secondary school	155 (18.1)	81 (23.0)			
Undergraduate or college	590 (68.8)	202 (57.4)			
Master's and above	82 (9.6)	19 (5.4)			
First pregnancy			8.8	1	0.003
Yes	531 (59.0)	182 (49.9)			
No	369 (41.0)	183 (50.1)			
Planned pregnancy			4.9	1	0.03
Yes	616 (68.0)	225 (61.5)			
No	290 (32.0)	141 (38.5)			
Previous children			12.2	1	<0.001
Yes	278 (30.2)	149 (40.3)			
No	664 (69.8)	221 (59.7)			
Opportunity to communicate with peer pregnant women			10.9	1	0.001
Have	774 (83.4)	280 (75.5)			
Did not have	154 (16.6)	91 (24.5)			

[a]*df*: degrees of freedom.

Table 3. Download channels of antenatal care apps.

Download variables	Value, n (%)
Emphasis when selecting antenatal care app	
Information content	490 (53.1)
User rating	385 (41.7)
Comprehensive ranking	178 (19.3)
Download times	137 (14.8)
Scoring	116 (12.6)
Other	19 (2.1)
Memory occupied	14 (1.5)
Download channel	
App stores	548 (59.7)
Official website search	333 (36.3)
Pop-up connection	19 (2.1)
Other	18 (2.0)

Table 4. Common antenatal care apps.

App name	Value, n (%)
Baobaoshu	426 (45.8)
Meiyou	139 (14.9)
Huaiyunguanjia	69 (7.4)
Mamabang	42 (4.5)
Yunqibanlv	21 (2.2)
Baobaozhidao	21 (2.2)
Yunqitixing	15 (1.6)
Others	45 (4.8)

Table 5. Common functions of antenatal care apps.

Functions	Common functions, n (%)
Health knowledge education	134 (33.3)
Changes during pregnancy record	104 (25.9)
Antenatal care tips	94 (23.5)
Diet recommendations	88 (21.9)
Answers and questions	37 (9.2)
Instant communication	21 (5.2)
Trading platform	17 (4.2)
Disease screening	11 (2.7)
Hospital related	9 (2.2)
Make friends	5 (1.2)

Table 6. Timing of using antenatal care apps.

Time	Value, n (%)
Start time	
Before pregnancy	189 (20.5)
Within 12 weeks of gestation	619 (67.1)
12-27 weeks of gestation	103 (11.2)
After 27 weeks of pregnancy	12 (1.3)
Frequency	
1-7 times/week	326 (35.2)
1-2 times/day	457 (49.4)
≥3 times/day	128 (13.8)
Other	15 (1.6)
Average duration of each use	
≤15 minute	455 (49.1)
15-30 minute	394 (42.5)
30-60 minute	56 (6.0)
≥1 hour	21 (2.3)
Longest duration of antenatal care app retainment	
<3 months	188 (20.4)
3-6 months	350 (38.0)
6-12 months	309 (33.5)
>1 year	75 (8)

Figure 2. Benefits of using antenatal care apps.

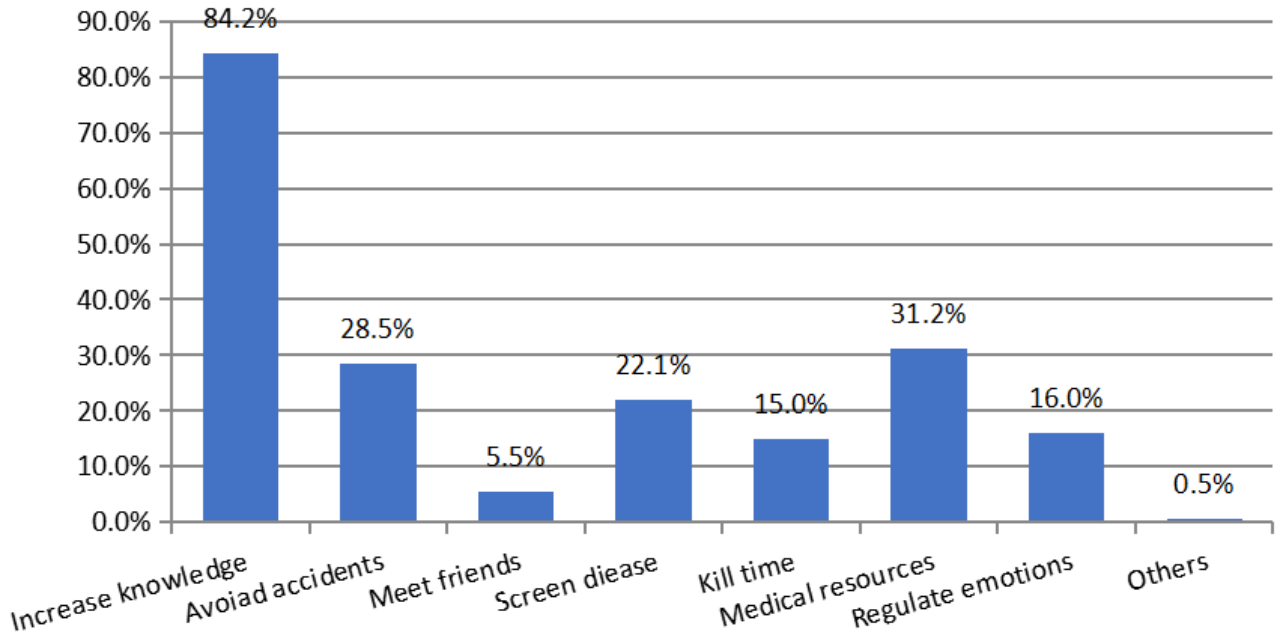


Table 7. Factors influencing antenatal depression.

Factors	B	Wals	*df*[a]	*P* value	Odds ratio	95% CI	
						Lower	Upper
Disease-screening function							
No	—	—	1	—	—	1.03	3.06
Yes	0.57	4.26	1	0.04	1.78		
Social function							
No	—	—	1	—	—	0.12	0.89
Yes	−1.10	4.76	1	0.03	0.33		
Average duration of each use							
<30 min	—	—	1	—	—	1.19	3.52
≥30 min	0.72	6.78	1	0.009	2.05		
Education level							
High school or secondary school	−0.96	4.28	1	0.04	0.04	0.15	0.95
Undergraduate or college	−1.04	5.61	1	0.02	0.35	0.15	0.84
Master's and above	−1.32	7.25	1	0.007	0.27	0.10	0.70
Constant	1.82	2.38	1	0.12	6.19	—	—

[a]*df*: degrees of freedom.

Discussion

This study was a cross-sectional survey carried out in Changsha, China, which revealed that the usage of acAPPs was very high (930/1304, 71.40% of the participants had used one) and frequent (457/930, 49.4% of the users used it 1-2 times per day) among pregnant women in urban areas. After controlling for several confounding factors, functions and duration of the use of acAPPs during pregnancy were found to be related to the incidence of antenatal depression. This reminds us that when considering pregnant women's health education, it is important to look beyond traditional pregnancy schools and maternal and child special education and recognize the usefulness of acAPPs as mobile medicine is expected to become a new means of health education and management during pregnancy [9,31,32]. At the same time, acAPP design can be optimized by adding functions, such as disease screening, peer communication, and time reminders, to increase population coverage for pregnant women and mental health care.

This study revealed 4 important aspects of the use of acAPPs. First, it confirmed that mobile apps are becoming major ways to obtain health information [33]. Research in Australia [12] involving 410 women has found that nearly three-quarters of women used pregnancy-related apps and that most of them used these apps at least once a week. A Chinese study [6] found that 78.81% of expectant mothers downloaded pregnancy-related apps. These results are similar to our findings in terms of acAPP usage rate. Compared with traditional methods of information acquisition, antenatal care apps make it possible for pregnant women to find pregnancy care knowledge at their own time, free from the restrictions of work, life, transportation, and family environment, via broad availability of smartphones and network coverage. This could help save health service costs [34,35].

Second, just as the literature [12] has shown, pregnant women with different characteristics use acAPPs differently. In this study, family income, place of residence, education, first pregnancy, planned pregnancy, having no previous children, and opportunity to communicate with pregnant women of the same age affected the usage of acAPPs. These findings suggest that we should tailor the design of acAPPs to meet the characteristics of pregnant women. Furthermore, as vulnerable groups (older, lower income and education, and unplanned pregnancies) in general tend to use acAPPs less, acAPPs would need to be better designed to reach these populations.

Third, the process of selecting an acAPP is different from routine health care decisions that are more influenced by health care professionals. Most pregnant women rely on user evaluations to select acAPPs rather than professional recommendations, which makes it particularly important for acAPPs to be professional and accurate in terms of information content. In contrast to other entertainment apps, user evaluations of acAPPs are often comments on user-friendliness, and they do not discuss professionalism, accuracy, or comprehensiveness. Studies have shown that if the pregnancy information shared by apps is inaccurate and unreliable, users of these apps face risks in terms of pregnancy protection [36]. The regulation of health acAPPs is an important issue raised in the existing research [37,38]. There are also studies [39,40] suggesting that the incorporation of a medical editorial team can increase the reliability of the knowledge provided in acAPPs in order to meet the requirements of information demanders.

Finally, acAPPs may be designed to monitor and intervene in perinatal depression [41]. This study found that scores for depression were higher in pregnant women who focused on using the acAPP's disease-screening functions. This may be because this group of pregnant women had underlying diseases

or anxieties about their own health, leading their depressive symptoms to be more pronounced. The pregnant women who preferred acApps with social functions had a desire for communication, and the use of acAPPs also increased the opportunities for communication with other pregnant women and channels of release. Pregnant women can receive mutual support and comfort [42], rendering them relatively less at risk of depressive symptoms. Studies have confirmed that peer communication is a protective factor for depressive symptoms [43]. Interestingly, the study also found that depressive symptoms were more pronounced in pregnant women who used acApps for more than half an hour. This may be because such pregnant women are isolated from the real world or have obstacles in their interpersonal relationships, leading to their immersion in the Internet [44]. This is consistent with the findings of Mansourian et al [45]. Augner et al [46,47] also found that overuse of smartphones is one of the predictors of adolescent depression. This suggests that when developers design acAPPs, proper tools, such as adding a time-use reminder function, may help users arrange rest time. AcAPPs with disease-screening functions that provide health consultations and psychological guidance could reduce the risk of antenatal depression. Adding related content and functions to acAPPs would also make them better tools for pregnancy management.

This study is a cross-sectional survey of all expecting mothers present at the Provincial Maternal and Child Care Hospital during a selected period of time; thus, data extrapolation was limited and no causal inference could be made. In addition, the self-made questionnaires used in this study were only perfected in the pre-experiment and had not been tested for reliability and validity. There is also heterogeneity in the types of antenatal care apps. In future studies, authors should consider quality evaluation and fine classification before conducting in-depth analyses. Our team used the mystery customer approach to evaluate the acAPPs' accuracy and reliability, the results of which will be described in another article.

Acknowledgments

We would like to thank the people who have contributed to the field implementation of this study, including Liao Nianquan from Hunan Provincial Maternal and Child Health Hospital, Tianlin Ma from Peking Union Medical College. This work was supported by the National Natural Science Foundation of China (Development of a dynamic predictive model of perinatal depression integrating social-environmental, psychological and biological factors: a prospective cohort study, grant number 81402690) and National Institutes of Health/Fogarty (eCapacity Development and Growth in the Aspire Network (eC.ASPIRE), grant #1R25TW010012).

Authors' Contributions

YM, as the first author, developed the initial manuscript. She observed the operation of the entire study. WG guided the overall design of the study, helped develop the instruments used in the study, and negotiated for program management of data and field access. DRX, JW, and XS contributed substantially to the revision and refinement of the final manuscript.

Conflicts of Interest

None declared.

References

1. Zenith. Smartphone penetration to reach 66% in 2018. URL: http://www.zenithmedia.com/smartphone-penetration-reach-66-2018/ [WebCite Cache ID 73Fwv7LBk]
2. Sydow L, Cheney S. 2017 Retrospective: A Monumental Year for the App Economy. 2018. URL: https://www.appannie.com/en/insights/market-data/app-annie-2017-retrospective/ [accessed 2018-10-18] [WebCite Cache ID 73FxEIWke]
3. The 39th China Statistical Report on Internet Development. China Policy Review 2017 Apr.
4. Yiguan: China Mobile Medical White Paper 2017 (with download). China Mobile Medical White Paper 2017 URL: http://www.199it.com/archives/574582.html [accessed 2017-03-21]
5. Pan G. The application research of pregnant m-health APP. Chongqing Medical University 2014.
6. Xiangzhou P. Research on influential factors of mom baby health management mobile APP service quality based on user experience theory. Beijing Forestry University 2015.
7. Acceptance TM, Barriers R. Patient Adoption of mHealth. China Digital Medicine 2015 [FREE Full text]
8. 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]
9. Kaewkungwal J, Singhasivanon P, Khamsiriwatchara A, Sawang S, Meankaew P, Wechsart A. Application of smart phone in 'Better Border Healthcare Program': a module for mother and child care. BMC Med Inform Decis Mak 2010 Nov 03;10:69 [FREE Full text] [doi: 10.1186/1472-6947-10-69] [Medline: 21047412]
10. Wenmin Y, Jie W, Hongyu Z, Jian YOU. Investigation on Health Demands of Health Website Visitors. Journal of Nursing Science 2009;24(9):86-87.
11. Lee Y, Moon M. Utilization and Content Evaluation of Mobile Applications for Pregnancy, Birth, and Child Care. Healthc Inform Res 2016 Apr;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]

12. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
13. 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 2018 Aug 21. [doi: 10.1002/da.22822] [Medline: 30129691]
14. Jiménez-Serrano S, Tortajada S, Garía-Gómez JM. A Mobile Health Application to Predict Postpartum Depression Based on Machine Learning. Telemed J E Health 2015 Jul;21(7):567-574. [doi: 10.1089/tmj.2014.0113] [Medline: 25734829]
15. Tsai AC, Tomlinson M, Dewing S, le Roux IM, Harwood JM, Chopra M, et al. Antenatal depression case finding by community health workers in South Africa: feasibility of a mobile phone application. Arch Womens Ment Health 2014 Oct;17(5):423-431 [FREE Full text] [doi: 10.1007/s00737-014-0426-7] [Medline: 24682529]
16. Marcano Belisario JS, Doherty K, O'Donoghue J, Ramchandani P, Majeed A, Doherty G, et al. A bespoke mobile application for the longitudinal assessment of depression and mood during pregnancy: protocol of a feasibility study. BMJ Open 2017 Dec 29;7(5):e014469 [FREE Full text] [doi: 10.1136/bmjopen-2016-014469] [Medline: 28554914]
17. Faherty LJ, Hantsoo L, Appleby D, Sammel MD, Bennett IM, Wiebe DJ. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. J Am Med Inform Assoc 2017 Jul 01;24(4):746-753. [doi: 10.1093/jamia/ocx005] [Medline: 28339686]
18. Wenhui B, Siyuan T, Jiarui C, Yao L, Chunxiang Q, Mei S. Application of Mobile Health Technology in the Research of Postpartum Depression. Nursing Journal of Chinese People's Liberation Army 2018;5:22-26 [FREE Full text]
19. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, 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]
20. Lu X, Chen X. Initial exploration of pregnancy decompression APP design based on emotion needs. Design 2015:142-143.
21. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
22. Sutter-Dallay AL, Cosnefroy O, Glatigny-Dallay E, Verdoux H, Rascle N. Evolution of perinatal depressive symptoms from pregnancy to two years postpartum in a low-risk sample: the MATQUID cohort. J Affect Disord 2012 Jun;139(1):23-29. [doi: 10.1016/j.jad.2011.08.018] [Medline: 22410506]
23. Juan H, Yu-qiong W. The relationship between antenatal depression and postnatal depression of women in Chengdu City. Chinese Journal of Nursing 2009;44(11):984-987.
24. Robertson E, Grace S, Wallington T, Stewart DE. Antenatal risk factors for postpartum depression: a synthesis of recent literature. Gen Hosp Psychiatry 2004;26(4):289-295. [doi: 10.1016/j.genhosppsych.2004.02.006] [Medline: 15234824]
25. Rahman A, Creed F. Outcome of prenatal depression and risk factors associated with persistence in the first postnatal year: prospective study from Rawalpindi, Pakistan. J Affect Disord 2007 Jun;100(1-3):115-121 [FREE Full text] [doi: 10.1016/j.jad.2006.10.004] [Medline: 17098291]
26. Wang Y, Guo X, Lau Y, Chan KS, Yin L, Chen J. Psychometric evaluation of the Mainland Chinese version of the Edinburgh Postnatal Depression Scale. Int J Nurs Stud 2009 Jun;46(6):813-823. [doi: 10.1016/j.ijnurstu.2009.01.010] [Medline: 19217107]
27. Weiqin L, Chuanlan C, Shimine C. Study on the correlation between depression and sleep quality in pregnant women. Maternal & Child Health Care of China 2015.
28. Deave T, Heron J, Evans J, Emond A. The impact of maternal depression in pregnancy on early child development. BJOG 2008 Jul;115(8):1043-1051 [FREE Full text] [doi: 10.1111/j.1471-0528.2008.01752.x] [Medline: 18651886]
29. Teixeira C, Figueiredo B, Conde A, Pacheco A, Costa R. Anxiety and depression during pregnancy in women and men. J Affect Disord 2009 Dec;119(1-3):142-148. [doi: 10.1016/j.jad.2009.03.005] [Medline: 19346001]
30. Eberhard-Gran M, Tambs K, Opjordsmoen S, Skrondal A, Eskild A. Depression during pregnancy and after delivery: a repeated measurement study. J Psychosom Obstet Gynaecol 2004 Mar;25(1):15-21. [Medline: 15376401]
31. Little A, Medhanyie A, Yebyo H, Spigt M, Dinant GJ, Blanco R. Meeting community health worker needs for maternal health care service delivery using appropriate mobile technologies in Ethiopia. PLoS One 2013;8(10):e77563 [FREE Full text] [doi: 10.1371/journal.pone.0077563] [Medline: 24204872]
32. Kahn JG, Yang JS, Kahn JS. 'Mobile' health needs and opportunities in developing countries. Health Aff (Millwood) 2010 Feb;29(2):252-258. [doi: 10.1377/hlthaff.2009.0965] [Medline: 20348069]
33. Daly LM, Horey D, Middleton PF, Boyle FM, Flenady V. The effect of mobile application interventions on influencing healthy maternal behaviour and improving perinatal health outcomes: a systematic review protocol. Syst Rev 2017 Dec 08;6(1):26 [FREE Full text] [doi: 10.1186/s13643-017-0424-8] [Medline: 28179012]
34. Hearn L, Miller M, Lester L. Reaching perinatal women online: the Healthy You, Healthy Baby website and app. J Obes 2014;2014:573928 [FREE Full text] [doi: 10.1155/2014/573928] [Medline: 24872891]
35. Moore J. The benefits of mobile apps for patients and providers. British Journal of Healthcare Management 2012 Sep;18(9):465-467. [doi: 10.12968/bjhc.2012.18.9.465]
36. Deng Y, Liu W, Tong-Xuan LI, Law DO. Discussion on the Current Running Status of Mobile Medical APP and Relevant Legal Regulatory Issues. China Medical Devices 2015.

37. Peng W, Hao W. Mobile Internet Medical Application Status and Future Development Trend of Domestic and Abroad. China Digital Medicine 2014:8-10.
38. Xue L, Jia YJ, Pang SL. A survey on knowledge of prenatal care of 300 pregnant women. Maternal & Child Health Care of China 2007.
39. Yi LI. Functions of domestic interrogation mobile medical APP. Chinese Journal of Medical Library & Information Science 2015.
40. Derbyshire E, Dancey D. Smartphone Medical Applications for Women's Health: What Is the Evidence-Base and Feedback? Int J Telemed Appl 2013;2013:782074 [FREE Full text] [doi: 10.1155/2013/782074] [Medline: 24454354]
41. Osma J, Barrera AZ, Ramphos E. Are Pregnant and Postpartum Women Interested in Health-Related Apps? Implications for the Prevention of Perinatal Depression. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):412-415. [doi: 10.1089/cyber.2015.0549] [Medline: 27327069]
42. Johnson SA. "Maternal Devices", Social Media and the Self-Management of Pregnancy, Mothering and Child Health. Societies 2014 Jun 13;4(2):330-350. [doi: 10.3390/soc4020330]
43. Chopra R, Punia S, Sangwan S. Depression and Peer-Relationship among Adolescents in Haryana. Studies on Home and Community Science 2016;10(1-3):21-25. [doi: 10.1080/09737189.2016.11885362]
44. Chen BF, Chen YJ, Zhu Y. Association of mobile phone addiction and functional impairment and depression among college students. Chinese Journal of School Doctor 2015.
45. Mansourian M, Solhi M, Adab Z, Latifi M. Relationship between dependence to mobile phone with loneliness and social support in University students. Razi Journal of Medical Sciences 2014.
46. Augner C, Hacker GW. Associations between problematic mobile phone use and psychological parameters in young adults. Int J Public Health 2012 Apr;57(2):437-441. [doi: 10.1007/s00038-011-0234-z] [Medline: 21290162]
47. Yen C, Tang T, Yen J, Lin H, Huang C, Liu S, et al. Symptoms of problematic cellular phone use, functional impairment and its association with depression among adolescents in Southern Taiwan. J Adolesc 2009 Aug;32(4):863-873. [doi: 10.1016/j.adolescence.2008.10.006] [Medline: 19027941]

Abbreviations

acAPPs: antenatal care apps
EPDS: Edinburgh Postnatal Depression Scale

Edited by G Eysenbach; submitted 06.07.18; peer-reviewed by P Akhtar, SFA Shah, I Adam; comments to author 15.08.18; accepted 28.08.18; published 29.11.18.

Please cite as:
Mo Y, Gong W, Wang J, Sheng X, Xu DR
The Association Between the Use of Antenatal Care Smartphone Apps in Pregnant Women and Antenatal Depression: Cross-Sectional Study
JMIR Mhealth Uhealth 2018;6(11):e11508
URL: http://mhealth.jmir.org/2018/11/e11508/
doi:10.2196/11508
PMID:30497996

Original Paper

Preferences of Underserved Chilean Women on a Mobile Technology Intervention for Cervical Cancer Screening: Qualitative Study

Mauricio Soto[1,2*], MD, MSc; Javiera Martinez-Gutierrez[1*], MD, MPH; McKenzie Momany[3*], BA; Daniel Capurro[4], MD, PhD; Francis Ciampi Spode[5], RN; Emilia Cea[6], RN; Tania Mergudich[7], MD; Klaus Puschel[1], MD, MPH, MBE

[1]Department of Family Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[2]École de Santé Publique, Université de Montréal, Montreal, QC, Canada

[3]School of Medicine, University of Washington, Seatlle, WA, United States

[4]Department of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[5]Centro de Salud Familiar El Roble, Santiago, Chile

[6]School of Nursing, Pontificia Universidad Católica de Chile, Santiago, Chile

[7]School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[*]these authors contributed equally

Corresponding Author:
Javiera Martinez-Gutierrez, MD, MPH
Department of Family Medicine
School of Medicine
Pontificia Universidad Católica de Chile
Vicuña Mackenna 4686
Macul
Santiago,
Chile
Phone: 56 223548688
Email: jmartigu@uc.cl

Abstract

Background: In Chile and Latin America, cervical cancer disproportionately affects women of low socioeconomic status. Mobile technology (mobile health, mHealth) may be able to address this disparity by targeting women in underserved populations. However, there is a lack of information regarding barriers to the implementation of mHealth interventions in underserved populations.

Objective: The objective of this study was to investigate the use of cell phones and text messaging (short message service, SMS) in Latina women from disadvantaged communities to design an mHealth intervention for improving cervical cancer screening rates.

Methods: We conducted 9 focus groups among women aged 25-64 years to better understand the implementation barriers and perceptions of a text message (SMS)–based intervention designed to improve cervical cancer screening rates. We used the PRECEDE-PROCEED model to categorize identified themes using template analysis.

Results: Focus group results indicated that older women use mobile phones to receive calls from family and friends but seldom send text messages. Furthermore, they prefer personal contact with their health care providers regarding Papanicolaou (Pap) testing. Younger women, on the other hand, find text messaging easy to use and frequently send texts to family and friends. Importantly, women of all ages mentioned they would like to receive text messages about Pap tests. Factors that facilitate the uptake of the intervention include ease of access to Pap testing, inclusion of family members, and reminder messaging. Potential barriers include cost and the impersonal nature of messaging. Health team members support an mHealth intervention even though they acknowledge the potential barriers to this strategy. Overall, these results support the implementation of an mHealth intervention to increase cervical cancer screening rates.

Conclusions: This study describes the opinions of women nonadherent to Pap testing on the potential use of mobile technologies for cervical cancer screening. Although the overall acceptance was positive, older women prefer personal contact and phone calls

over text messaging. Information surrounding these preferences will aid in the implementation of effective strategies to improve cancer screening in underserved populations.

(JMIR Mhealth Uhealth 2018;6(11):e196) doi:10.2196/mhealth.9494

KEYWORDS

mHealth; cancer screening; Latina women

Introduction

Cervical-uterine cancer is the fourth oncological cause of death in Chile with a mortality rate of 6/100,000 women in 2012. This rate is well above that in developed countries (Canada: 1.71/100,000 women; United States: 2.7/100,000 women) [1]. Studies in Chile have shown marked inequalities in cancer mortality rates between women with low and high educational levels [2]. In Chile, mortality from cervical cancer with respect to education level is 15 times higher in women with <4 years of education than that in women with >12 years of education [3]. Since 1965, the peak cervical cancer screening adherence rate in Chile has been 67%, far from the international recommendation of 80%. Furthermore, Papanicolaou (Pap) testing rates have markedly dropped in recent years, hovering at <60% [4]. There are currently no studies in Chile on inequality in adherence to cervical cancer screening. In Latin America, factors associated with screening adherence include access to health insurance and high socioeconomic status, while factors associated with nonadherence (in women with preinvasive cervical lesions) include concern about the procedure, time constraints, and lack of knowledge. Therefore, it is recommended to develop programs that address these barriers and change communication techniques in a manner that considers broader cultural frameworks [5]. Moreover, recent studies have shown that it is vital to involve community members throughout the program development process in order to maximize effectiveness [6-9].

Interventions based on short message service (SMS) text messaging have shown efficacy in encouraging behavior modification and improving health. Recently, evidence has showed that these interventions have been able to increase the cancer screening adherence [10-12]. A study on Korean American women using a week-long SMS text message intervention demonstrated increased acceptance to cervical cancer screening [13]. However, there is insufficient evidence to characterize maximally effective interventions, and none of these studies have been conducted in Latin American countries. An SMS text message intervention could fail for multiple reasons. Previously described barriers include SMS text message character restrictions, user illiteracy, poor coverage and signal quality, shared cell phones, confidentiality, and message appropriateness [14]. In the United States, the largest users of SMS text messaging are Hispanic people (83%), followed by African-American (76%) and white (70%) people [15], suggesting that cultural factors may influence the use of SMS text messaging. In Chile, there has been a marked growth in the number of mobile phone users in recent years. In December 2015, the number of subscribers reached 23,206,353 people, with a ratio of 128.22 mobile subscribers per 100 inhabitants [16]. According to the same report, an average of 109 million messages were generated per month in 2015, with 99.2% corresponding to SMS text messages and 0.8% to multimedia messaging service messages. These data support the use of SMS text messaging as a strategy to improve cervical cancer screening adherence in Chile. However, no study has yet described mobile phone use by Chilean women. Studying the use of mobile phones would allow us to understand the role they play in women's lives and enable the design of effective interventions to improve cancer screening adherence.

This study aims to investigate the use of cell phones and SMS text messaging in Latina women from disadvantaged communities in order to design a mobile health (mHealth) intervention to improve cervical cancer screening rates.

Methods

Inclusion Criteria

Women aged between 25 and 64 years who were nonadherent to cervical cancer screening guidelines (no Pap test within last 3 years) were included from three health care centers in underserved areas of Santiago. These health care centers are located in the outskirts of the city in the La Pintana and Puente Alto neighborhoods; 30% and 14% of the population, respectively, in these areas lives under the poverty line compared with 9% living under the poverty line in the entire metropolitan area [17].

Focus Groups

We conducted 9 focus groups in total—3 in each health center; one group of women aged 25-44 years, a second group of those aged 45-64 years, and a third group of health care personnel. Women were divided by age due to potential differences in cell phone use and generational barriers that could affect the use of mHealth interventions. Overall, 27 women participated in the interviews.

In the health professionals focus groups, 10 midwives and 1 paramedic participated. Focus groups were facilitated by 2 researchers trained in qualitative research methods. The interviews were recorded and transcribed verbatim for further analysis.

Topics were investigated through a semistructured guide and included the pattern of cell phones and other mobile use and barriers and facilitators to an SMS text message intervention to encourage Pap smear adherence. Each participant signed an informed consent document to participate in this study. This study was approved by the Ethics Committee of the Universidad Católica de Chile (ID: CEC Med UC 213).

Data Analysis

The focus group audiorecordings were transcribed and encoded using qualitative data analysis software (Atlas.ti version 6.2). An "analysis by templates" method was used for coding, based on a list of codes derived from the data analysis. This allowed for the creation of broader categories derived from emergent groups of code [18]. Coding was carried out by 2 researchers, using a consensus method to create new categories and resolve differences in the assessments. They identified first-order themes that corresponded to general cell phone and SMS text message use (main uses, barriers, and facilitators to use). In addition, we asked about preferences regarding cervical cancer prevention strategies based on SMS text messaging (usage, content, format, potential barriers, and facilitators). The identified themes were subsequently categorized according to the PRECEED-PROCEED model as factors that could serve as predisposers, facilitators, or reinforcers of SMS text messaging as a means to improve Pap smear adherence [19,20].

Results

We conducted 6 focus groups with 27 women aged between 25 and 64 years and 3 focus groups with 11 health care workers. Tables 1 and 2 present participants' demographics.

Table 1. Demographics of female focus group participants (N=27).

Characteristic	Value
Age (years), mean (SD)	41 (11.7)
People per household, n (%)	4 (15)
Children per household, n (%)	2 (7)
Marital status, n (%)	
Married	17 (63)
Single	8 (30)
Separated or divorced	2 (7)
Monthly income, n (%)	
<US $360	13 (48)
US $360-800	9 (33)
US $800-1500	5 (19)
Education level, n (%)	
Completed secondary education	18 (67)
Incomplete secondary education	6 (22)
Incomplete primary education	3 (11)
Occupation, n (%)	
Housewife	17 (63)
Student	3 (11)
Employed	7 (26)

Table 2. Demographics of health professionals (N=11).

Characteristic	Value
Age (years), mean (SD)	34 (9.1)
Gender, n (%)	
Male	4 (36)
Female	7 (64)
Occupation, n (%)	
Midwife	10 (91)
Technician	1 (9)
Education level, n (%)	
Postsecondary, nonuniversity	1 (9)
University	10 (91)

Textbox 1. Identification of the components of the PRECEED-PROCEED model regarding the use of mobile phones for cancer prevention.

Predisposing factors

- Use of technologies for personal and family contact.
- Economic barriers may limit access to mobile technology.
- Mobile phones are most frequently used for calls and messages.
- A significant proportion of older women report low use of phones for making calls and little use of other mobile functions.

Enabling factors

- Participants would like to be contacted by the health center, in general, and midwives, in particular.
- Most women have their own mobile phone.
- Theft, loss, or damage to the mobile phone would not be a barrier to maintaining the same cell phone number.

Reinforcing factors

- Use of multidimensional messages.
- Use of short and clear messages.

PRECEDE-PROCEED Model

SMS text messaging knowledge and attitudes, as well as practical knowledge of cell phones and messaging, were considered to be predisposing factors. Facilitating factors were considered to be those that will enable the use of SMS text messaging as a cervical cancer prevention strategy. In addition, the content of messages as well as their frequency and timing were considered to be reinforcing factors for the ongoing use of such SMS text messages as a preventive strategy. Textbox 1 presents an analysis summary.

Predisposing Factors

Cell Phone and Short Message Service Text Message Use

Cell phone usage varied by age. Women in the youngest age group (25-44 years) used their cell phones frequently (multiple times per day), while those in the older age group (45-64 years) reported less usage (the majority used their phones once every 2-3 days).

Communication With Family Members

The primary use of cell phones and messaging reported by women of all ages was communication with their families. Within this category, communication with their children was reported as the most frequent use, followed by communication with other relatives such as husband, mother, brothers, or sisters.

> *I use [my phone] only for communication with my children, and just with text messages. When I am at home with my two children, I don't use my phone. When my husband gets home, all the more reason [to not use it]. [I use my phone] only [to communicate] with my children.* [Participant from Group 1, 25-44-year olds]

Communication for Work

Communicating for work or school was another important but less frequent use that was brought up only among the younger women. Both groups reported other but less frequent uses, including communication with people outside their family and looking up information.

Barriers to Use

Financial considerations were the primary barrier to the implementation of an mHealth intervention. Most participants had prepaid cell phone plans and depended on money to have a functioning mobile phone. One group of women aged 25-44 years comprised women of higher socioeconomic status. These women had plans that included multimedia, allowing them to use additional mobile functions. In the 25 to 44-year-old age group, some women reported having Wi-Fi access at home or in public places. This allowed them to utilize all mobile phone functions more frequently.

> *It just so happens that in the middle of the month I can't make calls or send messages if I run out of money.* [Participant from Group 6, 45-64-year olds]

In the older group, some women indicated that mobile phones were a nuisance and took too much effort to use. In addition, they said that they did not have a thorough understanding of the various functions of cell phones, which was associated with a lower usage.

> *Look, my children always call me on my cell phone. I sometimes have 8 missed calls. You're not going to believe me, but I don't answer them. But if you call me on the landline, I answer right away. [I don't want to give you] false hope that I'm going to read [the messages, because] I won't.* [Participant from Group 2, 45-64-year olds]

Type of Use

Cell phones were most frequently used for calls, specifically receiving calls. Younger participants used phones for both making and receiving calls, while older participants mainly used them to receive calls. There was a group of women who, despite having a mobile phone and receiving calls, did not answer the calls because they did not notice them or were otherwise

occupied. The group of 45 to 64-year-old women expressed being more comfortable with landlines.

> *Yes, if I have a cell phone, I use it to communicate with my children, but it's always they who call me. I don't like using it. I don't like bringing it everywhere.* [Participant from Group 2, 45-64-year olds]

Younger participants also sent and received SMS text messages more frequently; older participants used SMS text messaging sparingly. In addition, younger participants appeared to use WhatsApp and other social media, particularly Facebook, for communication. The younger group included women who used WhatsApp rather than SMS text messaging for work and family communication. This same group also used Facebook to keep in contact with distant relatives.

> *I [use my cell phone] only for communication with my family via WhatsApp or to keep in contact with distant relatives via Facebook. There is very little you can do by making a call on a cell phone.* [Participant from Group 1, 25-44-year olds]

Those who used SMS text messaging perceived it as a useful form of communication because messages could be personalized and read at one's convenience.

> *You can read your messages whenever you want—you're in control.* [Participant from Group 1, 25-44-year olds]

Enabling Factors

Contact Preferences

Women from all age groups were favorable toward the idea of being contacted by their health center as part of an mHealth intervention.

> *I like the idea of messages. I get informational messages at the place I study at, and they're very good.* [Participant from Group 1, 25-44-year olds]
>
> *Yes, that message ["Go and get your PAP smear" or "did you know that by doing a PAP"...] would be good.* [Participant from Group 4, 45-64-year olds]

Women across all groups had different preferences and recommendations for the design of an mHealth intervention. Most of the women indicated that they would prefer a call from their health care centers as means of communication. They thought phone calls allowed for more detailed and personalized conversations and had received calls from their health center in the past, for reasons like rescheduling appointments.

> *I prefer phone calls, because I understand better what they're trying to tell me.* [Participant from Group 5, 25-44-year olds]
>
> *You can ask questions* [Participant from Group 5, 25-44-year olds]

Older women rated phone calls over other modes of communication like SMS text messaging and email.

> *For people like us who aren't so tech saavy, I think it would be better to use phone calls. For example, her text message got deleted, and others here don't read them.* [Participant from Group 4, 45-64-year olds]
>
> *I think it would be better if they called because it's more personal and you would be able to answer and say if you can't go [to the appointment], need to change the date, etc. With a message, you wouldn't be able to answer.* [Participant from Group 3, 25-44-year olds]

On the other hand, younger participants indicated a preference for the use of messages either via SMS text or WhatsApp, and email for more extensive messages.

> *I think we have more interaction with the midwives now that there is internet and we can contact with them through WhatsApp.* [Participant from Group 5, 25-44-year olds]

Participants in general did not feel strongly regarding which member of the health care team contacted them as long as he or she was an expert on the subject. However, some participants preferred to be contacted by a midwife, as midwives are well known and perceived to be knowledgeable.

> *As long as [the call/message] comes from the health center, it's fine.* [Participant from Group 1, 25-44-year olds]
>
> *[…the message] has to come from the midwife who knows the most. I would think she knows best.* [Participant from Group 3, 25-44-year olds]

This lack of consensus regarding the preferred contact method (whether via phone call or SMS text message, for example) may reflect different ways in which the interviewees use their mobile phones in daily life. For example, older women who prefer to be contacted via telephone typically communicate with their relatives via telephone calls.

Cell Phone Ownership

All women interviewed were the primary users of their cell phone. Some women in both age groups reported sharing their cell phone with children or grandchildren to allow them to play games. Having "dumbphones" (a basic cell phone without smartphone capabilities) allowed women to prevent their children from using them. One woman shared her cell phone with her husband.

> *Everyone has their own device, so no one needs to share.* [Participant from Group 1, 25-44-year olds]
>
> *Most of the time my children use my phone since it has games on it. They spend the most time with my cell phone.* [Participant from Group 3, 25-44-year olds]

Portability

Most women in the focus groups reported that they had not changed their mobile phone number in years. Some women replaced their cell phones after they were stolen or lost but still managed to keep the same number. This is possible due to a number portability law that allows them to keep their number even if they switch carriers. As such, people tend to believe that they can keep the same number even if their mobile is stolen,

lost, or damaged. This practice would facilitate communication between health centers and patients.

> *I have been robbed twice but kept my same number. They always steal my phone.* [Participant from Group 1, 25-44-year olds]

It is important to add that recruiting women via phone was one of the study's biggest difficulties, so the women interviewed here may represent a subgroup that changes its telephone numbers less frequently.

Enhancing Factors

Message Content

Two types of messaging content were identified as useful for women: informational messages and reminder messages.

Reminder Messages

Participants were very supportive of the use of messages to remind them if they are due for a Pap test.

> *I want a reminder because I forget.* [Participant from Group 3, 25-44-year olds]

> *It would be very good [to be sent reminder messages]…you would remember that you need to get a PAP and that you have to go tomorrow or the day after tomorrow [to the health care center].* [Participant from Group 4, 45-64-year olds]

Informative Messages

There was a consensus among the groups that it is useful to receive general information on cervical cancer prevention, Pap screening, and common myths.

> *More than anything, it should have information about PAP tests and emphasize that it doesn't hurt* [Participant from Group 5, 25-44-year olds]

> *[You should] not only call and tell them to get their PAP, but also give information on the consequences if they don't.* [Participant from Group 4, 45-64-year olds]

Messaging Format

Most women agree that the SMS text messages should be concise and use simple language.

> *And in simple words, please, in simple words, not technical language* [Participant from Group 1, 25-44-year olds]

There was less agreement regarding the tone of the messages. One group of women suggested that the messages should use the word *cancer* to have a more serious impact. They mentioned the notion of death should be included to emphasize the consequences of not having regular Pap tests for early detection of cervical cancer. On the other hand, a different group of women thought generating fear could be counterproductive.

> *PAP smears prevent cancer, and the word "cancer" scares people. It's good to say "cancer" because we need [to generate] awareness.* [Participant from Group 6, 45-64-year olds]

Opinions From Health Professionals Regarding an mHealth Strategy for Cervical Cancer Screening

Pap screening rates are part of Chile's performance evaluation system in Primary Health Care. Higher rates of screening are associated with more financial resources at the public primary health level. As a consequence, this screening test is highly prioritized in Primary Health Care. Therefore, it is not surprising that interviewees mentioned multiple strategies used by the three centers to improve Pap test adherence. The strategies included having exclusive schedules for Pap testing, daily walk-in appointments, Pap testing in the community, campaigns through posters, and radio announcements, among others. Nonadherent patients are identified when they come to the health center for other services, while in the waiting rooms, and even via telephone calls. In two centers, adherence information from the electronic medical record is used to schedule more efficiently.

Nevertheless, even with these resources allocated for Pap testing, screening rates have not reached the goals set by the Ministry of Health. Health professionals attributed nonadherence to several reasons that can be classified into three major domains: (1) lack of knowledge about the test; (2) negative predisposition toward the test; (3) and misconceptions about the test. Table 3 presents the reasons attributed by health professionals.

When asked about the potential applicability of mobile phones and SMS text messages in cancer screening, professionals of all three centers thought female patients would be receptive to it as the strategy demonstrates concern for their health. However, health professionals were reluctant to believe mobile technologies could help improve Pap test adherence rates unless several barriers were addressed.

In all centers, health professionals believed that women would prefer SMS text messaging as they text frequently and that messages do not need to be read at the same time they are sent. Professionals from one center thought that SMS text messages would be very useful among younger women.

Frequent cell phone number change was the most important barrier identified by professionals in these three centers. From developing scheduling strategies in the past for nonadherent women, the health professionals had determined that they lack a database with updated phone numbers of female patients. Health professionals from two centers mentioned that older women may not know how to use their phones or may have older models that do not properly display messages. In one center, the lack of availability of walk-in appointments for Pap tests was identified as an organizational barrier to addressing Pap nonadherence.

Table 3. Reasons for nonadherence to Pap testing according to health professionals (N=11).

Domain and reasons attributed	Health professionals mentioning the domain, n (%)
Lack of knowledge about Pap tests	
"They haven't had any health problems."	6 (55)
"They don't know when they need to do the test."	4 (36)
"They don't think the test needs to be performed if they aren't having current sexual relations."	4 (36)
"They feel healthy, so they think it is not necessary to carry out the test."	4 (36)
Negative predisposition toward the test	
"They're very busy/don't have time."	10 (91)
"They're always postponing the test."	5 (45)
Attitude toward the test	
"They think the test is too painful or unpleasant."	9 (82)
"They're too embarrassed by the test."	6 (55)

Regarding suggestions for an effective strategy using mobile phones and SMS text messages, health professionals recommended the following:

- Use of informative messages with helpful and clarifying information about Pap tests, reminders to get a Pap, available times to get a Pap, and messages that encourage getting a Pap (including information on health risks associated with not getting a Pap)
- Test availability during nonstandard work hours, and the ability to respond to messages to confirm receipt or schedule an appointment

Discussion

Principal Findings

This paper discusses the preferences of underserved Latina women on the use of mobile technology for cervical cancer prevention. Although study participants were generally favorable toward the idea of using these technologies, the following points are worth taking into consideration.

Participants primarily use mobile devices to stay in contact with family members. This is similar to what other comparable studies have found. In one study, Hernández et al [21] reported that family care represents one of the most important daily tasks for Mexican women. This finding can be extrapolated to Chilean women and is consistent with the fact that participants primarily use their mobile phones for family contact.

Challenges identified by a World Health Organization study [15] included security and privacy issues, especially in low- and middle-income countries where mobile phones were often shared with family and other community members. In our study, confidentiality did not arise as a barrier to mHealth interventions. Unlike other Latin American women, our participants did not typically share their phones. Regarding mobile usage, phone calls are used more than any other form of communication; this is consistent with how study participants prefer personalized calls from health care providers and also with the literature showing that Latina women use mobile devices for personal communication more than using them for receiving information [22]. In addition, the literature shows that Latina women value the opinions of their family members and health care providers, and these opinions heavily influence their attitudes on cancer screening [23]. This is consistent with the notion of personal contact being important, which was a predominant theme in this study.

Participants in this study preferred short SMS text messages as opposed to longer messages. This aligns with other studies that have identified such SMS text message characteristics as fundamental to the intervention's success [24,25]. We identified differences in mobile phone use according to participants' age, as well as variation in the preferred tone of messages. This is consistent with recommendations from the literature to avoid a one-dimensional approach and instead use a variety of messaging types to implement a persuasive health communication strategy [24].

Considering the perspective of both patients and health care teams may aid in the design of a more effective and comprehensive mHealth strategy. This study had many parallels between patients' and health care personnel's opinions regarding the potential usefulness, barriers, and recommendations on a mobile strategy to improve Pap screening rates. This may be due to the familiarity that health professionals have developed with women in their health centers.

Regarding health care professionals identifying the frequent change of cell phone numbers as a barrier, although the women in the focus groups did not refer to it as an obstacle, it is important to consider that this was a barrier during the recruitment phase of the study. This should be considered when designing strategies based on mobile phone use.

Conclusion

Low cervical cancer screening rates remain a public health problem worldwide. Mobile technologies have the potential to improve cancer screening rates in low-income countries, but there is a lack of information on the usefulness of mHealth for cervical cancer prevention in Latin America. Latina women use mobile devices regularly to communicate with their families

and value personal contact or phone calls from health personnel; therefore, any successful intervention using mHealth needs to take these factors into consideration. This study will provide valuable information in the design and implementation of a mobile intervention to increase cervical cancer screening rates.

Conflicts of Interest

None declared.

References

1. Ferlay JS, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. Lyon, France: International Agency for Research on Cancer; 2013. Cancer Incidence and Mortality Worldwide URL: http://gco.iarc.fr/ [accessed 2018-10-04] [WebCite Cache ID 72vSiXCCw]
2. Herrera RCA, Kuhn-Barrientos L, Rosso AR, Jiménez DLJJ. Tendencia de la mortalidad por cáncer en Chile según diferencias por nivel educacional, 2000-2010. Revista Panamericana de Salud Pública 2015;37:44-51 [FREE Full text]
3. Léniz Martelli J, Van De Wyngard V, Lagos M, Barriga MI, Puschel Illanes K, Ferreccio Readi C. Early detection of cervical cancer in Chile: time for change. Rev Med Chil 2014 Aug;142(8):1047-1055 [FREE Full text] [doi: 10.4067/S0034-98872014000800012] [Medline: 25424677]
4. Fica A. Cancer of cervix in Chile. Too much vaccine amid a neglected Papanicolau. Rev Chilena Infectol 2014 Apr;31(2):196-203 [FREE Full text] [doi: 10.4067/S0716-10182014000200010] [Medline: 24878908]
5. Agurto I, Bishop A, Sánchez G, Betancourt Z, Robles S. Perceived barriers and benefits to cervical cancer screening in Latin America. Prev Med 2004 Jul;39(1):91-98. [doi: 10.1016/j.ypmed.2004.03.040] [Medline: 15207990]
6. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004 Apr;16(2):107-123. [doi: 10.1093/intqhc/mzh030] [Medline: 15051705]
7. Lehoux P, Sicotte C, Denis JL, Berg M, Lacroix A. The theory of use behind telemedicine: how compatible with physicians' clinical routines? Soc Sci Med 2002 Mar;54(6):889-904. [Medline: 11996023]
8. Grol R. Improving Patient Care: The Implementation of Change in Clinical Practice. In: Improving Patient Care. USA: John Wiley & Sons Inc; 2018.
9. Lehoux P, Miller F, Hivon M, Demers-Payette O, Urbach D. Clinicians as health technology designers: Two contrasting tales about user involvement in innovation development. Health Policy and Technology 2013 Sep;2(3):122-130. [doi: 10.1016/j.hlpt.2013.05.003]
10. Hall A, Cole-Lewis H, Bernhardt J. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
11. Uy C, Lopez J, Trinh-Shevrin C, Kwon SC, Sherman SE, Liang PS. Text Messaging Interventions on Cancer Screening Rates: A Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
12. Mink J, Peterson C. MobileODT: a case study of a novel approach to an mHealth-based model of sustainable impact. Mhealth 2016 Apr;2:12 [FREE Full text] [doi: 10.21037/mhealth.2016.03.10] [Medline: 28293590]
13. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014 Aug 27;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
14. World Health Organization. 2011. 2nd Global Survey on eHealth URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2018-10-04] [WebCite Cache ID 72vh9ro85]
15. Zickuhr KS. Pew Research Center. 2012. Digital differences URL: http://www.pewinternet.org/2012/04/13/digital-differences/ [accessed 2018-10-04] [WebCite Cache ID 72vhEhvcA]
16. Subsecretaría de Telecomunicaciones. 2016. Estadísticas de abonados de telefonía móvil URL: https://www.subtel.gob.cl/estudios-y-estadisticas/telefonia/ [accessed 2018-10-04] [WebCite Cache ID 72vheWyOm]
17. Ministerio de Desarrollo Social. 2014. Serie Informes Comunales, Nº1. Reporte Comunal: La Pintana, Región Metropolitana URL: http://observatorio.ministeriodesarrollosocial.gob.cl/indicadores/reportes_com1_2.php [accessed 2018-10-04] [WebCite Cache ID 72viO8qao]
18. Marshall C, Rossman G. Designing Qualitative Research. Los Angeles: SAGE Publications; 2010.
19. Green L, Kreuter M. Health Program Planning: An Educational and Ecological Approach 4th Ed. New York: McGraw-Hill Higher Education; 2005.
20. Püschel K, Thompson B, Coronado G, Gonzalez K, Rain C, Rivera S. 'If I feel something wrong, then I will get a mammogram': understanding barriers and facilitators for mammography screening among Chilean women. Fam Pract 2010 Feb;27(1):85-92 [FREE Full text] [doi: 10.1093/fampra/cmp080] [Medline: 19897514]
21. Hernández TI, Arenas MM, Valde SR. El cuidado a la salud en el ámbito doméstico: interacción social y vida cotidiana. Revista de Saúde Pública 2001;35:443-450.

22. Pasick R, Burke N. A critical review of theory in breast cancer screening promotion across cultures. Annu Rev Public Health 2008;29:351-368. [doi: 10.1146/annurev.publhealth.29.020907.143420] [Medline: 17914932]
23. Molina Y, Martínez-Gutiérrez J, Püschel K, Thompson B. Plans to obtain a mammogram among Chilean women: the roles of recommendations and self-efficacy. Health Educ Res 2013 Oct;28(5):784-792 [FREE Full text] [doi: 10.1093/her/cyt047] [Medline: 23515116]
24. Rimer B, Kreuter M. Advancing Tailored Health Communication: A Persuasion and Message Effects Perspective. Journal of Communication 2006;56:S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
25. Curioso W, Quistberg D, Cabello R, Gozzer E, Garcia P, Holmes K, et al. "It's time for your life": How should we remind patients to take medicines using short text messages? Annu Symp Proc. 2009 Presented at: AMIA; Nov 14; San Francisco p. 129-133.

Abbreviations

mHealth: mobile health
PAP: Papanicolaou
SMS: short message service

Edited by G Eysenbach; submitted 26.11.17; peer-reviewed by S Lawpoolsri, Y Mejova; comments to author 11.01.18; revised version received 22.05.18; accepted 21.06.18; published 20.11.18.

Please cite as:
Soto M, Martinez-Gutierrez J, Momany M, Capurro D, Ciampi Spode F, Cea E, Mergudich T, Puschel K
Preferences of Underserved Chilean Women on a Mobile Technology Intervention for Cervical Cancer Screening: Qualitative Study
JMIR Mhealth Uhealth 2018;6(11):e196
URL: http://mhealth.jmir.org/2018/11/e196/
doi:10.2196/mhealth.9494
PMID:30459141

Original Paper

Insights From a Text Messaging–Based Sexual and Reproductive Health Information Program in Tanzania (m4RH): Retrospective Analysis

Patrick S Olsen[1], MPH; Kate F Plourde[2], MPH; Christine Lasway[3], MPH, MBA; Eric van Praag[4], MD, MPH

[1]Health Services Research, Global Health, Population, and Nutrition, FHI 360, Durham, NC, United States

[2]Research Utilization, Global Health, Population, and Nutrition, FHI 360, Durham, NC, United States

[3]Palladium: Make It Possible, Health Practice, Americas, Washington, DC, United States

[4]Public Health Consultant, Dar es Salaam, United Republic Of Tanzania

Corresponding Author:
Patrick S Olsen, MPH
Health Services Research, Global Health, Population, and Nutrition
FHI 360
359 Blackwell Street, Suite 200
Durham, NC, 27701
United States
Phone: 1 919 544 7040 ext 11713
Email: polsen@fhi360.org

Abstract

Background: Many mobile health (mHealth) interventions have the potential to generate and store vast amounts of system-generated participant interaction data that could provide insight into user engagement, programmatic strengths, and areas that need improvement to maximize efficacy. However, despite the popularity of mHealth interventions, there is little documentation on how to use these data to monitor and improve programming or to evaluate impact.

Objective: This study aimed to better understand how users of the Mobile for Reproductive Health (m4RH) mHealth intervention engaged with the program in Tanzania from September 2013 to August 2016.

Methods: We conducted secondary data analysis of longitudinal data captured by system logs of participant interactions with the m4RH program from 127 districts in Tanzania from September 2013 to August 2016. Data cleaning and analysis was conducted using Stata 13. The data were examined for completeness and "correctness." No missing data was imputed; respondents with missing or incorrect values were dropped from the analyses.

Results: The total population for analysis included 3,673,702 queries among 409,768 unique visitors. New users represented roughly 11.15% (409,768/3,673,702) of all queries. Among all system queries for new users, 46.10% (188,904/409,768) users accessed the m4RH main menu. Among these users, 89.58% (169,218/188,904) accessed specific m4RH content on family planning, contraceptive methods, adolescent-specific and youth-specific information, and clinic locations after first accessing the m4RH main menu. The majority of these users (216,422/409,768, 52.82%) requested information on contraceptive methods; fewer users (23,236/409,768, 5.67%) requested information on clinic location. The conversion rate was highest during the first and second years of the program when nearly all users (11,246/11,470, 98.05%, and 33,551/34,830, 96.33%, respectively) who accessed m4RH continued on to query more specific content from the system. The rate of users that accessed m4RH and became active users declined slightly from 98.05% (11,246/11,470) in 2013 to 87.54% (56,696/64,765) in 2016. Overall, slightly more than one-third of all new users accessing m4RH sent queries at least once per month for 2 or more months, and 67.86% (278,088/409,768) of new and returning users requested information multiple times per month. Promotional periods were present for 15 of 36 months during the study period.

Conclusions: The analysis of the rich data captured provides a useful framework with which to measure the degree and nature of user engagement utilizing routine system-generated data. It also contributes to knowledge of how users engage with text messaging (short message service)-based health promotion interventions and demonstrates how data generated on user interactions could inform improvements to the design and delivery of a service, thereby enhancing its effectiveness.

(JMIR Mhealth Uhealth 2018;6(11):e10190) doi:10.2196/10190

KEYWORDS

data analysis; mobile phone; mHealth; short message service; user engagement

Introduction

Mobile phones are a well-established platform for health education and behavior change, and as global levels of phone and internet access continue to rise, so does the application of mobile health (mHealth) approaches across multiple public health sectors [1-3]. Many mHealth interventions have the potential to generate and store vast amounts of system-generated data that could provide insight into user engagement, programmatic strengths, and areas that need improvement to maximize efficacy. However, despite the popularity of mHealth interventions, there is little documentation on how to use these data to monitor and improve programming or to evaluate impact.

The utilization of mHealth program data for process evaluation creates an opportunity to complement conventional impact evaluation approaches. The analysis of mobile phone program use data allows researchers to examine information from all participants compared with traditional process evaluation approaches (such as in-depth interviews or focus group discussions and questionnaires) that collect information from a subset of program participants. It is more cost effective than other methods, and because analytics are collected in real-time, the approach improves efficiency and eliminates the potential for recall or interview bias [4,5]. Research demonstrates that the analysis of these data does provide valuable information for program evaluation. For example, the examination of user experience data, including the number of menus accessed, from an mHealth intervention delivered to young and middle-aged male alcohol users in Scotland allowed researchers to measure the fidelity of the intervention delivery, user engagement, and message comprehensibility—all critical indicators of programmatic success [4].

As summarized by Perski et al, several conceptual models have been presented in the literature to depict the behavior change theory underlying digital behavior change interventions [6]. Across these, the level and quality of user engagement with digital health interventions are considered key factors in the intervention's effectiveness [6-8]. The literature presents a broad definition of user engagement that often encompasses the number of intervention interactions, the relevance and relatability of the content, and repeat use [9,10]. User engagement is influenced by multiple aspects of intervention design such as resonance of messages, message delivery pattern, and program interactivity [11].

To better understand how users of the Mobile for Reproductive Health (m4RH) mHealth intervention engaged with the program, we conducted a retrospective analysis of m4RH system use data from 1 country (Tanzania) from September 2013 to August 2016. m4RH is a short message service (SMS) text messaging-based sexual and reproductive health (SRH) information program. It was initially developed in 2009 to provide information about the full range of contraceptive methods available locally and was first implemented in Kenya and Tanzania. The program has since been adapted in multiple countries, including Uganda and Rwanda, and its content has been expanded to include additional SRH topics such as HIV, sexually transmitted infections (STIs), and puberty.

User engagement can be measured using both subjective (qualitative) and objective (quantitative) measures [6,10]. Previously published research describes the results of qualitative data collection efforts to understand the acceptability and quality of interactions with the m4RH system [12,13]. This publication focuses on the retrospective analysis of subjective measures of engagement: system use metrics. The concept of user engagement, for the purposes of this analysis, builds upon the work of previous authors and is characterized by the following attributes as defined in Table 1 [6,14].

Table 1. Definitions pertaining to user engagement attributes.

Dimensions	Indicator	Definition
Environment [15]: associated with factors thought to influence engagement such as access, social norms, and time use patterns	Location	Sum of unique requests for all wards per district
Interaction [6]: how often users interact with the system and over what period; a key dimension of engagement	New users	The number of unique users that accessed the system from September 2013 to August 2016
	Return users	The number of unique users that accessed the system at least once per month for 2 or more months
	Repeat users	The number of unique users that accessed the system more than once per month (does not include users who accessed the same menu items twice during the same month)
	Acquisition	Users (return and new) that accessed the system each month
Depth [6,8]: level of content consumed; a key dimension of engagement	Activation	Percent users who requested a submenu after receiving the main menu (indicates navigation through m4RH main menus)
	Active use	Percent of activated users who request content keywords after receiving submenu prompt
Loyalty [6,8,16]: Degree of involvement over time, retention—"stickiness"; a key dimension of engagement	Conversion	Percent of users who become active users (ie, request content)
	Churn	Percent users lost (uses that do not become active users)

m4RH is an on-demand system that requires users to navigate and request information before accessing content. Users can access content in 2 ways: by accessing the navigation menu, which takes an average of 3 steps to access content or directly entering a content code after the welcome menu (ie, 2 steps). In Tanzania, participants access the m4RH program by SMS text messaging "m4RH" to short code "15014" to receive a menu of choices for accessing information on a variety of SRH topics including contraceptive methods, family planning (FP) clinic locations, role model stories (story installments modeling positive health attitudes, norms, and behaviors), and so forth. To access any of the menu options, respondents reply with another SMS text message using a numerical code corresponding to each menu item. Respondents only receive SMS text messages as they request them; no follow-up or reminder SMS text messages are sent after the initial query. To encourage respondent participation, respondents incur no cost for either sending or receiving SMS text messages when interacting with the m4RH system.

Methods

Data Analysis

Secondary data analysis was conducted using longitudinal data collected by mobile SMS text messaging from respondents of the m4RH program from 127 districts in Tanzania from September 2013 to August 2016. Participant interactions were extracted from system logs that captured each query to the m4RH system, including respondent mobile phone number (used to identify unique system interactions), information requested (eg, contraceptive methods, HIV and STI prevention, sex and pregnancy, and puberty), and date and time of the query. The analysis utilized descriptive analysis (proportions and means) of participant interactions and adapted the conversion funnel framework and indicators (mainly used in e-commerce) to map out and assess the journey of a user in a Web-based system. Conversion funnels are grounded in consumer decision-making theory and lifecycle, which mirrors behavior change models such as the transtheoretical model [17].

This study was reviewed and approved by the federally registered institutional review board of FHI 360, the Protection of Human Subjects Committee, and the National Institute for Medical Research Tanzania Institutional Review Board. Respondents did not receive reimbursement for participating in any study activity.

Data Cleaning

From September 2013 to August 2016, there were 4,112,460 system queries. Among those queries, 35,978 were invalid menu selections or text, 402,780 were duplicated menu selections sent in succession, and 5663 were m4RH specified incorrectly (ie, mr4h, m4hr, mrh4, or mfrh). After removing invalid and duplicate queries, the total population for analysis included 3,673,702 queries among 409,768 unique visitors.

Data cleaning and analysis was conducted using Stata 13 (Stata Corp, College Station, TX, USA). For analysis, all data were combined into 1 dataset consisting of multiple data points per user. This repeated measures dataset was sorted by users' mobile number (unique ID) and date of system query. In this analysis, the proportions and frequencies of respondent queries are aggregated yearly unless specified, for ease of interpretability in user engagement over time. The data were examined for completeness (queries with missing time or date) and "correctness" (invalid menu option, duplicated menu selections sent in succession, and incorrectly specifying the program name, ie, mr4h, m4hr, mrh4, or mfrh). In addition, data were examined to identify illegitimate or implausible entries, for example, users sending duplicate queries in succession. Implausible entries identified during this process were excluded from analysis. No missing data were imputed. Respondents with missing or incorrect values were dropped from the analyses. To date, m4RH has reached 409,768 unique users in Tanzania living in 127 of the country's 129 districts.

Results

Environment

The environment is hypothesized to influence engagement as social norms about seeking health information and mobile phone use, access to digital technology, and time use patterns are often similar in like settings [6]. Setting or location is one dimension of context. Through this analysis, we were able to examine the location using clinic location as a proxy. Only 5.67% (23,236/409,768) of new users requested information on the location of nearby clinics, and among these users, 33.30% (7,737/23,236) requested information on clinics within a given district. The region with the most requests was Dar es Salaam, representing 50.91% (3939/7737) of all requests, followed by Dodoma, Mwanza, and Arusha, representing 9.65% (747/7737), 8.52% (659/7737), and 7.60% (588/7737), respectively (see Table 2). These are all urban districts.

Interaction

During the analysis period, 409,768 new users initiated m4RH in Tanzania. New users represented roughly 11.15% (409,768/3,673,702) of all queries. Table 3 shows total interactions for new and return users. Across all time points, on average 381 new users accessed the system per day and interacted with the system 5 times within 24 hours and 7 times per month. Among new, return, and repeat queries, the average duration between the first and last daily query was 64 minutes, and the duration between each query was 17 minutes.

Most daily queries occurred between 12 pm and 2 pm and between 6 pm and 8 pm (796,678/3,673,702, 21.69%, and 740,682/3,673,702, 20.16%, respectively). The fewest queries took place between 12 am and 5 am (see Figure 1). Access time shows similar trends across years (Figure 2).

Table 2. Frequency of wards requested per district in Tanzania from September 2013 to August 2016.

District	Value (N=7737), n (%)
Dar es Salaam	3939 (50.91)
Dodoma	747 (9.65)
Mwanza	659 (8.52)
Arusha	588 (7.60)
Mbeya	302 (3.90)
Kigoma	261 (3.37)
Tanga	193 (2.49)
Njombe	189 (2.44)
Iringa	174 (2.25)
Geita	170 (2.20)
Morogoro	158 (2.04)
Pwani	139 (1.80)
Simiyu	84 (1.09)
Ruvuma	68 (0.88)
Singida	52 (0.67)
Mjini Magharibi	13 (0.17)
Mara	1 (0.01)

Table 3. Total interactions by new and return users.

Users	Year, n (%)				Total (N=3,673,702), n (%)
	2013 (N=280,304)	2014 (N=923,661)	2015 (N=1,475,428)	2016 (N=994,309)	
New users	43,746 (15.61)	116,654 (12.63)	151,903 (10.30)	97,465 (9.80)	409,768 (11.15)
Return users	4731 (1.69)	36,887 (4.00)	70,703 (4.80)	38,594 (3.88)	150,915 (4.11)

Figure 1. Overall user access time in 3-hour increments.

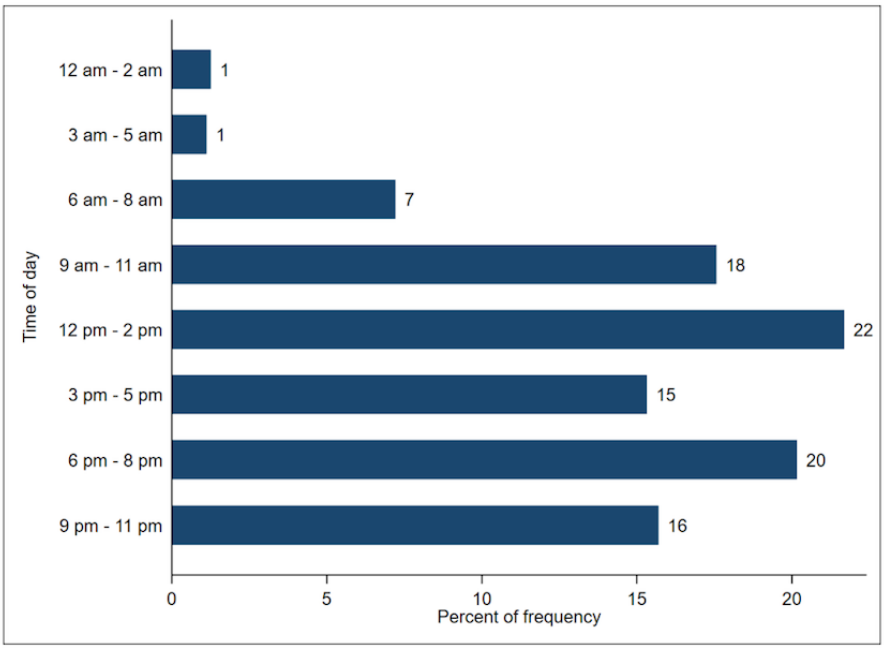
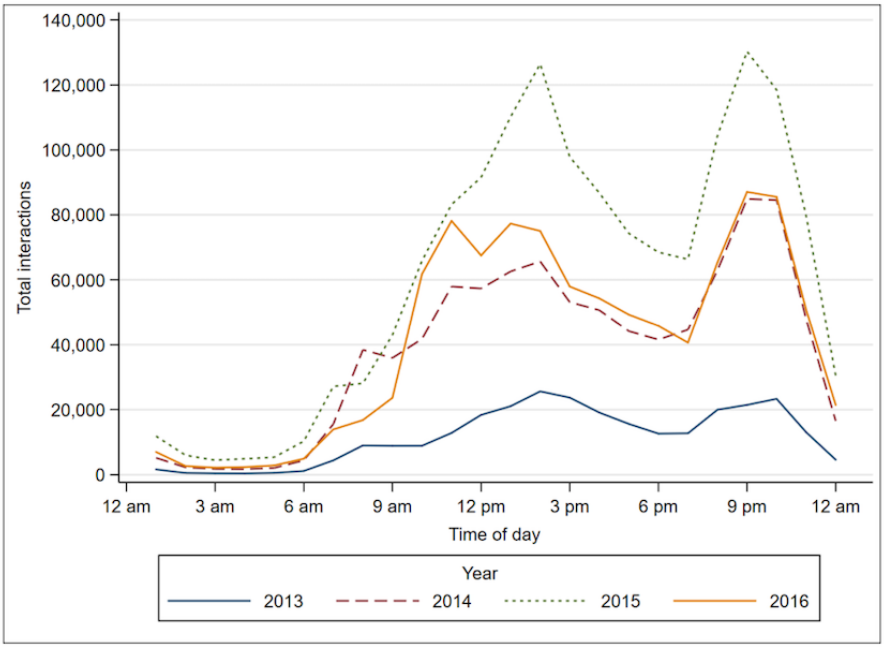


Figure 2. User access time, hourly by year.



Depth

After first accessing the m4RH main menu, 89.58% (169,218/188,904) of users requested FP, contraceptive methods, youth-specific information, role model stories, or clinic locations submenus. The majority of all users (216,422/409,768, 52.82%) requested the contraceptive methods submenu. Among those that requested the submenu, 24.15% (98,969/409,768) requested additional information on natural FP methods and only 9.02% (36,968/409,768) requested information on intrauterine contraceptive devices (IUCDs).Close to the same number of users accessed FP and role model stories.

The types of information requested on contraceptive methods over the 36-month period followed similar trends relative to the number of new users requesting contraceptive method information each year. Few users requested information on clinic location (see Table 4).

As shown in Table 5, across all years, users were most interested in FP methods. Within the pregnancy prevention menu, roughly one-quarter (98,969/409,768, 24.15%) of all queries were requests for information on natural FP, as shown in Table 6 below.

Queries for information on other pregnancy prevention methods were more equally distributed, ranging from roughly 10% to 17%. The pregnancy prevention method with the fewest queries was IUCD, representing only 9.02% (36,968/409,768) of all pregnancy prevention queries (Table 6). Among respondents seeking information on youth content (Table 7), the most commonly requested topic was About Sex (34,542/249,368, 13.85%), while requests for Puberty and Choices represented similar proportions, at 6.43% (16,043/249,368) and 5.31% (13,238/249,368), respectively.

Menu content for “General Family Planning” and “Youth Content” was not implemented in Tanzania until 2015 and can help explain the lower frequencies among the 2 menu options for 2014 and 2015.

On average, users queried the system 5.56 times per month on the m4RH content described above. The rate of users that accessed m4RH and become active users declined slightly from to 98.05% (11,246/11470) in 2013 to 87.54% (56,969/97,465) in 2016.

Table 4. Frequency of clinic menu requested from unique interactions.

Frequency	Unique visitors (N=409,768), n (%)
0	386,532 (94.33)
1	21,292 (5.20)
2	1635 (0.40)
3	246 (0.06)
≥4	63 (0.02)

Table 5. Activation and active use among unique users.

Activation and active use	Year				Total (N=409,768)
	2013 (N=43,746)	2014 (N=116,654)	2015 (N=151,903)	2016 (N=97,465)	
Activation, n (%)	11,470 (26.22)	34,830 (29.86)	77,839 (51.24)	64,765 (66.45)	188,904 (46.10)
Seek menu content, n (%)	11,246 (25.71)	33,551 (29.76)	67,725 (44.58)	56,696 (58.17)	169,218 (41.30)
Active use rate, n (%)	11,246 (98.05)	33,551 (96.33)	67,725 (87.01)	56,696 (87.54)	169,218 (89.58)
Keywords per month, mean (SD)	7.52 (6.09)	7.14 (5.96)	4.92 (6.23)	5.13 (6.49)	5.56 (6.34)
Menu engagement, n (%)					
About family planning	N/A[a]	N/A	36,067 (23.74)	32,380 (33.22)	68,447 (16.70)
Family planning methods	30,492 (69.70)	79,438 (68.10)	70,379 (46.33)	36,113 (37.05)	216,422 (52.82)
Youth	N/A	N/A	24,317 (16.01)	21,920 (22.49)	46,237 (11.28)
Role model stories	10,185 (23.28)	27,580 (23.64)	17,438 (11.48)	12,674 (13.00)	67,877 (16.56)
Clinic	402 (0.92)	8,835 (7.57)	7,474 (4.92)	6,525 (6.69)	23,236 (5.67)

[a]N/A=not applicable (“About family planning” and “Youth” menus not implemented until 2015).

Table 6. Information requested on different pregnancy prevention methods among unique users.

Pregnancy prevention methods, n (%)	Year				Total (N=409,768)
	2013 (N=43,746)	2014 (N=116,654)	2015 (N=151,903)	2016 (N=97,465)	
Natural family planning	13,451 (30.75)	33,827 (29.00)	33,126 (21.81)	18,565 (19.05)	98,969 (24.15)
Condom	9,332 (21.33)	24,807 (21.27)	21,263 (14.00)	10,586 (10.86)	65,988 (16.10)
Lactational amenorrhea method	8,263 (18.89)	22,009 (18.87)	23,258 (15.31)	13,163 (13.51)	66,693 (16.28)
Emergency contraception	8,655 (19.78)	21,778 (18.67)	18,264 (12.02)	9,167 (9.41)	57,864 (14.12)
Permanent	8,270 (18.90)	21,909 (18.78)	16,952 (11.16)	8,009 (8.22)	55,140 (13.46)
Implant	7,462 (17.06)	20,858 (17.88)	17,336 (11.41)	8,680 (8.91)	54,336 (13.26)
Injectable	6,786 (15.51)	18,903 (16.20)	17,050 (11.22)	8,593 (8.82)	51,332 (12.53)
Pills	5,865 (13.41)	16,208 (13.89)	13,183 (8.68)	6,344 (6.51)	41,600 (10.15)
Intrauterine contraceptive device	5,027 (11.49)	13,633 (11.69)	12,001 (7.90)	6,307 (6.47)	36,968 (9.02)

Table 7. Percent distribution of information requested on youth content among unique users.

Percent distribution, n (%)	Year		Total (N=249,368)
	2015 (N=151,903)	2016 (N=97,465)	
Puberty	8,366 (5.51)	7,677 (7.88)	16,043 (6.43)
About sex	18,144 (11.94)	16,398 (16.82)	34,542 (13.85)
Choices	6,756 (4.45)	6,482 (6.65)	13,238 (5.31)

Loyalty

The active use rate was highest during the first 2 years of the program, with nearly all users (11,246/11,470, 98.05%, and 33,551/34,830, 96.33%, respectively) who accessed m4RH continuing on to query more specific content from the system. An explanation for the decrease in conversion rates may be a result of the timing and frequency of promotions, for example, radio and magazine advertisements (Table 8).

Table 9 shows respondent retention and loyalty. Overall, slightly more than one-third (150,915/409,768) of all new users accessing m4RH sent queries at least once per month for 2 or more months and 67.86% (278,088/409,768) of new and return users requested information multiple times per month.

Table 8. Promotion periods per year and interactions among new and return users.

Promotion periods and interactions	Year				Total
	2013	2014	2015	2016	
Promotional months per year, n	4	10	1	0	15
Percent interactions, n (%)					
New users (N=409,768)	43,746 (100)	99,012 (84.88)	32,746 (21.56)	N/A[a]	175,504 (42.83)
Return users (N=150,915)	4,731 (100)	31,084 (84.27)	6,877 (9.73)	N/A	42,692 (28.29)

[a]N/A=not applicable (No promotional periods during 2016).

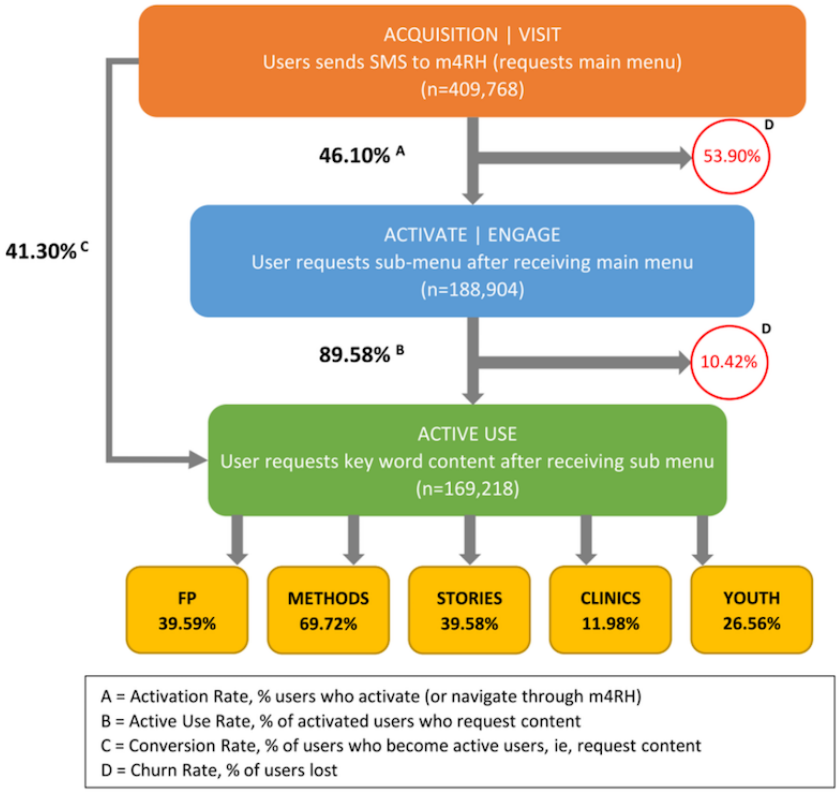
Table 9. Percentage of return and repeat users per year.

Rates, n (%)	Year				Total (N=409,768)
	2013 (N=43,746)	2014 (N=116,654)	2015 (N=151,903)	2016 (N=97,465)	
Return users	4,731 (10.81)	36,887 (31.62)	70,703 (46.54)	38,594 (39.60)	150,915 (36.83)
Repeat users	30,685 (70.14)	80,542 (69.04)	100,749 (67.64)	64,112 (65.78)	278,088 (67.86)
Conversion rate	11,470 (25.71)	33,551 (28.76)	67,725 (44.58)	56,696 (58.17)	169,218 (41.30)

On average, users accessed the system for 3 months over the 4-year period. This is also demonstrated by the average number of keywords queried per month presented in sections “Activate” and “Active Use” (Figure 3). Of the 409,768 new users acquired from September 2013 to October 2016, 41.30% (169,218/409,768) became active users seeking further content (Figure 3). At the end of the study period, there were a total of 169,218 active users (users who requested specific keyword content). Among these users, the majority (117,805/169,218, 69.62%) of them requested content on pregnancy prevention methods, and only 11.98% (20,277/169,218) sought information on FP clinics.

Promotional periods were present for 15/36 months during the study period. However, as shown in Table 8, nearly all promotional activities took place in 2013 and 2014, and no promotional activities were reported for 2016. Looking specifically at 2014, 84.88% (99,012/116,654) of new users and 84.27% (31,084/36,887) of return users were acquired during months when promotional activities were being conducted. The unequal distribution of promotional activities makes it difficult to establish the influence of promotional activities on user engagement.

Figure 3. Mobile for Reproductive Health respondent progression. m4RH: Mobile for Reproductive Health; SMS: short message service.



Discussion

Principal Findings

Our study shows that approximately 67.86% (278,088/409,768) of users accessed the system more than once within a month period, averaging 5.56 times a month. Since m4RH is an on-demand system requiring people to access the service voluntarily, our results may suggest that the users have a keen interest in information to enhance their contraceptive knowledge. Our results also suggest a high degree of demand and motivation among users who access the service. A conversion rate of 41.30% (169,218/409,768) was observed among all users who accessed m4RH in a 3-year period, and an active use rate of 89.58% (169,218/188,904) among all who chose to navigate through the service menu.

More users were lost in the first step than in the second step of the navigation menu, the churn rates being 53.90% (220,864/409,768) and 10.42% (19,686/188,904), respectively. The churn rate could be explained by several factors. First, the fact that most users drop during the first stage of the navigation may be reflective of curious users who are trying out the service but decide they are not interested in going further. Second, it could suggest problems with message delivery (technology challenges have been observed contributing to delayed, truncated, or lack of message delivery). Third, although m4RH was developed with considerable formative work that allowed end users to contribute to the design of the service [18], this dropoff could raise the question of whether our "pull" design across all stages, with its multiple navigation steps structure, discourages users from accessing content. Alternatively, the dropoff could signify that menu option descriptions are unclear to users. Despite these potential challenges, we observed a high level of repeat use across all years. Close attention to churn rate data could help the program implement measures to reduce user attrition to further increase active use. For example, a cluster-randomized trial using hybrid text messaging and face-to-face interventions to reduce teenage pregnancies in Denver, Colorado, USA, was able to reduce user drop-out rates through the use of follow-ups by the program facilitators [19]. Analysis of churn rate data can allow program managers to identify the need to implement innovative solutions such as those applied by the Denver team. Because few programs have reported their activation rate, we are unable to determine how our findings compare to what is observed in similar programs.

Our study examined user engagement with service promotion. While we were unable to determine a statistically significant relationship between promotion and increased use, we did observe the greatest number of new users during the months when promotion activities occurred. Our findings show that during 2014, 84.88% (99,012/116,654) of new users and 84.27% (31,084/36,887) of return users were acquired during months when promotional activities were being conducted. This finding supports the rationale for investing in mass media promotion to influence m4RH use. Information about patterns in the time of day when users access the system can provide important data for future promotional efforts. Peak times for users to access m4RH across all 4 years were 1 pm and 8 pm. This finding helps to inform the timing of radio promotions to get maximum responses, as studies have shown that people are likely to respond to cues to action soon after they receive them [20].

Behavioral experts recommend that health messages should be tailored to audiences so that they can be personally relevant and affect behavioral change [21]. Past surveys to assess the demographic profile of users repeatedly showed that m4RH is mostly utilized by young people. This finding influenced the addition of youth content in 2015 as a means to tailor messages to this audience segment. However, only 11.28% (46,237/409,768) of all users and 26.56% (44,943/169,218) of active users requested youth content. This could suggest a higher demand for specific contraceptive information compared with general adolescent and youth SRH information that is also offered through other communication channels, for example, through HIV or AIDS health promotion programs. However, because the m4RH system did not routinely collect demographic data from users, we are unable to disaggregate information-seeking behavior by age or sex. Our results show that m4RH users mostly requested content on contraceptive methods and role model stories. Relatively few users requested information about clinic locations, suggesting that this content may not be interesting to most users of the m4RH system or users may have been discouraged by the need to enter additional information (first 3 or 4 letters of their ward). It is notable that the 4 regions with the most requests were all urban areas. This may be a result of greater promotional activity or higher levels of mobile phone ownership in urban areas [22]. However, this finding may be skewed as we know certain populations (key populations and youth) will travel over long distances to visit clinics in urban settings to protect their confidentiality. The request for clinic information from these locations may reflect this trend.

Analysis of data about information-seeking behavior has important implications for future iterations of the m4RH program and other health promotion programs with similar target audiences and information-sharing goals.

Challenges and Limitations

The m4RH system design presented some challenges in understanding respondent progression through the menus. As described above, the system was designed so that a user first queried information from the main menu and then from submenus. However, it appears that over time, respondents would become familiar with numerical codes assigned to menu items and directly access content without first accessing the main menu or submenus. Initially, we intended to identify repeat users by summing the total number of unique phone numbers that sent "m4RH" more than once. However, we realized that this definition of repeat use did not account for users who re-entered the system without first SMS text messaging "m4RH." This required us to create a more sophisticated Stata code that instead summed the total number of unique phone numbers that sent any code to the system on >1 day. Respondent queries (either initial or follow-up), including invalid numerical codes, text, or a combination of both, also presented challenges for analysis. For example, a respondent interested in pregnancy prevention methods (menu code 66) might send a message

reading "m4RH 66," and although this query would be received by the m4RH system, the respondent would not receive any information on pregnancy prevention methods.

Information on use patterns alone does not paint a full picture of engagement [10]. Previously conducted and published research on the m4RH program in Tanzania and elsewhere sheds light on the program's feasibility and acceptability [18,23]. However, the m4RH program in Tanzania did not routinely collect qualitative and demographic data, information that would have helped us understand the nature of engagement. For example, questions about why users drop off the navigation menu before they access content or whether there are variations in information preferences across age and gender groups would inform further design improvements.

Conclusions

The vertical and horizontal scale-up and wide reach of m4RH in Tanzania has been a major success story [24,25]. The rich data captured over the 3-year timeframe of this analysis and the results of our analyses provide insights into a useful framework to measure the degree and nature of user engagement using routine system data. For example, the conversion rate suggests a need to identify approaches to improve user engagement beyond their first contact with the m4RH system, and our data on the clinic locater use suggests that it may not provide value in its current form. Additional data collection efforts could provide a deeper understanding of our findings. These analyses contribute to knowledge about how users engage with SMS text messaging-based health promotion interventions and demonstrate how data generated on user interactions could inform improvements to the design and delivery of a service, thereby enhancing its effectiveness.

Acknowledgments

This publication was made possible by the support of the American people through the United States Agency for International Development under contract number AID-A-OAA-12-00047. The contents of this manuscript are the sole responsibility of FHI 360 and do not necessarily reflect the views of the United States Agency for International Development or the US Government.

The authors also wish to wish to acknowledge the contributions of the m4RH Tanzania team; Trinity Zan Technical Advisor Research Utilization, Global Health, Population and Nutrition (GHPN) FHI 360; Sammy Musunga, Program Officer, Tanzania Country Office, FHI 360. This work would not have been possible without the support and involvement of various technological partners Text to Change and Rasello, the Tanzania Ministry of Health Community Development, Gender, Elderly and Children, and the United States Agency for International Development Tanzania Mission.

Conflicts of Interest

None declared.

References

1. High-Impact Practices in Family Planning (HIP). mHealth: Mobile technology to strengthen family planning programs. Washington, DC: USAID; 2013 Aug. URL: https://www.popline.org/sites/default/files/mHealth.pdf [accessed 2018-09-14] [WebCite Cache ID 72QlIwZWg]
2. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
3. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015 Dec;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
4. Irvine L, Falconer DW, Jones C, Ricketts IW, Williams B, Crombie IK. Can text messages reach the parts other process measures cannot reach: an evaluation of a behavior change intervention delivered by mobile phone? PLoS One 2012;7(12):e52621 [FREE Full text] [doi: 10.1371/journal.pone.0052621] [Medline: 23300726]
5. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
6. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
7. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting Engagement With a Digital Health Intervention (HeLP-Diabetes) Using Email and Text Message Prompts: Mixed-Methods Study. Interact J Med Res 2017 Aug 22;6(2):e14 [FREE Full text] [doi: 10.2196/ijmr.6952] [Medline: 28829328]
8. Couper MP, Alexander GL, Zhang N, Little RJA, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]

9. Johnson D, Juras R, Riley P, Chatterji M, Sloane P, Choi SK, et al. A randomized controlled trial of the impact of a family planning mHealth service on knowledge and use of contraception. Contraception 2017 Jan;95(1):90-97. [doi: 10.1016/j.contraception.2016.07.009] [Medline: 27421767]
10. Smith C, Vannak U, Sokhey L, Ngo TD, Gold J, Free C. Mobile Technology for Improved Family Planning (MOTIF): the development of a mobile phone-based (mHealth) intervention to support post-abortion family planning (PAFP) in Cambodia. Reprod Health 2016 Jan 05;13:1 [FREE Full text] [doi: 10.1186/s12978-015-0112-x] [Medline: 26728505]
11. McClure D. Startup Metrics for Pirates: AARRR !!! (Startup Metrics for Product Marketing & Product Management). 2007 Aug 08. URL: https://www.slideshare.net/dmc500hats/startup-metrics-for-pirates-long-version/7-Customer_Lifecycle_Conversion_Behavior_Websitecom [accessed 2018-01-22] [WebCite Cache ID 6wfPLzkBD]
12. Vahdat HL, L'Engle KL, Plourde KF, Magaria L, Olawo A. There are some questions you may not ask in a clinic: providing contraception information to young people in Kenya using SMS. Int J Gynaecol Obstet 2013 Nov;123 Suppl 1:e2-e6 [FREE Full text] [doi: 10.1016/j.ijgo.2013.07.009] [Medline: 24012514]
13. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
14. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
15. Short CE, Rebar AL, Plotnikoff RC, Vandelanotte C. Designing engaging online behaviour change interventions: A proposed model of user engagement. The European Health Psychologist 2015 Feb;17(1):32-38 [FREE Full text]
16. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
17. Ministry of Health, Community Development, Gender, Elderly and Children (MOHCDGEC) [Tanzania Mainland], Ministry of Health (MOH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF. Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015-16. 2016 Dec. URL: https://dhsprogram.com/pubs/pdf/FR321/FR321.pdf [accessed 2018-01-22] [WebCite Cache ID 6wfPl40QP]
18. Van Rossem R, Meekers D. The reach and impact of social marketing and reproductive health communication campaigns in Zambia. BMC Public Health 2007 Dec 18;7:352 [FREE Full text] [doi: 10.1186/1471-2458-7-352] [Medline: 18088437]
19. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
20. Devine S, Leeds C, Shlay JC, Leytem A, Beum R, Bull S. Methods to assess youth engagement in a text messaging supplement to an effective teen pregnancy program. J Biomed Inform 2015 Aug;56:379-386 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.003] [Medline: 26173038]
21. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
22. Tanzania Telecommunications Report Q4 2016. London, United Kingdom: BMI Research; 2016 Aug. URL: http://bestdialogue.antenna.nl/jspui/bitstream/20.500.12018/7271/1/Tanzania%20Telecommunications%20Report%20Q4%20-%202016.pdf [accessed 2018-05-15] [WebCite Cache ID 6zRCrOkPm]
23. High-Impact Practices in Family Planning (HIP). Health communication: enabling voluntary and informed decision-making. Washington, DC: USAID; 2012 Oct. URL: https://www.fphighimpactpractices.org/wp-content/uploads/2017/06/HealthCommunication.pdf [accessed 2018-01-22] [WebCite Cache ID 72QwygSrr]
24. World Health Organization. WHO Strategic Communications Framework for Effective Communications. Geneva, Switzerland: World Health Organization; 2017. URL: http://www.who.int/mediacentre/communication-framework.pdf [accessed 2018-07-02] [WebCite Cache ID 70btvMrh5]
25. L'Engle K, Plourde KF, Zan T. Evidence-based adaptation and scale-up of a mobile phone health information service. Mhealth 2017 Mar;3:11 [FREE Full text] [doi: 10.21037/mhealth.2017.02.06] [Medline: 28567408]

Abbreviations

FP: family planning
IUCD: intrauterine contraceptive device
m4RH: Mobile for Reproductive Health
mHealth: mobile health
SMS: short message service
SRH: sexual and reproductive health

STI: sexually transmitted infection

Edited by G Eysenbach; submitted 21.02.18; peer-reviewed by H Vahdat, S Bull; comments to author 22.03.18; revised version received 16.05.18; accepted 29.06.18; published 01.11.18.

Please cite as:
Olsen PS, Plourde KF, Lasway C, van Praag E
Insights From a Text Messaging–Based Sexual and Reproductive Health Information Program in Tanzania (m4RH): Retrospective Analysis
JMIR Mhealth Uhealth 2018;6(11):e10190
URL: https://mhealth.jmir.org/2018/11/e10190/
doi:10.2196/10190
PMID:30389651

Original Paper

Text Messages Sent to Household Tuberculosis Contacts in Kampala, Uganda: Process Evaluation

Amanda J Meyer[1,2], MPH; Diana Babirye[2], MSc; Mari Armstrong-Hough[1,2], MPH, PhD; David Mark[2], BA; Irene Ayakaka[2], MBChB; Achilles Katamba[2,3], MBChB, PhD; Jessica E Haberer[4], MD; J Lucian Davis[1,2,5], MD

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT, United States

[2]Uganda Tuberculosis Implementation Research Consortium, Makerere University, Kampala, Uganda

[3]Clinical Epidemiology Unit, Makerere University, Kampala, Uganda

[4]Harvard Medical School and Massachusetts General Hospital, Boston, MA, United States

[5]Pulmonary, Critical Care, and Sleep Medicine Section, Yale School of Medicine, Yale University, New Haven, CT, United States

Corresponding Author:
J Lucian Davis, MD
Department of Epidemiology of Microbial Diseases
Yale School of Public Health
Yale University
60 College Street
PO Box 208034
New Haven, CT, 06520
United States
Phone: 1 203 785 3665
Fax: 1 203 785 4782
Email: lucian.davis@yale.edu

Abstract

Background: Previous studies have reported the inconsistent effectiveness of text messaging (short message service, SMS) for improving health outcomes, but few have examined to what degree the quality, or "fidelity," of implementation may explain study results.

Objective: The aim of this study was to determine the fidelity of a one-time text messaging (SMS) intervention to promote the uptake of tuberculosis evaluation services among household contacts of index patients with tuberculosis.

Methods: From February to June 2017, we nested a process evaluation of text message (SMS) delivery within the intervention arm of a randomized controlled trial of tuberculosis contact investigation in Kampala, Uganda. Because mobile service providers in Uganda do not provide delivery confirmations, we asked household tuberculosis contacts to confirm the receipt of a one-time tuberculosis-related text message (SMS) by sending a text message (SMS) reply through a toll-free "short code." Two weeks later, a research officer followed up by telephone to confirm the receipt of the one-time text message (SMS) and administer a survey. We considered participants lost to follow-up after 3 unsuccessful call attempts on 3 separate days over a 1-week period.

Results: Of 206 consecutive household contacts, 119 had a text message (SMS) initiated from the server. While 33% (39/119) were children aged 5-14 years, including 20% (24/119) girls and 13% (15/119) boys, 18 % (21/119) were adolescents or young adults, including 12% (14/119) young women and 6% (7/119) young men. 50% (59/119) were adults, including 26% (31/119) women and 24% (28/119) men. Of 107 (90%) participants for whom we could ascertain text message (SMS) receipt status, 67% (72/107) confirmed text message (SMS) receipt, including 22% (24/107) by reply text message (SMS) and 45% (48/107) during the follow-up telephone survey. No significant clinical or demographic differences were observed between those who did and did not report receiving the text message (SMS). Furthermore, 52% (56/107) reported ever reading the SMS. The cumulative likelihood of a text message (SMS) reaching its target and being read and retained by a participant was 19%.

Conclusions: The fidelity of a one-time text message (SMS) intervention to increase the uptake of household tuberculosis contact investigation and linkage to care was extremely low, a fact only discoverable through detailed process evaluation. This study suggests the need for systematic process monitoring and reporting of implementation fidelity in both research studies and programmatic interventions using mobile communications to improve health.

(JMIR Mhealth Uhealth 2018;6(11):e10239) doi:10.2196/10239

KEYWORDS

Africa; fidelity; implementation; intervention; short message service; tuberculosis

Introduction

Mobile phone ownership in sub-Saharan Africa has increased exponentially over the last decade, becoming widespread even among those in the lowest strata of household income [1,2]. The concurrent emergence of low-cost, easy-to-use mobile health (mHealth) apps like telephone-based short message service (SMS) text messaging and chat apps like WhatsApp have facilitated a variety of new interventions to enhance communication between patients and health care providers [3-5]. While communicating health information through mobile phones seems to be acceptable [6-8], mHealth studies have reported varying levels of success at improving patient well-being and clinical outcomes [9-16]. We hypothesize that this variability may arise from a failure to plan for, collect, or report key process measures that would help differentiate between intervention failure and implementation failure, a distinction critical to understanding the feasibility and effectiveness of mHealth interventions [17]. There is a great need, both in research studies and routine practice, for carefully performed evaluations of the *fidelity* of mHealth technologies to better understand the processes and contexts that mediate their effects on patient-centered outcomes.

Process evaluations seek to understand the degree to which interventions are delivered as actually intended—also referred to as the "fidelity" of the intervention—to explain why they do or do not work and how they can be adapted to fit the local context [18,19]. Process evaluation studies may (1) measure the dose, frequency, and quality of interventions as actually delivered; (2) assess participants' responses to interventions; and (3) characterize the mechanisms through which interventions work to improve outcomes [20]. For a health-communication intervention, a multistage process evaluation can help determine (1) if the messages reach their intended recipient; (2) if they are delivered in an accessible and timely manner; (3) if recipients open and read the messages; (4) if they respond to the messages; and (5) if the messages achieve their desired effects on targeted health behaviors. We carried out such a process evaluation of the delivery of SMS text messages in Uganda within a randomized, controlled trial of an SMS text messaging intervention designed to communicate results of household-based tuberculosis evaluation and promote the completion of follow-up procedures.

Methods

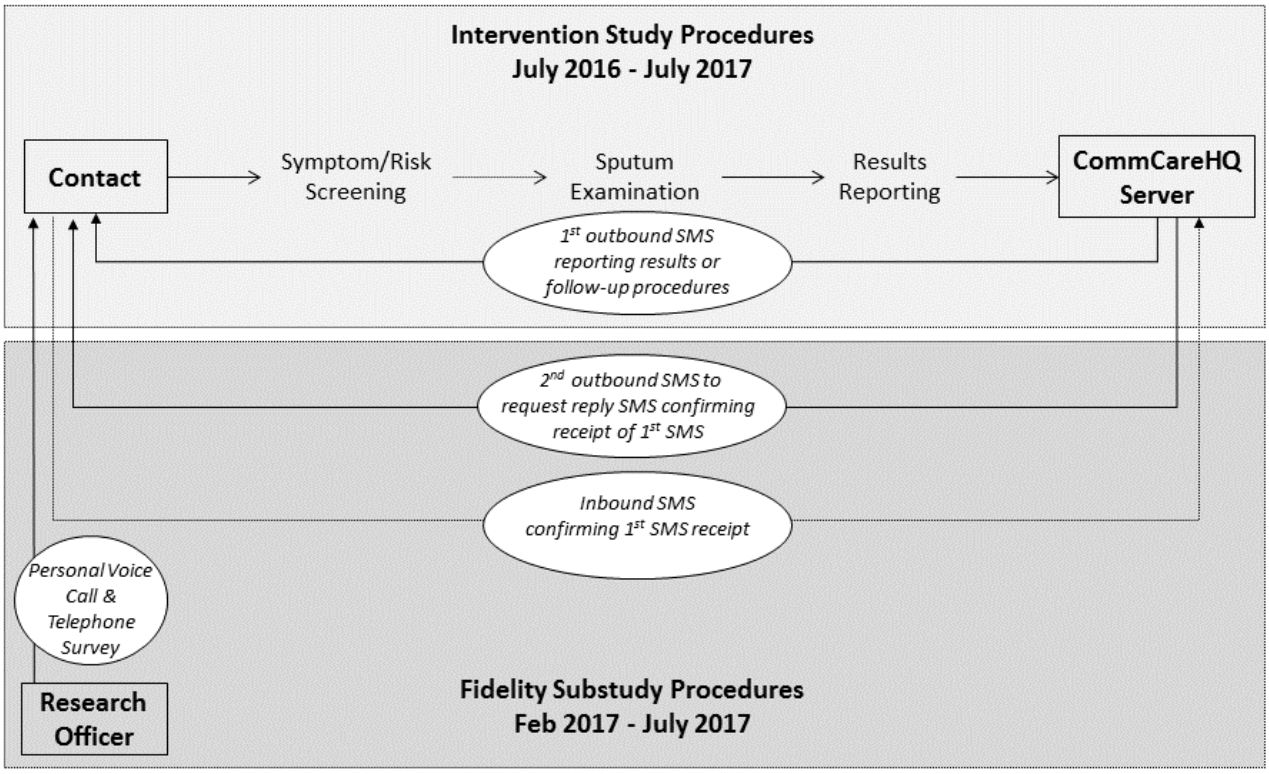
Study Population and Setting

From February to July 2017, we conducted a cross-sectional study to assess rates of the receipt of SMS, time to delivery of SMS, and retention of SMS text message content among household tuberculosis contacts (Figure 1) to evaluate the fidelity of a one-time SMS text messaging intervention. We nested this study within a randomized trial (called the "parent study") of an automated SMS text messaging intervention. The goal of the intervention was to promote the uptake of tuberculosis evaluation services among household contacts of index patients with tuberculosis at 7 public, primary care clinics in Kampala, Uganda, a setting with a high burden of undiagnosed tuberculosis [21]. In prior work in these communities, we have found that only 20% of eligible household contacts follow up in clinics to complete the tuberculosis evaluation; contacts report tuberculosis-related stigma, distrust of clinic staff, and concerns about the time and money needed to visit clinics as the main reasons for low rates of follow-up [22,23]. Although smartphones are uncommon in Uganda, we also found that almost all household members had access to mobile phones, and SMS text messages were deemed a highly acceptable way of transmitting personal health information [24]. Thus, we designed an intervention consisting of home sputum collection by community health workers (CHWs) and reporting of results by SMS text message to address these barriers to evaluation [25]. In the current process evaluation, we enrolled consecutive household contacts that were randomized into the intervention arm of the parent study.

Procedures for Screening Household Contacts for Tuberculosis in the Parent Study

According to the parent study protocol (Pan-African Clinical Trials Registration #201509000877140), participants were eligible if they were (1) household contacts of an index patient with tuberculosis; (2) able to provide informed consent; (3) not receiving tuberculosis treatment at the baseline; (4) able to access a mobile phone; (5) willing to receive SMS text messages containing personal health information; and (6) able to speak English or Luganda, the 2 most common languages in Kampala. After obtaining written informed consent from adults and parents or guardians of minors, as well as assent from minors aged 8-17 years, CHWs screened household contacts for symptoms and other indications for the evaluation for active tuberculosis. CHWs recorded clinical and demographic information using electronic tablets equipped with a customized survey app (CommCare, Dimagi, Boston, MA, USA) wirelessly linked to a remote, cloud-based server (CommCareHQ, Dimagi). Afterwards, CHWs helped contacts register each mobile phone number to be used for SMS text messaging (described below) on the remote server by entering a registration code sent to the handset. Based on the screening results, CHWs carried out additional procedures to evaluate contacts for tuberculosis and HIV [26].

Figure 1. Schema for the short message service (SMS) fidelity study.



Procedures for Short Message Service Text Messaging Within the Parent Study

CHWs recorded the results of sputum examination in the survey app, where we deployed an automated algorithm to process the relevant clinical data and assign each contact to 1-4, mutually exclusive clinical categories based on the required follow-up actions (Multimedia Appendix 1). Staff verified the logic underlying category assignments through systematic quality assurance testing of all input choices and outcomes using simulated data. We then programmed the mobile survey app to deliver a category-specific SMS text message of ≤145 characters (Multimedia Appendix 1), through integrated text messaging software (CommCare Messaging, Dimagi). The content of the messages was developed through household focus groups and interviews, as described previously [24]. Each message addressed participants by their name in English or Luganda according to their preference and provided the results of tuberculosis screening and testing with instructions for follow-up. All participants within the intervention arm of the parent study were eligible to receive SMS. All messages sent or received as part of the study were delivered free of charge to study participants.

Process Evaluation

For the process evaluation, we scheduled a second SMS text message to be sent to all household tuberculosis contacts eligible to receive SMS text message 5 minutes after the first SMS. The second SMS text message requested that participants confirm the receipt of the original SMS text message by sending an SMS text message reply through a toll-free, 4-digit "short code." Two weeks later, a research officer (DB) telephoned all participants (or parent or guardian for children aged <15 years), including those who had not responded to the SMS. After obtaining verbal consent, the research officer administered a short survey to confirm the SMS text message receipt and assess whether they recalled the information contained in the SMS. In addition, the research officer qualitatively recorded any unprompted comments made by participants about their overall experiences and interactions with SMS. We considered a participant a nonrespondent if he or she did not respond to the SMS text message within 2 weeks. We considered participants lost to follow-up after 3 unsuccessful call attempts on 3 separate days over a 1-week period.

Process Measurements

To measure the fidelity of the SMS text message intervention, we specified the following 4 steps in the delivery of an SMS, each measured either by direct observation or surveying participants: (1) SMS text message sent from the CommCare Messaging server; (2) SMS text message received at the mobile handset; (3) SMS text message read by the recipient; and (4) SMS text message understood and retained by the recipient. We measured the completion of Step 1 directly from server

logs; of Step 2, either by receipt of a reply SMS text message at the text messaging server or by surveying nonresponders by telephone; and of Steps 3 and 4, by surveying all participants by telephone (see Multimedia Appendices 2 and 3 for the survey content). A study epidemiologist (AJM) and a study social scientist (MAH) independently compared all responses to the retention questions about the SMS text message content. If a participant could report back any part of the message communicated within the SMS, he or she was considered to have retained the content. Furthermore, we measured the time to delivery of the SMS text message and time to reply among respondents. As a proxy measure for the time to complete Step 2, CHWs recorded the time to delivery of the SMS text message at the time of participant registration.

Statistical Analysis

We performed univariate analyses of participants' characteristics. We described the number of participants for whom SMS text messages successfully reached a given step in the delivery cascade as a proportion of the number of participants for whom SMS text messages successfully reached the previous step. We performed bivariate and multivariate analyses between participants' characteristics and completion of 3 process measures of interest—the proportion of participants who reported the SMS text message as received (Step 2), the proportion of participants from whom a reply SMS text message was sent, and the proportion of participants reporting that the SMS text message was understood and retained (Step 4). We assessed the significance of these comparisons using chi-square tests for categorical variables, and the Wilcoxon rank-sum test for continuous variables. We fit multivariate, mixed-effects logistic regression models for the same outcomes. In addition, we included all variables clinically significant in stepwise backward logistic regression models at $P<.2$ and, subsequently, adjusted for household-level clustering in a mixed-effects model. Although the sample size was based on convenience, we constructed 95% CIs to assess the precision of all study measurements. Finally, we estimated the cumulative reach of the SMS text messages (ie, of an SMS text message being sent, received, read, and retained) as the product of the proportions of participants, confirmed as reaching each subsequent step in the cascade, excluding those with unknown responses from the point of missingness. We performed all analyses using STATA version 14.2 (Stata Corporation).

Protection of Human Subjects

This study protocol was approved by the School of Medicine Research Ethics Committee at Makerere University, the Uganda National Council for Science and Technology, and the Yale University Human Investigation Committee.

Results

Step 1: Short Message Service Text Message Sent

Of 206 consecutive household contacts randomized into the intervention arm of the parent study, 58% (119/206) were sent an automated SMS text message containing tuberculosis-related information from the CommCare platform (Figure 2). A total of 42% (87/206) of participants did not have a message sent by the server, for 3 different reasons. First, 38% (79/206) of participants in the "tuberculosis visit pending" category did not have an SMS text message initiated because the delivery was erroneously posttimed owing to a programming error introduced while initiating this process evaluation substudy; this affected 25% (51/206) children aged <5 years, 7% (15/206) persons living with HIV, and 6% (13/206) individuals requiring a follow-up visit because of inadequate sputum collection. Second, 2% (4/206) of eligible participants did not have a message initiated because of missing data. Third, 2% (4/206) of participants did not have a message initiated because of server-related errors.

Study Population of the Process Evaluation

In this study, 33% (39/119) participants who were sent an SMS text message were children aged 5-14 years, including 20% (24/119) girls and 13% (15/119) boys, whereas 18% (21/119) were adolescents or young adults, 12% (14/119) young women and 6% (7/119) young men. In addition, 50% (59/119) were adults, 26% (31/119) women and 24% (28/119) men. The median age among adult participants was 27 years, interquartile range (IQR): 21-37. Of note, 6% (7/119) of the participants who had an SMS text message sent reported tuberculosis symptoms at the time of the interview. While 55% (66/119) preferred to receive SMS text message in English, the remainder preferred Luganda. While 60% (71/119) of the participants personally owned a mobile phone registered for the study, the remainder shared the phone with close relatives (32/119, 27%), other household members (13/119, 11%), or close friends (3/119, 3%).

We reached and interviewed 80% (95/119) of participants for follow-up telephone surveys; an additional 10% (12/119) of participants responded by SMS text message but could not be reached for the telephone survey. Furthermore, 10% (12/119) of participants did not respond by SMS text message and could not be reached by telephone, leaving 107 participants with information about SMS text message receipt.

Step 2: Short Message Service Text Message Received

Of 107 participants who had available information on SMS text message receipt, 67% (72/107) confirmed receiving an SMS, including 22% (24/107) by reply SMS text message and 45% (48/107) during the follow-up phone survey (Figure 2). No significant differences were noted in demographic characteristics between those who reported receiving the SMS text message and those who did not (Table 1). However, household contacts without tuberculosis symptoms reported receiving the SMS text message more frequently than those with tuberculosis symptoms when adjusting for household effects (69% vs 43%, cluster-adjusted odds ratio, OR, 2.9, 95% CI 0.61-13.9, P=.16), possibly because of chance.

Figure 2. Flow diagram showing process measures for short message service (SMS) delivery. *24 participants confirmed receipt by reply SMS and an additional 48 confirmed receipt during the telephone survey.

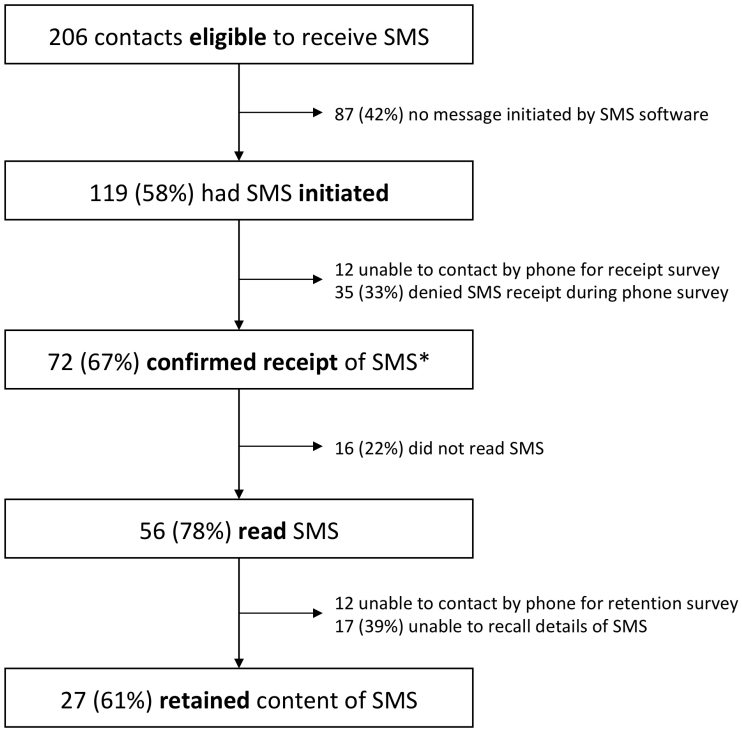


Table 1. Demographic characteristics of participants, stratified by whether the short message service was received by the participant.

Characteristic[a]	SMS[b] text message received by the participant (n=72)	SMS text message not received by the participant (n=35)	*P* value[c]
Female, n (%)	41 (57)	20 (57)	.98
Age, n (%)			.57
5-14 years	24 (33)	12 (34)	
15-21 years	14 (19)	4 (11)	
>21 years	34 (47)	19 (54)	
Tuberculosis symptoms present, n (%)	3 (4)	4 (11)	.15
Language of SMS, n (%)			.59
English	39 (54)	17 (49)	
Luganda	33 (46)	18 (51)	
Phone owner, n (%)	43 (60)	19 (54)	.59

[a]For 107 participants with definitive SMS text message receipt status; 87 participants were excluded because SMS text message was not sent and 12 were unreachable for phone survey.

[b]SMS: short message service.

[c]Chi-square tests of significance used unless otherwise noted.

Table 2. Quotations from participants on their experiences with short message service text messages.

Emergent themes	Quotations[a] from participants
Difficulty reading	• "[I] did not read it because [I] was busy but [I] will ask my daughter to read it for [me]." • "I could not have read it, I am old. However, I asked the CHW[b] and they told me I didn't have TB[c]." • "I am not good at reading. It asked me about TB symptoms."
Phone sharing	• "[I] only received the [registration] code but [I] am not the owner of the phone [my] wife is."
Interactions with phones	• "I didn't receive it because my phone has been having some problems." • "I did not pay much attention to the SMS[d] messages that come in." • "I don't usually check my SMS unless someone tells me they will be sending me a message." • "Maybe the message came but I just didn't notice."

[a]Quotes documented by a research officer during the phone survey.

[b]CHW: community health worker.

[c]TB: tuberculosis.

[d]SMS: short message service.

Short Message Service Text Message Reply Sent

No significant differences were observed in response to the SMS text message confirmation request by age, gender, or tuberculosis symptoms. However, those who preferred English as the language for SMS text messages were more likely to confirm receipt through an SMS text message than those who preferred Luganda (27% vs 11%, cluster-adjusted OR 3.8, 95% CI 1.03-14, *P*=.045). Similarly, those who personally owned a mobile phone were substantially more likely to respond with an SMS text message confirming the message receipt than those who shared a mobile phone with another individual (28% vs 8%, cluster-adjusted OR 6.7, 95% CI 1.3-34.9, *P*=.03). In a household-adjusted multivariate model (see Multimedia Appendix 4) including age, SMS text message language preference, and phone ownership, only phone ownership was significantly associated with increased odds of sending an SMS text message reply (adjusted OR 13.2, 95% CI 1.67-104, *P*=.01).

Time to Delivery of the Short Message Service Text Message

Of 119 participants who had a tuberculosis-related SMS text message sent from the CommCare server, all (n=119, 100%) were sent a registration SMS text message during the initial interview with CHWs. Most registration SMS text messages (72/119, 61%) took <5 minutes to arrive at the handset. An additional 10% (12/119) took from 5-10 minutes to arrive, while 20% (24/119) took >10 minutes. Finally, 9% (11/119) of registration messages were reported as never having arrived at the handset. For 24 individuals who sent a reply SMS text message in response to the first tuberculosis-related SMS, the median time between the SMS text message being sent and a participant sending a reply message was 35 (IQR 4-139) minutes, with all but one responding within 24 hours.

Step 3: Short Message Service Text Message Read

Of 107 participants for whom the message receipt could be determined, 52% (56/107) reported ever reading the tuberculosis-related SMS. No significant differences were observed by age, gender, tuberculosis symptoms, phone ownership, or SMS text message language preference between those who read and received and those who did not read or receive the SMS.

Step 4: Short Message Service Text Message Retained

Overall, 61% (27/44) of individuals who reported reading the SMS text message and who participated in the retention survey were able to accurately report the details of the message when prompted. No demographic or clinical characteristics were significantly associated with SMS text message retention. However, during the phone interviews, several individuals reported having difficulty or an aversion to reading as a reason for not having read the SMS text message (Table 2). In addition, several participants described how sharing a phone prevented the intended recipient from receiving the SMS text message. Finally, having a poorly functioning phone or lacking comfort with retrieving SMS text messages influenced both SMS text message receipt and the likelihood of a participant reading an SMS text message.

Cumulative Reach and Retention of the Short Message Service Text Message

The cumulative likelihood of an SMS text message reaching its target and being read and retained by participants was 19%. Among those for whom the text messaging server successfully initiated an SMS text message, the cumulative likelihood of receipt and retention was 32%.

Discussion

The potential for mobile phones to improve access to evaluation and treatment for tuberculosis by enhancing communication between at-risk individuals and health care workers has generated great enthusiasm for mHealth technologies, especially SMS text messages [27,28]. However, data about how these interventions actually work—or do not work—in routine practice are limited [29]. In this study, we applied the powerful approach of measuring the implementation fidelity of SMS text messages through cascade analysis, achieved through prospective cross-sectional surveys and other novel measures of SMS text message delivery, response, and mechanism. This study shows that multiple, frequently unobserved barriers exist to

implementing an SMS text messaging intervention in a way that ensures that participants receive and comprehend the messages. Our findings suggest the need for systematic monitoring and inclusion of detailed process evaluation in research studies and programmatic interventions using mobile communications to improve health.

The use of SMS text messages to provide tuberculosis-related health information is a complex intervention, as defined by having multiple, interacting components [30]. Key components of this intervention included health workers, computer servers and software, mobile telephone networks, mobile handsets, and community members; for such complex interventions, a detailed process evaluation is required to understand if all elements work together as intended [20]. We found that a significant proportion of participants never received the SMS text messages as intended, and that a large proportion of those who did receive the messages never read them. Even among those who did read the messages, a notable proportion were unable to accurately report the details of the message only 2 weeks later. Ultimately, less than a third of participants reported both receiving and retaining the tuberculosis-related information contained in the SMS text messages that were sent.

SMS text messaging interventions have been evaluated across sub-Saharan Africa for their capacity to improve medication adherence [31], support the dissemination of lab results [32], and reduce missed clinic visits [33]. However, with few exceptions [34], the existing literature does not address how often SMS text messages are received, understood, and retained by participants. In one study, which followed up on SMS text messages sent from a laboratory to 385 persons living with HIV in Uganda, only 72% of participants reported receiving the SMS text messages that were sent [34]. As in our study, participant literacy and the ability of participants to independently access SMS text messages on their phones at enrollment were associated with receiving SMS. Given these findings, future research should focus on improving accessibility to the behavioral components of the intervention through functionalities, such as automated voice calling, and, more generally, on embracing human-centered approaches to the design of mHealth interventions [35].

In our study, surprisingly, few individuals confirmed the message receipt through SMS text message reply. While a previous study in Uganda reported a 70% response rate to SMS text message containing health education quizzes [36], another large study conducted in northern Uganda involving a one-time SMS text messaging intervention found that only 23% responded, similar to what we observed in this study [37]. Although a widely cited systematic review has previously shown that 2-way SMS text messages are more effective than one-way SMS text messages in engaging patients [38], the focus of that review was on longitudinal SMS text messaging interventions for medication adherence, not on responses to a one-time communication. In a broader context that would include short-term SMS text messaging interventions to facilitate diagnostic evaluation and linkage to care, we hypothesize that SMS text message response rates—a key component of the implementation fidelity—likely also depend on other components of fidelity, including the dose, intensity, and behavioral mechanisms of the SMS text messaging intervention. Future studies should go beyond the simple process measures that we included in this study to describe these other important mediators.

The proportion of messages reported as received in this study was unexpectedly low given the widespread use of SMS text message in Uganda. Additional studies of how participants access SMS text messages are needed to understand the barriers between message initiation and participant receipt. Potential barriers at this step could include network outages and, as our qualitative data shows, malfunctioning of mobile devices and phone-sharing practices of participants. Furthermore, if mobile service providers do not provide delivery confirmations, innovative methods for assessing message receipt will be needed. In this study, simple SMS text message replies were an insufficient measure of receipt. Previous studies have shown higher levels of engagement when utilizing serial text messaging and 2-way communications [9,10,39,40] rather than one-time text messaging. Including quizzes or trivia may also improve participant response rates [36]. These strategies, along with more personalized SMS text message content [24], may increase participant engagement.

This study has a few limitations. First, we had a limited sample size that, combined with a high proportion of messages that did not reach their intended targets or convey the information intended, limited our ability to carry out stratified analyses to understand differences between subgroups. This limitation is partly moderated by our collection of qualitative responses from individuals about their experiences and interactions with the SMS text messages, which illustrate the types of barriers that participants face in engaging with mHealth interventions. Second, a programming error reduced the number of individuals who received SMS, including the subgroup for whom clinic follow-up would have been requested. This error prevented us from carrying out an analysis of the effect of different message types on participant interactions with SMS text messages and participant follow-up behaviors. In addition, it may have caused us to modestly underestimate response rates, as a previous study reported lower rates of response among those with normal results than among those with abnormal results [34]. Moreover, our programming error underscores the difficulties of ensuring successful SMS text message initiation, even with intensive quality assurance practices in place. Finally, we waited 2 weeks after triggering the initial SMS text message before attempting to contact participants to avoid interfering with parent study outcomes; this design feature may have biased the observed rate of retention of information downward, as recall error may increase with time.

This study also had several strengths. First, we had a low rate of participants lost to follow-up. We were able to interview 80% of participants by phone, and obtain, at least, some follow-up information from 90% of participants. Second, we applied innovative techniques to determine message receipt using reply SMS text messages, although the uptake of this method of message verification was extremely low. Finally, our study population included young residents of a crowded urban area with high rates of access to mobile phones [24], making our findings likely generalizable to many urban settings in

sub-Saharan Africa where health-related SMS text messaging interventions are being evaluated and implemented.

Overall, this study found lower than expected levels of SMS text message receipt and retention, and substantial delays in delivery. While SMS text messages have the potential to ease communication between health workers and patients, improving the delivery cascade of SMS text messages is imperative for the success of SMS text messaging interventions. If mobile text messaging interventions are to have their full impact, innovative process measures to confirm the receipt and comprehension must be developed and applied. With better monitoring and quality improvement strategies, SMS text messaging could reach more patients more effectively, enhancing communication between patients and practitioners and building more patient-responsive health systems.

Acknowledgments

We would like to thank the study participants, the staff of the participating health centers, and all community health workers who aided in this study, as well as the staff in research administration at the Makerere College of Health Sciences. This study was funded by National Institutes of Health R01AI104824 (JLD) and Nina Ireland Program in Lung Health at the University of California San Francisco (JLD).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tuberculosis evaluation categories and SMS message content.

[PDF File (Adobe PDF File), 16KB - mhealth_v6i11e10239_app1.pdf]

Multimedia Appendix 2

Follow-up phone survey for participants who did not send an SMS reply.

[PDF File (Adobe PDF File), 16KB - mhealth_v6i11e10239_app2.pdf]

Multimedia Appendix 3

Follow-up phone survey for participants who sent an SMS reply.

[PDF File (Adobe PDF File), 15KB - mhealth_v6i11e10239_app3.pdf]

Multimedia Appendix 4

Predictors of sending an SMS reply message.a.

[PDF File (Adobe PDF File), 13KB - mhealth_v6i11e10239_app4.pdf]

References

1. Poushter J. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Pew Research Center Global Attitudes & Trends: Pew Research Center; 2016. URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2018-02-01] [WebCite Cache ID 6xXooBQrq]
2. Pew Research Center. 2015. Cell Phones in Africa: Communication Lifeline URL: http://www.pewglobal.org/2015/04/15/cell-phones-in-africa-communication-lifeline/ [accessed 2018-02-01] [WebCite Cache ID 6xXp3aIik]
3. Lukhele BW, Musumari P, El-Saaidi C, Techasrivichien T, Suguimoto SP, Ono Kihara M, et al. Efficacy of Mobile Serious Games in Increasing HIV Risk Perception in Swaziland: A Randomized Control Trial (SGprev Trial) Research Protocol. JMIR Res Protoc 2016 Nov 22;5(4):e224 [FREE Full text] [doi: 10.2196/resprot.6543] [Medline: 27876685]
4. Martinez R, Rogers AD, Numanoglu A, Rode H. The value of WhatsApp communication in paediatric burn care. Burns 2018 Jun;44(4):947-955. [doi: 10.1016/j.burns.2017.11.005] [Medline: 29395403]
5. Henwood R, Patten G, Barnett W, Hwang B, Metcalf C, Hacking D, et al. Acceptability and use of a virtual support group for HIV-positive youth in Khayelitsha, Cape Town using the MXit social networking platform. AIDS Care 2016 Dec;28(7):898-903. [doi: 10.1080/09540121.2016.1173638] [Medline: 27098208]
6. Mitchell KJ, Bull S, Kiwanuka J, Ybarra ML. Cell phone usage among adolescents in Uganda: acceptability for relaying health information. Health Educ Res 2011 Oct;26(5):770-781 [FREE Full text] [doi: 10.1093/her/cyr022] [Medline: 21536715]

7. Rana Y, Haberer J, Huang H, Kambugu A, Mukasa B, Thirumurthy H, et al. Short message service (SMS)-based intervention to improve treatment adherence among HIV-positive youth in Uganda: focus group findings. PLoS One 2015;10(4):e0125187 [FREE Full text] [doi: 10.1371/journal.pone.0125187] [Medline: 25881059]
8. Kim J, Zhang W, Nyonyitono M, Lourenco L, Nanfuka M, Okoboi S, et al. Feasibility and acceptability of mobile phone short message service as a support for patients receiving antiretroviral therapy in rural Uganda: a cross-sectional study. J Int AIDS Soc 2015;18:20311 [FREE Full text] [Medline: 26654029]
9. Pop-Eleches C, Thirumurthy H, Habyarimana J, Zivin J, Goldstein M, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
10. Lester R, Ritvo P, Mills E, Kariri A, Karanja S, Chung M, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
11. 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 Dec 15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
12. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester R, Mills E, Smieja M, et al. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. PLoS One 2012;7(12):e46909 [FREE Full text] [doi: 10.1371/journal.pone.0046909] [Medline: 23236345]
13. Georgette N, Siedner M, Petty C, Zanoni B, Carpenter S, Haberer J. Impact of a clinical program using weekly Short Message Service (SMS) on antiretroviral therapy adherence support in South Africa: a retrospective cohort study. BMC Med Inform Decis Mak 2017 Feb 20;17(1):18 [FREE Full text] [doi: 10.1186/s12911-017-0413-9] [Medline: 28219372]
14. Dryden-Peterson S, Bennett K, Hughes M, Veres A, John O, Pradhananga R, et al. An augmented SMS intervention to improve access to antenatal CD4 testing and ART initiation in HIV-infected pregnant women: a cluster randomized trial. PLoS One 2015;10(2):e0117181 [FREE Full text] [doi: 10.1371/journal.pone.0117181] [Medline: 25693050]
15. Reid M, Steenhoff A, Thompson J, Gabaitiri L, Cary M, Steele K, et al. Evaluation of the effect of cellular SMS reminders on consistency of antiretroviral therapy pharmacy pickups in HIV-infected adults in Botswana: a randomized controlled trial. Health Psychol Behav Med 2017;5(1):101-109 [FREE Full text] [doi: 10.1080/21642850.2016.1271333] [Medline: 28966882]
16. Giordano V, Koch H, Godoy-Santos A, Dias Belangero W, Esteves Santos Pires R, Labronici P. WhatsApp Messenger as an Adjunctive Tool for Telemedicine: An Overview. Interact J Med Res 2017 Jul 21;6(2):e11 [FREE Full text] [doi: 10.2196/ijmr.6214] [Medline: 28733273]
17. Rabin BR. Terminology for Dissemination and Implementation Research. In: Brownson R, Colditz G, Proctor E, editors. Dissemination and Implementation Research in Health. New York: Oxford University Press; 2012:23-51.
18. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007 Nov 30;2:40 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
19. Pérez D, Van der Stuyft P, Zabala MC, Castro M, Lefèvre P. A modified theoretical framework to assess implementation fidelity of adaptive public health interventions. Implement Sci 2016 Dec 08;11(1):91 [FREE Full text] [doi: 10.1186/s13012-016-0457-8] [Medline: 27391959]
20. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258-h1258. [doi: 10.1136/bmj.h1258]
21. World Health Organization. Global Tuberculosis control: WHO Report. Geneva: World Health Organization; 2017. URL: http://apps.who.int/iris/bitstream/10665/259366/1/9789241565516-eng.pdf?ua=1 [accessed 2018-02-01] [WebCite Cache ID 725klLRyZ]
22. Armstrong-Hough M, Turimumahoro P, Meyer AJ, Ochom E, Babirye D, Ayakaka I, et al. Drop-out from the tuberculosis contact investigation cascade in a routine public health setting in urban Uganda: A prospective, multi-center study. PLoS One 2017;12(11):e0187145 [FREE Full text] [doi: 10.1371/journal.pone.0187145] [Medline: 29108007]
23. Ayakaka I, Ackerman S, Ggita J, Kajubi P, Dowdy D, Haberer J, et al. Identifying barriers to and facilitators of tuberculosis contact investigation in Kampala, Uganda: a behavioral approach. Implement Sci 2017 Dec 09;12(1):33 [FREE Full text] [doi: 10.1186/s13012-017-0561-4] [Medline: 28274245]
24. Ggita J, Ojok C, Meyer AJ, Farr K, Shete P, Ochom E, et al. Patterns of usage and preferences of users for tuberculosis-related text messages and voice calls in Uganda. Int J Tuberc Lung Dis 2018 May 01;22(5):530-536. [doi: 10.5588/ijtld.17.0521] [Medline: 29663958]
25. Ayakaka I, Ackerman S, Ggita J, Kaiubi P, Haguma P, Haberer J. Designing an intervention to improve household TB contact investigation in Uganda. 2016 Presented at: 9th Annual Conference on the Science of Dissemination and Implementation; December 14-15, 2016; Washington DC.
26. Ochom E, Meyer AJ, Armstrong-Hough M, Kizito S, Ayakaka I, Turimumahoro P, et al. Integrating home HIV counselling and testing into household TB contact investigation: a mixed-methods study. Public Health Action 2018 Jun 21;8(2):72-78 [FREE Full text] [doi: 10.5588/pha.18.0014] [Medline: 29946523]

27. Denkinger C, Grenier J, Stratis A, Akkihal A, Pant-Pai N, Pai M. Mobile health to improve tuberculosis care and control: a call worth making. Int J Tuberc Lung Dis 2013 Jun;17(6):719-727. [doi: 10.5588/ijtld.12.0638] [Medline: 23541232]
28. Stop TB Partnership. mHealth to Improve TB Care.: Interactive Research and Development; 2012 May 27. URL: http://www.stoptb.org/assets/documents/resources/publications/acsm/mHealth%20to%20Improve%20TB%20Care.pdf [accessed 2018-02-01] [WebCite Cache ID 73PO5xD1z]
29. Labrique A, Vasudevan L, Chang L, Mehl G. H_pe for mHealth: more "y" or "o" on the horizon? Int J Med Inform 2013 May;82(5):467-469 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.016] [Medline: 23279850]
30. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [Medline: 18824488]
31. Nhavoto JA, Grönlund Å, Klein GO. Mobile health treatment support intervention for HIV and tuberculosis in Mozambique: Perspectives of patients and healthcare workers. PLoS One 2017;12(4):e0176051 [FREE Full text] [doi: 10.1371/journal.pone.0176051] [Medline: 28419149]
32. Sutcliffe C, Thuma P, van Dijk JH, Sinywimaanzi K, Mweetwa S, Hamahuwa M, et al. Use of mobile phones and text messaging to decrease the turnaround time for early infant HIV diagnosis and notification in rural Zambia: an observational study. BMC Pediatr 2017 Mar 08;17(1):66 [FREE Full text] [doi: 10.1186/s12887-017-0822-z] [Medline: 28270134]
33. Coleman J, Bohlin KC, Thorson A, Black V, Mechael P, Mangxaba J, et al. Effectiveness of an SMS-based maternal mHealth intervention to improve clinical outcomes of HIV-positive pregnant women. AIDS Care 2017 Dec;29(7):890-897. [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]
34. Siedner MJ, Santorino D, Haberer JE, Bangsberg DR. Know your audience: predictors of success for a patient-centered texting app to augment linkage to HIV care in rural Uganda. J Med Internet Res 2015 Mar 24;17(3):e78 [FREE Full text] [doi: 10.2196/jmir.3859] [Medline: 25831269]
35. International Organization for Standardization (ISO). 2010 Mar. Ergonomics of human-system interaction--Part 210: Human-centered design for interactive systems2010; ISO 9241-210 URL: https://www.iso.org/standard/52075.html [accessed 2018-10-24] [WebCite Cache ID 73PORxuPk]
36. de Lepper AM, Eijkemans M, van Beijma H, Loggers J, Tuijn C, Oskam L. Response patterns to interactive SMS health education quizzes at two sites in Uganda: a cohort study. Trop Med Int Health 2013 Apr;18(4):516-521 [FREE Full text] [doi: 10.1111/tmi.12059] [Medline: 23297650]
37. Chib A, Wilkin H, Ling LX, Hoefman B, Van Biejma H. You have an important message! Evaluating the effectiveness of a text message HIV/AIDS campaign in Northwest Uganda. J Health Commun 2012;17 Suppl 1:146-157. [doi: 10.1080/10810730.2011.649104] [Medline: 22548607]
38. Wald D, Butt S, Bestwick J. One-way versus two-way text messaging on improving medication adherence: meta-analysis of randomized trials. Am J Med 2015 Oct;128(10):1139.e1-1139.e5. [doi: 10.1016/j.amjmed.2015.05.035] [Medline: 26087045]
39. Odeny T, Bukusi E, Cohen C, Yuhas K, Camlin C, McClelland R. Texting improves testing: a randomized trial of two-way SMS to increase postpartum prevention of mother-to-child transmission retention and infant HIV testing. AIDS 2014 Sep 24;28(15):2307-2312 [FREE Full text] [doi: 10.1097/QAD.0000000000000409] [Medline: 25313586]
40. Georgette N, Siedner MJ, Zanoni B, Sibaya T, Petty CR, Carpenter S, et al. The Acceptability and Perceived Usefulness of a Weekly Clinical SMS Program to Promote HIV Antiretroviral Medication Adherence in KwaZulu-Natal, South Africa. AIDS Behav 2016 Dec;20(11):2629-2638 [FREE Full text] [doi: 10.1007/s10461-016-1287-z] [Medline: 26781866]

Abbreviations

CHW: community health worker
IQR: interquartile range
mHealth: mobile health
OR: odds ratio
SMS: short message service
TB: tuberculosis

Edited by G Eysenbach; submitted 27.02.18; peer-reviewed by A DeLuca, L Li; comments to author 30.08.18; revised version received 31.08.18; accepted 10.09.18; published 20.11.18.

Please cite as:
Meyer AJ, Babirye D, Armstrong-Hough M, Mark D, Ayakaka I, Katamba A, Haberer JE, Davis JL
Text Messages Sent to Household Tuberculosis Contacts in Kampala, Uganda: Process Evaluation
JMIR Mhealth Uhealth 2018;6(11):e10239
URL: http://mhealth.jmir.org/2018/11/e10239/
doi:10.2196/10239
PMID:30459147

Original Paper

Medical Correctness and User Friendliness of Available Apps for Cardiopulmonary Resuscitation: Systematic Search Combined With Guideline Adherence and Usability Evaluation

Bibiana Metelmann[1], MD; Camilla Metelmann[1], MD; Louisa Schuffert[1]; Klaus Hahnenkamp[1], MD, PhD; Peter Brinkrolf[1], BA, MD

Department of Anaesthesiology, University Medicine Greifswald, Greifswald, Germany

Corresponding Author:
Bibiana Metelmann, MD
Department of Anaesthesiology
University Medicine Greifswald
Ferdinand-Sauerbruch-Straße
Greifswald, 17489
Germany
Phone: 49 38348760 ext 2848
Fax: 49 3834876092848
Email: bibiana.metelmann@uni-greifswald.de

Abstract

Background: In case of a cardiac arrest, start of cardiopulmonary resuscitation by a bystander before the arrival of the emergency personnel increases the probability of survival. However, the steps of high-quality resuscitation are not known by every bystander or might be forgotten in this complex and time-critical situation. Mobile phone apps offering real-time step-by-step instructions might be a valuable source of information.

Objective: The aim of this study was to examine mobile phone apps offering real-time instructions in German or English in case of a cardiac arrest, to evaluate their adherence to current resuscitation guidelines, and to test their usability.

Methods: Our 3-step approach combines a systematic review of currently available apps guiding a medical layperson through a resuscitation situation, an adherence testing to medical guidelines, and a usability evaluation of the determined apps. The systematic review followed an adapted preferred reporting items for systematic reviews and meta-analyses flow diagram, the guideline adherence was tested by applying a conformity checklist, and the usability was evaluated by a group of mobile phone frequent users and emergency physicians with the system usability scale (SUS) tool.

Results: The structured search in Google Play Store and Apple App Store resulted in 3890 hits. After removing redundant ones, 2640 hits were checked for fulfilling the inclusion criteria. As a result, 34 apps meeting all inclusion criteria were identified. These included apps were analyzed to determine medical accuracy as defined by the European Resuscitation Council's guidelines. Only 5 out of 34 apps (15%, 5/34) fulfilled all criteria chosen to determine guideline adherence. All other apps provided no or wrong information on at least one relevant topic. The usability of 3 apps was evaluated by 10 mobile phone frequent users and 9 emergency physicians. Of these 3 apps, solely the app "HELP Notfall" (median=87.5) was ranked with an SUS score above the published average of 68. This app was rated significantly superior to "HAMBURG SCHOCKT" (median=55; asymptotic Wilcoxon test: z=−3.63, P<.01, n=19) and "Mein DRK" (median=32.5; asymptotic Wilcoxon test: z=−3.83, P<.01, n=19).

Conclusions: Implementing a systematic quality control for health-related apps should be enforced to ensure that all products provide medically accurate content and sufficient usability in complex situations. This is of exceptional importance for apps dealing with the treatment of life-threatening events such as cardiac arrest.

(JMIR Mhealth Uhealth 2018;6(11):e190) doi:10.2196/mhealth.9651

KEYWORDS

mHealth; resuscitation; review; guidelines; mobile phones; health care information systems; health informatics

Introduction

Background

In case of a cardiac arrest, cardiopulmonary resuscitation (CPR) has to start as soon as possible [1]. However, even in the most advanced emergency systems, the emergency personnel needs a medium time of 5-8 min to arrive at the emergency site [2]. Therefore, it is crucial that a bystander—that is, a person noticing the cardiac arrest—starts CPR [3,4]. In the majority of cases, the bystander will be a medical layperson [5]. It has been shown in multiple studies that bystander CPR increases the rate of survival [1,6-9]. Nevertheless, the rate of bystander CPR is still relatively low [10]. Various reasons for this gap are discussed [11,12]. One of them might be that the bystander has probably never experienced a similar situation before and is therefore unsure what to do and fears to make mistakes [13]. The situation being highly time-critical further increases the cognitive workload. Decisions have to be made fast, leaving no time for elaborate reflections. It has been shown that cognitive aids can help reduce stress in these types of situations [14]. However, because cardiac arrests can happen anytime and at any place, it is unlikely that the bystander carries a traditional cognitive aid such as a textbook or leaflet with the required information.

Objectives

A possible solution to allow immediate assistance might be a mobile phone app offering real-time step-by-step instructions. Mobile electronic devices such as mobile phones are ubiquitously available and a bystander is likely to have them available at site [15]. Mobile phone apps have become a part of everyday culture and have changed daily life in nearly all aspects [16]. Especially digital natives are used to being able to receive information immediately; internet research and apps are their first choice in cases of questions and often the main source of information [16,17]. Likewise, the market of mHealth apps has grown exponentially over the last years [18-20]. Cognitive aids, which are based on medical guidelines, are increasingly accepted in health care [21]. Ahn et al described that the total number of downloads for CPR training apps is about several hundred thousand [22]. However, this field is getting increasingly complex and unmanageable [23]. Kumar et al raised concerns toward untested apps already at the mHealth Evidence Workshop at the US National Institutes of Health in 2011 [24]. They demanded rigorous research to examine the potentially negative consequences of ineffective mHealth apps or apps based on incorrect facts. This could lead to patient harm and higher medical costs [24]. It cannot be expected from medical laypersons to analyze all possible apps, evaluate the content, and decide whether it conforms to current medical guidelines.

The aim of this study was to systematically detect apps giving German or English step-by-step instructions to perform CPR by an adult bystander and determine both their adherence to current medical guidelines and their usability.

Methods

Study Setup

Our 3-step approach combined (1) a systematic review of currently available apps guiding a medical layperson through a resuscitation situation, (2) an adherence testing to medical guidelines, and (3) a usability evaluation of the determined apps. The study was approved by the institutional review board of Universitätsmedizin Greifswald with the case number BB 055/17.

Systematic Review of Available Apps

To date, there is no standardized search method for identifying mobile health apps. We used an approach similar to other studies [25-27]. The search was structured to an adapted preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram [28].

We focused on the 2 largest and most popular stores for mobile apps—Apple App Store (for Apple iOS apps) and the Google Play Store (for Android apps). Smaller stores such as Amazon App store, Windows Store, Samsung Apps, or Blackberry World were not included in this study [29-31]. The apps offered in these 2 app stores are automatically preselected depending on the region from where the search is conducted. Google Play Store identifies the location based on the users' IP address, whereas Apple App Store uses Apple ID. This default country setting can be changed [32,33]. An extensive search with country settings of all English-speaking countries would have led to an unmanageable amount of apps; whereas a restriction to just a few selected English-speaking countries would have been arbitrary. Therefore, we decided to restrict the search to the country setting of Germany. We defined 16 keywords and hand-searched each term separately. These keywords were the English words "CPR," "resuscitation," "chest compression," "basic life support," "BLS," "first aid," "cardiac arrest," "112," and similar German expressions ("Reanimation," "Wiederbelebung," "Thoraxkompression," "Herzdruckmassage," "Erste Hilfe," "Herzstillstand," "Kreislaufstillstand," and "Notfall"). The systematic search was carried out on a MacBook Pro between May 26, 2017, and June 23, 2017. For Apple App Store, the iTunes search configuration was set to "all" (Mac, iPad, iPhone, and Apple Watch). Google Play Store was searched on the same MacBook Pro [34] accessed via Safari internet browser. Consistent with other studies, apps were identified if the keyword was either part of the title or the description of the app [27].

All identified apps were screened. Apps that were found under different keywords but had the same name and were developed by the same company were considered redundant.

For assessment of eligibility, the remaining apps were examined regarding the study's inclusion criteria. Apps not coherent with 1 or more of these criteria were excluded from further evaluation. The following inclusion criteria were used in a descending rank order: availability on both Google Play Store and Apple App Store, language of the app either German or English, free of charge, covered the topic resuscitation, resuscitation of human beings, provides real-time step-by-step

instructions, no duplicate under different names, and no technical problems. The inclusion criterion "availability in both stores" was chosen to find an app that can be recommended in, for example, basic life support trainings and is usable by the majority of mobile phone users without being restricted to a subgroup. Apps were classified as duplicates if they were developed by the same company and had the same interface but had different names. All remaining apps were included for evaluation of guideline adherence.

Conformity to Guideline

We analyzed the quality of content based on the adherence to the European Resuscitation Council Guidelines for Resuscitation 2015 [2], the American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care 2015 [35], and the 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations guidelines of the International Liaison Committee on Resuscitation [36,37]. A conformity checklist was developed in 2 successive brainstorming sessions of 8 emergency physicians and paramedics containing the following 9 items: the app should request the user to "check responsiveness," "open the airway," "assess whether the person is breathing normally (see, hear, and feel)," "consider no breathing or abnormal, ie, agonal breathing," "call 112 (or 911) or ask somebody to call 112 (or 911)," "start chest compression," "pay attention to correct positioning of hands," "compress the chest at a rate of 100-120 bpm," and "compress to a depth of at least 5 cm but not more than 6 cm." If the app explained ventilation, it was expected to include "opening of the airway" and "verify rising of the chest." To be rated as guideline-conform, all criteria needed to be covered in substance; verbatim coverage was not required.

Usability Evaluation

Remaining apps were evaluated using the system usability scale (SUS) developed by John Brooke [38]. This tool is based on the 3 categories of International Organization for Standardization norm 9241-11 for usability: "effectiveness," "efficiency," and "satisfaction" [38,39]. The SUS was rated as a highly robust and versatile tool to evaluate usability [40]. It is the most widely used scale to evaluate usability and has been cited in more than 600 research publications [41]. The questionnaire of the SUS consists of 10 statements on a 5-point Likert scale with 5 positive statements (item 1, 3, 5, 7, and 9) and 5 negative statements (item 2, 4, 6, 8, and 10). The user is asked to rate their level of agreement with these statements concerning the software under review. To get a total score between 0 and 100, the individual scores are calculated as follows: each item's score ranges from 1 to 5 depending on the position. In case of the uneven items, the scale position minus 1 contributes to the total score. In case of the even items, the contribution to the total score is 5 minus the scale position. In the next step, the sum of these results is multiplied by 2.5. This results in a value on a scale between 0 and 100 [38]. This value does not represent a percentage of usability [42]. The German version of the SUS used is attached in Multimedia Appendix 1. When translating the SUS questionnaire from its original version, we changed the general term "system" into the more specific word "app."

Sauro and Lewis detected a rate of 11% of coding errors when working with a conventional SUS score with positive and negative statements [43]. Therefore, in our evaluation, 2 researchers calculated the SUS score independently to diminish the rate of mistakes in calculation. If their results did not match, a third researcher calculated the SUS score.

Usability Evaluators

Barnum recommends asking multiple groups to evaluate with SUS to emphasize different aspects of a given system [44]. Regarding this topic, 3 groups, whose SUS evaluation could show different perspectives, were identified: (1) people with a higher chance of having to use the app, (2) experienced app users, and (3) people with high experience regarding the medical content of the app. The highest risk of being confronted with cardiac arrest can be attributed to elderly people [45] as well as professionals of the medical field. However, the percentage of individuals owning a mobile phone is by far smaller among people aged 65 years and older than found in the average population [46]. People working in medical environments are educated and trained in basic life support and unlikely to need the help of an app to perform the basic steps of resuscitation. Thus, we decided not to interview these groups and focus on the remaining 2: those who frequently use related products (apps) and those whose work is relevant to the content of the product. Consequently, 1 group evaluating the app consisted of mobile phone frequent users, whereas the other group consisted of emergency physicians. Mobile phone frequent users were defined as individuals who had owned a mobile phone for more than 3 years, currently have more than 15 apps installed on their device, and use these for more than 1 hour per day. The emergency physicians are all currently employed in the German emergency system. The emergency physicians were asked to keep in mind that the apps were designed to teach the steps of basic life support to medical laypersons.

In addition, the emergency physicians were asked to rank the apps according to the quality of teaching different aspects of high-quality CPR: 7 aspects were developed in 2 successive brainstorming sessions of 8 emergency physicians and paramedics based on the German translation of the European Resuscitation Council guidelines. Aspects, all researchers involved associated with high quality CPR were collected and evaluated. The criteria were as follows: "The app comprehensively explains the opening of the airway," "The app points out the problem of agonal breathing," "The app emphasizes the importance of complete recoil of the chest after each compression," "The app indicates that pauses in chest compressions should be minimized," "The app helps the medical layperson to find the correct frequency for chest compression," "The design and user-interface supports an optimal execution of cardiopulmonary resuscitation," and "The app requests the medical layperson to continue with cardiopulmonary resuscitation until the arrival of emergency service."

Statistical processing of the data was carried out using IBM SPSS Statistics, version 26.0 (IBM Corporation, Armonk, New York, USA), and Microsoft Excel 2010 (Microsoft Corporation,

Redmond, Washington, USA). We assessed normal distribution by the Shapiro-Wilk test; median and interquartile range were calculated. In case of normal distribution, *t* test was used to assess significance levels. To assess significance level in the absence of normal distribution, Mann-Whitney *U* test was used between the 2 groups testing the same app and Wilcoxon test between the same people testing different apps.

Results

Systematic Review of Apps

The results of the systematic review are depicted in Figure 1 as an adapted PRISMA flow diagram. The search of the 16 German and English keywords identified 3146 search results in Google Play Store and 744 in Apple App Store. After the exclusion process, 34 apps remained for the evaluation of guideline adherence (see Multimedia Appendix 2).

Adherence to Guideline

The results of the analysis of guideline adherence are depicted in Figure 2. A total of 7 apps taught hands-only CPR, whereas 27 also explained ventilation. Out of the 34 apps analyzed, 18 (53%, 15/34) did not indicate to consider "no breathing and abnormal, ie, agonal breathing," 17 (50%, 17/34) did not request to assess whether the person is breathing normally, 18 (53%, 18/34) did not explain to compress to the recommended depth of at least 5 cm but not more than 6 cm, and 17 (50%, 17/34) did not recommend to open the airway. In our evaluation, only 5 out of the 34 (15%, 5/34) apps met all criteria of guideline adherence tested; these apps are listed in Table 1.

Usability Evaluation

The usability evaluation with the SUS tool was conducted in October 2017. Of the 5 apps that met the criteria, 2 had to be excluded before starting the usability evaluation: 1 app was no longer available in both app stores ("St John Wales First Aid") and the other app ("Notfall-Hilfe") was excluded because it showed fundamental differences between the version of the Google Play Store and Apple App Store. The version of the Google Play Store contained pictures and movies explaining all steps, and the text was read aloud, if desired. None of these features were available in the Apple App Store version. This gap would profoundly influence the results of the SUS, leading to the decision to exclude this app in the usability evaluation.

The group of mobile phone frequent users consisted of 10 participants (7 females and 3 males) with a median age of 23 years (minimum=20 years and maximum=25 years) and the group of emergency physicians of 9 participants (4 females and 5 males) with a median age of 37 years (minimum=32 years and maximum=56 years). The SUS participants assessed the apps on their own mobile phone. iPhone 6Plus, provided by the researchers, was used by 2 emergency physicians.

Figure 1. Results of the systematic review of apps providing step-by-step instructions for cardiopulmonary resuscitation (CPR) in case of a cardiac arrest.

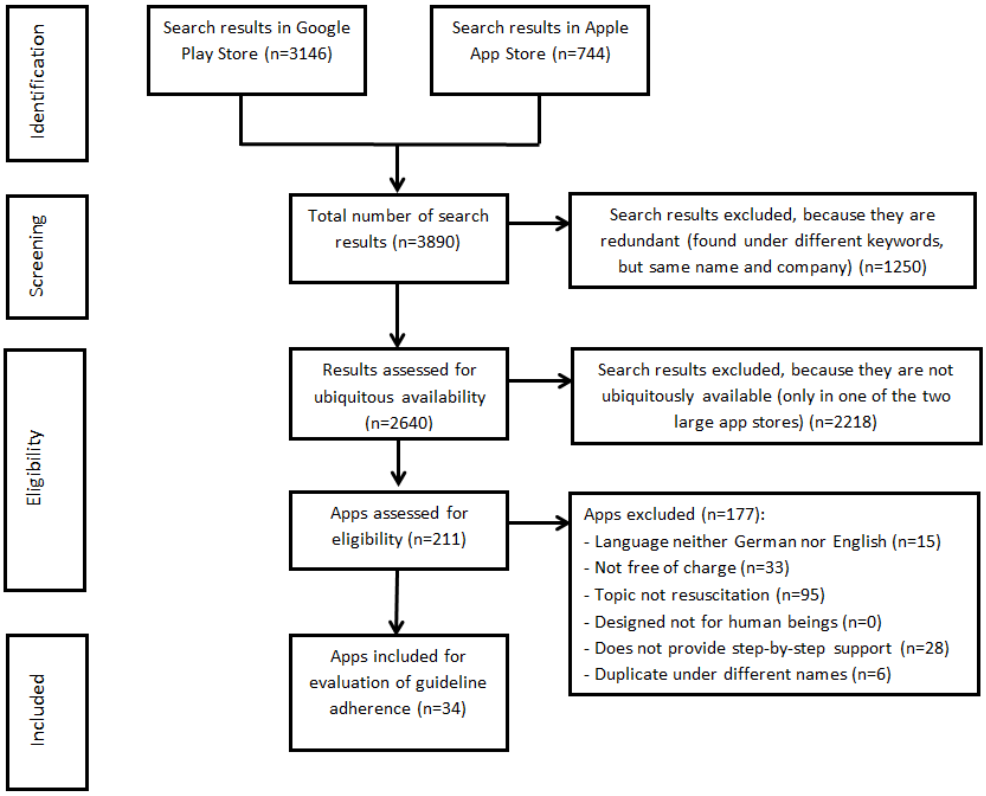


Figure 2. Apps in accordance with the criteria used to evaluate guideline adherence.

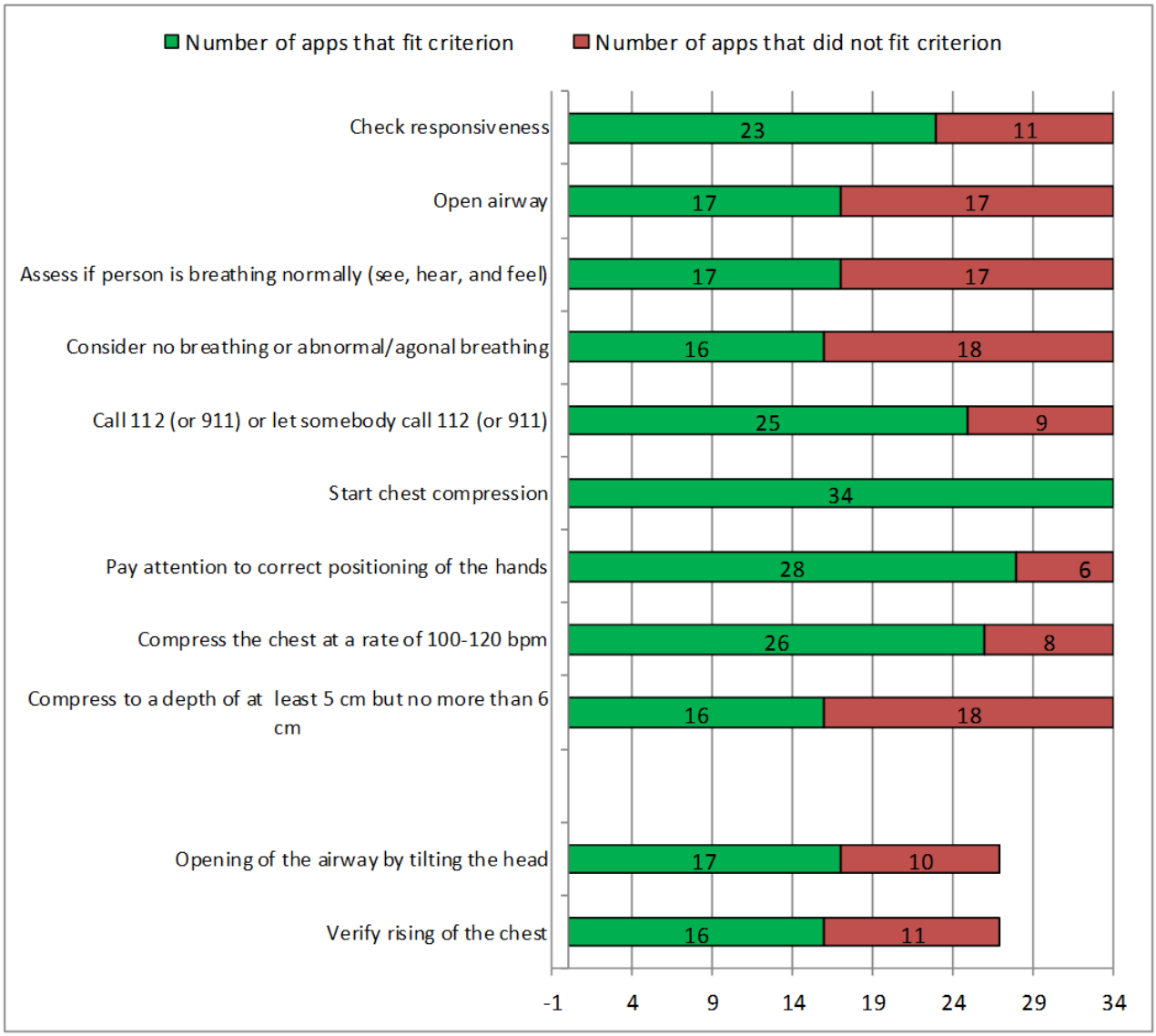


Table 1. Apps that met all 9 of our criteria for guideline adherence.

Name	Company	Google Play Store version number	Apple App Store version number
HELP Notfall	Schweizerische Herzstiftung	1.0	1.0
HAMBURG SCHOCKT	Arbeiter-Samariter-Bund Hamburg	1.5.0	1.5.0
Mein DRK-Die Rotkreuz-App des DRK e.V	Deutsches Rotes Kreuz e.V	2.5.5	2.8.3
St John Wales First Aid	St John Cymru Wales	1.03	Unknown
Notfall-Hilfe	PASS Consulting Group	4.0	3.9.3

The median SUS of “HELP Notfall” was significantly higher than “HAMBURG SCHOCKT” (87.5 vs 55; asymptotic Wilcoxon test: z=−3.63, *P*<.01, n=19) and also significantly higher than “Mein DRK” (87.5 vs 32.5; asymptotic Wilcoxon test: z=−3.83, *P*<.01, n=19). The median SUS of “HAMBURG SCHOCKT” was significantly higher than “Mein DRK” (55 vs 32.5; asymptotic Wilcoxon test: z=−2.81, *P*<.01, n=19). The median SUS scores did not differ significantly between the group of mobile phone frequent users and emergency physicians. The SUS results are shown in Figure 3.

Table 2 depicts how the emergency physicians ranked the apps according to the quality of teaching different aspects of high-quality CPR. Of the 9 emergency physicians, 1 did not complete this part of the questionnaire. The participating emergency physicians rated the app “HELP Notfall” as the one, teaching the majority of relevant aspects (6 out of 7) best. There was no clear result for the aspect “The app emphasizes the importance of complete recoil of the chest after each compression.”

Figure 3. Usability evaluation of the apps with system usability scale (SUS) score. IQR: interquartile range.

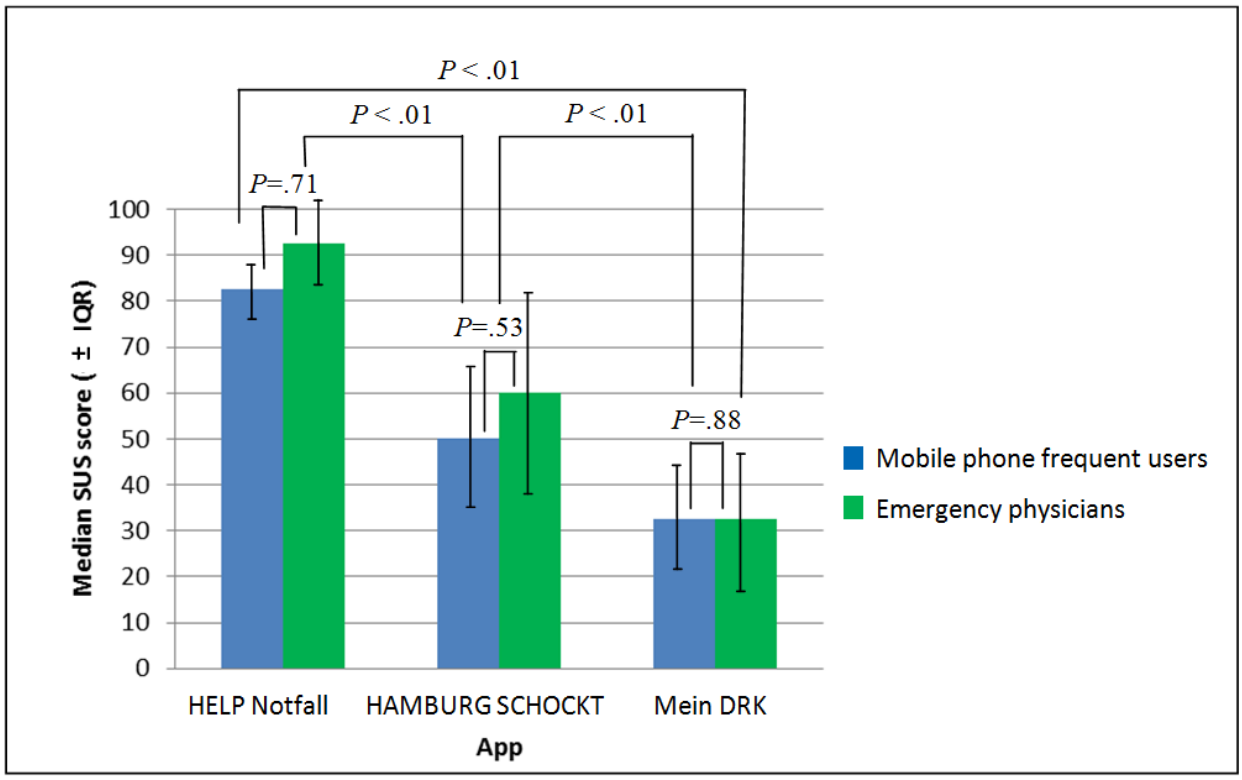


Table 2. Number of emergency physicians rating the app listed as the one teaching a specific aspect of high-quality cardiopulmonary resuscitation best. A total of 8 emergency physicians evaluated the apps.

Specific aspects of high-quality cardiopulmonary resuscitation	HELP Notfall, n (%)	HAMBURG SCHOCKT, n (%)	Mein DRK, n (%)
The app comprehensively explains the opening of the airway	7 (88)	1 (12)	0 (0)
The app points out the problem of agonal breathing	8 (100)	0 (0)	0 (0)
The app emphasizes the importance of complete recoil of the chest after each compression	3 (38)	4 (50)	1 (12)
The app indicates that pauses in chest compressions should be minimized	7 (88)	1 (12)	0 (0)
The app helps the medical layperson to find the correct frequency for chest compression	8 (100)	0 (0)	0 (0)
The design and user interface supports an optimal execution of cardiopulmonary resuscitation	8 (100)	0 (0)	0 (0)
The app requests the medical layperson to continue with cardiopulmonary resuscitation until the arrival of emergency service	8 (100)	0 (0)	0 (0)

Discussion

Principal Findings

The structured search in 2 app stores resulted in 3890 hits. After removing redundant ones, 2640 hits were checked for fulfilling the inclusion criteria. Hereby, 34 apps were identified, meeting all inclusion criteria, of which only 5 (15%, 5/34) fulfilled all defined criteria of adherence to the guidelines of the European Resuscitation Council and the American Heart Association. All other apps gave no or incorrect information on at least one relevant topic. Regarding the usability, only 1 out of the 3 apps was evaluated with an SUS score above the published average of 68 [41].

Systematic Review of Apps

The systematic review of apps available on Google Play Store and Apple App Store took place between May 26, 2017, and June 23, 2017. After excluding redundant hits; duplicates apps; and apps that were not ubiquitously available, free of charge, or did not provide step-by-step instructions in English or German for resuscitation of human beings, 34 apps remained.

Similar to other studies, the search in Google Play Store yielded far more results than the Apple App Store [47]. One reason for this striking difference might be the different submission systems. Although there are no admission requirements in the Google Play Store, Apple tests each submitted app for technical

compatibility and conducts a content verification review [27,48]. However, the main goal of this content verification seems to be to ensure that the name and description of the app match with the content.

Various aspects increase the difficulties for a medical layperson to find a suitable app: the sheer volume of apps to choose from can overwhelm the user [27]. Which app will be downloaded by the user depends on a number of factors such as user ratings, appealing of screenshots, keywords, and number of downloads [49]. Therefore, in the last years, the term "app store optimization" was coined describing strategies to increase the likelihood of an app being downloaded [49,50]. Moreover, the availability of apps differs not only between operators but also between countries, which influences and limits the user's choice.

Adherence to Guidelines

Of the 34 examined apps, only 5 (15%, 5/34) apps fit all of our criteria of guideline-adherent resuscitation. This alarming result is concordant with that of other studies examining the content of mHealth apps [51-54]. The "European Resuscitation Council Guidelines for Resuscitation 2015" [55] and the "American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care 2015" [35] are well known and highly respected international medical guidelines, which are evidence based on current literature. These guidelines offer clear and easy-to-understand advice on which steps should be taken to resuscitate a person. All flowcharts, pictures, and movies are freely available and translated into many different languages [56,57]. Nevertheless, only few apps implemented these recommendations. This might possibly lead to reduced probability of survival of the victim resuscitated.

It has been shown in multiple studies [58-60] that a cardiac arrest victim showing abnormal, ie, agonal breathing has an increased chance of survival compared with cardiac arrest patients suffering from apnea. However, agonal breathing is often misjudged by medical laypersons not realizing the need for performing CPR in these cases. Therefore, it is crucial that CPR apps point out that patients presenting with agonal breathing are also in need of CPR. More than half of the examined apps (53%, 18/34) did not consider this important fact. Furthermore, 50% (17/34) of the apps did not even guide the user to open the airway and 50% (17/34) did not recommend to assess whether the person is breathing (see, hear, and feel). Of the 34 apps studied, 18 (53%, 18/34) recommended no or wrong compression depth, despite numerous studies suggesting to compress to a depth of at least 5 cm but not more than 6 cm to increase the chance of a positive medical outcome [61-65]. Of the 34 apps, 8 (24%, 8/34) apps did not show the correct compression rate of 100-120 bpm that has been proven to improve survival rates [66]. Of the 34 apps, 27 additionally explained ventilation, although the European Resuscitation Council as well as the American Heart Association recommend to teach medical laypersons hands-only CPR [35,55]. The 2 additional criteria for apps explaining ventilation as well were not met by all of these apps. Although this further diminishes the number of apps meeting all criteria, it is not the sole reason for the low adherence rate. A long conformity checklist certainly increases the risk of an app not meeting every single criterion. However, all chosen criteria for guideline adherence are important evidence-based aspects, which are taught in resuscitation courses worldwide.

Determining a global quality management for apps is complicated by different legislations in different countries and multiple concerned governmental agencies (eg, health and privacy legislation) [67,68]. Although institutions and authorities from different states suggested approaches to a quality control of apps, there is no universally accepted procedure [69-73]. However, in our opinion, all efforts to increase the medical accuracy of mHealth products should be made. The aspiration of all persons teaching CPR should be that the general public is provided with correct information on resuscitation, independent on how the information is spread (by textbooks, movie clips, leaflets, or apps).

Usability Evaluation

Of the 3 apps examined with regard to usability, the app "HELP Notfall" had the highest median SUS score (87.5), followed by "HAMBURG SCHOCKT" (55) and "Mein DRK" (32.5). This difference was seen in both evaluating groups (mobile phone frequent users and emergency physicians). There was no significant difference between the evaluating groups.

On the basis of a work by Sauro et al, an SUS score above 68 is rated as a value above average [41]. Such a score was solely achieved by "HELP Notfall." The other 2 apps tested performed below average. The usability of an app is crucial for its implementation and usage. Sauro reported that products with an SUS score higher than 82 have a considerable chance of being recommended to a friend or colleague [41], which was reached by "HELP Notfall." In his famous technology acceptance model, Davis (1986) stated that systems will only be used if they are perceived as useful [74]. If the user can see a clear advantage in comparison with their previous approach, users will utilize an app [75]. In the case of an app designed for the use in a time-critical and extremely challenging situation, it is even more important that the operation is highly intuitive. If a patient is in cardiac arrest and resuscitation becomes necessary, there is no time to first become acquainted with the software. Otherwise, there seems to be a relevant risk of apps not helping the user but leading him astray from doing best for the patient. Thus, a high usability of the app is crucial.

Limitations

The systematic review of apps was conducted in the 2 main app stores Apple App Store and Google Play Store, whereas smaller stores such as Amazon App Store, Windows Store, Samsung Apps, or Blackberry World were not included. The review was completed by only 1 researcher (LS), leaving a possibility of misjudging inclusion or exclusion criteria. However, the criteria were phrased distinctly and clearly to diminish this risk, and 2 other researchers (BM and CM) made spot checks. If an app was available either in Google Play Store or in Apple App Store but not in both stores, it was excluded because a recommendable CPR app should be usable by the majority of the population. This criterion led to the exclusion of 2218 search results, which was the majority of all search hits. Furthermore, we decided to choose "free of charge" as an inclusion criterion to enlarge the

group of possible future users. A study by Lim et al conducted in different countries worldwide showed that the most important factor influencing people in the process of downloading an app was the price of the app and that 57% of users will not download apps they have to pay for [76]. We do not know how many apps explain the topic of resuscitation according to the medical guidelines in a user-friendly way but are not free of charge and have to be purchased or are available in just 1 app store. These inclusion criteria certainly influenced the amount and choice of apps analyzed. Furthermore, the world of mobile phone apps is fast moving with new apps entering the market and other ones vanishing. Hence, a review of apps always reflects availability at a certain time. We conducted the search with keywords in English and German language. We cannot say whether a search in other languages might lead to different results. The search was carried out with the app stores' default country setting of Germany. This certainly further reduced the number of possible apps.

The SUS was evaluated by mobile phone frequent users and emergency physicians. As described in the Methods section, groups at high risk of witnessing a cardiac arrest were not included. This might bias the results.

To date, it is not known whether the use of an app providing step-by-step instructions in a CPR situation increases the rate or quality of bystander resuscitation and leads to higher survival rates among the victims. To broadly recommend the use of such apps, further studies are needed to evaluate positive and negative effects.

Comparison With Prior Work

Only few studies systematically evaluating CPR apps exist [22,47,51-54,77-79]. In contrast to our work, some studies did not have a structured search but evaluated only a representative sample of apps [47,77] or searched only in 1 operating system [51,78]. Only a few previous studies evaluated the adherence of an app to an existing medical guideline [53,54,79]. These 3 studies covered weight loss and pain management. The study of Kalz et al included both resuscitation-teaching apps as well as apps providing guidance in a resuscitation situation in real time [52]. However, teaching a topic in a classroom and giving step-by-step instructions in a real situation are different purposes and call for different designs of the app. We decided to focus only on apps offering real-time support. This is in contrast to the study of Ahn et al, concentrating solely on CPR-training apps [22].

Contrary to all other studies, we did not select a reduced number of apps for the SUS evaluation but did a comprehensive evaluation of all apps that fit the inclusion criteria.

Conclusions

This work combined a systematic review of currently available resuscitation apps with an assessment of guideline adherence and an evaluation of usability. The search resulted in 3890 hits. Of 34 apps that met the inclusion criteria, only 5 (15%, 5/34) fulfilled all of the criteria applied to determine guideline adherence. All other apps gave no or incorrect information on at least one relevant topic. Furthermore, our evaluation of usability revealed that only 1 of the 3 apps tested had an above average usability rate according to SUS. Implementing a systematic quality control for health-related apps should be enforced to ensure medical accuracy and sufficient usability. This is of superior importance for apps focusing on the treatment of life-threatening events such as cardiac arrest.

Acknowledgments

The authors would like to thank Berthold Henkel, Anne Schiller, Stefan Mockler, and Jan Bartels for their ongoing support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The German version of the system usability scale used.

[PDF File (Adobe PDF File), 61KB - mhealth_v6i11e190_app1.pdf]

Multimedia Appendix 2

Apps analyzed for guideline adherence.

[PDF File (Adobe PDF File), 26KB - mhealth_v6i11e190_app2.pdf]

References

1. Sasson C, Rogers MA, Dahl J, Kellermann AL. Predictors of survival from out-of-hospital cardiac arrest: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes 2010 Jan;3(1):63-81 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.109.889576] [Medline: 20123673]
2. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, Adult basic life supportautomated external defibrillation section Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult

basic life support and automated external defibrillation. Resuscitation 2015 Oct;95:81-99. [doi: 10.1016/j.resuscitation.2015.07.015] [Medline: 26477420]
3. Park GJ, Song KJ, Shin SD, Lee KW, Ahn KO, Lee EJ, et al. Timely bystander CPR improves outcomes despite longer EMS times. Am J Emerg Med 2017 Aug;35(8):1049-1055. [doi: 10.1016/j.ajem.2017.02.033] [Medline: 28237384]
4. Krammel M, Schnaubelt S, Weidenauer D, Winnisch M, Steininger M, Eichelter J, et al. Gender and age-specific aspects of awareness and knowledge in basic life support. PLoS One 2018;13(6):e0198918 [FREE Full text] [doi: 10.1371/journal.pone.0198918] [Medline: 29894491]
5. Nord A, Svensson L, Karlsson T, Claesson A, Herlitz J, Nilsson L. Increased survival from out-of-hospital cardiac arrest when off duty medically educated personnel perform CPR compared with laymen. Resuscitation 2017 Nov;120:88-94. [doi: 10.1016/j.resuscitation.2017.08.234] [Medline: 28870719]
6. Fordyce CB. Reduced critical care utilization: another victory for effective bystander interventions in cardiac arrest. Resuscitation 2017 Oct;119:A4-A5. [doi: 10.1016/j.resuscitation.2017.08.009] [Medline: 28818522]
7. Riddersholm S, Kragholm K, Mortensen RN, Pape M, Hansen CM, Lippert FK, et al. Association of bystander interventions and hospital length of stay and admission to intensive care unit in out-of-hospital cardiac arrest survivors. Resuscitation 2017 Oct;119:99-106. [doi: 10.1016/j.resuscitation.2017.07.014] [Medline: 28716539]
8. Hasselqvist-Ax I, Riva G, Herlitz J, Rosenqvist M, Hollenberg J, Nordberg P, et al. Early cardiopulmonary resuscitation in out-of-hospital cardiac arrest. N Engl J Med 2015 Jun 11;372(24):2307-2315. [doi: 10.1056/NEJMoa1405796] [Medline: 26061835]
9. Malta HC, Kragholm K, Pearson DA, Tyson C, Monk L, Myers B, et al. Association of bystander and first-responder intervention with survival after out-of-hospital cardiac arrest in North Carolina, 2010-2013. J Am Med Assoc 2015 Jul 21;314(3):255-264. [doi: 10.1001/jama.2015.7938] [Medline: 26197186]
10. Brinkrolf P, Bohn A, Lukas RP, Heyse M, Dierschke T, Van Aken HK, et al. Senior citizens as rescuers: is reduced knowledge the reason for omitted lay-resuscitation-attempts? Results from a representative survey with 2004 interviews. PLoS One 2017;12(6):e0178938 [FREE Full text] [doi: 10.1371/journal.pone.0178938] [Medline: 28604793]
11. Starks MA, Schmicker RH, Peterson ED, May S, Buick JE, Kudenchuk PJ, Resuscitation Outcomes Consortium (ROC). Association of neighborhood demographics with out-of-hospital cardiac arrest treatment and outcomes: where you live may matter. J Am Med Assoc Cardiol 2017 Oct 1;2(10):1110-1118. [doi: 10.1001/jamacardio.2017.2671] [Medline: 28854308]
12. Axelsson C, Herlitz J, Ekström L, Holmberg S. Bystander-initiated cardiopulmonary resuscitation out-of-hospital. A first description of the bystanders and their experiences. Resuscitation 1996 Nov;33(1):3-11. [Medline: 8959767]
13. Özbilgin Ş, Akan M, Hancı V, Aygün C, Kuvaki B. Evaluation of public awareness, knowledge and attitudes about cardiopulmonary resuscitation: report of İzmir. Turk J Anaesthesiol Reanim 2015 Dec;43(6):396-405. [Medline: 27366536]
14. Marshall S. The use of cognitive aids during emergencies in anesthesia: a review of the literature. Anesth Analg 2013 Nov;117(5):1162-1171. [doi: 10.1213/ANE.0b013e31829c397b] [Medline: 24029855]
15. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3. [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
16. Catharine RB. Educational use of smart phone technology: a survey of mobile phone application use by undergraduate university students. Program 2013 Sep 23;47(4):424-436. [doi: 10.1108/PROG-01-2013-0003]
17. Prensky M. Digital natives, digital immigrants part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
18. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth enhanced chronic care model: a theory derivation approach. J Med Internet Res 2015;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
19. Wiederhold BK. mHealth apps empower individuals. Cyberpsychol Behav Soc Netw 2015 Aug;18(8):429-430. [doi: 10.1089/cyber.2015.29006.bkw] [Medline: 26252926]
20. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
21. Gálvez JA, Lockman JL, Schleelein LE, Simpao AF, Ahumada LM, Wolf BA, et al. Interactive pediatric emergency checklists to the palm of your hand-How the Pedi Crisis app traveled around the world. Paediatr Anaesth 2017 Aug;27(8):835-840. [doi: 10.1111/pan.13173] [Medline: 28593682]
22. Ahn C, Cho Y, Oh J, Song Y, Lim TH, Kang H, et al. Evaluation of smartphone applications for cardiopulmonary resuscitation training in South Korea. Biomed Res Int 2016;2016:6418710 [FREE Full text] [doi: 10.1155/2016/6418710] [Medline: 27668257]
23. van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013 Feb 11;13:23 [FREE Full text] [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]
24. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]

25. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
26. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
27. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
28. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
29. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
30. Müller RM, Kijl B, Martens JK. A comparison of inter-organizational business models of mobile app stores: there is more than open vs. closed. J Theor Appl Electron Commer Res 2011;6(2):13-14. [doi: 10.4067/S0718-18762011000200007]
31. Jansen S, Bloemendal E. Defining app stores: the role of curated marketplaces in software ecosystems. Springer 2013:195-206. [doi: 10.1007/978-3-642-39336-5_19]
32. Igeeksblog. How to Change App Store Country Region in iPhone or iPad! URL: https://www.igeeksblog.com/how-to-change-app-store-country-region-in-iphone-or-ipad/ [accessed 2018-04-18] [WebCite Cache ID 6ylt2A6xX]
33. Thevpn.guru. How to Change Google Play Store Region/Country URL: https://thevpn.guru/change-google-play-store-region-country-vpn-help [accessed 2018-04-18] [WebCite Cache ID 6yltLQUdK]
34. Apple. MacBook Pro URL: https://www.apple.com/macbook-pro/ [accessed 2017-12-08] [WebCite Cache ID 6vYIOC5ti]
35. Kleinman ME, Brennan EE, Goldberger ZD, Swor RA, Terry M, Bobrow BJ, et al. Part 5: adult basic life support and cardiopulmonary resuscitation quality: 2015 American Heart Association Guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2015 Nov 3;132(18 Suppl 2):S414-S435 [FREE Full text] [doi: 10.1161/CIR.0000000000000259] [Medline: 26472993]
36. Perkins GD, Travers AH, Berg RA, Castren M, Considine J, Escalante R, Basic Life Support Chapter Collaborators. Part 3: adult basic life support and automated external defibrillation: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. Resuscitation 2015 Oct;95:e43-e69. [doi: 10.1016/j.resuscitation.2015.07.041] [Medline: 26477428]
37. Travers AH, Perkins GD, Berg RA, Castren M, Considine J, Escalante R, Basic Life Support Chapter Collaborators. Part 3: adult basic life support and automated external defibrillation: 2015 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Circulation 2015 Oct 20;132(16 Suppl 1):S51-S83 [FREE Full text] [doi: 10.1161/CIR.0000000000000272] [Medline: 26472859]
38. Jordan PW, Thomas B, McClelland IL, Weerdmeester B. Usability Evaluation in Industry. London: Taylor & Francis, Inc; 1996.
39. Bevan N. Quality and usability: a new framework. Usability Services 1997:25-34 [FREE Full text]
40. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
41. Sauro J. A Practical Guide to the System Usability Scale. Denver, CO: CreateSpace Independent Publishing Platform; 2011.
42. Brooke J. SUS: a retrospective. J Usability Stud 2013;8:29-40 [FREE Full text]
43. Sauro J, Lewis JR. When designing usability questionnaires, does it hurt to be positive? 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7, 2011; Vancouver p. 2215-2224. [doi: 10.1145/1978942.1979266]
44. Barnum CM. Usability Testing Essentials. Burlington: Elsevier; 2010.
45. Szczerbinski S, Ratajczak J, Lach P, Rzeszuto J, Paciorek P, Karlowska-Pik J, et al. Epidemiology and chronobiology of out-of-hospital cardiac arrest in a subpopulation of southern Poland: a two-year observation. Cardiol J 2018 Apr 3:1174 [FREE Full text] [doi: 10.5603/CJ.a2018.0025] [Medline: 29611174]
46. Pewinternet. Nearly half of American adults are smartphone owners URL: http://www.pewinternet.org/2012/03/01/nearly-half-of-american-adults-are-smartphone-owners/ [accessed 2018-10-10] [WebCite Cache ID 7347QUdGh]
47. Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. Pediatr Diabetes 2013 Jun;14(4):231-238 [FREE Full text] [doi: 10.1111/pedi.12034] [Medline: 23627878]
48. BinDhim N, Trevena L. There's an app for that: a guide for healthcare practitioners and researchers on smartphone technology. Online J Public Health Inform 2015;7(2):e218 [FREE Full text] [doi: 10.5210/ojphi.v7i2.5522] [Medline: 26392848]
49. Meatti. App Store Optimization Tips & Checklist URL: https://www.meatti.com/blog/app-store-optimization-tips/ [accessed 2018-04-16] [WebCite Cache ID 6yikcTXQb]

50. En.wikipedia. App store optimization URL: https://en.wikipedia.org/wiki/App_store_optimization [accessed 2018-04-16] [WebCite Cache ID 6yimeQjns]
51. Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
52. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014;16(3):e89 [FREE Full text] [doi: 10.2196/jmir.2951] [Medline: 24647361]
53. Alnasser AA, Amalraj RE, Sathiaseelan A, Al-Khalifa AS, Marais D. Do Arabic weight-loss apps adhere to evidence-informed practices? Transl Behav Med 2016 Dec;6(3):396-402 [FREE Full text] [doi: 10.1007/s13142-015-0347-7] [Medline: 27528528]
54. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
55. Monsieurs KG, Nolan JP, Bossaert LL, Greif R, Maconochie IK, Nikolaou NI, ERC Guidelines 2015 Writing Group. European Resuscitation Council Guidelines for Resuscitation 2015: Section 1. Executive summary. Resuscitation 2015 Oct;95:1-80. [doi: 10.1016/j.resuscitation.2015.07.038] [Medline: 26477410]
56. Cprguidelines. ERC Guidelines 2015 have arrived! URL: https://cprguidelines.eu/ [accessed 2017-12-07] [WebCite Cache ID 6vWmeoQfE]
57. Eccguidelines.heart. CPR & ECC Guidelines URL: https://eccguidelines.heart.org/index.php/circulation/cpr-ecc-guidelines-2/ [accessed 2018-04-11] [WebCite Cache ID 6yauntmgc]
58. Perkins GD, Walker G, Christensen K, Hulme J, Monsieurs KG. Teaching recognition of agonal breathing improves accuracy of diagnosing cardiac arrest. Resuscitation 2006 Sep;70(3):432-437. [doi: 10.1016/j.resuscitation.2006.01.015] [Medline: 16831503]
59. Perkins GD, Stephenson B, Hulme J, Monsieurs KG. Birmingham assessment of breathing study (BABS). Resuscitation 2005 Jan;64(1):109-113. [doi: 10.1016/j.resuscitation.2004.09.007] [Medline: 15629563]
60. Bobrow BJ, Zuercher M, Ewy GA, Clark L, Chikani V, Donahue D, et al. Gasping during cardiac arrest in humans is frequent and associated with improved survival. Circulation 2008 Dec 9;118(24):2550-2554 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.799940] [Medline: 19029463]
61. Vadeboncoeur T, Stolz U, Panchal A, Silver A, Venuti M, Tobin J, et al. Chest compression depth and survival in out-of-hospital cardiac arrest. Resuscitation 2014 Feb;85(2):182-188. [doi: 10.1016/j.resuscitation.2013.10.002] [Medline: 24125742]
62. Hellevuo H, Sainio M, Nevalainen R, Huhtala H, Olkkola KT, Tenhunen J, et al. Deeper chest compression- more complications for cardiac arrest patients? Resuscitation 2013 Jun;84(6):760-765. [doi: 10.1016/j.resuscitation.2013.02.015] [Medline: 23474390]
63. Stiell IG, Brown SP, Nichol G, Cheskes S, Vaillancourt C, Callaway CW, Resuscitation Outcomes Consortium Investigators. What is the optimal chest compression depth during out-of-hospital cardiac arrest resuscitation of adult patients? Circulation 2014 Nov 25;130(22):1962-1970 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008671] [Medline: 25252721]
64. Hostler D, Everson-Stewart S, Rea TD, Stiell IG, Callaway CW, Kudenchuk PJ, Resuscitation Outcomes Consortium Investigators. Effect of real-time feedback during cardiopulmonary resuscitation outside hospital: prospective, cluster-randomised trial. Br Med J 2011 Feb 4;342:d512 [FREE Full text] [doi: 10.1136/bmj.d512] [Medline: 21296838]
65. Stiell IG, Brown SP, Christenson J, Cheskes S, Nichol G, Powell J, Resuscitation Outcomes Consortium (ROC) Investigators. What is the role of chest compression depth during out-of-hospital cardiac arrest resuscitation? Crit Care Med 2012 Apr;40(4):1192-1198 [FREE Full text] [doi: 10.1097/CCM.0b013e31823bc8bb] [Medline: 22202708]
66. Idris AH, Guffey D, Pepe PE, Brown SP, Brooks SC, Callaway CW, Resuscitation Outcomes Consortium Investigators. Chest compression rates and survival following out-of-hospital cardiac arrest. Crit Care Med 2015 Apr;43(4):840-848. [doi: 10.1097/CCM.0000000000000824] [Medline: 25565457]
67. Yang YT, Silverman RD. Mobile health applications: the patchwork of legal and liability issues suggests strategies to improve oversight. Health Aff (Millwood) 2014 Feb;33(2):222-227. [doi: 10.1377/hlthaff.2013.0958] [Medline: 24493764]
68. Parker L, Karliychuk T, Gillies D, Mintzes B, Raven M, Grundy Q. A health app developer's guide to law and policy: a multi-sector policy analysis. BMC Med Inform Decis Mak 2017 Oct 2;17(1):141 [FREE Full text] [doi: 10.1186/s12911-017-0535-0] [Medline: 28969704]
69. Charani E, Castro-Sánchez E, Moore LS, Holmes A. Do smartphone applications in healthcare require a governance and legal framework? It depends on the application!. BMC Med 2014 Feb 14;12:29 [FREE Full text] [doi: 10.1186/1741-7015-12-29] [Medline: 24524344]
70. Barton AJ. The regulation of mobile health applications. BMC Med 2012 May 8;10:46 [FREE Full text] [doi: 10.1186/1741-7015-10-46] [Medline: 22569114]
71. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]

72. Cortez NG, Cohen IG, Kesselheim AS. FDA regulation of mobile health technologies. N Engl J Med 2014 Jul 24;371(4):372-379. [doi: 10.1056/NEJMhle1403384] [Medline: 25054722]
73. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
74. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
75. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
76. Lim SL, Bentley PJ, Kanakam N, Ishikawa F, Honiden S. Investigating country differences in mobile app user behavior and challenges for software engineering. IIEEE Trans Software Eng 2015 Jan 1;41(1):40-64. [doi: 10.1109/TSE.2014.2360674]
77. Rossi MG, Bigi S. mHealth for diabetes support: a systematic review of apps available on the Italian market. Mhealth 2017;3:16 [FREE Full text] [doi: 10.21037/mhealth.2017.04.06] [Medline: 28567412]
78. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. J Telemed Telecare 2012 Jun;18(4):235-238. [doi: 10.1258/jtt.2012.111002] [Medline: 22604278]
79. 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]

Abbreviations

CPR: cardiopulmonary resuscitation
PRISMA: preferred reporting items for systematic reviews and meta-analyses
SUS: system usability scale

Edited by G Eysenbach; submitted 13.12.17; peer-reviewed by J Creutzfeldt, M Leary, S Nimbalkar; comments to author 20.03.18; revised version received 15.05.18; accepted 09.08.18; published 06.11.18.

Please cite as:
Metelmann B, Metelmann C, Schuffert L, Hahnenkamp K, Brinkrolf P
Medical Correctness and User Friendliness of Available Apps for Cardiopulmonary Resuscitation: Systematic Search Combined With Guideline Adherence and Usability Evaluation
JMIR Mhealth Uhealth 2018;6(11):e190
URL: http://mhealth.jmir.org/2018/11/e190/
doi:10.2196/mhealth.9651
PMID:30401673

Review

Mobile Phone Apps for Low-Income Participants in a Public Health Nutrition Program for Women, Infants, and Children (WIC): Review and Analysis of Features

Summer J Weber[1], PhD, RD; Daniela Dawson[2], BS; Haley Greene[3], BS; Pamela C Hull[1], PhD

[1]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Leighton School of Nursing, Marian University, Nashville, TN, United States

[3]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Pamela C Hull, PhD
Division of Epidemiology
Department of Medicine
Vanderbilt University Medical Center
2525 West End Avenue, Suite 800
Nashville, TN,
United States
Phone: 1 615 936 3241
Email: pam.hull@vanderbilt.edu

Abstract

Background: Since 1972, the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) has been proven to improve the health of participating low-income women and children in the United States. Despite positive nutritional outcomes associated with WIC, the program needs updated tools to help future generations. Improving technology in federal nutrition programs is crucial for keeping nutrition resources accessible and easy for low-income families to use.

Objective: This review aimed to analyze the main features of publicly available mobile phone apps for WIC participants.

Methods: Keyword searches were performed in the app stores for the 2 most commonly used mobile phone operating systems between December 2017 and June 2018. Apps were included if they were relevant to WIC and excluded if the target users were not WIC participants. App features were reviewed and classified according to type and function. User reviews from the app stores were examined, including ratings and categorization of user review comments.

Results: A total of 17 apps met selection criteria. Most apps (n=12) contained features that required verified access available only to WIC participants. Apps features were classified into categories: (1) shopping management (eg, finding and redeeming food benefits), (2) clinic appointment management (eg, appointment reminders and scheduling), (3) informational resources (eg, recipes, general food list, tips about how to use WIC, links to other resources), (4) WIC-required nutrition education modules, and (5) other user input. Positive user reviews indicated that apps with shopping management features were very useful.

Conclusions: WIC apps are becoming increasingly prevalent, especially in states that have implemented electronic benefits transfer for WIC. This review offers new contributions to the literature and practice, as practitioners, software developers, and health researchers seek to improve and expand technology in the program.

(JMIR Mhealth Uhealth 2018;6(11):e12261) doi:10.2196/12261

KEYWORDS

WIC; low-income; mobile phone; mHealth

Introduction

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) provides nutritious foods, nutrition education, and health care referrals to low-income pregnant, postpartum, and breastfeeding women and children from birth until their 5th birthday in the United States. Since its inception in 1972, WIC has been associated with improving health outcomes of participants, including but not limited to positive impacts on nutrient and dietary intakes [1,2], childhood obesity rates [3], household food security [4], immunization rates [5],

and birth outcomes [6]. As one of the only federal nutrition programs designed to monitor the health of participants, WIC has been a fundamental contributor toward research surrounding maternal and child health for decades and has been proven to be one of the most successful and cost-effective nutrition intervention programs in the United States [7].

The benefits of WIC participation are clear; however, participation in WIC has declined since 2010 [8], and the program is in need of updated tools to help future generations. Barriers to participating in WIC exist, especially when shopping for WIC foods at the store. Participants often face difficulty finding the correct foods on their family's WIC food list and problems with redeeming benefits during checkout [9,10]. Clinic appointment wait times and scheduling issues [9,11] have also been reported as program barriers. To receive food benefits and remain enrolled, WIC participants are required to complete periodic nutrition education. Some participants face difficulties attending these classes in person [9]. Many WIC agencies have attempted to reduce this burden by allowing participants to complete some nutrition education requirements online. As of 2016, 34 states and 5 intertribal councils used some form of online nutrition education [12].

Electronic benefits transfer (EBT) for WIC is currently being implemented in several states and is mandated for all states by 2020 [13]. EBT is a mechanism for using WIC benefits like a debit card to make WIC transactions less cumbersome for participants and vendors alike [14]. By eliminating the need for paper vouchers, participants no longer need to redeem all benefits on a voucher in a single transaction, wait for the cashier to verify each food, and sign each voucher like a check. This paper voucher process has been known to cause holdups at checkout, which adds stigma and embarrassment [11].

Although EBT has been found to improve redemption and ease at checkout [15], it has not been associated with an increase in enrollment [16] and cannot alleviate all the problems associated with shopping for and choosing the correct WIC foods. Mobile phone apps for WIC, therefore, are becoming increasingly prevalent and necessary to assist participants with the WIC shopping experience [17,18]. Several WIC apps have been designed to help participants navigate the program and its benefits; however, these apps offer a variety of features, and few have been assessed in the scientific literature [17,18]. As part of our team's efforts to develop a new app for the Tennessee WIC program called Children Eating Well (CHEW), we conducted a review of existing WIC apps. Specifically, this review aimed to identify publicly available apps for WIC participants, analyze their main features, and examine user review ratings and comments. This review will be useful for practitioners, software developers, and researchers to inform them about existing tools for WIC and guide planning for the incorporation of new technologies to improve the program.

Methods

Keyword searches were performed between December 2017 and June 2018 in the largest online stores for the two most commonly used operating systems: the App Store for Apple iOS devices and Google Play for Android devices. Keywords used in the search were my WIC, WIC, and WIC app. The inclusion criteria for apps was being related in some capacity to the WIC program. The exclusion criteria were apps that were not targeted for use to program participants (eg, designed for vendors). Two coauthors screened the apps for eligibility.

The range of number of installations was documented for apps found in Google Play. The App Store, however, does not list the number of installations. The app rating and number of reviews were documented for each app for each store, if available. Each store displays a rating out of 5 possible points. The App Store does not provide a user rating if fewer than 5 users have submitted a review. The states where each app was available for use was also documented.

To review the features of the included apps, publicly available information was gathered from the App Store and Google Play store via preview images, descriptions, user ratings, and review comments. In addition, each app was installed on both iOS and Android devices to view features when possible. Two coauthors reviewed the apps and extracted data, and any differences were resolved through discussion with a third coauthor. Most WIC apps required user verification as an enrolled participant in the respective state to fully access the features. Therefore, it was not possible to test usability of the apps.

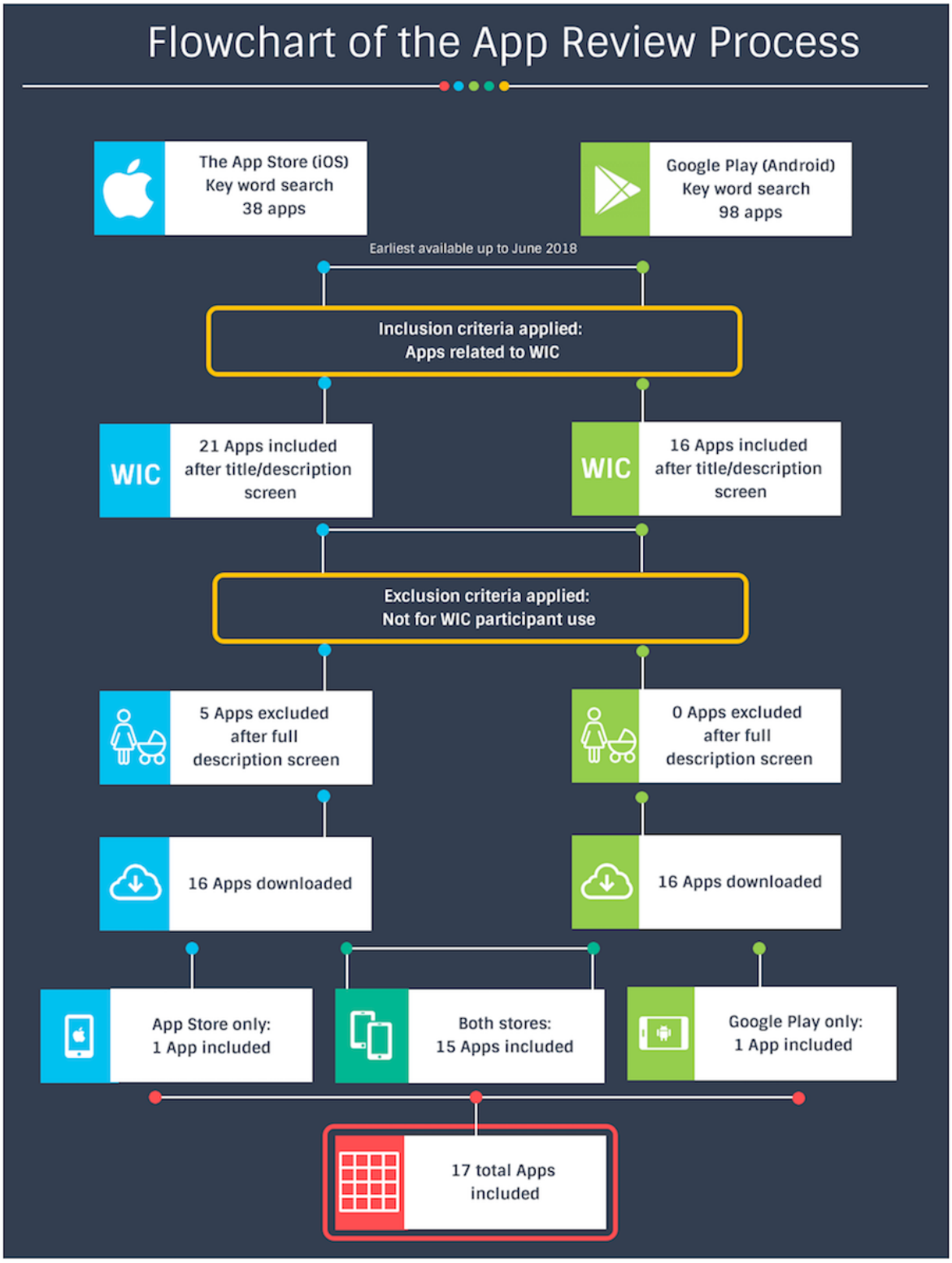
App features were then organized into major categories. To our knowledge, this is the first review of WIC-related mobile phone apps and features. Determining the major categories, therefore, was an inductive process supported by the authors' familiarity and experience with the WIC program at the participant, clinic, vendor, and program administration levels.

Results

Search Results

A total of 38 apps were found from the keyword search within the App Store, 21 of which were related to WIC (see Figure 1). Five WIC apps were excluded from the App Store results because they were not targeted for users who are WIC participants; instead, they targeted WIC vendors and agents to view and manage eligible products. In the Google Play store, 98 apps were found from the keyword search, 16 of which were related to WIC. Fifteen WIC apps were duplicates in both online app stores, while 1 app (SAC WIC) was available for iOS only and 1 app (AZ WIC Clinic Search) was available for Android only. This resulted in a final total of 17 apps targeted toward WIC participants included in the review. Two apps, SAC WIC and AZ WIC Clinic Search, however, were removed from the online stores during the time of this review.

Figure 1. Flowchart of the app review process.



States and User Ratings

Although WIC is a federal program used nationwide, it is administered at the state and local level. Therefore, WIC apps were designed for specified states, territories, or tribal nations. Currently 30 states and Washington DC, 3 US territories (American Samoa, Guam, and the Commonwealth of Northern Mariana Islands), and 3 tribal nations have WIC-related apps. Several states (n=6) had more than one WIC-related app; however, these did not overlap in content or features (eg, nutrition education vs shopping). All apps were free to download; however, most apps (12/17, 70%) required the user to register with the app to verify WIC participation so the user could view and use features specific to his or her own WIC ID number (eg, food benefit balance or appointment times). Some apps (5/17, 29%) contained features that could be readily accessed by any public user, regardless of WIC enrollment or residence in the state for which it was developed. User

verification requirements for each app are displayed in Table 1.

WIC apps and their corresponding states, ratings, number of ratings, and range of installations (for Android only) are also displayed in Table 1. A total of 41% (7/17) of WIC apps fell into the 10,000+ downloads category in Google Play. The most frequently installed (500,000+ downloads in Google Play) WIC-related app was WIC Shopper. WIC Shopper's relatively high install rate may be due to the fact that it was used by 18 states and agencies, all of which have begun to implement EBT for WIC. Only 2 other apps were being used by multiple states and agencies (WIC Smart and EzWIC). User ratings varied widely for both iOS and Android apps, and many were not available due to a lack of ratings. However, for those that had available data, Maryland WIC received the lowest rating in both stores (3.0 and 2.2, respectively).

Most apps received some negative reviews. According to user review comments in both app stores, many of the low ratings were due to the app crashing or freezing or an inability to log in due to system maintenance. User review commenters who gave low ratings also mentioned experiencing issues with registering their accounts, password reset, internet connectivity issues inside the grocery store, or using the app in a state for which the app was not designed to work. App developer responses to these negative reviews typically indicated they were working on resolving these issues or that the user needed to update their app for bug fixes and performance enhancements. For example, one user pointed out a challenge with the log-in process being too cumbersome and time consuming.

> *My only complaint is that the app won't keep the login information saved. I keep having to type all the card info in whenever I want to use it. Not a terrible thing, just really inconvenient when I'm balancing a crying baby and want to take a quick look while grocery shopping.* [Wisconsin MyWIC (iOS version) user review, 4 out of 5 stars]

Women, Infants, and Children Program App Features

App features fell into 5 main categories: shopping management features, clinic appointment management features, informational resources, WIC-required nutrition education modules, and other user input. Specific feature types under these main categories varied across the apps and are summarized in Multimedia Appendix 1.

Shopping Management Features

Most of the WIC apps requiring user verification (11/12, 92%) offered features to assist the WIC participant with real-time shopping for WIC foods in the grocery store using EBT. These shopping management features included real-time food benefit balance checking (10/17, 59%), barcode scanning (10/17, 59%) and manual PLU/UPC (price look-up and universal product code) entry (3/17, 18%) to check if an item was WIC eligible, benefit expiration reminders (9/17, 53%), future benefit viewing (5/17, 29%), and transaction dispute functions for if the user disagreed or did not understand why an item could not be redeemed with WIC (1/17, 6%). Seven apps also included shopping management features that allowed participants to view and navigate to the locations of WIC-approved vendors.

> *This is a great app. I like how everything is going paperless, like I do not have to keep the balance receipts. I can check on the app and scan the items to see if they are approved [and] also find stores that take WIC near me.* [Wisconsin MyWIC (Android version) user, 5 out of 5 stars]

Negative user reviews regarding shopping management features included comments about an app's barcode scanner not working or not correctly determining WIC items. According to positive user reviews, benefit balance checking and bar code scanning features were highly valuable to users trying to complete their WIC shopping.

> *It's time saving. You don't have to stay for long at the cashier anymore!* [EzWIC (Android version) user, 5 out of 5 stars]
>
> *I love this app! It makes using WIC so much easier. Instead of getting to the register and realizing I picked the wrong thing, I can scan it right there and check it. It lets me know exactly how much I have left of each item. And I can look up a list of what I can buy. It's so much nicer than having to carry around the bright orange folder and stop in the middle of the store and search and search to see if I was buying the right thing. I love that I can see what I have left and that it updates immediately...* [WIC Shopper (Android version) user, 5 out of 5 stars]
>
> *Everything about this app is awesome! I like the fact that I don't have to carry paper around. Lets you be discreet about your business.* [Wisconsin MyWIC (Android version) user, 5 out of 5 stars]
>
> *Very helpful being a busy mom with young children to just check the app to see what I have left.* [Bnft (Android version) user, 5 out of 5 stars]

Only one app, Bnft, allowed users to not only manage their WIC benefits but also their Supplemental Nutrition Assistance Program (SNAP, formerly known as Food Stamps) benefits.

Apps with features that enabled benefit balance viewing accessed this information by syncing with the EBT vendor contracted through each state. If EBT access was not provided to the app developer in specific states, users could not access this feature, even if it existed in other states.

> *I like the version of this app for other states but not for Texas. I can't see my benefits; I have to upload a photo of my WIC receipt... [WIC Shopper (Android version) user review, 2 out of 5 stars] Reply from JPMA, Inc: Unfortunately, [that] is a limitation of Texas WIC and not the app. If they could support card registration, we surely would add it.*

Table 1. Mobile phone apps for Women, Infants, and Children programs available for public download.

App name (developer)	State or WIC[a] agency	Range of installations (Android only)	App rating out of 5 (Android, iOS)	Number of ratings (Android, iOS)	Requires user verification to access features
Bnft (Soltran, Inc)[b]	North Carolina	10,000+	4.4, 3.6	18, 9	x
EzWIC (Arizona Department of Health Services)[b]	Arizona, American Samoa, CNMI[c], Guam, Navajo Nation	50,000+	4.0, 4.8	69, 20	x
Indiana WIC (Indiana Office of Technology)	Indiana	10,000+	3.4, 4.8	112, 19	x
Maryland WIC (3 Sigma Software Inc)	Maryland	10,000+	3.0, 2.3	70, 29	x
My Minnesota WIC App (Minnesota Development Team)[d]	Minnesota	10,000+	4.2, —	29, —	x
My Oklahoma WIC (My WIC Development Team)	Oklahoma	1000+	5.0, —	1, —	x
MyWIC (Mobile Benefits Inc)	Chickasaw Nation	100+	Not available; beta version	Not available; beta version	x
WIC Connect (State of Michigan)	Michigan	10,000+	3.9, —	11, —	x
WIC Shopper (JPMA)	Connecticut, Colorado, Florida, Iowa, ITC[e] Arizona, Kansas, Kentucky, Massechusetts, Montana, Nevada, New Mexico, Oregon, Texas, Vermont, West Virginia, Washington DC, Wyoming	500,000+	3.9, 4.6	2966, 442	x
WICSmart[f] (JPMA)	Arkansas, Choctaw Nation, Connecticut, ITC Arizona, Massechusetts, Montana, Rhode Island, West Virginia, Pennsylvania, Washington DC	10,000+	3.3, —	26, —	x
WIC2Go (3 Sigma Software Inc)	New York	500+	1.0, 4.7	1, 7	x
Wisconsin MyWIC (Wisconsin Department of Health Services)	Wisconsin	10,000+	4.7, 2.4	97, 19	x
Alabama WIC (OCV, LLC)	Alabama	5000+	4.6, —	19, —	
Arizona WIC Clinic Search (Arizona Department of Health Services)	Arizona	No longer publicly available	No longer publicly available	No longer publicly available	
SAC WIC (Social Interest Solutions)	Sacramento County, California	No longer publicly available	No longer publicly available	No longer publicly available	
WIC Food Shopping Guide (WYWICAPP)	Wyoming	1000+	4.3, —	6, —	
WIC San Diego (SDSU WIC)	San Diego County, California	500+	4.5, —	6, —	

[a]WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

[b]Contains shopping features only.

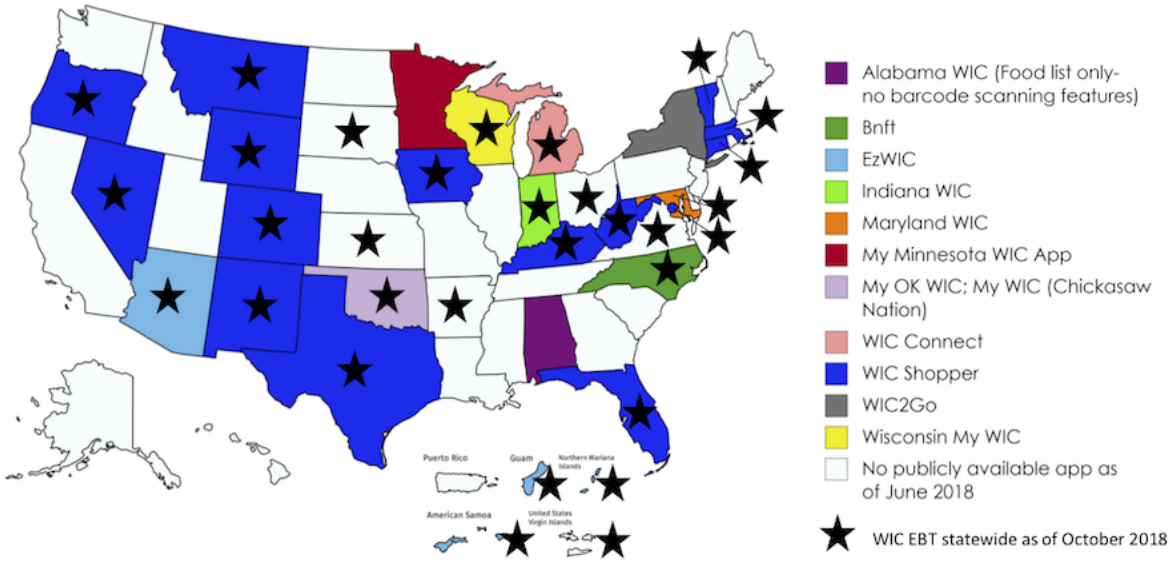
[c]Commonwealth of the Northern Mariana Islands.

[d]Requires user verification for appointment reminder features only.

[e]Intertribal Council.

[f]Contains WIC-required nutrition education features only.

Figure 2. States and territories with shopping apps and electronic benefits transfer for Women, Infants, and Children program participants.



As of June 2018, 24 states had not yet fully implemented EBT for WIC [13] and were still using paper-based vouchers; therefore, apps with shopping management features were not available to participants in most non-EBT states. Minnesota, however, is the only one of these paper-based voucher–based states with a shopping app available for public download with features including barcode scanning and PLU/UPC entry to determine WIC eligible items but without the ability to verify if it is available on a specific user's balance. Like the apps from other states that do not share EBT information with the app developer, My Minnesota WIC App is unable to display participants' own real-time benefit balance since the state does not manage benefits electronically. Figure 2 shows available WIC apps with shopping management features in the United States.

Clinic Appointment Management Features

Most WIC apps requiring user verification (9/12, 75%) also included at least one feature to help participants manage and navigate their WIC clinic visits. Eight apps had features that allowed participant users to view their future appointment date and provided a list of required documents needed for the upcoming appointment. In 3 of these apps, users could also receive notifications for appointment reminders based on information that is synced with the clinic. Alabama WIC also had an appointment reminder feature. However, this app requires no user verification, and the user must manually enter the information about their visit to set up a reminder, similar to a standard reminder feature included on mobile phones; therefore, this feature was not counted as a clinic management feature for the app. User review comments for several apps indicated desire for an app feature that would allow appointment scheduling and viewing.

> *I like the app but I would give 5 stars if you were [...] able to change your appointment thru the app...* [Indiana WIC (Android version) user, 3 out of 5 stars]

Although no app in this review included a feature to schedule or reschedule an appointment, Michigan's WIC Connect had a feature that allowed participants to submit a request to the WIC office for a specific date and time for their next appointment, and the WIC office would contact them to schedule the appointment. Michigan's app also allowed participants to update their contact information (address, phone number, email address), which could potentially assist WIC staff in contacting the participant for new or missed appointments.

Several apps (5/17, 29%) provided clinic locations with contact information and hours. AZ WIC Clinic Search was an app that solely provided clinic location features and clinic photos for potential and current participants; however, during the time of this review, EZ WIC Arizona added both vendor and clinic location features to its app, which eliminated the need for Arizona to have 2 separate apps. Thus, the AZ WIC Clinic Search app was removed from the app store. SAC WIC and WIC San Diego did not offer clinic management features; however, these apps did contain Web links to the California WIC Web portal (wic.ca.gov), which allowed verified WIC participants to view their upcoming appointments and voucher balance. User reviews did not provide positive or negative feedback about the clinic management features.

Informational Resource Features

General information resources were provided in 4 of the 5 WIC apps that did not require user verification and three-fourths (8/12, 67%) of the apps requiring user verification. The most common (10/17, 59%) was enabling participants to view a general WIC food list or simply providing a link to a webpage that contained their state's general WIC food list. Other informational resource features included external links to videos or social media (6/17, 35%), breastfeeding resources (5/17, 29%), WIC eligibility information (3/17, 18%), community resources (2/17, 12%), and nonrequired nutrition education (2/17, 12%). Examples of nonrequired nutrition education

included WIC tips (eg, tips on how to use WIC efficiently, general nutrition, and food preparation) and recipes with WIC foods. User reviews did not provide positive or negative feedback about these informational resource features. Users of apps that did not contain these features, however, did state that they would enjoy them.

> *I like the app but I would give 5 stars if you were able to add [...] a nutrition/ healthy eating section with recipes and ideas to help people learn to eat better with healthy choices not just the WIC-approved foods but any healthy option.* [Indiana WIC (Android version) user, 3 out of 5 stars]

Required Nutrition Education Modules as App Features

Alabama WIC, Indiana WIC, WIC Connect, SAC WIC, and WIC San Diego apps contained Web links to portals where participants could log in and complete required nutrition education modules on their phones instead of going in person to the WIC clinic. Only one app, WIC Smart, was designed to allow participants to complete these modules within the app itself and enabled the corresponding WIC state or agency to tailor these modules. According to comments in user reviews, apps that included features or links to WIC nutrition education requirements were perceived as useful to participants who could access them.

> *Needs updates its locked [...] please fix it asap lots of people use this to save time and gas money please help.* [WICSmart (Android version) user, 3 out of 5 stars]

Other User Input Features

SAC WIC and WIC San Diego were the only 2 that contained features that allowed any (nonverified) user to submit feedback about the app to the WIC program through the app. Wyoming's WIC Food Shopping Guide was the only app that contained input features that allowed any (nonverified) user to complete an app usefulness poll or report WIC fraud.

Discussion

Principal Findings

During the time of this review, 17 mobile phone apps for WIC participants existed in 37 states, US territories, and tribal nations. WIC apps that assisted participants with real-time shopping management received the most positive user ratings and reviews. The most common features of these apps included WIC benefit balance checking and barcode scanning features. These advanced features were not available on apps that required no user verification. Two of these apps were removed from the app stores during the time of this review. Although WIC apps varied in user ratings and features, the ongoing development of technology to help low-income families navigate government nutrition assistance programs is promising and continues to grow as EBT for WIC becomes implemented throughout the United States.

Despite the recent expansion of EBT in WIC, EBT technology is not new in federal nutrition programming. SNAP first started implementing EBT for its participants and vendors in 1984, and by 2004, EBT for SNAP was nationwide [19]. The transition from paper to EBT in WIC, however, has been more challenging due to the complex nature of WIC food restrictions. WIC-authorized foods must meet specific nutrition criteria, whereas foods purchased with SNAP do not need to meet any nutrition-related guidelines [20]. Each of the WIC participant eligibility categories is issued a unique food package, or "food prescription," that varies in the specific type and amount of allowed food items based on age and special dietary needs, whereas allotted SNAP benefits vary only by dollar amount based on household size and income level of the recipients.

As families continue to face stigma in using government nutrition assistance programs [21,22], updated technology such as EBT and mobile phone apps for SNAP and WIC could potentially make access to program benefits more discreet and acceptable. Despite the need for technological updates within these programs [23], barriers to innovation exist due to the complex nature of contractual agreements between the private and public sectors. WIC app developers who wish to give users the ability to view their benefit balance in real time must gain access to data from the EBT vendor contracted though the WIC state agency for which the app is designed to work. This process requires permission to access these Web-based services and can potentially be a hurdle for app developers, as was found for the app in Texas.

Web-based resources to help participants navigate WIC are not new to the program; however, existence of these resources is variable depending on the state agency. These tools include state-based webpages by which users can learn about WIC, check benefit balances, determine eligibility, locate clinics/vendors [24], and access nutrition education modules that can be completed online (rather than in person at the clinic) [25,26]. A recent study assessing preferences for Web-based technology apps within WIC has shown that most participants had access to the internet and own cell phones [27]. The study's findings aligned with the current review in that most WIC participants reported that it would be very useful to access EBT balance (6678/8144, 82%), UPC scanning (5782/8144,71%), appointment scheduling (5212/8144,64%), recipes and cooking demos (5130/8144,63%), store locations (5049/8144,62%), and nutrition education (4804/8144,59%) through online technology [27].

Although low-income families' access to technology continues to pose a limitation for Web-based health interventions, the digital divide is not the same as it was 10 or even 5 years ago due to the surge of mobile phone access and app technology [28]. Many low-income families do have internet access but are underconnected, with mobile-only access [29]. According to Pew Research 2018 data, over 90% of 18- to 49-year-olds and two-thirds of all low-income adults owned a mobile phone. Only 45% of the lowest income category (<$30,000) had broadband internet access at home, yet low-income adults were the most likely to report only accessing the internet via a mobile phone (31%) [30]. This helps to highlight the need to provide mobile-based apps and mobile-friendly websites for the WIC population since they may not have a computer with internet at home. As health-related resources continue to move online [29], Healthy People 2020 has outlined objectives to improve internet

access [31]. Programmatic technology developments through mobile phone apps are, therefore, increasingly important for these families trying to use public services for which they qualify.

Limitations

This review includes only those apps for WIC that were publicly available as of June 2018 in app stores for Android and iOS. As EBT for WIC continues to roll out nationwide, more state agencies are projected to either develop or adopt existing apps for participant use. App ratings cannot reflect comprehensive preferences of all users since there is no requirement to rate the apps that are downloaded. Publicly displayed reviews could potentially reflect a high negativity bias, especially for newer apps that are still working out bug fixes. Since users can rate any app they download, negative bias may also have influenced ratings by users who tried to use an app in a nondesignated state. Finally, since most apps required WIC user verification to gain full access to the app functionality, it was not possible to test usability of the apps directly for this review.

Conclusions

Mobile phone apps for families using federal nutrition programs are becoming increasingly prevalent, especially in states that have implemented EBT for WIC. Based on user reviews of the included WIC apps, developers of future app versions may consider including shopping management features that were mentioned as especially useful (eg, benefit balance viewing and barcode scanning) and expanding on clinic management features (eg, appointment scheduling) and nutrition-related informational features (eg, healthy recipe demos) that were suggested by users. Collaboration between WIC state agencies, contracted private sector EBT vendors, app developers, and researchers is necessary to create and evaluate apps that can help low-income families with children access healthy foods and nutrition services.

Acknowledgments

This study was funded by a grant from the United States Department of Agriculture (award number 2017-68001-26352).

Authors' Contributions

SJW designed and conducted the review and wrote and formatted the article. DD and HG assisted with the app search, documentation of findings, and review. PCH contributed to the methods, edited the article, and oversaw the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Classification of features in mobile phone apps for Women, Infants, and Children program participants.

[PNG File, 86KB - mhealth_v6i11e12261_app1.png]

References

1. Siega-Riz AM, Kranz S, Blanchette D, Haines PS, Guilkey DK, Popkin BM. The effect of participation in the WIC program on preschoolers' diets. J Pediatr 2004 Feb;144(2):229-234. [doi: 10.1016/j.jpeds.2003.10.052] [Medline: 14760267]
2. Whaley SE, Ritchie LD, Spector P, Gomez J. Revised WIC food package improves diets of WIC families. J Nutr Educ Behav 2012 May;44(3):204-209. [doi: 10.1016/j.jneb.2011.09.011] [Medline: 22406013]
3. Chiasson MA, Findley SE, Sekhobo JP, Scheinmann R, Edmunds LS, Faly AS, et al. Changing WIC changes what children eat. Obesity (Silver Spring) 2013 Jul;21(7):1423-1429 [FREE Full text] [doi: 10.1002/oby.20295] [Medline: 23703806]
4. Metallinos-Katsaras E, Gorman KS, Wilde P, Kallio J. A longitudinal study of WIC participation on household food insecurity. Matern Child Health J 2011 Jul;15(5):627-633. [doi: 10.1007/s10995-010-0616-5] [Medline: 20455015]
5. Cortese MM, Diaz PS, Samala U, Mennone JZ, Mihalek EF, Matuck MJ, et al. Underimmunization in Chicago children who dropped out of WIC. Am J Prev Med 2004 Jan;26(1):29-33. [Medline: 14700709]
6. Bitler MP, Currie J. Does WIC work? The effects of WIC on pregnancy and birth outcomes. J Policy Anal Manage 2005;24(1):73-91 [FREE Full text] [doi: 10.1002/pam.20070] [Medline: 15584177]
7. US Dept of Agriculture, Food and Nutrition Services. About WIC: how WIC helps URL: https://www.fns.usda.gov/wic/about-wic-how-wic-helps [WebCite Cache ID 73Mc5lKFY]
8. Jocknowitz A, Tiehen L. Participation patterns: an investigation of delayed entry and early exit. URL: https://permanent.access.gpo.gov/websites/www.ers.usda.gov/media/134411/err109.pdf [accessed 2018-10-22] [WebCite Cache ID 71zWTUH3b]
9. Woelfel ML, Abusabha R, Pruzek R, Stratton H, Chen SG, Edmunds LS. Barriers to the use of WIC services. J Am Diet Assoc 2004 May;104(5):736-743. [doi: 10.1016/j.jada.2004.02.028] [Medline: 15127057]
10. Ritchie LD, Whaley SE, Crocker NJ. Satisfaction of California WIC participants with food package changes. J Nutr Educ Behav 2014 May;46(3 Suppl):S71-578. [doi: 10.1016/j.jneb.2014.01.009] [Medline: 24810000]

11. Summary of WIC state agency strategies for increasing child retention.: United States Department of Agriculture, Food and Nutrition Service URL: https://wicworks.fns.usda.gov/sites/default/files/uploads/ChildRetentionStrategiesReport.pdf [accessed 2018-10-22] [WebCite Cache ID 73Mte4rVz]
12. Bensley R. Selecting an online nutrition education program: a comparison of existing sites. URL: http://www.joinwichealth.org/about [accessed 2018-08-27] [WebCite Cache ID 71zXb5Vox]
13. United States Department of Agriculture, Food and Nutrition Service. State agency detail EBT status report URL: https://fns-prod.azureedge.net/sites/default/files/wic/August2018WICEBTDetailStatusReport_.pdf [WebCite Cache ID 71zYdj0WJ]
14. The WIC program: background, trends, and economic issues, 2015 edition. United States: United States Department of Agriculture, Economic Research Service URL: https://www.ers.usda.gov/publications/pub-details/?pubid=43927 [accessed 2018-10-22] [WebCite Cache ID 71zWTUH3b]
15. Oliveira V. The WIC program: background, trends, and economic issues, 2015 Edition. United States: Department of Agriculture, Economic Research Service URL: https://www.ers.usda.gov/webdocs/publications/43925/50999_eib134.pdf?v=0 [accessed 2018-10-22] [WebCite Cache ID 73MtYyGjT]
16. Hanks A, Gunther C, Scharff R. Understanding the impact of EBT on WIC recipient behavior. 2017 Presented at: Proceedings of the Allied Social Sciences Association Annual Meeting; 2017; Chicago.
17. Hull P, Emerson JS, Quirk ME, Canedo JR, Jones JL, Vylegzhanina V, et al. A smartphone app for families with preschool-aged children in a public nutrition program: prototype development and beta-testing. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e102 [FREE Full text] [doi: 10.2196/mhealth.7477] [Medline: 28768611]
18. Biediger-Friedman L, Crixell SH, Silva M, Markides BR, Smith KS. User-centered design of a Texas WIC app: a focus group investigation. Am J Health Behav 2016 Jul;40(4):461-471. [doi: 10.5993/AJHB.40.4.8] [Medline: 27338993]
19. United States Department of Agriculture, Food and Nutrition Service. A short history SNAP URL: https://www.fns.usda.gov/snap/short-history-snap [WebCite Cache ID 73McZFJhR]
20. Pomeranz JL, Chriqui JF. The Supplemental Nutrition Assistance Program: analysis of program administration and food law definitions. Am J Prev Med 2015 Sep;49(3):428-436 [FREE Full text] [doi: 10.1016/j.amepre.2015.02.027] [Medline: 26091926]
21. Liu CH, Liu H. Concerns and structural barriers associated with WIC participation among WIC-eligible women. Public Health Nurs 2016 Dec;33(5):395-402 [FREE Full text] [doi: 10.1111/phn.12259] [Medline: 26956356]
22. Kaiser L. Why do low-income women not use food stamps? Findings from the California Women's Health Survey. Public Health Nutr 2008 Dec;11(12):1288-1295. [doi: 10.1017/S1368980008002528] [Medline: 18507889]
23. Institute of Medicine. Planning a WIC Research Agenda: Workshop Summary. Washington: National Academies Press; 2010.
24. California Department of Public Health. Welcome to California WIC Mobile URL: https://m.wic.ca.gov/ [WebCite Cache ID 71za0i65D]
25. WIChealth. WIC online nutrition education URL: https://www.wichealth.org/ [WebCite Cache ID 720eqejdX]
26. Bensley RJ, Brusk JJ, Anderson JV, Mercer N, Rivas J, Broadbent LN. wichealth.org: impact of a stages of change-based Internet nutrition education program. J Nutr Educ Behav 2006;38(4):222-229. [doi: 10.1016/j.jneb.2006.03.008] [Medline: 16785091]
27. Bensley RJ, Hovis A, Horton KD, Loyo JJ, Bensley KM, Phillips D, et al. Accessibility and preferred use of online Web applications among WIC participants with Internet access. J Nutr Educ Behav 2014;46(3 Suppl):S87-S92. [doi: 10.1016/j.jneb.2014.02.007] [Medline: 24810002]
28. Serrano KJ, Thai CL, Greenberg AJ, Blake KD, Moser RP, Hesse BW. Progress on broadband access to the Internet and use of mobile devices in the United States. Public Health Rep 2017 Dec;132(1):27-31 [FREE Full text] [doi: 10.1177/0033354916679365] [Medline: 28005473]
29. Katz VS, Gonzalez C, Clark K. Digital inequality and developmental trajectories of low-income, immigrant, and minority children. Pediatrics 2017 Nov;140(Suppl 2):S132-S136. [doi: 10.1542/peds.2016-1758R] [Medline: 29093048]
30. Demographics of mobile device ownership and adoption in the United States. Washington: Pew Internet and American Life Project URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-10-22] [WebCite Cache ID 71zafHllw]
31. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Healthy People 2020 URL: https://www.healthypeople.gov/sites/default/files/HP2020_brochure_with_LHI_508_FNL.pdf [WebCite Cache ID 71zaaKaCJ]

Abbreviations

CHEW: Children Eating Well
EBT: electronic benefits transfer
PLU: price look-up
SNAP: Supplemental Nutrition Assistance Program
UPC: universal product code
WIC: Special Supplemental Nutrition Program for Women, Infants, and Children

Edited by G Eysenbach; submitted 20.09.18; peer-reviewed by JC Shin, A Benis; comments to author 12.10.18; revised version received 15.10.18; accepted 15.10.18; published 19.11.18.

Please cite as:
Weber SJ, Dawson D, Greene H, Hull PC
Mobile Phone Apps for Low-Income Participants in a Public Health Nutrition Program for Women, Infants, and Children (WIC): Review and Analysis of Features
JMIR Mhealth Uhealth 2018;6(11):e12261
URL: http://mhealth.jmir.org/2018/11/e12261/
doi:10.2196/12261
PMID:30455172

Original Paper

Description of Cardiological Apps From the German App Store: Semiautomated Retrospective App Store Analysis

Urs-Vito Albrecht[1*], MD, PhD, MPH; Gerd Hasenfuß[2*], MD, PhD; Ute von Jan[1*], Dr rer biol hum

[1]Peter L Reichertz Institute for Medical Informatics, Hannover Medical School, Hannover, Germany

[2]Department of Cardiology and Pneumology, University Medical Center Göttingen, Göttingen, Germany

[*]all authors contributed equally

Corresponding Author:
Urs-Vito Albrecht, MD, PhD, MPH
Peter L Reichertz Institute for Medical Informatics
Hannover Medical School
Carl-Neuberg-Str 1
Hannover, 30625
Germany
Phone: 49 511532 ext 3508
Fax: 49 5115322517
Email: albrecht.urs-vito@mh-hannover.de

Abstract

Background: In the app stores of mobile platforms, consumers are confronted with an enormous number of mobile apps. Over the past few years, considerable research has been undertaken into to identifying, characterizing, and evaluating apps, be it in health-related or other contexts. However, many of these projects are restricted to specific areas of application and offer little flexibility in adapting the applied criteria.

Objective: This paper presents an adaptable method for selecting and characterizing mobile apps listed in a mobile App Store (the Apple App Store). The method is based on filtering processes using predefined criteria, through a semiautomated retrospective App Store analysis (SARASA).

Methods: To illustrate the SARASA process, keyword-based filtering and metadata-based description, review, and ranking steps were applied to a dataset, more specifically, an April 2018 readout of the Medical category of the German App Store, with the aim of obtaining a list of cardiology-related apps.

Results: From the original list of 39,427 apps within the "Medical" category of the App Store on April 14, 2018, 34,382 apps with store descriptions in languages other than German were removed. For the remaining 5045 apps, keywords related to cardiology were applied to filter the output, obtaining a final total of 335 subject-specific apps for further analysis and description.

Conclusions: SARASA provides an easy to use method for applying filtering processes to identify apps matching predefined, formal criteria from app stores. The criteria can be well adapted to the needs of users. Automatic and manual analyses are easily combined when using SARASA. In the future, additional features, such as algorithmic topic analyses, may supplement the process. Although the area of application is currently limited to Apple's App Store, expansion to other stores is planned. The method stands or falls with the transparency of the app store providers and the manufacturers to make relevant meta-information available. It is up to them to liberalize information and restrict censorship to provide clients, customers, and users truly fair circumstances finding their way around the app market.

(JMIR Mhealth Uhealth 2018;6(11):e11753) doi:10.2196/11753

KEYWORDS

mHealth; mobile health; mobile apps; retrospective app store analysis; cardiology

Introduction

Background

Analyses of software repositories predate modern distribution channels for mobile software [1-3]. In the context of mobile apps, the demand for methods to identify, characterize, and evaluate health-related apps has led to considerable research activity in recent years [4], be it in health-related fields or other areas of application. For many of these projects, there is an emphasis on collecting metadata for apps and evaluating and preparing meaningful information for users [5] to help them select apps that meet their needs. In addition to commercial resources, for example, those that contain (preselected) apps evaluated by experts or user communities, independent scientific analyses (based on peer review) can provide valuable information [6]. In the former case, interested parties must rely on the qualifications and thoroughness of the evaluators, whereas in the latter case, the delay between the point in time when the evaluation was performed and the publication of the results—as it is often customary in science—can lead to incongruence between the information published and reality. Evaluations and ratings from official bodies are only infrequently available because of the sheer number of apps that are listed on the stores [7].

However, there is often only limited information about the methodologies used by those performing the analyses. Scientific approaches try to describe and classify apps based on available information [8-10]. For example, lists of apps may be assigned a ranking [1] or apps may be evaluated in terms of their suitability for specific user groups or specific areas of application [11]. There are also approaches aimed at providing developers with information [12], for example, related to measures they could implement to increase the reach of their apps and thus increase their success [13].

Respective analyses are based on various data sources and data types. For example, metadata about an app can be retrieved directly from the stores, for example, using query interfaces provided by the store operators themselves or Web crawlers. It may not only consist of somewhat unstructured store descriptions but also structured information used for organization and management purposes in the stores. Often, factors such as user ratings [14], update frequency, and other attributes [12] are used. Sometimes data from web-based services or other sources available online, for example, search engine results [9] or Twitter posts [15], are also included in the analyses, depending on the context. The evaluations of the recorded data range from simple metadata evaluations to machine learning–based approaches, which rely on the evaluation or classification of apps, requiring time-consuming, prior training.

Objectives

This paper presents a method for identifying and describing health apps based on formal criteria. To illustrate the process, we present an evaluation of our proposed semiautomated retrospective App Store analysis (SARASA) [16,17] applied to the German Apple App Store, using manufacturer-provided app descriptions and other metadata, shown in an exemplary manner for apps related to the field of cardiology. Compared with other approaches, SARASA emphasizes the aggregation and filtering of apps, rather than a possible qualitative evaluation. The latter would constitute additional processing based on appropriate methods, exceeding the scope of the work presented here. This study introduces the filtering methodology for apps listed in the App Store and demonstrates an exemplary descriptive evaluation using cardiology-related apps, listed within the "Medical" store category, for which German app descriptions are available. In the final part, the proposed methods are discussed critically.

Methods

Principles of the Semiautomated Retrospective App Store Analysis

SARASA describes a multistep procedure, consisting of automated extraction and analysis and manual review and assessment processes, which are described in further detail in the following paragraphs. An example of a viable application of SARASA, as shown in the following sections, is a descriptive evaluation of cardiology-related, German-language apps in the "Medical" category of the German storefront of the App Store.

Fundamentally, the SARASA method consists of 7 steps (Figure 1): (1) first, the base data are collected automatically through a total data collection for the desired store categories at a specific time point (step 1); (2) following this, search terms, inclusion and exclusion criteria, and ranking criteria (to which weightings are assigned) are defined, in this example, as mentioned before, for cardiology-related apps (step 2); (3) in the next step, an automated text analysis and filtering is performed to extract a selection of apps from the database. For this purpose, the manufacturer-provided app description texts are used (step 3); (4) the results are then manually validated (step 4) and (5) manually categorized (step 5); (6) following this, the extract is automatically sorted according to the defined ranking criteria (step 6); (7) finally, an app selection with accompanying summary descriptions of the extract is available and may be used for further evaluations as desired, for example, content-based evaluations for assessing app quality (step 7).

Data Collection

Due to the lack of publicly available, comprehensive, and readily accessible app inventory lists, which also provide full access to metadata, such a list was created for Apple's German App Store on April 14, 2018, using specifically developed, R-based scripts (R Version 3.4.4 [18] with the following libraries: rvest [19], httr [20], jsonlite [21], RSQLite [22], DBI [23], and stringr [24]). The starting points for data acquisition were the German Web pages for the 2 chosen App Store categories: "Medical" and "Health & Fitness." This allowed for the collection of information about apps provided for health-related purposes. The initial R-based script was used to read the names and unique app IDs of a total of 103,364 apps in the 2 store categories. The associated meta-information (see Table 1) was acquired in the following 24 hours using a second script, which was based on the "iTunes Search Application Interface (API)," provided by the App Store provider. Results were stored in an SQLite-based database to be perused for later evaluations. Varying slightly

to this, in the evaluation shown here, only data for apps listed in the "Medical" category were used. In addition, following the initial data collection process, automated language detection processes, based on algorithms published by Google [25], were applied to the store description texts to support filtering by language.

Definition of Cardiology-Related Search Terms, Inclusion and Exclusion Criteria

SARASA provides a method for identifying apps related to a desired topic based on various criteria that are applied to the available app data. In addition to utilizing the apps' properties for characterization, for example, based on attributes derived directly from the available metadata, an inclusion or exclusion criterion based on suitable keywords is also possible. To this end, lists of search terms for the desired subject area need to be created before the data can be filtered according to further criteria.

Defining Appropriate Keyword Lists

A key element of the selection process is the list of keywords that ultimately influences the app selection for later (manual) fine-tuning. In the sample run of SARASA as presented here, the definition of cardiology-related keywords was based on a list of terms commonly used in cardiology and cardiology-related areas, established through consensus of the authors. This initial list was extended and validated by means of an iterative procedure. For this purpose, functions of the R-package "wordVectors" [26] were used. These make it possible to identify words frequently associated with given terms. The texts of the store descriptions for all 103,364 apps in the "Medical" and "Health & Fitness" categories that were initially selected were processed as follows: Initially, there was a restriction to apps for which a German-language store description could be confirmed using automated algorithms [25]. This analysis was necessary because otherwise, a correct language classification would not have been possible: For a number of apps, the language of the available store descriptions does not match purported language in the corresponding, manufacturer-provided metadata field that denotes the languages in which the app is provided. Filtering the texts by keywords would thus have failed.

The initial list of search terms was extended on the basis of the existing app description texts. For this purpose, the descriptions of apps with recognized German texts that matched the initial list of search terms were preprocessed. Formatting and punctuation marks, digits not enclosed in a word or acronym, and filling words such as articles, number words, and pronouns were removed, and particularly, frequently combined terms were combined to form so-called N-grams (eg, "hoher Blutdruck" = "high blood pressure" and "externer Defibrillator" = external defibrillator). The app descriptions prepared in this manner were then used to fine tune the keyword lists. It is only during this process that we chose to also include app descriptions assigned to the "Health & Fitness" category of the App Store.

Figure 1. The 7 steps of the semiautomated retrospective app store analysis (SARASA) workflow.

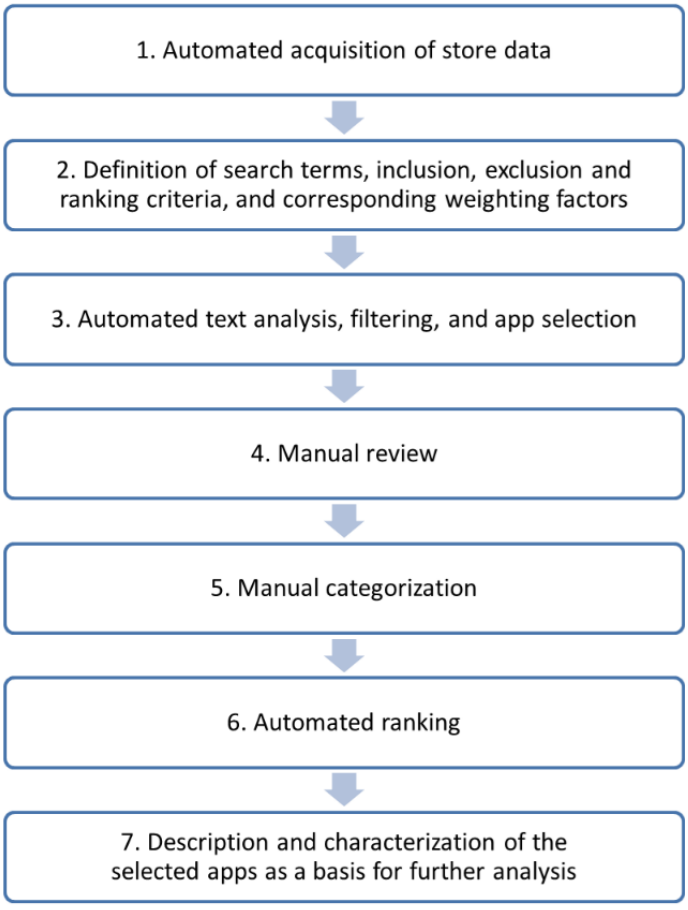


Table 1. Metadata fields provided by Apple with relevance to our evaluation.

Data field[a]	Description
trackId	Numeric identifier of the app
trackName	App name
features	This is set to “iOSUniversal” if the app will work on all iOS-based devices independent of form factor
supportedDevices	List of possible devices on which the app runs
fileSizeBytes	File size in bytes
artistId	Numeric identifier for the manufacturer
sellerUrl	Link to a website provided by the manufacturer (if available)
price	Price of the app (numeric value without currency specification), example, 1.99
currency	Currency for the “ price” field
genreIds	Numeric identifiers of the store categories assigned to the app
primaryGenreId	Numeric identifiers for the primary store category
minimumOsVersion	Minimum iOS version required
releaseDate	Date the app was first deployed; corresponds to field “currentVersionReleaseDate” for apps that have not yet been updated
currentVersionReleaseDate	Release date of the currently available, most recent app version
releaseNotes	Additional information provided by the manufacturer, if the app has been updated at least once
description	Full text of the store description
averageUserRating	The average star rating of all versions of an app (if sufficient numbers of ratings are available); may be empty
userRatingCount	The number of ratings of all versions of an app (if there are sufficient ratings); may be empty.

[a]Data fields with differing identifiers but identical content were merged.

This was done to obtain a more comprehensive set of keywords, for example, cardiology-related search terms more commonly used in an amateur context, which would have been more likely to be listed under the store category “Health & Fitness” and could possibly have been overlooked if we had solely restricted ourselves to apps listed in the “Medical” category. For each search term in the initial keyword list, the words most commonly associated with the respective term were recorded. To compile this list, for each occurrence of a keyword, the 10 words or N-grams in its direct proximity (± 5 words) were appended to a list, which was then ordered by the number of occurrences of each word or word combination. For words with an obvious cardiology reference (established through consensus of the authors), again, a list of the 10 most frequently associated terms was compiled. These final word lists were then manually checked by the authors for their potential to extend the keyword list. However, apart from different spellings (eg, words with or without hyphen for compound terms and common typographical errors), there were no significant changes observed for cardiology-related keywords.

For the subsequent automated filtering of all apps with German descriptions, the identified search keywords were then converted into regular expressions (in Perl notation) and optimized (see Textbox 1).

We chose not to use case sensitivity. For example, the partial term “blut[hoc]*druck” thus matched terms “Bluthochdruck” (high blood pressure) and “Blutdruck” (blood pressure) in the filtering process.

Definition of Inclusion and Exclusion Criteria

For the final analysis, only apps with German-language store descriptions, for which at least one of the predefined, cardiology-related keywords matched, were retained. Furthermore, for these apps, “Medical” had to be set as either as the primary or secondary category (assigned by the manufacturer). All apps not meeting these criteria were excluded from further analysis.

Textbox 1. Conversion of search keywords into regular expressions in Perl notation. Spaces around the vertical bar characters were only inserted to improve readability and were not part of the actual search string used.

a[r]{1,2}hythmie[n]* | atrioventrikularklapp|bikuspidalklapp | blut[hoc]*druck | blutgefä[sß]{1,2} | bradykard|cardiol | defibrillat[orin]* | elektrokardiogra[phien]* | erregungsleitungssyst | extrasystol | herzanalys | herzanf[aä]+ll | herzbeschwerd | herz[druck]*massag | herz[er]*krank | herzfehler | herzfit | herzfrequenz | herzfunktion | herzgesund | herzgeweb | herzinfarkt | herzinsuffizien | herzkamm | herzkatheter | herzklapp | herz[kranz]*gefä[sß]* | herz[-]*kreislauf[-]* | herz[minute-]*volum | herz[-]*monitor | herz[-]*patient | herzprobl | herzras | herzrhythmus | herzschl[aä]+g | herzschrittmach | herzschw[aä]+ch | herzspezialist | herzstiftung | herzstillstand | herztagebuch | herztest | herztod | herztransplantation | herzzyklus | hypertens | hyperton | kardial | kardiol | klappenprolaps | koronar | kreislaufforsch | kreislaufstillstand | kreislaufsystem | mitralklapp | myo[ck]+ard | pulmonalarterie | pulmonalklapp | schlagader | systol | trikuspidalklapp | ventrikel | vorhof | diastol | bekg[s]* | bgefä[sß]{1,2}w* | bw*aort[aen]+w*

Added Information: Readability Indices

For each app remaining in the analysis process, the readability index according to Flesch, that is, the Flesch Readability Ease ([27,28], with adaptations by Amstad for German language texts [29]), was determined using the functions provided in R (Version 3.4.4, [18]) via the koRpus package [30]. The aim was to offer an additional descriptive feature related to text difficulty for assessing the suitability and comprehensiveness of the store descriptions for specific target groups. In preparation, the descriptions were edited with regard to possible misinterpretations of sentence lengths (number of words), especially in the context of bullet lists; for example, we used regular expression-based search and replace operations matching commonly used bullet point characters to identify and replace these with periods if the previous bullet point had not been ended by appropriate punctuation. Similar to approaches described elsewhere [31], this preparation was done to avoid a grossly incorrect calculation of the readability index, which includes, among other things, recognizing sentence lengths. As the readability index developed by Flesch [27,28] has—in addition to the above-mentioned adaptation to the German language—been adapted to various other languages, we chose it over several other candidates, for example, for German-language texts, the "Wiener Sachtextformeln" by Bamberger and Vanecek [32] or the "Läsbarhetsindex" LIX, which was originally developed for Swedish-language texts [33] but is also applicable to those in German language. Furthermore, despite literature about the score's validity being somewhat limited, it is nevertheless widely employed (sometimes even as a legal requirement, eg, with insurance contracts being required to have a Flesch readability index of at least 45 in Florida state law [34]). Although not being fully comparable across languages, the diverse availability of the Flesch index provides a possible avenue to adapt the implemented methods to analyses of apps in languages other than German: This would only require adjusting 1 parameter in the analysis pipeline, without demanding additional changes to the code base.

Manual Review of the Cardiology-Related Set of Apps

The cardiology-related apps, as determined in the previous steps, were manually validated by the authors, with any uncertainties being resolved by discussion. As the aim was to identify apps relevant to the field of cardiology, all apps remotely addressing cardiological issues were included. As we did not want to limit ourselves to a specific target audience, apps deemed acceptable included those for cardiologists or other medical specialties as well as apps for patients or health conscious users, the latter also including apps one might use in a preventive or rehabilitative context. With this in mind, with the exception of 1 app, in which the search term "Vorhof" (atrium) was not employed in a cardiology-related context but as a part of a term related to other anatomical structures ("Kehlkopfvorhof", literal translation: atrium of the larynx), there were no obvious mismatches to terms not used in cardiology-related contexts. Some apps related to cardiology but intended for use in a veterinary environment could have been excluded, and there were also apps trying to influence their users' heart rate, for example, by means of meditation or other apps with a rather alternative approach to the subject. For cardiologists applying SARASA to identify apps for their specific professional needs, these apps would of course not be acceptable and would be eliminated. However, in our evaluation, these apps were not removed to create a realistic application scenario that can reproduce a manual keyword-based search within the app store, and we chose instead to differentiate them via the manual categorization process described in the following sections.

Manual Categorization and (Metadata-Based) Evaluation

The remaining apps were then classified manually by the authors, according to function types and subject areas.

App Categorization by Function Type

The 22 function types developed in the CHARISMHA (Chances and Risks of Mobile Health Apps) study [35], which can be grouped into 6 superordinate categories (Table 2), were used to subdivide the apps. These function types are to be seen independently of the "cardiology" application case considered here and should generally be applicable to apps that are used in health contexts.

Classification of the Apps by Topics

The groups of topics used for classifying the apps with respect to their subject areas were developed in a discussion between the authors. In addition to the function types mentioned in the previous step, which allow for a subdivision independent of the app's application area, a classification method focusing on the respective subject area and its facets, in this case, cardiology, is of advantage. For the example shown here, the apps were manually assigned to a set of defined topic groups (Table 3). Disputed topic assignments were clarified by discussion between the authors. Especially in cases with potentially overlapping topics, emphasis was placed on reaching a consensus about the main topic of each app.

Ranking the Apps by Predefined Criteria

In addition to the aforementioned filtering and review procedures, SARASA also provides a ranking mechanism with the intent to support presorting for manual processing. This mechanism presents apps that best match a manually determined and adaptable set of criteria in an order that displays apps that conform better to certain characteristics more prominently. To this end, predefined (and weighted) ranking criteria are used, relying on attributes either directly deducible from the metadata or calculated by various means.

Table 2. Function-related types as defined in the CHARISMHA study, including their superordinate categories, which were used in our manual classification [35].

Category	Description
Category: provision of information	
News	News apps, for example, for professional newspapers or news portals or for patient organizations
Reference	Apps that provide users with knowledge on health-related topics (eg, reference material)
Learning material	Apps that provide learning and teaching materials for education and training
Player/viewer	An app that permits playing or viewing media (eg, music, image data, and videos)
Broker	Apps that provide targeted information based on collected data (eg, location-based services)
Category: data acquisition, processing, and evaluation	
Decision support	Apps that support decision making based on collected data, based on the definition of decision support according to the study by Shortliffe and Cimino [36]
Calculator	Apps that perform calculations
Meter	Apps for immediate measurement of phenomena and characteristics not immediately accessible otherwise, for example, pulse measurement via the camera of the mobile device
Monitor	An app that may either serve as a measuring tool of its own or connects to a measuring device that is designed for multiple measurements of vital functions and stores them in a diary
Surveillance/tracker	Apps that automatically and continuously capture certain parameters in the background but do not interpret the data in a medical sense
Category: administrative use	
Administration	Apps for managing administrative data
Category: calendar and appointment-related apps	
Diary	Apps used for detailed data collection and tracking
Reminder	Apps that remind you of specific tasks
Calendar	Apps that are used to display and manage health-related events, for example, appointments, in the form of daily, weekly, or monthly overviews
Category: support	
Utility/aid	Apps that can be used as aids and help users to compensate for existing personal limitations (eg, hearing or vision problems)
Coach	Apps that teach users an activity and help them to carry it out
Health manager	Apps that are designed to continuously support users in health matters. A combination of several function types is required for an app to be assigned to this function type
Category: other	
Actuator	Apps that produce a direct physical impact in the form of mechanical motion or other physical effects
Communicator	Apps that are used for communication and getting into contact with others
Game	Apps that are used for pleasure, relaxation, and enjoyment
Store	Apps that offer opportunities to buy or sell goods and services
Other	All apps that cannot be assigned to any of the aforementioned function types

The selection of ranking criteria used in the example evaluation (Table 4) was chosen with the intention, among other things, of taking into account both transparency on the part of the manufacturer as well as (available) user evaluations. The ranking criteria and their assigned weighting factors are, however, freely adaptable, depending on the chosen topic and the objective of the evaluation being performed. As presented here, the inclusion of references to a possible medical device (or the explicit exclusion of this) in the store description, which we also used as a ranking criterion, may at least reflect a basic understanding of the associated problems on the manufacturer's part. Similar to a comprehensive store description and the provision of adequate information about the manufacturers themselves (eg, availability of an associated website for the app or the manufacturer), this examines the transparency of information provided to users. If available, user ratings are also included. It should be noted, however, that user ratings were only available for relatively few apps in the store, and even for these, there were only a limited number of apps that had obtained a significant number of ratings; this may be the reason to assign lesser weight to these factors in the future.

Table 3. Definition of the app-related topics, specifically for a cardiology context.

Topic	Description
Atlases	Dedicated teaching, learning, and reference works, for example, anatomy atlases
Blood pressure	Apps containing content and functions that can help with blood pressure management, for example
Conferences	Apps related to organizing one's conference visit or obtaining information about a conference
ECG[a]	Contents and functions relating to ECG
Nutrition	Nutritional content, for example, on diets (including nutrients) and nutrition-dependent health aspects or disorders
Fitness	Apps that promote fitness content and functions
Women	Apps specifically targeting women
Health data	Apps for the recording, monitoring, and analyzing of health-related data (eg, vital signs)
Communication	Apps with a communicative character, for example, for the exchange of information between medical staff and patients, within patient groups, and online communities
Medication	Medication-related apps
Complementary medicine	Apps that adopt an alternative medical approach (eg, acupuncture, acupressure, meditation, complementary medicine)
Neurology	Apps containing neurology-related content and functionalities
Emergencies	Apps for emergency medical or first aid use
Medical practice or hospital	Apps to be used in medical practice or hospital settings
Psyche	Apps covering psychological and psychiatric issues
Sleep	Apps to be used in sleep-related contexts
Metabolism	Apps specifically designed for use in managing metabolic disorders (eg, diabetes or other metabolic diseases)
Animals	Apps that have a cardiology reference but are intended for use in the field of veterinary medicine

[a]ECG: electrocardiography.

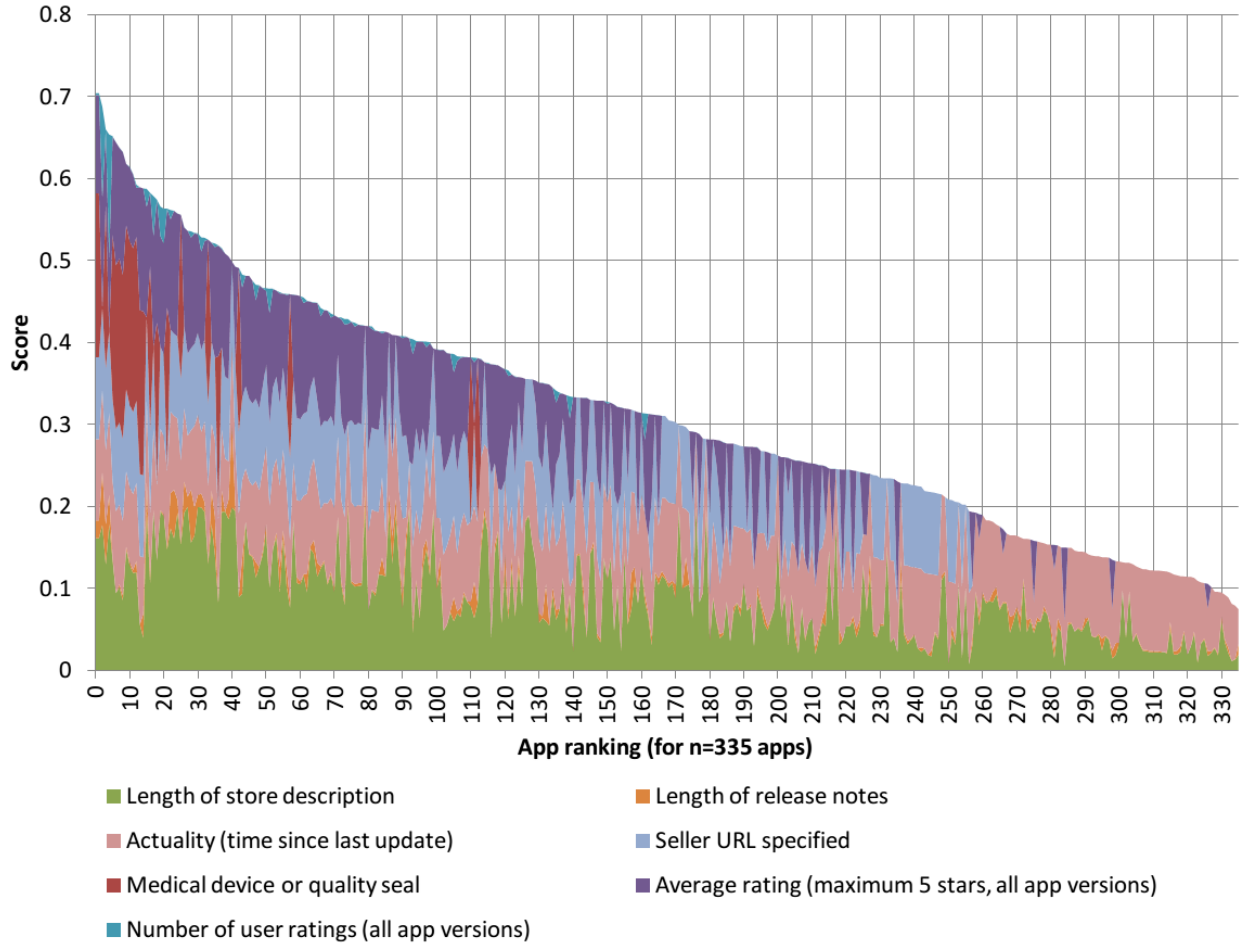
Table 4. Ranking criteria with weighting factor (percentage of the overall score), item name, description and condition to be fulfilled, or their explanation.

Variable	Source	Description	Score (%)
medicalDevice	Keyword-based evaluation of the store description (0: no keywords mentioned, 1: entry made)	The regulatory status (medical device) of the app is mentioned or explicitly excluded in the store description or there is mention of a seal of approval (reference to CE[a], FDA[b], medical device, or a seal of approval)	20
descriptionLength	Calculated value (number of characters)	Length of the store description	20
averageUserRating	Store metadata (from (0, maximum scoring reached) normalized to value range [0,1])	Average rating of all versions of the app	15
userRatingCount	Store metadata (from (0, maximum number of ratings) normalized to value range [0,1])	Overall number of user ratings that were obtained (for all versions)	15
sellerUrl	Evaluation of the metadata field sellerUrl for a valid URL (defined as nonempty and unequal to *http://* or *https://* without any further information); 0: invalid or empty, 1: valid URL	Link to a website (eg, a manufacturer's homepage or a Web page for the app) has been provided	10
releaseNotesLength	Calculated value (between (0, maximum number of characters) normalized to [0,1])	Length of the release notes, if available (prerequisite: at least one update, as only then must the field be set)	10
actuality	Calculated value including the time span between publication and readout time (from (0, maximum time span) normalized to 1-[0,1])	Whether the app is up to date	10

[a]CE: Conformité Européenne.

[b]FDA: Food and Drug Administration.

Figure 2. Actual contribution of the currently used ranking factors to the ranking of 335 cardiology-related apps.



Ranking the apps, in this case, based on a score calculated using the attributes and weighting factors defined in Table 4, is intended to support those interested in making an app selection with the help of SARASA. This is particularly useful in the case of a large number of possible results by first drawing attention to apps that are deemed to be particularly relevant depending on conformity to the chosen characteristics. An app presenting with an ideal rating for all attributes contributing to the score would—theoretically—achieve a score of 1.0. In reality, however, the maximum score achieved will usually be much less (eg, 0.7 for our cardiology-related sample evaluation in Figure 2). However, the calculated score and the factors contributing to it are not meant to be used for automatically evaluating app quality- or content-related aspects of the apps.

Results

Automated Filtering and App Selection

As basis for the descriptive statistics presented here, German-language, cardiology-related apps were selected using the aforementioned processes. Originally, there were 39,427 apps listed within the "Medical" category on April 14, 2018. First, 34,382 apps with store descriptions in languages other than German were removed, and for the remaining 5045 apps, the selected keywords were used to further filter the output, obtaining a final total of 335 apps, related to cardiology, for further analysis (see Figure 3).

Cardiology-Related Apps: Descriptive Statistics

General App Demographics

For an initial overview and comparison, descriptive statistics were first calculated for all 39,427 apps of the "Medical" category as well as for the 5045 apps with German-language store descriptions and the 335 apps with matches for the cardiology keywords (Table 5).

German-language apps, as well as those related to cardiology, were on the market slightly longer on average at 32.58 (interquartile range [IQR] 33.35) and 39.25 months (IQR 48.39), respectively, than all apps in the "Medical" category (median 28.22, IQR 34.89).

German-language apps in the "Medical" category (median 11.07 months, IQR 22.51) or with a cardiology reference (median 7.73, IQR 20.20) were updated more than usual, compared with the "Medical" store category (median 12.98 months, IQR 22.32).

Figure 3. Acquisition and keyword-based selection process for the 335 cardiology-related apps.

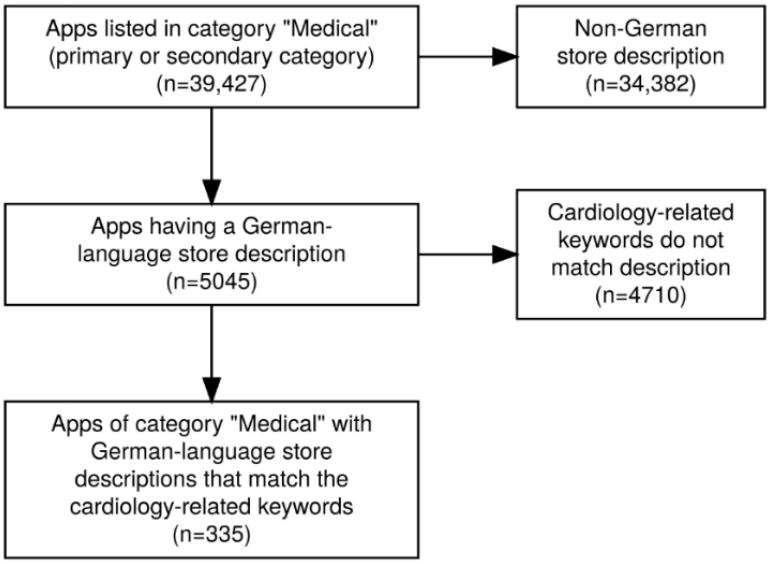


Table 5. App demography in comparison: apps within the "Medical" category versus those for which a German-language store description was provided versus those having a cardiology reference.

App demographics	All apps assigned to the "Medical" category (N=39,427)	Apps assigned to the "Medical" category that also have a German-language store description (N=5045)	Apps selected via the cardiology keywords (N=335)
Overall age of the apps in months, median (IQR[a])	28.22 (34.89)	32.58 (33.35)	39.25 (48.39)
Age in months (current version only), median (IQR)	12.98 (22.32)	11.07 (22.51)	7.73 (20.20)
File size in megabytes, median (IQR)	22.56 (36.08)	24.65 (30.37)	30.25 (53.58)
Price in Euros (€)			
Number of paid apps and percentage of total, n (%)	6838 (17.34)	846 (16.77)	91 (27.2)
Price, median (IQR)	3.99 (7.70)	3.49 (3.20)	3.49 (4.70)
Price range (€)	0.49-1099.99	0.49-499.99	0.49-249.99
Length of the store description (number of characters), median (IQR)	757 (1048.50)	921 (1502.00)	1630 (1585.50)
Star ratings (current version)			
Rated apps, n (%)	2072 (5.26)	1408 (27.91)	144 (43.0)
Median rating (IQR)	4.50 (2.00)	4.50 (2.00)	4.50 (2.00)
Maximum number of ratings (n)	6900	6900	645
Number of ratings, median (IQR)	2.00 (3.00)	2.00 (5.00)	3.00 (8.00)
Overall star ratings (all versions)			
Rated apps, n (%)	2581 (6.55)	1681 (33.32)	173 (51.6)
Median rating (IQR)	4.00 (2.00)	4.00 (2.00)	4.00 (1.50)
Maximum number of ratings (n)	22,153	22,153	6881
Number of ratings, median (IQR)	6.00 (18.00)	7.00 (22.00)	14.00 (49.00)

[a]IQR: interquartile range.

With regard to the file sizes, cardiological apps seem to be somewhat larger in median (median 30.25 megabytes, IQR 53.58) than those without restriction of the field of application (all apps in category "Medical": median 22.56 megabytes, IQR 36.08; German-language store description: median 24.65 megabytes, IQR 30.37). It is conceivable that this is influenced by the contents included for apps in the field of cardiology. For example, a higher proportion of reference works with somewhat larger amounts of texts and multimedia content may exert an influence here.

With regard to app pricing, a higher proportion of paid apps (27.2%, 91/335) is seen in those related to cardiology than in other medical apps (all medical apps: 17.34%, 6838/39,427;

apps with German description: 17.13%, 864/5045). Although only about every sixth app in the medical category requires a purchase, this is necessary for slightly more than every fourth app in cardiology-related apps with a German store description.

It is also noteworthy that apps in the "Medical" category (with a median of 757 characters, IQR 1048.5) have shorter description texts overall than those in German (median 921 characters, IQR 1502) and, in particular, cardiology-related apps (median 1630, IQR 1585.5). Although German-language texts are generally known for being longer compared with, for example, English-language texts, this does not explain the much greater length of descriptions of cardiology-related apps; it is also in this instance that specific cardiology-related peculiarities are speculated as a potential cause.

With regard to the app ratings given by users for the most recent versions available through the store, there were only a small number of apps that had received any ratings at all. That being said, there was a much larger proportion of current ratings for German-language apps (27.91%, 1408/5045) and apps found using the cardiology-related keywords (43,0%, 144/335) compared with all apps (5.26%, 2072/39,427) listed in the "Medical" category. Looking at the ratings given for all versions of the apps, the percentage of apps rated for all medical apps was 6.55% (2581/39,427) compared with 33.32% (1681/5045) for German-language apps and 51.6% (173/335) for those relating to cardiology. Median ratings differ only marginally.

Manufacturers

The vast majority of the vendors of the 335 selected apps are represented with only a single app in the app selection (Table 6). This can be determined based on the manufacturers' names and identification numbers. Some vendors provide more than 1 app with cardiology-related content and German-language store description in the store. Especially for manufacturers who offer several apps within this field, there are often only small variations in the content of the apps provided. These can be separately listed as lite or full versions of apps or versions for different form factors (iPhone and iPad) that are also shown separately. Also noticeable in this context are manufacturers who provide several apps covering cardiology conferences or various reference books or atlases on cardiology-related topics.

Hardware and Software Requirements Specified by the Manufacturers

The extent to which an app can be used depends, among other things, on the technical requirements it demands from the devices on which it should be run. Devices with iOS 9 or 10 are still represented in relevant figures. According to Apple, the (at the time of this writing) current iOS 11 version was installed on 81% of all devices at the end of May 2018. This does, however, mean that approximately one-fifth of the devices in use were not yet equipped for apps requiring this version. Apps that only require iOS 6 or older versions can hardly be expected to have been updated (see Table 7; the "Cumulative percentage" column indicates the percentage of apps listed under category "Medical" that can be used with a device running the corresponding iOS version).

As to usability on different form factors, 175 of the 335 apps stated that they could be used universally, that is, on all device types. The remaining 160 apps, on the other hand, require specific device types.

Automated Text Complexity Analysis of Store Description Texts

For about one-quarter of the apps (26.0%, 87/335), based on the automatically derived readability scores, only relatively low educational standards were required for potential users: On the basis of the available description texts for the corresponding apps, a maximum of 10 school years was required to comprehend the texts. However, just over half of the apps (53.7%, 180/335) required a high school diploma level, and for about one-fifth of the apps (20.3%, 68/335), the results of the text complexity analysis according to Flesch [27,28] (with adaptations for German-language texts according to Amstad [29]) indicated a level of difficulty going beyond that (Figure 4). It should be noted that the text complexity analysis as described here does not allow any statement as to whether or not the texts were actually grammatically correct. During a manual check, a few apps were identified whose description texts had obviously been automatically translated from other languages.

Table 6. The number of apps per manufacturer.

Apps provided by a single manufacturer (n)	Manufacturers with n apps in the store (n)	Percentage of 335 apps
1	178	53.1
2	43	25.7
3	4	3.6
4	3	3.6
6	2	3.6
8	1	2.4
9	3	8.1

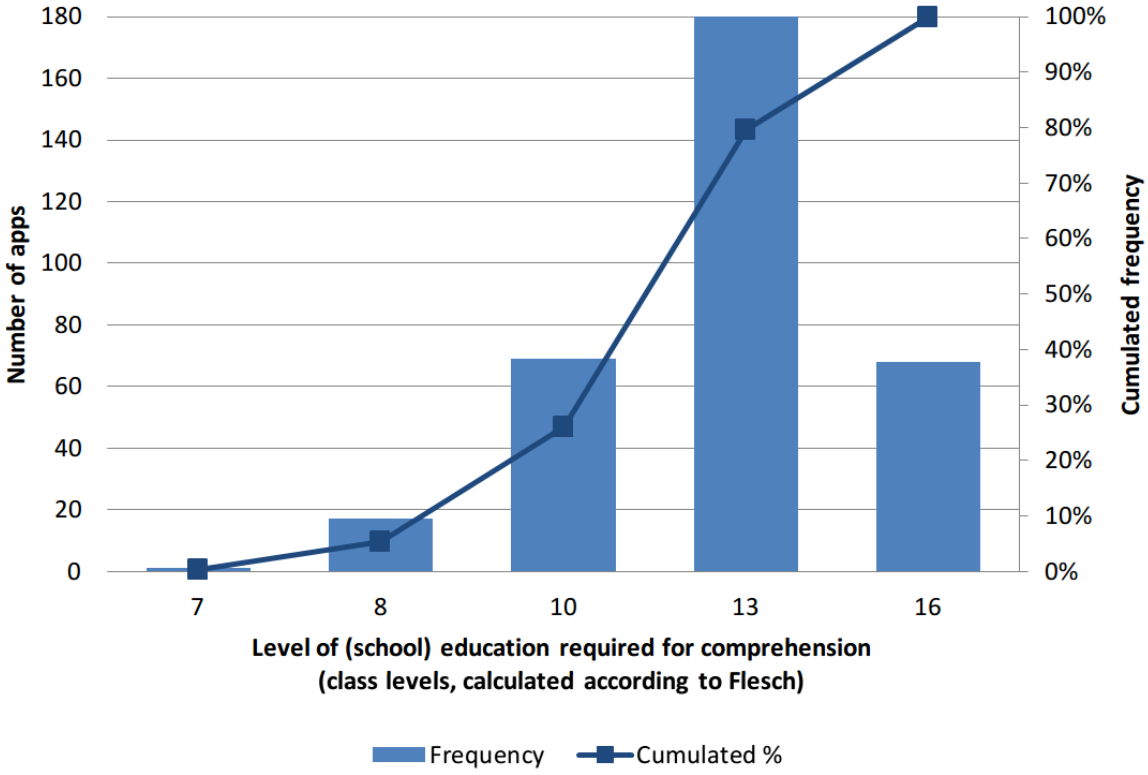
Table 7. Description of the 335 cardiology-related apps stratified by their minimally required iOS versions.

iOS version[a]	First date of release [37]	End of life for the following devices [37]	Apps (n)	Proportion (%)	Age (in days) on the readout date		Cumulative percentage (%)
					Minimum age	Median (IQR[a])	
3.x	June 17, 2009	iOS 3.1.3: iPhone 1st generation, iPod touch 1	2	0.6	2744.9	2892.11 (147.17)	0.6
4.x	June 21, 2010	iOS 4.2.1: iPhone 3G, iPod touch 2	6	1.8	1575.9	2076.26 (216.79)	2.4
5.x	October 12, 2011	iOS 5.1.1: iPad 1st generation, iPod touch 3	11	3.3	113.6	1599.42 (474.23)	5.7
6.x	September 19, 2012	iOS 6.1.6: iPhone 3GS, iPod touch 4	31	9.3	222.4	998.73 (495.18)	14.9
7.x	September 18, 2013	iOS 7.1.2: iPhone 4	47	14.0	5.7	586.39 (627.97)	29,0
8.x	September 17, 2014	N/A[b]	119	35.5	0.9	276.84 (329.45)	64.5
9.x	September 16, 2015	iOS 9.3.5: iPad 2 and 3, iPad Mini 1, iPhone 4S, iPod touch 5	79	23.6	0.1	68.25 (148.08)	88.1
10.x	September 13, 2016	iOS 10.3.3: iPad 4, iPhone 5 and 5C	33	9.9	2.8	50.84 (130.24)	97.9
11.x	September 19, 2017	N/A	7	2.1	33.2	83.69 (89.56)	100

[a]Minimum version. For clarity, the information is summarized according to the main iOS versions.

[b]N/A: not applicable.

Figure 4. Distribution of the educational levels required for comprehending the description texts of apps related to cardiology.



Comparisons of Function Type Groups and Subject Areas as Well as Cardiology-Related Keyword Groups

With regard to function types (as defined in Table 2), the cardiology-related apps were predominately apps that provide information (n=130), support users (n=72), or collect, provide, or evaluate data (n=60, Multimedia Appendix 1). Apps that are designed to provide information often do this by instructing users on how to behave or take action in emergencies (n=32). Apps for medical conferences (n=14), medical reference works

(eg, anatomy atlases, n=10) and other apps that cannot be assigned to a specific topic are also frequently found in the context of information provision. In the cardiology-related app group, health data are primarily collected and processed by apps falling into the functional categories of *support* (n=22) and *data acquisition, processing, and evaluation* (n=28). Support apps offer functions that go beyond the mere recording and processing of data and strive to achieve an added value for the user. Of these, apps belonging to the field of complementary medicine (n=24) are noteworthy. For example, although not exactly targeting cardiology in a strict sense, these often try to influence parameters such as heart rate or blood pressure with acupressure or mediation and were therefore also included on the basis of hits when employing the corresponding keywords.

Among other things, apps classified under *calendar and appointment related* often offer functions that remind users to record their data or help with adherence to prescribed medication. These can include blood pressure and diabetes diaries that record cardiology-relevant parameters or remind users of other measurements or medications.

Categorization of Apps by Function Type, Topics, and Associated Metadata

When looking at app demography stratified by manually assigned function types, some particularities are noticeable for apps related to cardiology (Multimedia Appendix 2). Based on the median number of months since the original publication of the apps, apps that deal with data acquisition, provision, and evaluation (44.59, IQR 41.07) or provide calendar- and time-related functions (53.38, IQR 36.63) have been on the market the longest. With median updates dating back 7.73 months (IQR: 20.20), updates to the analyzed cardiology-related apps are significantly more recent than for other German-language apps in the "Medical" category (12.98, IQR 22.32, Table 5). Updates to apps that serve to provide information are slightly less recent with a median of 9.47 months (IQR 20.30), despite the fact that, for these apps, it would be especially desirable to regularly check that the content is up to date.

As expected, the size (in megabytes) of apps that provide information (median 28.79, IQR 54.58) or help guide users to do exercises, for example coaching apps (function type *support*, median 52.59, IQR 51.28), and thus include additional text and multimedia content, exceeds the size of apps that only evaluate and process recorded data (16.66, IQR 36.27) or offer management functions (9.28, IQR 53.79).

The available ratings of the cardiology-related apps are unremarkable. This is regardless of their assigned function type and whether only the current or all versions provided in the store are considered. Notable is the significantly higher proportion of apps rated (between 33.8% of the current versions of apps providing information and 78% of all versions of apps of the "Support" function type), when compared with the values otherwise usual in the medical category (proportion of apps with ratings for the current version: 5.26%, all versions: 6.55%); the median ratings differ only marginally in comparison.

Similarly, differences between the groups with respect to the distribution of prices within the groups are unremarkable. With the exception of a larger proportion of paid cardiology-related apps (27.2%, 91/335)—compared with all apps in the medical category (17.34% paid apps, 6838/39,427) and those with German-language descriptions (17.13% paid apps, 864/5045)—there are only minor differences in median prices, which can mainly be explained by some particularly expensive "outliers."

In terms of the 19 manually assigned subject areas, apps that use health data in the broadest sense are the most common (20.0%, 67/335). Apps for use in emergencies (13.1%, 44/335) and those where blood pressure (10.7%, 36/335) or complementary medicine (9.3%, 31/335,) are also seen frequently. Overall, 13.4% (45/335) of the apps could not be assigned to a specific topic (Multimedia Appendix 3).

Observing the median, the sample apps originally appeared 4 to 5 years ago, and apps belonging to the field of complementary medicine had been published approximately 6 to 8 month (58.49, IQR 47.48) before the blood pressure apps (52.68, IQR 37.49), metabolic apps (52.11, IQR 48.71), and ECG apps (50.74, IQR 31.85) contained in the sample.

Apps are updated relatively frequently (median age of the currently available version in month 7.73, IQR 20.20). Exceptions are apps for organizing one's conference participation or apps that provide information about such events after the conference has taken place.

Apps from the ECG domain exhibit noticeable differences in terms of pricing. Almost every second app (47%, 7/15) is subject to a fee. With a median price of €20.99 (IQR 119.85) and a maximum price of €249.99, these apps are also significantly more expensive when compared with the other thematic areas.

With respect to the length of provided store descriptions, apps aimed at laymen and patients, in particular, for example, for complementary medicine (2600, IQR 736.50), metabolism-related apps (2003.50, IQR 1268.00), or those to be used in conjunction with medication (2023.00, IQR 1829.50), tend to have more extensive descriptions (represented by the number of characters) than other cardiology apps (1630, IQR 1585.50).

Apps for cardiological issues were rated more frequently than other apps in the "Medical" category (6.55%, 2581/39,427, Table 5) and apps with German descriptions (33.32%, 1681/5045, Table 5). More than every second (51.6%, 173/335) app has a star rating. In some areas (blood pressure and metabolism), as many as 3 of 4 apps were rated by users.

Discussion

Principal Findings

The aim of this study was to present a low-threshold solution for store analysis, which provides flexible support in the selection of apps, despite changing requirement profiles (previous knowledge of the interested parties, variability of use cases, or application scenarios) and without additional effort. The manual assignment of function types and topics for

descriptive purposes is to be understood as an optional step until it can be carried out automatically. The analysis not only takes into account all available metadata, especially app description texts, but also other attributes such as average user ratings. Although others such as Berardi et al [8], for example, strive for app classification using a somewhat similar approach, SARASA is currently designed as a tool to prepare for further manual evaluations. SARASA enables filtering of larger sets of apps, for example, entire store categories, based on the entirety of available data. To this end, it also evaluates available store descriptions and creates a ranking of the apps based on selected attributes (multiplied by weighting factors), making it possible to sort apps according to subjective, analyst-defined relevance. Depending on the problem, both the attributes to be evaluated and the weighting factors can be adjusted as required when using SARASA. Additions, such as the inclusion of an automated text complexity analysis for app descriptions in the ranking, for example, to create a focus on specific target groups, are easily possible. In particular, with regard to urgent questions of quality and security and regarding the suitability [7,38-40] and benefits [41] of apps, additions such as source code analyses or the recording of API calls [42] and requested authorizations [43] are to be integrated in future.

At the current stage of development, the approach presented is subject to several limitations, the knowledge of which is essential for assessing the method. These are explained in further details below. Much of future work on SARASA will have to take these aspects into account. In the course of this, additional modules may be included into SARASA.

Limitations

Platforms, Manufacturers, and Their Commitment to Transparency

The incorporation of various characteristics into analyses on a larger scale is strongly dependent on the willingness of the respective bodies to be transparent. For example, analyzing app installation archives may require the ability to download and analyze installation archives, without actually installing them on a physical device, but this may not be possible for all mobile platforms. In addition, not all app stores provide the same amount of access to desired attributes, for example, regarding required app permissions and numbers of downloads, the willingness of store providers to be transparent is also somewhat limited. For this reason, the analysis according to the SARASA scheme is currently limited to a single App Store (Apple). In our example evaluation, the German-language storefront was used to obtain the data. Evaluating other regional storefronts might have led to larger numbers of apps for which a more time-consuming manual evaluation would have been necessary. Using the ranking methodology, possibly with further adaptations to the attributes and ranking factors used, may be essential to still keep filtered results manageable, for example, by only evaluating a specific proportion of the top-ranked apps (based on the calculated score).

For stores of other platforms, for example the Play Store provided by Google for Android, an adaptation of the readout and—to a lesser extent—the evaluation routines will be necessary. For example, there is no official interface available for Google's Play Store that would allow full capture of the store or individual categories. In addition, some of the attributes provided in the stores differ between platforms, making it difficult to compare results of the SARASA method when applied to apps on different platforms.

Sampling Bias

The method presented in this study, a complete survey of the desired store categories, demonstrates a substantial reduction in, often criticized, sampling bias [44] in the evaluation of app-related data. It is conceivable, however, that manufacturers inappropriately assign other categories to their apps (eg, combinations of the store categories “Reference,” “Book,” or “Lifestyle” as primary and secondary categories for a reference work aimed at laymen)—because of uncertainties—and thus, some apps may be wrongly excluded from a SARASA-based analysis. Despite this, such cases would still yield a matching result based on keywords, if applied to data from other categories.

Regardless, the SARASA method is prone to underrepresentation of certain apps if the keywords chosen for selecting the apps do not adequately cover the desired subject. It could be argued that the proposed approach may not guarantee the identification of all suitable apps: For example, although nutritional and many other types of apps may also exert influence on cardiological parameters and thus be relevant in cardiology-related use cases, these will not be returned if their descriptions do not match any of the chosen search terms; adding corresponding terms to our search would, however, have been outside the scope of our presented work, as our aim was not to even identify apps for which the manufacturers had failed to specify a corresponding connection or purpose. In this instance, the limitations are comparable with those of a systematic literature search in review articles. Here, a strategy is used, searching for potentially relevant literature in databases; following the initial search, the results of course need to be evaluated manually. These reviews, however, do not commonly aim at determining whether or not there were potential matches that were missed, and which exactly were these; measures such as sensitivity and specificity, which are indispensable in diagnostic studies, are not common in literature searches, and this also holds true for the SARASA method. We believe that the comprehensiveness of the results for both literature and app searches, as they were described in this paper, can be derived from the comprehensive and easily verifiable search strategy, which includes a transparent specification of the search keywords, inclusion and exclusion criteria, and so on.

Selecting apps based on our methodology may also be favorable when compared with solely searching for apps based on search APIs or Web interfaces, as they are services provided by the respective app stores. In Apple’s case, for example, there is currently a maximum of 200 search results (in this case, apps) for keyword-based searches when using the provided search API [45], and it is somewhat tedious to perform searches for multiple keywords; our approach, at least for iOS-based apps, does not suffer from such restrictions.

Still, a bias may be introduced when applying the methodology to data acquired from other App Stores in the future. Much will depend on whether readout routines for these stores allow access to all apps in the desired store categories, rather than restricting access to so-called top apps. A complete survey of an entire app store's content, which could, for example, counteract the aforementioned bias of incorrect category allocation on the part of the manufacturers, will not be expedient simply because of the scope. Here, the platform providers' desire for transparency also plays a decisive role.

Language-Dependent Aspects

When selecting storefronts provided for other (language) regions (eg, this is possible when reading out the apps via the API provided by Apple), a significant variation of the results is to be expected. For example, if English-language apps are included in the evaluation, the number of apps selected increases many times over. As the purpose of the study was purely to illustrate the filtering and classification of apps using SARASA, we decided to confine ourselves to the described restrictions in database acquisition. In the future, the processes for other app stores or linguistic and geographical regions will have to be adapted so that universal statements beyond geographical or language borders can be made. In addition, changing the language or adding additional languages to SARASA-based evaluations will require adjustments that go beyond simply translating the search terms. For example, although the German language is known for the use of, often rather lengthy, compound terms, N-grams will be more relevant in other languages. Aspects such as these must then be taken into account within the search.

Machine Learning

At present, the assignment to define function types and subject areas was done manually to classify prefiltered apps based on language and keywords. For topics or inquiries that lead to a number of hits that significantly exceed the number demonstrated here, an automation of these assignments would be desirable. In an initial attempt to achieve this by means of keyword-based assignment, only little correlation with the manually defined assignments was observed. This is why the strategy was not pursued further in the context of the work presented here.

In spite of advantages, a more efficient procedure, for example, via a machine learning–based assignment, would have initially increased the work required (eg, due to the need for manually preclassifying training data). Nevertheless, it is planned to implement natural language processing (NLP)-based methods (specifically, topic analysis) in the future to enable at least a basic assignment. This idea seems particularly promising for the manual definition and assignment of subject areas. These would otherwise have to be redefined and discussed when using SARASA if the selected area of application changes. A topic analysis that would be an automatic definition of certain thematic subareas from the initially filtered apps and the ability to reliably assign the apps to these subareas would be helpful and should therefore be a goal of future developments of SARASA. The extent to which a successful assignment to the function types, known from the study by Albrecht et al [35], by means of such an approach is possible or whether this will indeed remain a meaningful part of SARASA-based analyses in the future must also be part of future investigations. A preliminary topic analysis [46,47] carried out for testing purposes, based on the available store descriptions, indicated that at least a mixed assignment of topics and function types is possible; this may already be sufficient.

Ranking

The ranking is strongly dependent on the analyst-defined values and factors used. Depending on the desired filtering objective or target group, it may make sense to adjust the ranking factors and/or to include the score calculated for the ranking as an additional filter criterion in the SARASA process. This could exclude apps that fall below a minimum score, defined from the outset. For example, Berardi et al [8] also use attributes such as the size of an app, for which they assume significance in relation to the complexity of an app or the scope of its content. Depending on the objective—should the analysis apply, for example, to learning apps or reference works—larger apps could then be ranked more prominently.

Text Complexity Analysis

The automated text complexity analysis, outlined only briefly in this paper, is subject to certain limitations as well, especially for nonstandard texts. Only in a few cases do the authors of the description texts follow the conventions commonly used for continuous texts in scientific or journalistic fields. Algorithms for determining text complexity, however, are usually specially standardized for such texts, expecting, among other things, a certain minimum length. Particularly with regard to punctuation and formatting, but also addressing texts that are (too) short in length, peculiarities or deviations are to be noted in app descriptions, which can have a negative influence on the automated analysis. Our analysis tried to eliminate the most common problems relating to punctuation and so on. Nevertheless, it cannot be ruled out that, for example, missing sentence points or formatting characters recognized as punctuation elements may cause the results to be skewed because the (recognized) sentence length (eg, represented by the number of words in a sentence) plays an essential role in the calculation of many readability formulas. It is, however, not possible to counteract the widespread problem of texts being too short. In the future, it may therefore make sense to additionally resort to other measures such as lexical diversity (calculated based on the number of different words/terms in a text) in addition to a pure text complexity analysis, if corresponding statements on text complexity or comprehensibility that can be derived from the texts are to remain part of SARASA.

Usability Aspects of the Semiautomated Retrospective App Store Analysis

For the future, it is planned to evolve the filtering process. At present, filtering can only be adjusted by parameterization in the R-based scripts. The aim would be to create a shiny frontend [48], that is, a user interface that can be operated via Web browser, even without programming knowledge, thus making the filtering process available to a wider audience.

Conclusions

SARASA is a method for filtering app store data according to formal criteria and accompanying description of the extract using common statistical measures. The filter results contain a selection of apps that can be passed through subsequent processing steps, which can, for example, following manual review of the list, consist of content-based quality assessments. SARASA allows the implementation of a flexible filter strategy, adaptable to the needs of the user. Automatic and manual analyses are easily combined when using SARASA. In the future, current functions will be supplemented by additional features, such as algorithmic topic analyses or sentiment analyses of user-provided comments (whenever user ratings are to be included as part of the analysis pipeline). The area of application is currently only limited to Apple's App Store, although expansion to other stores is planned. The method stands or falls with the transparency of app store providers and manufacturers, and their will to make relevant meta-information available. It is up to them to liberalize information and restrict censorship to provide clients, customers, and users truly fair circumstances finding their way around the app market. However, based on the available information, a fully automated selection, assessment, and recommendation of apps is not the aim of the SARASA method: The final decision about whether an app really has desired characteristics can only be made by reviewing and analyzing the metadata provided on the store as well as the apps themselves, which, for the time being, is not feasible without human intervention.

Acknowledgments

The authors are grateful to the Deutsche Gesellschaft für Innere Medizin (DGIM) e.V., Task Force Mobile Health, for their support of this project. No external funding was obtained to conduct the presented work. The authors thank Ms Sarah Long for proofreading.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Assignment of CHARISMHA function type groups versus subject areas (both assigned manually, one hit per dimension allowed). CHARISMHA: Chances and Risks of Mobile Health Apps.

[PDF File (Adobe PDF File), 26KB - mhealth_v6i11e11753_app1.pdf]

Multimedia Appendix 2

Descriptive breakdown by manually assigned function type group (as defined in the CHARISMHA study). CHARISMHA: Chances and Risks of Mobile Health Apps.

[PDF File (Adobe PDF File), 25KB - mhealth_v6i11e11753_app2.pdf]

Multimedia Appendix 3

Description of cardiology-related apps broken down by manually defined topics. Topics with 10 or less apps were not listed.

[PDF File (Adobe PDF File), 28KB - mhealth_v6i11e11753_app3.pdf]

References

1. Woerndl W, Schueller C, Wojtech R. A Hybrid Recommender System for Context-aware Recommendations of Mobile Applications. : IEEE; 2007 Presented at: IEEE 23rd International Conference on Data Engineering Workshop; April 17-20, 2007; Istanbul, Turkey p. 871-878. [doi: 10.1109/ICDEW.2007.4401078]
2. Hassan AE. The road ahead for Mining Software Repositories. In: Proceedings of the 2008 Frontiers of Software Maintenance.: IEEE; 2008 Presented at: 2008 Frontiers of Software Maintenance; Sept 28-October 4, 2008; Beijing, China p. 48-57. [doi: 10.1109/FOSM.2008.4659248]
3. Grechanik M, Conroy KM, Probst KA. Finding Relevant Applications for Prototyping. In: IEEE Xplore Digital Library.: IEEE; 2007 Presented at: Fourth International Workshop on Mining Software Repositories (MSR'07: ICSE Workshops 2007); May 20-26, 2007; Minneapolis, MN, USA p. 12. [doi: 10.1109/MSR.2007.9]
4. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
5. Paglialonga A, Pinciroli F, Tognola G, Barbieri R, Caiani EG, Riboldi M. e-Health solutions for better care: Characterization of health apps to extract meaningful information and support users' choices. In: IEEE Xplore Digital Library.: IEEE; 2017 Presented at: 2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI); October 12, 2017; Modena, Italy. [doi: 10.1109/RTSI.2017.8065915]

6. Paglialonga A, Lugo A, Santoro E. An overview on the emerging area of identification, characterization, and assessment of health apps. J Biomed Inform 2018 Jul;83:97-102. [doi: 10.1016/j.jbi.2018.05.017] [Medline: 29852318]
7. Hillebrand U, von Jan U, Albrecht UV. Concepts for quality assurance of health related apps. Stud Health Technol Inform 2016;226:209-212. [doi: 10.3233/978-1-61499-664-4-209] [Medline: 27350506]
8. Berardi G, Esuli A, Fagni T, Sebastiani F. Multi-store Metadata-based Supervised Mobile App Classification. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing. New York, NY, USA: ACM; 2015 Presented at: 30th Annual ACM Symposium on Applied Computing; April 13-17, 2015; Salamanca, Spain p. 585-588. [doi: 10.1145/2695664.2695997]
9. Zhu H, Chen E, Xiong H, Cao H, Tian J. Mobile app classification with enriched contextual information. IEEE Trans on Mobile Comput 2014 Jul;13(7):1550-1563. [doi: 10.1109/tmc.2013.113]
10. Lindorfer M, Neugschwandtner M, Platzer C. MARVIN: Efficient and Comprehensive Mobile App Classification through Static and Dynamic Analysis. 2015 Presented at: 2015 IEEE 39th Annual Computer Software and Applications Conference; September 24, 2015; Taichung, Taiwan p. 422-433. [doi: 10.1109/COMPSAC.2015.103]
11. Kahng M, Lee S, Lee SG. Ranking in Context-aware Recommender Systems. In: Proceedings of the 20th International Conference Companion on World Wide Web. New York, NY, USA: ACM; 2011 Presented at: WWW '11; March 28 - April 01, 2011; Hyderabad, India p. 65-66. [doi: 10.1145/1963192.1963226]
12. Martin W, Sarro F, Jia Y, Zhang Y, Harman M. A survey of app store analysis for software engineering. IIEEE Trans Software Eng 2017 Sep 1;43(9):817-847. [doi: 10.1109/tse.2016.2630689]
13. Lee G, Raghu TS. Determinants of mobile apps' success: evidence from the app store market. J Manag Inf Syst 2014 Nov 10;31(2):133-170. [doi: 10.2753/MIS0742-1222310206]
14. Johann T, Stanik C, Alizadeh B AM, Maalej W. SAFE: A Simple Approach for Feature Extraction from App Descriptions and App Reviews. : IEEE; 2017 Presented at: 2017 IEEE 25th International Requirements Engineering Conference (RE); September 4-8, 2017; Lisbon, Portugal p. 21-30. [doi: 10.1109/RE.2017.71]
15. Nayebi M, Cho H, Farrahi H, Ruhe G. App store mining is not enough. In: Proceedings of the 39th International Conference on Software Engineering Companion.: IEEE; 2017 Presented at: 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C); 20-28 May, 2017; Buenos Aires, Argentina p. 152-154. [doi: 10.1109/icse-c.2017.77]
16. Deutscher Ärzteverlag GmbH RDÄ. 2018. Internists sift and classify health apps URL: https://www.aerzteblatt.de/nachrichten/92003/Internisten-sichten-und-klassifizieren-Gesundheits-Apps [accessed 2018-07-09]
17. Albrecht UV, Hillebrand U, von Jan U. Relevance of trust marks and CE labels in German-language store descriptions of health apps: analysis. JMIR Mhealth Uhealth 2018 Apr 25;6(4):e10394 [FREE Full text] [doi: 10.2196/10394] [Medline: 29695374]
18. R Core Team. R-Project. Vienna, Austria: R Foundation for Statistical Computing; 2018. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2018-09-23]
19. Wickham H. CRAN. 2016. rvest: Easily Harvest (Scrape) Web Pages URL: https://CRAN.R-project.org/package=rvest [accessed 2018-09-23]
20. Wickham H. CRAN. 2017 Aug 20. httr: Tools for Working with URLs and HTTP URL: https://cran.r-project.org/web/packages/httr/index.html [accessed 2018-09-23]
21. Ooms J. Arxiv. 2014. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects URL: https://arxiv.org/abs/1403.2805 [accessed 2018-11-07]
22. Müller K, Wickham H, James DA, Falcon S. CRAN. 2018 May 6. RSQLite: 'SQLite' Interface for R URL: https://cran.r-project.org/web/packages/RSQLite/index.html [accessed 2018-09-23]
23. R Special Interest Group on Databases (R-SIG-DB), Wickham H, Müller K. CRAN. 2018 May 2. DBI: R Database Interface URL: https://cran.r-project.org/web/packages/DBI/index.html [accessed 2018-09-23]
24. Wickham H. CRAN. 2018 May 10. stringr: Simple, Consistent Wrappers for Common String Operations URL: https://cran.r-project.org/web/packages/stringr/index.html [accessed 2018-09-23]
25. Ooms J. CRAN. 2018 May 11. cld2: Google's Compact Language Detector 2 URL: https://CRAN.R-project.org/package=cld2 [accessed 2018-11-07]
26. Schmidt B, Li J. GitHub. 2017. wordVectors: Tools for creating and analyzing vector-space models of texts URL: https://github.com/bmschmidt/wordVectors [accessed 2018-09-23]
27. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [Medline: 18867058]
28. Flesch R. New facts about readability. Coll English 1949 Jan;10(4):225-226. [doi: 10.2307/372603]
29. Amstad T. In: Burger H, editor. How understandable are our newspapers?. Zürich: Universität Zürich; 1978.
30. Michalke M. Reaktanz. 2017. koRpus: An R Package for Text Analysis URL: https://reaktanz.de/?c=hacking&s=koRpus [accessed 2018-09-23]
31. Palotti JD, Zuccon G, Hanbury A. The Influence of Pre-processing on the Estimation of Readability of Web Documents. In: Proceedings of the 24th ACM International on Conference on Information Knowledge Management. New York, NY, USA: ACM; 2015 Presented at: CIKM '15; October 18-23, 2015; Melbourne, Australia p. 1763-1766. [doi: 10.1145/2806416.2806613]
32. Bamberger R, Vanecek E. Read, understand, learn, write the difficulty levels of texts in German. Wien: Jugend und Volk Sauerlaender; 1984.

33. Bjornsson CH. Readability. Stockholm: Liber; 1968.
34. FindLaw. 2018. Florida Statutes Title XXXVII. Insurance § 627.4145. Readable language in insurance policies URL: https://codes.findlaw.com/fl/title-xxxvii-insurance/fl-st-sect-627-4145.html [accessed 2018-09-23] [WebCite Cache ID 73kRbBumD]
35. Albrecht UV, Höhn M, von Jan U. Chapter 2: Health Apps and Market. In: Albrecht UV, editor. Chances and Risks of Mobile Health Apps (CHARISMHA). Hannover: Medizinische Hochschule Hannover; 2016:62-82.
36. Shortliffe EH, Cimino JJ. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Springer, New York, NY: Springer; 2006.
37. Wikipedia. 2018 Sep 23. iOS version history URL: https://en.wikipedia.org/wiki/IOS_version_history [accessed 2018-09-23]
38. Al-Subaihin A, Finkelstein A, Harman M, Jia Y, Martin W, Sarro F, et al. App store mining and analysis. In: Proceedings of the 3rd International Workshop on Software Development Lifecycle for Mobile. 2015 Presented at: DeMobile 2015; August 31-31, 2015; Bergamo, Italy p. 1-2. [doi: 10.1145/2804345.2804346]
39. Hartz T, von Jan U, Albrecht UV. Chapter 14: Orientation for Professional Users of Health Apps. In: Albrecht UV, editor. Chances and Risks of Mobile Health Apps (CHARISMHA) Hannover. Hannover: Medizinische Hochschule Hannover; 2016:302-319.
40. Albrecht UV. Chapter 13: Orientation for users of health apps. In: Albrecht UV, editor. Chances and Risks of Mobile Health Apps (CHARISMHA) Hannover. Hannover: Medizinische Hochschule Hannover; 2016:282-300.
41. Albrecht UV, Kuhn B, Land J, Amelung VE, von Jan U. Assessing the benefits of digital health solutions in the societal reimbursement context. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar;61(3):340-348. [doi: 10.1007/s00103-018-2696-0] [Medline: 29368121]
42. Linares-Vásquez M, Holtzhauer A, Bernal-Cárdenas C, Poshyvanyk D. Revisiting Android Reuse Studies in the Context of Code Obfuscation Library Usages. In: Proceedings of the 11th Working Conference on Mining Software Repositories. New York, NY, USA: ACM; 2014 Presented at: 11th Working Conference on Mining Software Repositories; May 31-June 01, 2014; Hyderabad, India p. 242-251. [doi: 10.1145/2597073.2597109]
43. Pustozerov E, von Jan U, Albrecht UV. Evaluation of mHealth applications security based on application permissions. Stud Health Technol Inform 2016;226:241-244. [Medline: 27350515]
44. Martin W, Harman M, Jia Y, Sarro F, Zhang Y. The app sampling problem for app store mining. : IEEE; 2015 Presented at: 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories; 16-17 May, 2015; Florence, Italy p. 123-133. [doi: 10.1109/msr.2015.19]
45. Apple Inc. iTunes. 2018. iTunes Affiliate Resources URL: https://affiliate.itunes.apple.com/resources/documentation/itunes-store-web-service-search-api/ [accessed 2018-11-07]
46. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003 Jan 3;3:933-1022 [FREE Full text]
47. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015;16(Suppl 13):S8 [FREE Full text] [doi: 10.1186/1471-2105-16-S13-S8] [Medline: 26424364]
48. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. CRAN. 2018 Nov 2. shiny: Web Application Framework for R URL: https://cran.r-project.org/web/packages/shiny/index.html [accessed 2018-11-07]

Abbreviations

API: application interface
CE: Conformité Européenne
CHARISMHA: Chances and Risks of Mobile Health Apps
ECG: electrocardiography
FDA: Food and Drug Administration
IQR: interquartile range
NLP: natural language processing
SARASA: semiautomated retrospective App Store analysis

Edited by G Eysenbach; submitted 31.07.18; peer-reviewed by A Paglialonga, C Jacob, L Zhou; comments to author 17.09.18; revised version received 27.09.18; accepted 25.10.18; published 20.11.18.

Please cite as:
Albrecht UV, Hasenfuß G, von Jan U
Description of Cardiological Apps From the German App Store: Semiautomated Retrospective App Store Analysis
JMIR Mhealth Uhealth 2018;6(11):e11753
URL: http://mhealth.jmir.org/2018/11/e11753/
doi:10.2196/11753
PMID:30459140

Review

Smartphone Apps About Crystal Methamphetamine ("Ice"): Systematic Search in App Stores and Assessment of Composition and Quality

Cath Chapman[1], BA (Hons), PhD; Katrina Elizabeth Champion[1,2], BA Psych (Hons), BHealth, PhD; Louise Birrell[1], BSocSc, BPsych (Hons), PhD; Hannah Deen[1], BPsych (Hons); Mary-Ellen Brierley[1], BPsych (Hons); Lexine A Stapinski[1], BPsych (Hons), MClinPsych, PhD; Frances Kay-Lambkin[1,3], BScPsych (Hons), PhD; Nicola C Newton[1], BPsych (Hons), PhD; Maree Teesson[1], BSc Psych (Hons), PhD

[1]National Health and Medical Research Council Centre of Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia

[2]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Priority Research Centre for Brain and Mental Health, The University of Newcastle, Newcastle, Australia

Corresponding Author:
Cath Chapman, BA (Hons), PhD
National Health and Medical Research Council Centre of Research Excellence in Mental Health and Substance Use
National Drug and Alcohol Research Centre
University of New South Wales
Building R1
22–32 King Street
Sydney, 2031
Australia
Phone: 61 2 9385 ext 0317
Fax: 61 2 9385 0222
Email: c.chapman@unsw.edu.au

Abstract

Background: Amid considerable community concern about the prevalence and harms associated with the use of crystal methamphetamine ("ice"), the increased use of smartphones to access health information and a growing number of available smartphone apps related to crystal methamphetamine, no previous reviews have examined the content and quality of these apps.

Objective: This study aims to systematically review existing apps in the iTunes and Google Play Stores to determine the existence, composition, and quality of educational smartphone apps about methamphetamines, including ice.

Methods: The iTunes and Google Play Stores were systematically searched in April 2017 for iOS Apple and Android apps, respectively. English-language apps that provided educational content or information about methamphetamine were eligible for inclusion. Eligible apps were downloaded and independently evaluated for quality by 2 reviewers using the Mobile Application Rating Scale (MARS).

Results: A total of 2205 apps were initially identified, of which 18 were eligible and rated using the MARS. The mean MARS quality total score for all rated apps was 3.0 (SD 0.6), indicating poor to acceptable quality. Overall, mean scores were the highest for functionality (mean 4.0, SD 0.5) and lowest for engagement (mean 2.3, SD 0.7).

Conclusions: This study demonstrates a shortage of high-quality educational and engaging smartphone apps specifically related to methamphetamine. The findings from this review highlight a need for further development of engaging and evidence-based apps that provide educational information about crystal methamphetamine.

(JMIR Mhealth Uhealth 2018;6(11):e10442) doi:10.2196/10442

KEYWORDS

internet; methamphetamine; mobile phone; review; substance-related disorder

Introduction

In recent years, there has been a marked concern about the use of methamphetamine, particularly crystal methamphetamine (or "ice"), and the considerable harms associated with its use for both individuals, their loved ones, and communities. According to the latest National Drug Strategy Household Survey [1], approximately 1.4% of the Australian population (aged ≥14 years) reported past year use of methamphetamine (including ice) in 2016. Although these data indicate that rates of methamphetamine use in the general population have remained fairly stable over the last decade, data from several sources suggest that there have been considerable changes in the patterns of use and harmful use; these include increases in the number of regular users who report using crystal methamphetamine (ice), as opposed to powder (speed) as their main form of methamphetamine [1,2], regular and dependent users [3], and harms associated with use [4,5]. Data from several sources suggest that rates of crystal methamphetamine use in regional and rural areas of Australia are of particular concern [1,6,7], and in 2016, 40% of Australians rated methamphetamine (including ice) as the drug of most concern, compared with 16% in 2013 [1]. In addition, international data indicate increasing use and harms associated with methamphetamines. Methamphetamines now account for 11% of overdose deaths in the United States [8], and market analyses in several parts of the globe indicate an increase in the use of methamphetamines (including ice) in recent years [9].

A key component of addressing community concern around illicit drugs, including ice, and preventing use and related harms, is the provision of accurate and evidence-based information, resources, and support. The use of the internet, smartphone apps, and mobile technology is a key means of disseminating public health information to the community and facilitating broad reach and engagement. Smartphone devices are now widely used, with 64% of the United States and 74% of the Australian population owning a device in 2015-2016 [10,11], and 62% of smartphone owners reporting that they used their phone to access health-related information in the past year [11]. Like internet-based and Web-based interventions, smartphone apps offer numerous advantages in terms of addressing public health issues [12,13] such as increased accessibility, portability of information, low costs, anonymity, and the ability to provide tailored feedback and support.

Over the past decade, there has been a dramatic increase in the number of smartphone apps designed to address health-related issues [14], with >165,000 health apps available for download in 2015 [15]. Systematic reviews of smartphone apps have been conducted in a range of health domains, including depression [16], anxiety [17], bipolar disorder [18], smoking cessation [19,20], nutrition [21], diabetes management [22], suicide prevention [23], health information seeking for cancer [24], and psychology or general mental health [25-27]. The majority of these reviews focused on the quality and content of apps and concluded that the existing apps vary in quality, with few grounded in scientific evidence. Nonetheless, the results from these reviews provide useful information about the features and functionality of high-quality apps and an increased understanding of community information and support needs. Thus, they serve to guide the future development of apps, as well as future research.

When looking at the substance use field, systematic reviews and content analyses of available apps that specifically target illicit drug use are sparse. We identified several reviews of apps that target alcohol use and related behaviors [28-30], or addiction and addictive behaviors in general [31]. We found one review that specifically reviewed apps that *promote* illicit drug use [32], and one review that analyzed illicit drug overdose apps [33]. In addition, we identified 2 systematic reviews of Web-based and mobile interventions targeting problematic substance use [12,13], both of which reviewed published intervention studies rather than available apps *per se*. However, we were unable to identify an existing systematic review of publicly available apps targeting either illicit drug use in general or crystal methamphetamine specifically. This is despite the fact that a marked number of apps related to crystal methamphetamine exist, amid increasing community concern about ice use and related harms. Therefore, this study aims to systematically review existing apps in the iTunes and Google Play Stores to determine the existence, composition, and quality of educational smartphone apps about methamphetamines, including ice. We anticipate that the results from this review will inform future development and research in this area.

Methods

Search Strategy

Adopting similar methodology to that used in previous reviews of smartphone apps [20,23], the Australian Google Play Store for Android phone apps and Australian iOS iTunes Store for Apple iPhone apps were searched in April 2017. A comprehensive list of keywords, including common street names for crystal methamphetamine, were used including "crystal methamphetamine," "methamphetamine," "crystal meth," "ice drug," "meth," "shabu," "tina," "glass," "illegal drugs," and "illicit drugs." Given the changes in app availability that occurs day-to-day, all searches were undertaken on the same day.

Eligibility Criteria and App Selection

Free and paid apps containing educational or information-based content related to methamphetamines, including ice, were included if they could be downloaded through the official Android or iOS store. Apps did not have to exclusively focus on crystal methamphetamine to be eligible for inclusion, rather they had to include some relevant educational content about ice. Apps were excluded if they were in a language other than English, if their content did not relate to crystal methamphetamine at all, if they did not include educational content about methamphetamine (eg, gaming apps and apps promoting drug use), or they were an audiobook (ie, voice recordings). After removing duplicate apps, initial screening of the titles and descriptions of identified apps was conducted by one author (KEC) using the eligibility criteria mentioned above to identify potentially relevant apps. Potentially eligible apps were then downloaded onto their respective devices (iPhone or Android) and independently assessed by 2 reviewers (LB and HD) to confirm eligibility. Eligibility for inclusion in the review

was assessed by both raters in relation to predetermined criteria. Apps were excluded if they were no longer available for download; did not include educational content; were not accessible once downloaded; or were not able to be downloaded. Any discrepancies in eligibility screening between the 2 raters were resolved by consensus.

App Classification and Quality Rating Tool

The quality of the apps was evaluated using the Mobile Application Rating Scale (MARS) [34]. Increasingly used in the mobile health field [20,35], the MARS is designed to collect descriptive and technical information about the app (App Classification) and assess the quality of the app (App Quality Ratings). The app classification information includes name, brief description, version, developer, costs, platform, focus of the app, theoretical background, affiliations (commercial, government, nongovernment organization, or university), target age group, and technical aspects (eg, ability to send reminders or share on social media). A 23-item Quality Rating scale assesses the app quality across 5 dimensions—engagement (interactivity, interest, and suitability for target audience; 5-items), functionality (ease of use and navigation; 4-items), esthetics (layout, graphics, and visual appeal; 3-items), information quality (accuracy, evidence base, and credibility; 7-items), and subjective quality (likelihood to use and recommend; 4-items). Responses are made on a 5-point scale (1="inadequate" to 5="excellent"), with mean scores calculated for each dimension. An overall mean quality rating score is calculated by combining the mean scores for the first 4 subscales, excluding the subjective quality. In addition, the mean of the subjective quality items is calculated to produce a separate subjective quality total score. The subjective ratings provide a measure of the raters' perceptions of the apps guided by objective anchors and focused on the relevance for people who might benefit from the app (rather than on themselves).

A total of 18 eligible apps were rated independently by 2 assessors (LB and HD) according to the MARS. Both assessors watched a Web-based training video provided by the scale's developers. The training video comprehensively explained the purpose of each of the subscales, as well as demonstrating how to rate items (with reference to real-world examples), and scoring instructions. Prior to rating the apps, assessors engaged with each app for, at least, 10 minutes, as recommended by the scale developers [34]. The information subscale relies on the raters' expertise and the availability of published studies to judge the accuracy and evidence base of the provided information. Hence, both raters held, at least, an honors degree in psychology and had relevant experience in drug and alcohol research, allowing them to assess whether the information provided was relevant, accurate, came from a reliable source, and was not potentially harmful to users. Moreover, each rater searched for any published studies to establish whether or not there was an evidence base for each app. Furthermore, high interrater agreement was observed between the assessors (intraclass correlation coefficient .904), indicating strong agreement between raters.

Results

App Selection

Figure 1 displays the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the full search strategy and the app selection process. A total of 2205 apps were identified through the iTunes and Google Play Stores. Of these, 223 were duplicate apps and 1953 were ineligible (1919 were not about crystal methamphetamine; 23 were about crystal methamphetamine, but not educational; 11 were not in English). A total of 30 potentially relevant apps were downloaded for further screening, and 18 apps were selected for the final data extraction and quality ratings.

App Classification

Table 1 presents descriptive information about each of the 18 included apps, and Table 2 presents a summary of the app classification across included apps. Just over half of the apps (10/18, 56%), were designed to be used on an iPhone or iPad, with just under half designed for use on the Android platform (8/18, 44%). The majority of apps were freely available for download, and 6 apps (6/18, 33%) required payment before download, with costs ranging from Aus $1.49 to Aus $42.99 per app. The majority of apps were affiliated with a commercial organization (10/18, 56%), 3 of 18 (17%) were affiliated with a university or government department, and 1 app was affiliated with a nongovernment organization. The affiliations of 4 of 18 apps (22%) were unable to be determined because of insufficient information provided by the app developers.

All rated apps included information on methamphetamines (including ice), and the majority (14/18, 78%) also included content about other drugs; only 4 apps exclusively focused on crystal methamphetamine. The most common app feature was the provision of factual information and educational materials (17/18, 94%), followed by offering advice, tips, strategies, or skills training to reduce drug use or related harms (6/18, 33%), assessment of drug use (2/18, 11%), feedback on drug use (2/18, 11%), monitoring or tracking (2/18, 11%), and goal setting (1/18, 5%). None of the identified apps explicitly mentioned a specific theoretical background or utilized evidence-based theory. In terms of technical requirements, 1 app (Street Drugs Organisation) explicitly stated that it needed internet access to function, another reported using automatic sensing (eg, global positioning system), which only functioned in the United States (Drug Sign), and 1 app required a password to log in (ASSIST app). Nearly all apps (17/18, 94%) were targeted to the general population; however, 1 app (Pure Rush) was specifically designed for use by young people aged ≥12 years.

Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the search strategy and app selection.

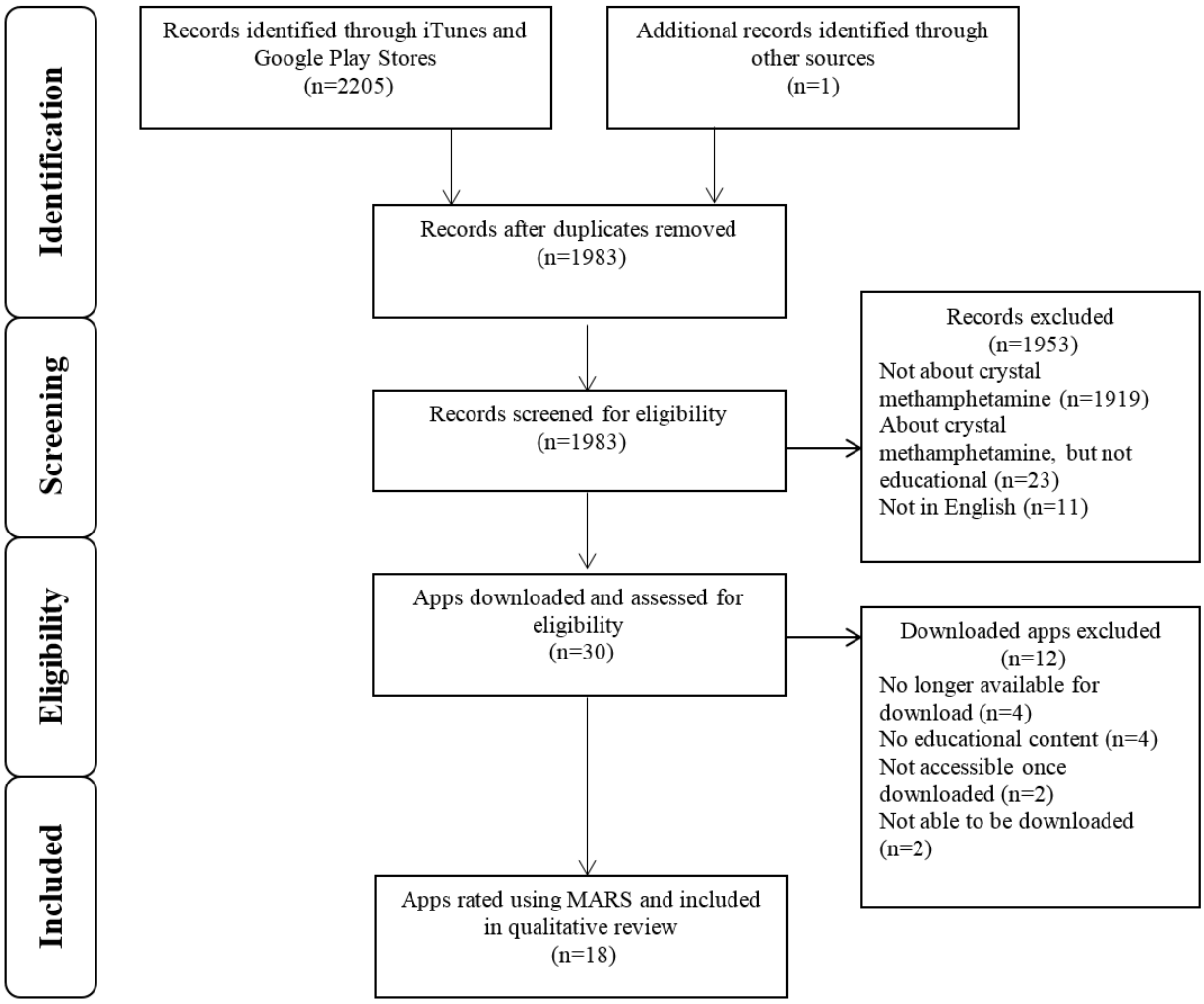


Table 1. The classification and content of included apps.

App name	Platform	Cost (Aus $)	Focus: what the app targets	Theoretical background and strategies	Affiliations	Target age groups	Technical aspects
Alcohol + Drugs e-Learning Pro	Android	3.15	Methamphetamines (including ice),other drugs	Information or education	Commercial	General	None listed
ASSIST App	iPhone	Free	Methamphetamines (including ice), other drugs	Assessment; feedback; information or education; monitoring or tracking; advice, tips, strategies, or skills training	Government/university	General	Allows password protection; Requires log-in
Drug Addiction	Android	Free	Methamphetamines (including ice), other drugs	Information or education; Advice, tips, strategies, or skills training	Unknown	General	None listed
Drug Addiction—How to Stop Your Dependence on Drugs	iPhone	1.49	Methamphetamines (including ice), other drugs	Information or education; Advice, tips, strategies, or skills training	Commercial	General	None listed
Drug Addiction: Drugs Handbook	iPhone	Free	Methamphetamines (including ice), other drugs	Information or education	Unknown	General	None listed
Drug Detection App	iPhone	Free	Methamphetamines (including ice), other drugs	Information or education	Commercial	General	None listed
Drug Detection App—Family and Home	iPhone	5.99	Methamphetamines (including ice), other drugs	Information or education	Commercial	General	None listed
Drug Effects Guide & Quiz Game	Android	Free	Methamphetamines (including ice), other drugs	Information or education	Unknown	General	None listed
Drug Sign	iPhone	42.99	Methamphetamines (including ice), other drugs	Assessment; Information or education; Advice, tips, strategies, or skills training	Commercial	General	Uses automatic sensing (eg, global positioning system)
Emergency Responder	Android	2.43	Methamphetamines (including ice)	Information or education; Advice, tips, strategies, or skills training	Commercial	General	None listed
Ice Your Body Belongs to You	iPhone	Free	Methamphetamines (including ice)	Information or education	Commercial	General	None listed
Meth Ice (methamphetamine)	Android	Free	Methamphetamines (including ice)	Information or education	Unknown	General	None listed
Meth Streetdrugs.org	Android	2.54	Methamphetamines (including ice)	Information or education	Commercial	General	None listed
National Drugs Campaign	iPhone	Free	Methamphetamines (including ice), other drugs	Information or education	Government	General	None listed
Overdose Aware	iPhone	Free	Methamphetamines (including ice), other drugs	Information or education	Nongovernment organization	General	None listed
Pure Rush	Android	Free	Methamphetamines (including ice), other drugs	Information or education	Government/university	12 years+	None listed
Street Drugs Organisation	Android	Free	Methamphetamines (including ice), other drugs	Information or education	Commercial	General	Needs Web access to function
Triggr Health—Support for Reducing Drinking/Using	iPhone	Free	Methamphetamines (including ice), other drugs	Feedback; Monitoring or tracking; goal setting; advice, tips, strategies, or skills training	Commercial	General	None listed

Table 2. The summary of features of included apps.

Feature	Apps (N=18), n (%)
Platform	
Android	8 (44)
iPhone	10 (56)
Cost (Aus $)	
Free	12 (67)
<6	5 (28)
>40	1 (6)
Focus	
Methamphetamines (including ice)	4 (22)
Methamphetamines (including ice) and other drugs	14 (78)
Theoretical background and strategies	
Information or education	17 (94)
Advice, tips, strategies, or skills training	6 (33)
Assessment of drug use	2 (11)
Feedback on drug use	2 (11)
Monitoring or tracking	2 (11)
Goal setting	1 (6)
Affiliations	
Commercial	10 (56)
University or government	3 (17)
Nongovernment organization	1 (6)
Unknown	4 (22)
Target age groups	
General population	17 (94)
>12 y	1 (6)
Technical aspects	
None	15 (83)
Password protection or requires log-in	1 (6)
Automatic sensing (eg, global positioning system)	1 (6)
Web access required	1 (6)

App Quality Ratings

Table 3 presents the mean subscale scores and the overall mean quality ratings on the MARS for each included app. Figure 2 presents the overall mean scores on the MARS subscales for all included apps. The mean MARS quality total score for all rated apps was 3.0 (SD 0.6), indicating *poor* to *acceptable* quality. A cutoff of 3.0 has been established as a minimum acceptability score [34], and nearly half (8/18, 44%) of the rated apps failed to meet this threshold. Only 2 apps (*Pure Rush* and *Triggr Health)* achieved an overall quality score >4, indicating they were of *good* quality. Similarly, the mean subjective quality total was *poor* to *inadequate* at 1.8 (SD 0.8). No apps received a rating of ≥4 on both the overall quality and the subjective quality, indicating there were no apps of overall *good* or *excellent* quality when both scales were considered together.

When examining the mean MARS subscale scores for included apps, we observed considerable variability in mean ratings of engagement, functionality, esthetics, and information quality. The overall mean scores were the highest for functionality (mean 4.0, SD 0.5), with 44% of apps (8/18) achieving a score of ≥4, indicating *good* functionality and reflecting high scores in particular on the subscale item ease of use, where only 1 app failed to achieve a rating of ≥4. Mean scores were the lowest for engagement (mean 2.3, SD 0.7), with only 1 app scoring ≥3 and none scoring ≥4. Scores were particularly low for interactivity and customization with the majority of apps rated as 1 (*inadequate*) on one (6/18) or both (9/18) of these items.

Table 3. Mobile Application Rating Scale ratings.

App name	Engagement	Functionality	Esthetics	Information	Overall quality total	Subjective quality total
Alcohol + Drugs e-Learning Pro	3.5	4.1	3.5	3.6	3.7	2.1
ASSIST App	2.3	3.5	2.3	4.0	3.0	2.1
Drug Addiction	1.6	4.1	2.0	1.8	2.4	1.0
Drug Addiction—How to Stop Your Dependence on Drugs	2.3	4.4	2.2	1.8	2.7	1.3
Drug Addiction: Drugs Handbook	1.3	4.3	1.3	2.0	2.2	1.0
Drug Detection App	2.0	3.4	3.0	3.2	2.9	1.9
Drug Detection App—Family and Home	2.0	3.9	3.0	3.2	3.0	2.1
Drug Effects Guide & Quiz Game	2.6	2.9	2.8	1.4	2.4	1.0
Drug Sign	2.1	3.5	3.7	2.7	3.0	1.9
Emergency Responder	2.0	4.0	2.7	3.2	3.0	1.8
Ice Your Body Belongs to You	2.9	4.0	3.0	2.8	3.2	1.5
Meth Ice (methamphetamine)	1.8	4.3	2.3	2.2	2.6	1.3
Meth Streetdrugs.org	1.8	4.4	2.3	2.1	2.6	1.5
National Drugs Campaign	1.8	3.5	1.8	4.1	2.8	2.1
Overdose Aware	2.2	4.5	3.5	4.0	3.6	2.9
Pure Rush	3.6	4.6	5.0	3.8	4.2	3
Street Drugs Organisation	2.1	4.3	3.0	2.7	3.0	1.0
Triggr Health—Support for Reducing Drinking/Using	4.0	4.8	4.5	2.9	4.0	3.6

Figure 2. The means and SDs for the overall quality rating and Mobile Application Rating Scale subscales for the included apps.

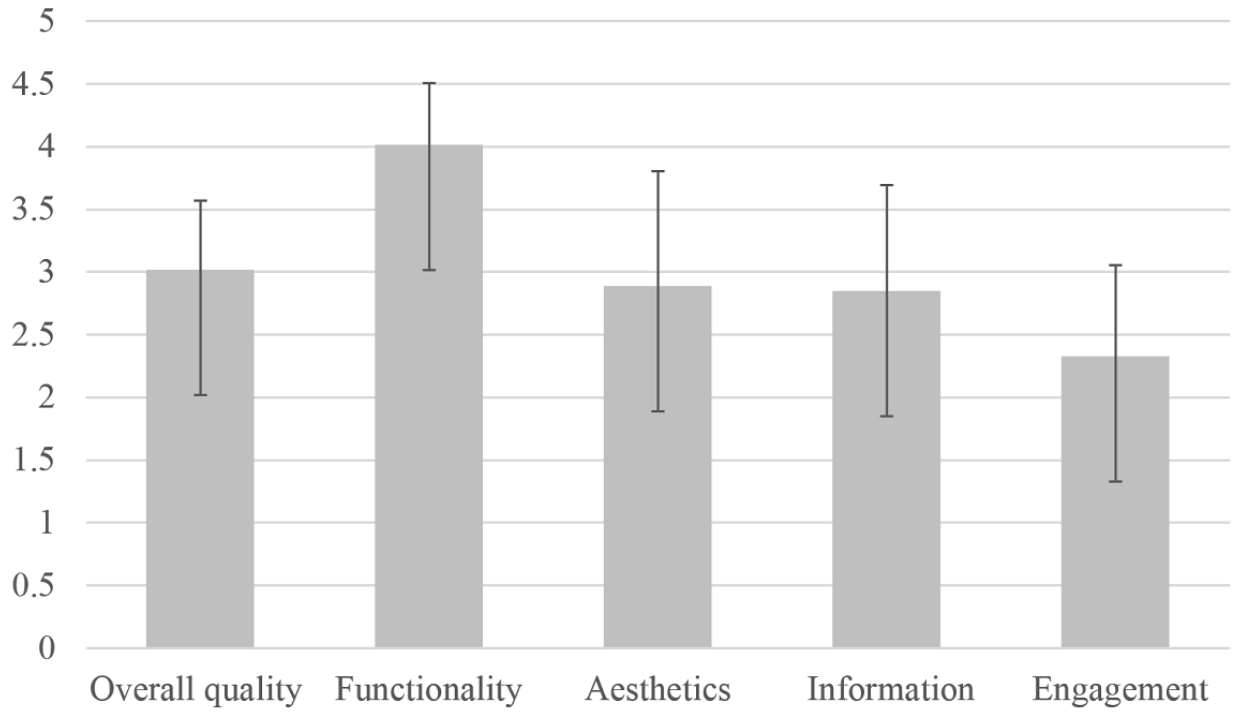
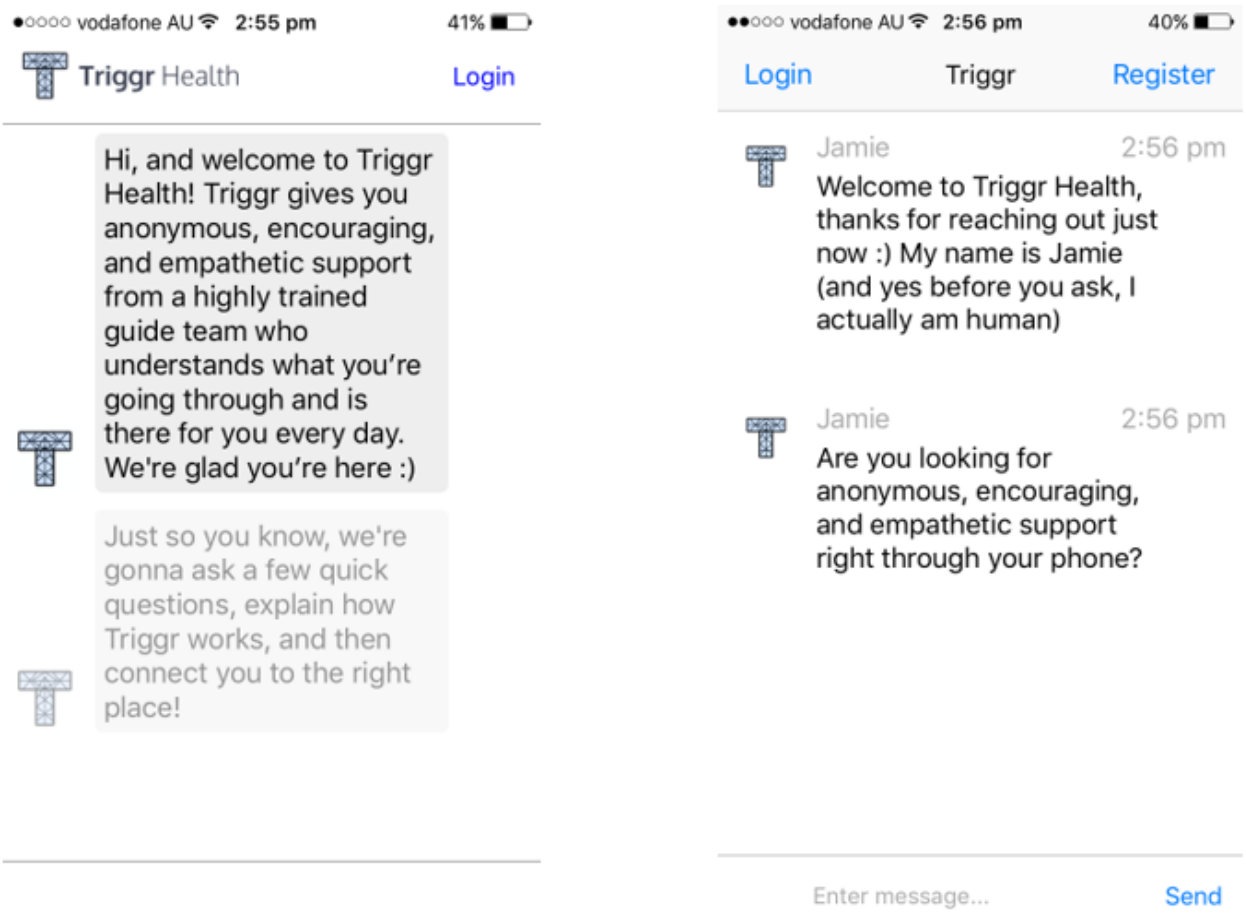


Figure 3. Screenshots of the Triggr Health App. (Source: Triggr Health - Support for Reducing Drinking/Using iPhone App Version 2.19.4. Developer: Triggr LLC. Licensed under fair use).



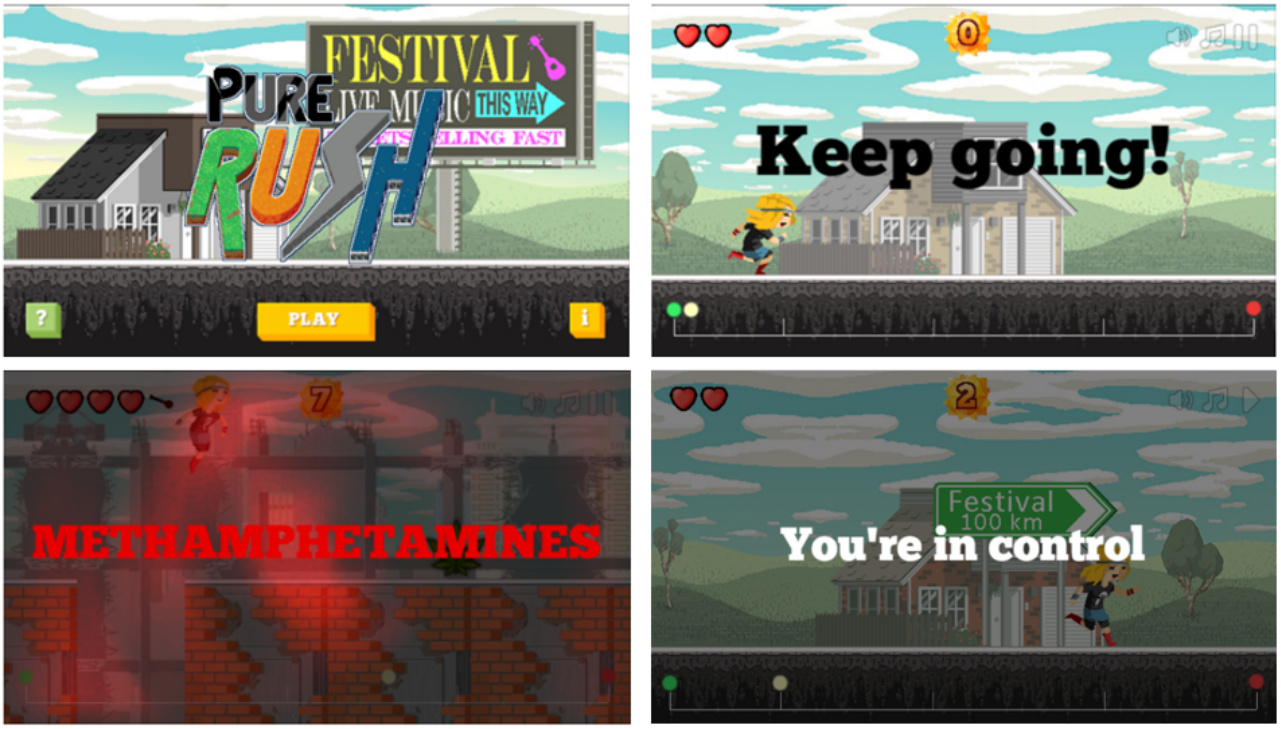
In addition, mean scores for esthetics were poor with only 2 apps scoring ≥4, indicating a rating of *good*. The layout was rated as, at least, acceptable for 67% of apps (12/18); however, graphics and visual appeal were rated as *poor* to *inadequate* in 6 and 9 out of the 18 included apps, respectively. While the overall mean score for the information quality was low for included apps (mean 2.8, SD 0.8), there was considerable variability across apps, with 2 apps scoring <2 and 3 apps scoring ≥4. Apps that achieved lower scores on the information quality subscale received low scores across subscale items, including inaccurate descriptions in the app store, poorly defined goals, poor-quality content, poor quantity of information provided (too much or too little), the app was unlikely to come from a credible source (ie, was most likely developed by a source with a vested interest, for example, commercial business), visual information was incorrect or confusing, and there was no available evidence base for these apps in the form of published evaluations.

Features of the Top-Ranked Apps

Although no apps were identified as having *excellent* overall quality, 2 apps achieved a *good* overall quality rating, with *acceptable* subjective quality ratings (Triggr Health—Support for Reducing Drinking/Using and Pure Rush). Triggr Health is a commercial app developed by a US company Triggr Health (Figure 3). The app targets addiction recovery and is focused on reducing substance dependence for a range of drugs, including crystal methamphetamine. The app connects users to a real-time behavioral change "guide" who interacts with users to set personalized goals and recovery plans. In addition, it utilizes predictive machine learning algorithms to identify smartphone use patterns and "check in" with users about their goals and recent activity. To date, no efficacy or effectiveness data exist to indicate that the Triggr Health app can prevent or reduce crystal methamphetamine use and harms.

Pure Rush is a serious educational game [36] for young people including information about cannabis, hallucinogens, crystal methamphetamine, and 3,4-methylenedioxy-methamphetamine (MDMA or ecstasy; Figure 4). In the app, users select a character and navigate through a music festival, with the goal of the game to avoid "running into" drugs. The app provides educational information about the negative effects of different drugs in an engaging manner. Pure Rush has been evaluated in one published peer-reviewed study [37], in which 281 young people (aged 13-16 years) were randomly allocated to receive a lesson involving Pure Rush or to an active control lesson; this evaluation found that the app was enjoyable to use, with both conditions associated with a marked increase in drug knowledge from pre- to posttest. There was some evidence that females who received Pure Rush showed greater knowledge gains compared with those in the control.

Figure 4. Screenshots of the Pure Rush app. (Source: Pure Rush Android App Version 2. Developers: National Health and Medical Research Council Centre of Research Excellence in Mental Health and Substance Use and 2and2. Licensed under fair use).



While this evaluation is promising, it focused on overall illicit drug knowledge and its effects on drug usage and, specifically, crystal methamphetamine usage, is not known. Furthermore, it should be noted that the app was designed to be implemented alongside a companion booklet that provides additional information (normative education, assertiveness skills, and corrections of common misperceptions), rather than as a standalone educational tool [37].

Other than for functionality, very few apps achieved ratings of *good* on any of the MARS subscales (information quality, engagement, and esthetics). Three apps achieved ratings of *good* on the information quality subscale—the ASSIST app, National Drugs Campaign app, and Overdose Aware app. Both the ASSIST and National Drugs Campaign apps are affiliated with a university or government department, and the Overdose Aware app is affiliated with a nongovernment organization. The ASSIST app was developed by the University of Adelaide and provides information, self-assessment, and links to support services for a range of drugs including amphetamines. Information covers the physical and mental effects of different drugs, harm reduction information, and links to relevant support services. The information provided is of high quality, but focused on a wide range of drugs and amphetamines as a broad class of drugs. The National Drugs Campaign app is a companion app to the Australian Government National Drugs Campaign website. It provides information about a range of drugs, including crystal methamphetamine, and links to external support sites. Overdose Aware was developed by the Pennington Institute and targets overdose education, including information about what an overdose is and recognizing overdose symptoms. It focuses on 4 classes of drugs—stimulants, depressants, opioids, and alcohol. None of these apps achieved a score of *acceptable* or above for engagement.

Discussion

Principal Findings

This study aimed to systematically review the existing apps to determine the existence, composition, and quality of educational smartphone apps about methamphetamines, including ice. We used the MARS to identify descriptive features and content of the apps and systematically rate quality in terms of engagement, functionality, esthetics, and quality of information. We examined app store descriptions for 1983 apps, downloaded and assessed 30 apps for eligibility, and rated 18 for content and quality. Only 4 identified apps focused exclusively on crystal methamphetamine, with the majority focusing on illicit drugs in general. The majority of apps reviewed focused on providing information or education, one-third provided advice, tips, strategies, or skills training, with few offering other features such as self-assessment of drug use, feedback, and monitoring or tracking of drug use. Functionality was rated as *acceptable* to *good* on the majority of apps; however, the overall quality was low, with only 2 apps achieving a rating of *good* on overall objective quality and *acceptable* on subjective quality according to the MARS. Ratings of the information quality varied considerably with only 3 apps achieving a rating of *good* in this domain, despite 94% of apps aiming to provide accurate information and education.

Comparison With Prior Work

Several findings warrant further comment. First, the mean overall MARS quality rating of 3.0 is commensurate with other apps of this scale in the health field [35,38,39]. Similarly, the finding that few apps consistently scored highly across all MARS dimensions also mirrors findings of previous studies that have assessed the quality of apps using this scale [35,38-40].

Bardus et al [38] used the MARS to assess the quality of commercial apps to assist with weight management and found that the information quality was the lowest scoring subscale for included apps and noted the lack of evidence-based content as a key driver of poor quality. In this study, the information quality was the second lowest scoring dimension after the engagement dimension. Notably, of apps that achieved either an overall quality rating of *good* (n=2) or a rating of *good* in terms of the information quality (n=3), only one carried a commercial affiliation (Triggr Health). The remainder were affiliated with either a university or government department (Pure Rush, ASSIST app, and National Drugs Campaign), or a nongovernment organization (Overdose Aware). Previous reviews have noted that the involvement of health professionals or other experts in health-related app development is often lacking or difficult to assess [29,33]. It is likely that this has contributed to low quality, particularly in terms of the evidence base of the apps and points to the need to involve experts in the development of good-quality apps in this area. In addition, it highlights the need to clearly identify affiliations and the evidence base used in the development of apps so that consumers can make informed choices [33]. Reviewers of apps in the substance use field have suggested clearer guidelines for the public [25], more stringent regulations [32], or more visible means of identifying quality apps [33] to improve consumer choice. In the process of conducting this review, >1900 irrelevant apps were identified in the searches; this has significant implications for people who are genuinely seeking help or information about crystal methamphetamine or other drugs and underscores the need for both better-quality apps on the market and better ways of guiding consumer choice about these apps.

Even within the context of crystal methamphetamine apps that do provide high-quality and evidence-based information, the lack of interactivity and engaging features represents a significant lost opportunity for reach and impact [31]. Only 3 apps offered interactive features such as monitoring, tracking, or goal setting. A previous review of smartphone apps to manage alcohol use, found that features such as tracking and tailoring were markedly associated with the app popularity (in terms of downloads) and user-rated quality [30], providing support for the inclusion of these features in the future development of apps. Similarly, a previous review of apps to support app users' weight management using the MARS also found that the overall app quality correlated with the number of different techniques or interactive features available [38]. One of the challenges in developing engaging, evidence-based apps is that despite the fact that government or university developed apps may be more likely to be based on evidence and involve health professional and other expert input, they are often competing with higher budget apps in the commercial space [30]. In this review, Triggr Health was the only app to achieve a score of *good* in terms of engagement, and this was also the only commercial app to achieve an overall quality rating above *acceptable*. It did, however, achieve its lowest rating in the area of information quality, further highlighting the need to balance credibility and accuracy of information with highly specialized technical features and interactivity.

The final point worthy of discussion is a large number of entertainment-based apps about crystal methamphetamine that were identified in the initial searches of the app stores. For example, several apps capitalized on the popularity of the television show, *Breaking Bad*, in which the central character manufactures and deals crystal methamphetamine. Although some of the available apps were seemingly harmless, for example, trivia about the episodes, downloadable artworks, and ringtones, others included game functionality where users could virtually cook methamphetamine or search for methamphetamine crystals. These latter apps highlight the influence of the media and popular culture on commercial app development as a potential public health concern, especially for young people or vulnerable groups. Gaming technology does hold great potential for the app development industry, especially for apps related to addiction and other health issues [31,37]; however, it is critical that app developers achieve the right balance among gamification, evidence, and quality of information.

Limitations

There are several limitations to this review that warrant discussion. First, the app market is highly dynamic. The availability of apps changes regularly, and this review can only offer a snapshot at one point in time. Within this context, it should also be noted that app stores do allow publishers to restrict distributions to particular countries. A previous systematic review of suicide apps available in Australia found 100% concordance between available apps across iOS app stores in a number of countries [23]. Nonetheless, this review focused only on apps available for download in Australia. Second, although this review focused on systematically rating the content and quality of apps, it does not provide information about the effectiveness of the apps in promoting accurate messages about crystal methamphetamine or preventing or reducing use and harms. One of the apps reviewed had been subject to a randomized controlled trial [37]; however, rigorous evaluations of app effectiveness in the illicit substance use field are rare [12,13]. While scientific evaluation of the effectiveness of apps is important, the speed with which the app market and the technology on which it is based changes rapidly and the length of time needed to conduct and publish randomized controlled trials presents a considerable challenge to researchers [12]. Some reviewers have suggested it might be useful for the mobile health field more broadly to focus on more pragmatic and less traditional modes of evaluation to assess the effectiveness of apps and other mobile health interventions to enable the field to build the evidence base more quickly [20,25].

Third, a relatively small number of apps met the inclusion criteria for the review, limiting the generalizability of the findings and indicating a need for further research in this area. For example, although the review points to general principles to guide the future development of evidence-based, high-quality apps, this study is not able to inform the specific circumstances under which these features might best apply. Along similar lines, very few of the included apps exclusively targeted crystal methamphetamine; this reflects the current app landscape and demonstrates a gap in the app market for apps specifically targeting crystal methamphetamine. The large number of apps

targeting substances such as cannabis, on the other hand, points to a need for future systematic reviews of apps that target substances other than methamphetamine to inform a changing and fast-growing area. Finally, the minimum time period of a single 10-minute session per app may not be sufficiently long enough to comprehensively evaluate app features, such as tracking use over time and behavioral change, which were present in 2 of the included apps.

Conclusions

Crystal methamphetamine is a high-impact drug that is associated with considerable harms and high levels of community concern. Importantly, the majority of people using crystal methamphetamine do not want to engage with traditional treatment or support services for fear of stigma and concerns about relevance [41]. This makes the increasing availability of app-based information and support for crystal methamphetamine (and other substances) critically important. This study was the first to systematically review the quality of available apps focusing on methamphetamine, including "ice." Despite the fact that many available apps purport to be about crystal methamphetamine, most do not offer educational content. Of those that do, most have not been subject to rigorous evaluations, they vary in quality, and despite having good functionality, few are likely to engage the public. Given the enormous potential of smartphone apps to promote positive and accurate public health messages and to prevent use and harms, this represents a significant opportunity for future development.

Acknowledgments

This study was supported by funding from the Australian Government Department of Health, under the Cracks in the Ice online community toolkit project, which aims to provide the Australian community with evidence-based information about ice [42]. This study was also supported by funding from the National Health and Medical Research Council through a National Health and Medical Research Council Centre of Research Excellence (APP1041129).

Conflicts of Interest

CC, LAS, NCN, and MT are 4 of the developers of the Pure Rush app. They derive no financial interest from this program.

References

1. Australian Institute of Health and Welfare. National Drug Strategy Household Survey: Detailed Report 2016. 2017. URL: https://www.aihw.gov.au/getmedia/15db8c15-7062-4cde-bfa4-3c2079f30af3/21028a.pdf.aspx?inline=true [accessed 2018-10-11] [WebCite Cache ID 734muUR0x]
2. Stafford J, Breen C. Australian Drug Trends 2015: Findings from the Illicit Drug Reporting System (IDRS). 2016. URL: https://ndarc.med.unsw.edu.au/sites/default/files/ndarc/resources/AUSTRALIAN_IDRS_2015_0.pdf [accessed 2018-03-16] [WebCite Cache ID 6xx4cgiDR]
3. Degenhardt L, Larney S, Chan G, Dobbins T, Weier M, Roxburgh A, et al. Estimating the number of regular and dependent methamphetamine users in Australia, 2002-2014. Med J Aust 2016 Mar 07;204(4):153. [Medline: 26937668]
4. Degenhardt L, Sara G, McKetin R, Roxburgh A, Dobbins T, Farrell M, et al. Crystalline methamphetamine use and methamphetamine-related harms in Australia. Drug Alcohol Rev 2017 Mar;36(2):160-170. [doi: 10.1111/dar.12426] [Medline: 27286742]
5. Darke S, Kaye S, Duflou J. Rates, characteristics and circumstances of methamphetamine-related death in Australia: a national 7-year study. Addiction 2017 Jun 11. [doi: 10.1111/add.13897] [Medline: 28603836]
6. Roche A, McEntee A. Ice and the outback: Patterns and prevalence of methamphetamine use in rural Australia. Aust J Rural Health 2017 Aug;25(4):200-209. [doi: 10.1111/ajr.12331] [Medline: 27868256]
7. Australian Criminal Intelligence Commission. National Wastewater Drug Monitoring Program, Report 1. 2017. URL: https://www.acic.gov.au/sites/g/files/net1491/f/national_wastewater_drug_monitoring_program_report_1_0.pdf?v=1490333695 [accessed 2018-10-11] [WebCite Cache ID 734nrFgh2]
8. The Lancet. Opioids and methamphetamine: a tale of two crises. The Lancet 2018 Feb 24;391(10122):713. [doi: 10.1016/S0140-6736(18)30319-2] [Medline: 29486928]
9. United Nations Office on Drugs and Crime. World Drug Report 2017: Market Analysis of Synthetic Drugs. 2017. URL: https://www.unodc.org/wdr2017/field/Booklet_4_ATSNPS.pdf [accessed 2018-03-12] [WebCite Cache ID 6xs7EWS7g]
10. Smith A, Rainie L, McGeeney K, Keeter S, Duggan M. U.S. Smartphone Use in 2015. 2015. URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/03/PI_Smartphones_0401151.pdf [accessed 2018-03-16] [WebCite Cache ID 6xx4BRI5A]
11. Sensis. Sensis Social Media Report 2016: How Australian people and businesses are using social media. 2016. URL: https://www.sensis.com.au/asset/PDFdirectory/Sensis_Social_Media_Report_2016.PDF [accessed 2018-03-05] [WebCite Cache ID 6xha4TfTC]
12. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A Systematic Review of the mHealth Interventions to Prevent Alcohol and Substance Abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]

13. Giroux I, Goulet A, Mercier J, Jacques C, Bouchard S. Online and Mobile Interventions for Problem Gambling, Alcohol, and Drugs: A Systematic Review. Front Psychol 2017;8:954 [FREE Full text] [doi: 10.3389/fpsyg.2017.00954] [Medline: 28649211]
14. Martínez-Pérez B, de la Torre-Diez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
15. Aitken M, Lyle J. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. 2015 URL: http://microsites.schott.com/d/studentchallenge/346d3edb-327f-42ad-b3b5-d859b8c4cf6f/1.0/ims_report_patient_adoption_of_mhealth.pdf [accessed 2018-03-06] [WebCite Cache ID 6xhoDP9ZR]
16. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
17. 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 May;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
18. 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;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
19. Bennett ME, Toffey K, Dickerson F, Himelhoch S, Katsafanas E, Savage CL. A Review of Android Apps for Smoking Cessation. J. Smok Cessat 2014 Feb 4;10(02):106-115. [doi: 10.1017/jsc.2014.1]
20. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Jul;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]
21. DiFilippo KN, Huang W, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: A systematic literature review. J Telemed Telecare 2015 Feb 12. [doi: 10.1177/1357633X15572203] [Medline: 25680388]
22. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
23. Larsen ME, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
24. Pandey A, Hasan S, Dubey D, Sarangi S. Smartphone apps as a source of cancer information: changing trends in health information-seeking behavior. J Cancer Educ 2013 Mar;28(1):138-142. [doi: 10.1007/s13187-012-0446-9] [Medline: 23275239]
25. 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;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
26. Harrison AM, Goozee R. Psych-related iPhone apps. J Ment Health 2014 Feb;23(1):48-50. [doi: 10.3109/09638237.2013.869575] [Medline: 24484194]
27. Turvey CL, Roberts LJ. Recent developments in the use of online resources and mobile technologies to support mental health care. Int Rev Psychiatry 2015;27(6):547-557. [doi: 10.3109/09540261.2015.1087975] [Medline: 26523397]
28. Meredith SE, Alessi SM, Petry NM. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Technol 2015;1:47-54 [FREE Full text] [doi: 10.2147/AHCT.S65791] [Medline: 27478863]
29. Penzenstadler L, Chatton A, Van Singer M, Khazaal Y. Quality of Smartphone Apps Related to Alcohol Use Disorder. Eur Addict Res 2016;22(6):329-338. [doi: 10.1159/000449097] [Medline: 27598779]
30. Hoeppner BB, Schick MR, Kelly LM, Hoeppner SS, Bergman B, Kelly JF. There is an app for that - Or is there? A content analysis of publicly available smartphone apps for managing alcohol use. J Subst Abuse Treat 2017 Nov;82:67-73. [doi: 10.1016/j.jsat.2017.09.006] [Medline: 29021117]
31. Savic M, Best D, Rodda S, Lubman DI. Exploring the focus and experiences of smartphone applications for addiction recovery. J Addict Dis 2013;32(3):310-319. [doi: 10.1080/10550887.2013.824331] [Medline: 24074196]
32. Bindhim NF, Naicker S, Freeman B, McGeechan K, Trevena L. Apps Promoting Illicit Drugs—A Need for Tighter Regulation? Journal of Consumer Health On the Internet 2014 Feb 24;18(1):31-43. [doi: 10.1080/15398285.2014.869166]
33. Photiou AJ. Smartphone Applications for Use in Drug Overdose: A review. J Clin Toxicol 2015;05(04). [doi: 10.4172/2161-0495.1000262]
34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
35. Regmi K, Kassim N, Ahmad N, Tuah N. Assessment of content, quality and compliance of the STaR mobile application for smoking cessation. Tob. Prev. Cessation 2017;3(July). [doi: 10.18332/tpc/75226]
36. Rodriguez DM, Teesson M, Newton NC. A systematic review of computerised serious educational games about alcohol and other drugs for adolescents. Drug Alcohol Rev 2014 Mar;33(2):129-135. [doi: 10.1111/dar.12102] [Medline: 24329810]

37. Stapinski LA, Reda B, Newton NC, Lawler S, Rodriguez D, Chapman C, et al. Development and evaluation of 'Pure Rush': An online serious game for drug education. Drug Alcohol Rev 2017 Oct 06. [doi: 10.1111/dar.12611] [Medline: 28986963]
38. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
39. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
40. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. JMIR Mhealth Uhealth 2017 Feb 21;5(2):e7 [FREE Full text] [doi: 10.2196/mhealth.6956] [Medline: 28223263]
41. Tait RJ, McKetin R, Kay-Lambkin F, Carron-Arthur B, Bennett A, Bennett K, et al. A Web-Based Intervention for Users of Amphetamine-Type Stimulants: 3-Month Outcomes of a Randomized Controlled Trial. JMIR Ment Health 2014;1(1):e1 [FREE Full text] [doi: 10.2196/mental.3278] [Medline: 26543901]
42. NHMRC Centre of Research Excellence in Mental Health and Substance Use, National Drug and Alcohol Research Centre, University of New South Wales, National Drug Research Institute, Curtin University. Cracks in the Ice Online Toolkit. URL: https://cracksintheice.org.au/ [accessed 2018-03-14] [WebCite Cache ID 6xtqdx2P9]

Abbreviations

MARS: Mobile Application Rating Scale

Edited by C Dias; submitted 19.03.18; peer-reviewed by M Hilgart, A Hammond, N Shen; comments to author 14.05.18; revised version received 01.07.18; accepted 10.08.18; published 21.11.18.

Please cite as:
Chapman C, Champion KE, Birrell L, Deen H, Brierley ME, Stapinski LA, Kay-Lambkin F, Newton NC, Teesson M
Smartphone Apps About Crystal Methamphetamine ("Ice"): Systematic Search in App Stores and Assessment of Composition and Quality
JMIR Mhealth Uhealth 2018;6(11):e10442
URL: http://mhealth.jmir.org/2018/11/e10442/
doi:10.2196/10442
PMID:30463834

Original Paper

What Drives Young Vietnamese to Use Mobile Health Innovations? Implications for Health Communication and Behavioral Interventions

Bach Xuan Tran[1,2], PhD; Melvyn WB Zhang[3], MBBS, MRCPsych; Huong Thi Le[1], PhD; Hinh Duc Nguyen[1], PhD; Long Hoang Nguyen[4], MSc; Quyen Le Thi Nguyen[5], MD, MPH; Tho Dinh Tran[6], PhD; Carl A Latkin[2], PhD; Roger CM Ho[7], MBBS, MRCPsych, FRCPC

[1]Institute for Preventive Medicine and Public Health, Hanoi Medical University, Hanoi, Viet Nam

[2]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[3]Biomedical Global Institute of Healthcare Research & Technology, National University of Singapore, Singapore, Singapore

[4]Centre of Excellence in Behavioural Medicine, Nguyen Tat Thanh University, Ho Chi Minh City, Viet Nam

[5]Institute for Global Health Innovations, Duy Tan University, Da Nang, Viet Nam

[6]Department of Hepatobiliary Surgery, Vietnam-Germany Hospital, Hanoi, Viet Nam

[7]Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Bach Xuan Tran, PhD
Institute for Preventive Medicine and Public Health
Hanoi Medical University
1 Ton That Tung
Hanoi,
Viet Nam
Phone: 84 982228662
Email: bach.ipmph@gmail.com

Abstract

Background: Mobile phone use in Vietnam has become increasingly popular in recent years, with youth (people aged 15-24 years) being one of the groups with the heaviest use. Health-related apps on mobile phones (mobile health [mHealth] apps) appear to be a feasible approach for disease and health management, especially for self-management. However, there has been a scarcity of research on mobile phone usage for health care among youth and adolescents.

Objective: This study aims to identify the patterns of usage of mobile phone apps and the preferences for functionalities of mobile phone-based health-related apps among Vietnamese youth.

Methods: An online cross-sectional study was conducted in Vietnam in August to October 2015. Web-based respondent-driven sampling technique was adopted to recruit participants. The online questionnaire was developed and distributed using Google Forms. Chi square and Mann-Whitney tests were used to investigate the difference in attitude and preference for mobile phone apps between the two genders.

Results: Among 356 youths (age from 15 to 25 years) sampled, low prevalence was found of using mHealth apps such as beauty counseling (6.5%, 23/356), nutrition counseling (7.9%, 28/356), disease prevention (9.8%, 35/356), and disease treatment (7.6%, 27/356). The majority of users found the app(s) they used to be useful (72.7%, 48/356) and reported satisfaction with these apps (61.9%, 39/356). No significant differences were found between the genders in their perception of the usefulness of apps and their satisfaction with mobile health apps. Most of the participants (68.2%, 238/356) preferred apps which are conceptualized and designed to run on a mobile phone compared to Web-based apps, and 50% (176/356) preferred visual materials. Approximately 53.9% (188/356) reported that it was integral for the mobile phone apps to have a sharing/social network functionality. Participants with a higher perceived stress score and EuroQol-5 Dimensions (EQ-5D) index were significantly less likely to use mHealth apps.

Conclusions: This study found a low proportion using mHealth-related mobile phone apps, but a high level of receptiveness and satisfaction among Vietnamese youth. Acceptance level and preferences toward mHealth apps as well as specifically preferred functionalities discovered in this study are essential not only in conceptualizing and developing appropriate mobile phone

interventions targeting youth and adolescents, but also in the application of technically advanced solutions in disease prevention and health management.

(JMIR Mhealth Uhealth 2018;6(11):e194) doi:10.2196/mhealth.6490

KEYWORDS

youth; adolescent; Vietnam; mHealth; mobile phone; app

Introduction

The advancement of the mobile phone has greatly enhanced and expanded the computing capabilities and functionalities of mobile phones and mobile devices [1]. Individuals using mobile devices are no longer confined to calling or text-messaging functionalities, but now they can perform a myriad of computing tasks with mobile phone apps. In 2017, it was estimated that more than 2.3 billion people owned a mobile phone and this figure will increase to 2.9 billion at the end of 2020 [2]. Mobile health interventions have taken advantage of this expansion and have been increasingly employed in clinical and community settings to educate and support people to protect, monitor, and manage their health [3].

Young adults and adolescents are among the heaviest mobile phone users. In Sweden, 93% of adolescents have a mobile phone [4], while in the United States and China, this rate is approximately 73% and 82%, respectively [5,6]. Some previous literature emphasized the health problems related to mobile phone use and abuse among these groups, such as poor sleep quality [7], sedentary lifestyle [8], and increased risk of mental health issues [4,9]. However, the popularity of the mobile phone provides opportunities to engage youth and adolescents in innovative mobile health (mHealth) interventions that facilitate personal health care and self-management [10,11]. Sexual health, for example, can be taught effectively to adolescents via mobile phone apps [12]. Majeed-Ariss et al [13] suggested that mobile phone apps could be considered a new approach to support adolescents in their management of chronic conditions. Apps can be customized and individualized by applying behavioral theories and interactive platforms; therefore, youth and adolescents can find appropriate apps based on their specific health care needs [14].

Despite a growing body of literature about mHealth, research on young adults and adolescents has been constrained [13]. Most trials have targeted adult populations, which may not be applicable to young adults and adolescents, who possibly have different patterns of mobile phone usage [5]. Previous literature argued that low attrition and adherence rates among these populations might be due to problems with designing apps, but it was not clear how the youth and adolescents adopt new mobile phone apps [15]. Hermawati and Lawson [16] reviewed and found that only 22.5% of mHealth trials involved targeted populations in the development phase. Therefore, an understanding of the usage patterns and preferences for mobile phone apps among youth and adolescents is critical to developing mHealth interventions effectively [17].

In recent years, mobile phone use in Vietnam has become increasingly common thanks to their affordability [18,19]. In 2017, the number of mobile phone users was estimated at 28.9 million people and this was projected to reach 42.7 million in 2022 [20]. Among 90% of mobile phone users, mobile phones are the main devices that individuals use to access the internet. Prior research conducted in Vietnam has demonstrated that those aged between 15 and 35 years spend an average of 169 minutes daily on their mobile devices [21], mostly on accessing social networking sites [19]. Moreover, Vietnamese adolescents and youth now face various health problems, such as risk behaviors (eg, alcohol abuse, smoking, unprotected sexual activities), overweight/obesity, and mental health issues [22-26], which can be intervened by using health-related mobile phone apps. Nonetheless, there has been a scarcity of research on mobile phone usage for health care among youth and adolescents in Vietnam. Therefore, this study aimed to explore their patterns of usage of mobile phone apps, attitudes toward health-related apps, and preferences for the functionalities of apps.

Methods

Study Setting and Population

An online cross-sectional study was conducted in Vietnam from August to October 2015. People who met the following criteria were invited to enroll in the study: (1) age from 15 to 25 years, (2) currently living in Vietnam, (3) have an email or a social network account to invite their peers, (4) able to consent and participate in the study, and 5) have a mobile phone device. There were no other specific exclusion criteria.

Web-Based Respondent-Driven Sampling Technique and Sample Size

The Web-based respondent-driven sampling (WebRDS) technique was utilized to recruit participants. WebRDS holds the potential to apply in public health studies. This method has demonstrated to be able to recruit a representative sample among youth in the United States [27]. Moreover, in a study among college students, it was found to boost the recruitment process up to 20 times compared to traditional RDS [28].

In this study, we selected the first waves (ie, core seeds) as representative of the diversity of the sampled population by taking into consideration their age, gender, and level of education. This strategy potentially initiates long recruitment chains of multiple waves of recruits that can ensure sample equilibrium being reached [29]. Firstly, 30 seeds from Hanoi Medical University, Vietnam National University, Hung Yen High School, and Phan Boi Chau High School were invited to participate in the study. The first two universities were used to recruit young adults, whereas the others were used for enrolling adolescents. We selected a respondent as a seed of the WebRDS if he/she also was a high-energy sociometric star and committed to being generative in recruiting their peers in the study [29].

The response rate was 100%. Participants in the seed group were informed that they needed to assist in the recruitment of other participants through their individual social network. They could not only invite people from their schools, but also from other schools. There was no predetermined duration with regard to the duration of recruitment. The recruitment was terminated only when the recruitment network was deemed to be no longer able to expand in size. A total of seven cases that were duplicated and three cases that did not meet the inclusion criteria (ie, did not answer at least 23 questions) were excluded. The total size of the sample for our study was 356 youths.

Study Procedure

After being invited to enroll in the study, the seeds were sent a Web link that contained an eligibility screening on the first page, information about the study and the electronic informed consent on the second page, and the Web-based questionnaire on the last page. The average time to complete this questionnaire was 20 minutes. Those participants who had difficulties accessing the original Web link were provided with an alternative Web link to access the survey.

After finishing the survey, each seed was asked to recruit at least five other members to participate in the survey, and the incentives were topped up by the number of peers they referred to the survey (US $0.50 per peer). We allowed copying of the Web link to text messages and social network sites to refer peers. Double entries were identified and removed via the email address that they have entered or IP address from the internet network that they logged in from.

Web Survey Design

The online questionnaire was developed and distributed using Google Forms, which met the security requirements as set forth by the ethical approval board for this study. The Web-based survey consisted of a total of 40 questions; the minimum number of questions participants needed to answer to be included was 23. We developed the questionnaire based on previous studies on preferences for functionalities of mHealth apps [30-33]. The researchers also included a logic check of the survey questionnaire to ensure that the data captured corresponded to the theme of the questions and was accurately captured in the backend database. To determine the feasibility and the reliability of the platform, the Web-based questionnaire survey was piloted among a group of 20 youths of varying ages and genders. The pilot group provided the investigators with recommendations to further optimize the online survey platform. Because only several minor changes were raised that did not affect the answers of people participating in the pilot study, we decided to include those youths in the final sample.

The Web-based questionnaire consisted of the following parts:

1. Sociodemographics and health status: including age, gender, educational status, marital status, and current living location. We also asked respondents to report their height and weight to compute body mass index (BMI). Their health-related quality of life (HRQOL) and perceived stress were measured using EuroQol-5 dimensions-5 levels (EQ-5D-5L) with EQ-5D index (ranging from –0.452 to 1; higher index indicates higher HRQOL) and the 4-item Perceived Stress Scale (PSS) with perceived stress score (ranging from 0 to 16 points; higher score indicates a higher level of stress).
2. Mobile phone usage pattern: we asked questions about whether they had a mobile phone with access to a wireless network (with or without 3G), operating system of their existing mobile phone devices, duration of ownership of a mobile phone device, time to use mobile phone per day (for calling, texting, and other purposes), self-reported level of proficiency in the usage of a mobile phone, and types of mobile phone apps they have ever downloaded and used.
3. Attitudes toward using health-related apps: information of interest included types of health-related apps they currently used (beauty counseling, nutrition counseling, disease prevention/treatment counseling, or others) and levels of usefulness and satisfaction regarding these apps. Beauty counseling apps refer to apps that provide advice, for instance, for using functional foods or dermatological interventions (not related to specific cosmetics).
4. Preferences for a mobile phone app: to determine preferences for functionalities of mobile phone health-related apps, the following questions were asked: how the mobile phone app was conceptualized (ie, with an addition Web-based version or mobile phone-based only), whether the app was comprised predominantly of text or images, whether the app allowed for any forms of sharing/social network, and whether the app had advertisements within it.

Statistical Analysis

STATA software version 12.0 was used to analyze the data. Chi square, Fisher exact, and Mann-Whitney tests were used to explore the differences between means and proportions of characteristics of interest by gender. Multivariate logistic regression was used combined with a stepwise backward strategy to build a reduced model. A *P* value of less than .05 was set as the level of statistical significance.

Results

A total of 356 individuals completed the Web-based questionnaires, of which 32.0% (n=114) were males. Most of the participants had a university education (83.1%, 289/356) and were single (80.6%, 286/356). Approximately 45.2% (160/356) lived as tenants (rent a house/room) whereas 28.5% (101/356) lived with their families. The mean BMI was 19.7 (SD 2.0) kg/m^2; the mean EQ-5D index and stress score were 0.76 (SD 0.16) and 6.5 (SD 2.1), respectively (Table 1).

Table 2 highlights that the Android operating system was dominant (62.9%, 224/356). Approximately 49.4% (176/356) of participants had used mobile phones for less than 24 months and 58.1% (202/356) rated themselves as being intermediate proficiency. Music players and social networks were the most frequently downloaded and used apps with 61.5% (219/356) and 51.1% (182/356), respectively. The participants used their mobile phone devices mostly for texting and other activities such as gaming or watching movies with means of 2.1 (SD 10.5) and 3.1 (SD 9.3) hours daily, respectively.

Table 1. Sociodemographics and health status of respondents (N=356).

Sociodemographics and health status	Participants
Male, n (%)	114 (32.0)
Age groups, n (%)	
<18	6 (1.7)
18-22	196 (55.4)
>22	152 (42.9)
Education attainment, n (%)	
≤High school	12 (3.5)
Vocation training	7 (2.0)
College	30 (8.6)
University	289 (83.1)
Postgraduate	10 (2.9)
Marital status, n (%)	
Single	286 (80.6)
Living with spouse/partner	69 (19.4)
Current living location, n (%)	
Rent	160 (45.2)
Dormitory	55 (15.5)
Living with family	101 (28.5)
Living with relatives	35 (9.9)
Others	3 (0.9)
Body mass index, mean (SD)	19.7 (2.0)
EuroQol-5 dimensions (EQ-5D) index, mean (SD)	0.76 (0.16)
Perceived stress score, mean (SD)	6.5 (2.1)

Table 2. Mobile phone use patterns among respondents.

Mobile phone use patterns	Female (n=242)	Male (n=114)	Total (N=356)	*P* value
Operating system, n (%)				
Android	158 (65.3)	66 (57.9)	224 (62.9)	.18
iOS	45 (18.6)	33 (29.0)	78 (21.9)	.03
Window Phone	32 (13.2)	18 (15.8)	50 (14.0)	.52
Blackberry	4 (1.7)	2 (1.8)	6 (1.7)	.95
Duration of using a mobile phone (months), n (%)				
<12 months	60 (24.8)	29 (25.4)	89 (25.0)	.19
12 to <24 months	61 (25.2)	26 (22.8)	87 (24.4)	
24 to <36 months	59 (24.4)	19 (16.7)	78 (21.9)	
≥36 months	62 (25.6)	40 (35.1)	102 (28.7)	
Level of proficiency in using the mobile phone, n (%)				
Novice	22 (9.3)	7 (6.3)	29 (8.3)	.008
Intermediate	146 (61.6)	56 (50.5)	202 (58.1)	
Advance	69 (29.1)	45 (40.5)	114 (32.8)	
Expert	0 (0.0)	3 (2.7)	3 (0.9)	
Apps downloaded and used (except mHealth apps), n (%)				
Games	98 (40.5)	67 (58.8)	165 (46.4)	.001
Weather forecast	47 (19.4)	25 (21.9)	72 (20.2)	.58
Music player	149 (61.6)	70 (61.4)	219 (61.5)	.98
Movie player	58 (24.0)	37 (32.5)	95 (26.7)	.09
Geographic Information system	49 (20.3)	34 (29.8)	83 (23.3)	.04
Social networks	116 (47.9)	66 (57.9)	182 (51.1)	.08
Financial management	11 (4.6)	11 (9.7)	22 (6.2)	.06
News	78 (32.2)	49 (43.0)	127 (35.7)	.05
Entertainment	81 (33.5)	43 (37.7)	124 (34.8)	.43
Education	71 (29.3)	40 (35.1)	111 (31.2)	.28
Sport	13 (5.4)	41 (36.0)	54 (15.2)	<.001
Book	105 (43.4)	44 (38.6)	149 (41.9)	.39
Shopping	29 (12.0)	6 (5.3)	35 (9.8)	.05
Time to use mobile phone per day (hours), mean (SD)				
For calling	0.7 (0.7)	0.9 (1.5)	0.8 (1.0)	.34
For texting	2.2 (11.0)	2.1 (9.4)	2.1 (10.5)	.19
For others (eg, game, movies)	2.1 (2.5)	5.2 (16.0)	3.1 (9.3)	.07

Table 3. Usage, attitudes, and preferences for mobile phone health care apps.

Usage, attitudes, and preferences for mobile phone health care apps	Female (n=242), n (%)	Male (n=114), n (%)	Total (N=356), n (%)	*P* value
Using any health care apps	47 (19.4)	20 (17.5)	67 (18.8)	.67
Types of health care apps				
Beauty counseling	20 (8.3)	3 (2.6)	23 (6.5)	.04
Nutrition counseling	19 (7.9)	9 (7.9)	28 (7.9)	.99
Disease prevention counseling	25 (10.3)	10 (8.8)	35 (9.8)	.65
Disease treatment counseling	17 (7.0)	10 (8.8)	27 (7.6)	.56
Useful for health				
Not useful	13 (28.3)	5 (25.0)	18 (27.3)	.79
Useful	33 (71.7)	15 (75.0)	48 (72.7)	
Satisfaction with mobile health apps				
Dissatisfied	19 (43.2)	5 (26.3)	24 (38.1)	.21
Satisfied	25 (56.8)	14 (73.7)	39 (61.9)	
Preferences				
Type of apps				
Web-based apps	69 (29.2)	42 (37.2)	111 (31.8)	.14
Mobile phone apps	167 (70.8)	71 (62.8)	238 (68.2)	
Information and contents of apps				
Visuals	124 (51.9)	52 (46.0)	176 (50.0)	.30
Text	6 (2.5)	1 (0.9)	7 (2.0)	
Combination of visuals and text	109 (45.6)	60 (53.1)	169 (48.0)	
Areas of apps				
Specific area (focus on one topic)	115 (48.7)	67 (58.8)	182 (52.0)	.08
Integrative areas (focus on multiple topics)	121 (51.3)	47 (41.2)	168 (48.0)	
Sharing/social network functionalities				
Yes	131 (55.0)	57 (51.3)	188 (53.9)	.52
No	107 (45.0)	54 (48.7)	161 (46.1)	
Feeling toward advertisements within apps				
Neutral	49 (20.5)	24 (21.6)	73 (20.9)	.70
Negative	188 (78.7)	85 (76.6)	273 (78.0)	
Positive	2 (0.8)	2 (1.8)	4 (1.1)	

Table 3 illustrates that 18.8% (67/356) had ever used any health care apps. Rates of using mHealth apps such as beauty counseling (6.5%, 23/356), nutrition counseling (7.9%, 28/356), disease prevention (9.8%, 35/356), and disease treatment (7.6%, 27/356) were also observed. The majority of users found the app(s) they used to be useful (72.7%, 48/356) and reported satisfaction with such app(s) (61.9%, 39/356). There were no significant differences between the genders in their perception of the usefulness of apps and their satisfaction with mobile health apps. There was a significant difference between genders in their usage of beauty counseling apps, with there being a predominance of females over males using beauty counseling apps (*P*=.04). Most participants (68.2%, 238/356) preferred apps which were conceptualized and designed to run on a mobile phone as compared to Web-based apps, and 50.0% (176/356) preferred apps with a dominance of visuals. Approximately 53.9% (188/356) reported that it was integral for the mobile phone apps to have a sharing/social network functionality, and 78.0% (273/356) felt negatively about the advertising within apps.

Table 4 depicts that those studying in college were more likely to use health-related apps (OR 4.16, 95% CI 1.76-9.82) compared to those having a high school education or less. Meanwhile, people having a higher stress score or EQ-5D index and an intermediate level of using the mobile phone were less likely to download this type of app.

Table 4. Associated factors with the use of mobile phone health-related apps.

Using any health-related apps	Odds ratio (95% CI)	*P* value
Current living location (vs homestay)		
Living with relatives	0.30 (0.08-1.12)	.07
Education (vs ≤high school)		
College	4.16 (1.76-9.82)	<.001
Duration of mobile phone use (vs <12 months)		
12 to <24 months	1.61 (0.81-3.19)	.18
Level of proficiency in using the mobile phone (vs novice)		
Intermediate	0.26 (0.10-0.69)	.01
Advanced	0.48 (0.18-1.29)	.15
Perceived stress score	0.86 (0.74-0.99)	.04
EuroQol-5 dimensions (EQ-5D) index	0.13 (0.02-0.84)	.03

Discussion

This study critically points out the mobile phone app usage patterns and preferences that could direct further mHealth interventions among Vietnamese youth. Although we only observed a low proportion of respondents using mHealth apps, we found a positive premise to develop such apps since the majority of youths using mHealth apps expressed their high degree of acceptance and satisfaction. This study also implies some important features regarding app conceptualization and design, which could help to optimize the receptiveness of mHealth apps in Vietnamese youth.

Among a sample of Vietnamese youths, we found a low percentage of people using mHealth-related mobile phone apps (18.8%, 67/356), from 6.5% (23/356) for beauty counseling apps to 9.8% (35/356) for disease prevention counseling apps. These figures were similar to results reported in previous reports in Vietnam and other countries, in which health-related apps only accounted for a minor percentage of apps downloaded [13,15,21]. To our knowledge, there are several reasons that can be used to explain this phenomenon. First, the youth might be more attracted by the apps that can entertain them after learning or working time rather than health-related apps. Second, there is a lack of mHealth apps in the Vietnamese language, which would cause difficulties for youth and adolescents in navigating the content (in a foreign language) and deciding which apps are the most suitable for them. Meanwhile, the few existing health-related apps developed by Vietnamese firms available in app stores such as eDoctor and Bacsi24 mainly have been used to connect health care facilities and patients, and lack specific content for youth and adolescents. Finally, most existing mHealth apps have not had their effectiveness evaluated nor have been officially approved by health care authorities, which could undermine the acceptability of the public. Currently, the Vietnam Ministry of Health has drafted a policy to regulate the implementation of distance health care (including the dissemination of mHealth apps). Quality assurance from a government body would be expected to increase the demand for health-related apps among Vietnamese youth and adolescents.

Despite the low download rates of health-related apps, the majority of people who actually downloaded and used these apps reported finding the apps useful (72.7%, 48/356) and were satisfied with their decision (61.9%, 39/356). This result is more positive than what was indicated in an American study, which reported that most participants who downloaded mHealth apps condemned them as irrelevant and/or user unfriendly and did not utilize the downloaded apps [15]. Such high satisfaction found in our study may be the result of a number of factors—the relevance of existing health apps, open-mindedness and acceptability of these particular youth and adolescents regarding mHealth apps, or possibly a high level of concern toward health among these particular individuals—but it suggests that mHealth apps can be accepted and helpful for users, perhaps with content enhanced to be more specific and design improved to be more youth appealing.

Enhancing the content of the mHealth apps would be of great importance in changing the behavior of people toward using these apps. Our study found an inverse relationship between perceived stress score and the EQ-5D index of participants against the likelihood of mHealth app utilization, implying a possible lack of relevant and perceived useful content and function within apps for both perceived healthy and stress-suffering individuals. Existing studies on mHealth have highlighted a major issue of mHealth apps not having been built based on evidence from sufficiently large empirical studies nor having health professionals involved in the development process [13]. Utilizing the knowledge and expertise of health professionals in building the mHealth apps would not only ensure the quality of app content, but also likely increase the acceptability of users knowing that the apps have been backed by more reliable professional know-how.

Incorporating other content into mHealth platforms such as games, music players, and social networks, which we found in our study to be the three most downloaded types of apps, would also likely help increase app acceptability and usage. Primack et al [34] found that electronic games could provide educational health messages and motivate adolescents and young adults to do physical activity and self-manage their health conditions. Meanwhile, social network/sharing functionalities would enable

users to share information/achievements onto their own personal social networks and facilitate learning from healthy lifestyle / keeping fit experiences, shared by their peers through similar platforms. Previous literature has shown that social networks and social influences have large impacts on behavior changes of Vietnamese youth [35] and are significant facilitators of mHealth app usage among adolescents and youths [15]. Indeed, the majority of our participants preferred having sharing/social network functions in their mHealth apps. Additionally, with regard to interface design, preference was found for visuals, which suggests the integration of multimedia with videos and images to elaborate apps would appeal to youth and adolescents [15].

Several implications can be drawn from this study. First, for researchers and developers, the involvement of end users and health professionals in developing and evaluating mHealth apps is essential. End users could inform their preferences regarding content, format, and display, as well as express their experiences in using and perception toward the apps, which are crucial for ensuring the relevance and appeal of the apps developed [36,37]. Meanwhile, health care professionals could facilitate the development and approval of these health-related apps, in particular those targeting disease prevention apps—the most downloaded types of mHealth apps as indicated by our results. Furthermore, the promising acceptability and usage of mHealth apps, coupled with the current popularity of mobile phone devices, social networks, and connection through the internet, suggests that physicians and health management authorities (eg, the Ministry of Health) should continue to pay more attention to this technology-enabled solution and consider it a key solution in health management and disease prevention. Support in terms of legal platforms and policies would encourage the creation and enhancement of apps by developers, while campaigns on benefits of the correct use of these health apps would help promote usage among the population, assisting self-management, early detection, and prevention of diseases.

The strength of this research is that we provided evidence about the usage patterns of mobile phone apps and identified user attitudes and preferences to conceptualize further mHealth apps. This type of study has not been conducted previously in Vietnam. However, there are several methodological limitations that should be acknowledged. First, due to the recruitment-via-internet method, our study might not be able to recruit people who do not have access to the internet, thus our sample cohort might not be entirely representative of the general population, which could undermine the ability to generalize our results. Moreover, our sample seems to be homogeneous, which makes it impossible to stratify our analysis into other demographic characteristics such as age and education. The WebRDS method produced shorter recruitment chains than expected, hence we could not establish equilibrium with this sample. Moreover, we did not develop a Web-based questionnaire that fits on tablets or mobile phones, which possibly impacted the ability to recruit more participants since these devices are increasingly popular. Furthermore, some features, namely input method, earning rewards, or connection between app users and health care providers, were not included in this study, suggesting further research is needed to provide comprehensive evidence about the preference for mHealth apps.

Finally, because we only researched general features of apps rather than particular content, such as physical activity [30], HIV [31], diabetes [32], or cystic fibrosis [33], some tailored features for these topics could not be studied. For example, patients with HIV required apps with automated reminders, motivational messages, mental support, and password protection [31]. Meanwhile, for apps promoting physical activity, Rabin and Bock [30] revealed that automated progress tracking, problem-solving messages, and user-friendly interfaces were those features most preferred. Most of the studies also highlighted some features that should be included for mHealth apps, such as communicating with health care providers and location tracking for contextualized messages [31-33]. Therefore, it appears that specific features for each condition and each population should be considered independently during development of mHealth apps.

This study found low proportions using mHealth-related mobile phone apps, but a high level of receptiveness and satisfaction among Vietnamese youth. Acceptance level and preferences toward mHealth apps as well as specifically preferred functionalities discovered in this study are essential not only in conceptualizing and developing appropriate mobile phone interventions targeting youth and adolescents, but also in the application of technically advanced solutions in disease prevention and health management.

Conflicts of Interest

None declared.

References

1. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
2. Statisticta. 2017 Dec 31. Number of smartphone users worldwide from 2014 to 2020 (in billions) URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2018-10-16] [WebCite Cache ID 73DIyL48a]
3. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2018-10-28] [WebCite Cache ID 73VaBnOeV]

4. Roser K, Schoeni A, Foerster M, Röösli M. Problematic mobile phone use of Swiss adolescents: is it linked with mental health or behaviour? Int J Public Health 2016 Apr;61(3):307-315. [doi: 10.1007/s00038-015-0751-2] [Medline: 26450576]
5. Pew Research Center. 2015 Apr 08. 73% of Teens gave access to a smartphone; 15% have only a basic phone URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/pi_2015-04-09_teensandtech_06/ [accessed 2018-10-17] [WebCite Cache ID 73Elwe9Hx]
6. Go-Globe. 2013. Smartphone usage in China - statistics and trends [blog] URL: https://www.go-globe.com/blog/smartphone-usage-in-china/ [accessed 2018-10-17] [WebCite Cache ID 73Em1rTVK]
7. Schweizer A, Berchtold A, Barrense-Dias Y, Akre C, Suris J. Adolescents with a smartphone sleep less than their peers. Eur J Pediatr 2017 Jan;176(1):131-136. [doi: 10.1007/s00431-016-2823-6] [Medline: 27981353]
8. Hardell L. Effects of mobile phones on children's and adolescents' health: a commentary. Child Dev 2018 Jan;89(1):137-140. [doi: 10.1111/cdev.12831] [Medline: 28504422]
9. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
10. Sage A, Roberts C, Geryk L, Sleath B, Tate D, Carpenter D. A self-regulation theory-based asthma management mobile app for adolescents: a usability assessment. JMIR Hum Factors 2017 Feb 01;4(1):e5 [FREE Full text] [doi: 10.2196/humanfactors.7133] [Medline: 28148471]
11. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
12. Brayboy L, Sepolen A, Mezoian T, Schultz L, Landgren-Mills BS, Spencer N, et al. Girl Talk: a smartphone application to teach sexual health education to adolescent girls. J Pediatr Adolesc Gynecol 2017 Feb;30(1):23-28 [FREE Full text] [doi: 10.1016/j.jpag.2016.06.011] [Medline: 27393638]
13. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
14. Dute D, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 05;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
15. Chan A, Kow R, Cheng J. Adolescents' perceptions on smartphone applications (apps) for health management. JournalMTM 2017 Aug;6(2):47-55. [doi: 10.7309/jmtm.6.2.6]
16. Hermawati S, Lawson G. Managing obesity through mobile phone applications: a state-of-the-art review from a user-centred design perspective. Pers Ubiquit Comput 2014 Jan 20;18(8):2003-2023. [doi: 10.1007/s00779-014-0757-4]
17. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
18. FinTechNews Singapore. 2015 Oct 21. Studies Vietnam: mobile phone continues to be Vietnam's no. 1 device for internet use URL: http://fintechnews.sg/855/studies/mobile-phone-vietnams-no-1-device-internet-use [accessed 2018-10-17] [WebCite Cache ID 73Emjw8Gh]
19. Tran BX, Houston S. Mobile phone-based antiretroviral adherence support in Vietnam: feasibility, patient's preference, and willingness-to-pay. AIDS Behav 2012 Oct;16(7):1988-1992. [doi: 10.1007/s10461-012-0271-5] [Medline: 22814571]
20. Statisticta. 2017. Number of smartphone users in Vietnam from 2015 to 2022 (in millions) URL: https://www.statista.com/statistics/467739/forecast-of-smartphone-users-in-vietnam/ [accessed 2018-10-17] [WebCite Cache ID 73Emnnjjw]
21. Davis B. Forbes. 2016 Feb 18. Growing smartphone ownership in Vietnam opens door for mobile market URL: https://www.forbes.com/sites/davisbrett/2016/02/18/growing-smartphone-ownership-in-vietnam-opens-door-for-mobile-marketers/ [accessed 2018-10-17] [WebCite Cache ID 73EmsQ34f]
22. Nguyen D, Dedding C, Pham TT, Wright P, Bunders J. Depression, anxiety, and suicidal ideation among Vietnamese secondary school students and proposed solutions: a cross-sectional study. BMC Public Health 2013 Dec 17;13:1195 [FREE Full text] [doi: 10.1186/1471-2458-13-1195] [Medline: 24341792]
23. Do TH, Le LC, Burgess JA, Bui DS. Determinants of condom use at sexual debut among young Vietnamese. Int J Adolesc Med Health 2014;26(3):403-410. [doi: 10.1515/ijamh-2013-0322] [Medline: 24451069]
24. Jordan LP, Graham E, Vinh ND. Alcohol use among very early adolescents in Vietnam: what difference does parental migration make? Asian Pac Migr J 2013 Dec 01;22(3):401-419 [FREE Full text] [Medline: 24954964]
25. Nguyen PV, Hong TK, Hoang T, Nguyen DT, Robert AR. High prevalence of overweight among adolescents in Ho Chi Minh City, Vietnam. BMC Public Health 2013 Feb 15;13:141 [FREE Full text] [doi: 10.1186/1471-2458-13-141] [Medline: 23414441]
26. Huong LT, Vu NT, Dung NN, Xuan LTT, Giang KB, Hai PT, et al. Cigarette smoking among adolescents aged 13-15 in Viet Nam and correlates of current cigarette smoking: results from GYTS 2014 data. Asian Pac J Cancer Prev 2016;17(S1):17-23 [FREE Full text] [Medline: 27087178]

27. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838 [FREE Full text] [Medline: 22846248]
28. Wejnert C, Heckathorn DD. Web-based network sampling. Sociol Method Res 2008 Aug;37(1):105-134. [doi: 10.1177/0049124108318333]
29. Kubal A, Shvab I, Wojtynska A. Initiation of the RDS recruitment process: seed selection and role. In: Tyldum G, Johnston LG, editors. Applying Respondent Driven Sampling to Migrant Populations: Lessons from the Field. London: Palgrave Pivot; 2014:37-48.
30. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
31. Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers. Int J Med Inform 2013 Apr;82(4):e38-e46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.05.009] [Medline: 22704234]
32. King D, Toobert DJ, Portz JD, Strycker LA, Doty A, Martin C, et al. What patients want: relevant health information technology for diabetes self-management. Health Technol 2012 Mar 5;2(3):147-157. [doi: 10.1007/s12553-012-0022-7]
33. Hilliard M, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mHealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
34. Primack B, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
35. Tran BX, Huong LT, Hinh ND, Nguyen LH, Le BN, Nong VM, et al. A study on the influence of internet addiction and online interpersonal influences on health-related quality of life in young Vietnamese. BMC Public Health 2017 Dec 31;17(1):138 [FREE Full text] [doi: 10.1186/s12889-016-3983-z] [Medline: 28143462]
36. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
37. Arsand E, Demiris G. User-centered methods for designing patient-centric self-help tools. Inform Health Soc Care 2008 Sep;33(3):158-169. [doi: 10.1080/17538150802457562] [Medline: 18850399]

Abbreviations

BMI: body mass index
EQ-5D: EuroQol-5 dimensions
HRQOL: health-related quality of life
mHealth: mobile health
PSS: Perceived Stress Scale
RDS: respondent-driven sampling
WebRDS: Web-based respondent-driven sampling

Edited by G Eysenbach; submitted 12.08.16; peer-reviewed by A Mosa, K D'Souza, J Wang; comments to author 05.12.16; revised version received 31.07.18; accepted 28.08.18; published 30.11.18.

Please cite as:
Tran BX, Zhang MWB, Le HT, Nguyen HD, Nguyen LH, Nguyen QLT, Tran TD, Latkin CA, Ho RCM
What Drives Young Vietnamese to Use Mobile Health Innovations? Implications for Health Communication and Behavioral Interventions
JMIR Mhealth Uhealth 2018;6(11):e194
URL: http://mhealth.jmir.org/2018/11/e194/
doi:10.2196/mhealth.6490
PMID:30251704

Original Paper

The Perceived Benefits of an Artificial Intelligence–Embedded Mobile App Implementing Evidence-Based Guidelines for the Self-Management of Chronic Neck and Back Pain: Observational Study

Wai Leung Ambrose Lo[1,2], MSc, PhD; Di Lei[1], MSc; Le Li[1], PhD; Dong Feng Huang[1], MD; Kin-Fai Tong[2], PhD

[1]Guangdong Engineering and Technology Research Center for Rehabilitation Medicine and Translation, Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

[2]Department of Electronic and Electrical Engineering, University College London, London, United Kingdom

Corresponding Author:
Wai Leung Ambrose Lo, MSc, PhD
Guangdong Engineering and Technology Research Center for Rehabilitation Medicine and Translation
Department of Rehabilitation Medicine
The First Affiliated Hospital, Sun Yat-sen University
58 Zhongshan Second Road
Guangzhou, 510000
China
Phone: 86 87332200 ext 8536
Email: ambroselo0726@outlook.com

Abstract

Background: Chronic musculoskeletal neck and back pain are disabling conditions among adults. Use of technology has been suggested as an alternative way to increase adherence to exercise therapy, which may improve clinical outcomes.

Objective: The aim was to investigate the self-perceived benefits of an artificial intelligence (AI)–embedded mobile app to self-manage chronic neck and back pain.

Methods: A total of 161 participants responded to the invitation. The evaluation questionnaire included 14 questions that were intended to explore if using the AI rehabilitation system may (1) increase time spent on therapeutic exercise, (2) affect pain level (assessed by the 0-10 Numerical Pain Rating Scale), and (3) reduce the need for other interventions.

Results: An increase in time spent on therapeutic exercise per day was observed. The median Numerical Pain Rating Scale scores were 6 (interquartile range [IQR] 5-8) before and 4 (IQR 3-6) after using the AI-embedded mobile app (95% CI 1.18-1.81). A 3-point reduction was reported by the participants who used the AI-embedded mobile app for more than 6 months. Reduction in the usage of other interventions while using the AI-embedded mobile app was also reported.

Conclusions: This study demonstrated the positive self-perceived beneficiary effect of using the AI-embedded mobile app to provide a personalized therapeutic exercise program. The positive results suggest that it at least warrants further study to investigate the physiological effect of the AI-embedded mobile app and how it compares with routine clinical care.

(JMIR Mhealth Uhealth 2018;6(11):e198) doi:10.2196/mhealth.8127

KEYWORDS

low back pain; neck pain; mobile app; exercise therapy; mHealth

Introduction

Chronic musculoskeletal neck and back pain are prevalent and disabling conditions among adults [1]. The report on global economic burden of disease published in 2010 indicated low back pain and neck pain ranked sixth and 21st, respectively, out of 291 diseases and injuries [2] for medical cost. The studies published in 2016 indicated that combined low back and neck pain ranked fourth out of 315 diseases and injuries [3] for causes of years lived with disability. Thus, there is an urgent need to identify an effective means to promote self-management of these conditions.

Recent advancement in information and communication technology (ICT) places strong emphasis on supporting self-management programs for chronic pain through mobile software apps [4,5]. Mobile apps have been identified as a way to reduce the financial burden of chronic neck and back pain [6]. Use of ICT as part of rehabilitation programs has been suggested to increase adherence to self-management programs [7]. However, published studies have indicated that approximately 65% of the neck and back pain-related mobile apps have no evidence of the involvement of health care professionals [8,9] in their development.

To date, only one study was found that investigated the benefits of a professionally designed mobile app that had been trialed. This study tested the effect of using a mobile app on and its benefit in self-management of low back pain [10]. The app was designed with a group of pain professionals and utilized a cognitive behavior approach based on recommendations from the American Pain Society [11]. A total of 597 participants with chronic low back pain who had not received any intervention for the condition were recruited. They were randomly assigned to intervention, routine care, and control group. Outcome measures included level of pain and pain frequency. Results indicated greater improvement at 4-month follow-up in both parameters compared to the control group. This study demonstrated that a mobile app may be an effective tool in self-management of low back pain. Although this study provided positive results for the use of mobile app, it excluded the majority of people with chronic back pain that had some sort of prior intervention. In addition, the exercise program appeared to be generic exercises rather than ones specific to the users' symptoms.

Other studies are also trialing different mobile apps that focus on relaxation therapy [12], with a combination of mobile app and clinician offline support [6]. Results of these studies are yet to be published. Currently, there are a large number of health-related mobile apps on the market, but the majority of the neck and back pain apps have focused on pain management education. Other mobile apps claim to offer exercises and education for chronic neck and back pain but provide limited evidence that a health care professional was involved in their development [13].

Our institute has taken part in developing an artificial intelligence (AI) system embedded into a mobile app called "Well Health." The AI-embedded mobile app was developed in response to the "World Report on Disability" [14], which suggested that electronic health (eHealth) or telerehabilitation techniques are effective means to enable people to receive appropriate intervention. However, it also stated that telerehabilitation should be tested to assess its feasibility within the local culture. As documented in a survey of the Chinese population with chronic low back pain (N=113) in 2016, it was found that self-management behaviors were poor [15]. The survey identified that the contributing factors to poor self-management behavior included a lack of disease knowledge, a lack of understanding of the benefits of exercise, and a lack of communication with health professionals. The AI-embedded mobile app was designed to address some of these factors. It allows users to input their symptoms and generates an exercise program accordingly. The exercise program is delivered in the form of a video that plays in real time.

The aim of this study was to evaluate the feasibility of the AI-embedded mobile app Well Health to assist people with chronic neck and back pain to self-manage their condition, and the perceived benefits in symptom improvement by its users. This is the first study that evaluates an AI-embedded mobile app for neck and back pain rehabilitation among the Chinese population.

Methods

Ethical Consideration

The study was approved by the Medical Ethical Committee for Clinical Research and Animal Trials of the First Affiliated Hospital of Sun Yat-sen University (reg # [2016] 185). Informed consent was obtained electronically. Once the users clicked on the survey invitation icon, they were directed to an informed consent form. Informed consent was given by selecting "Agree to take part" and "Submit." Each participant was assigned a number to maintain anonymity.

Recruitment

This study was a retrospective evaluation study of an AI-embedded mobile app that was available to the general public. All existing users of the mobile app were invited to take part via the in-app "Information" section. The invitation also appeared on the home page where it would be displayed among five rolling pictures of different newsfeeds when the app was loaded each time. Interested users could click on the link which directed them to the evaluation questionnaire. All the participants in this study completed the questionnaire one time, with retrospective recall of preintervention pain level.

Sample Population

The inclusion criteria for the study were adults (1) aged between 18 and 65 years, (2) who had experienced neck and low back pain within the past 3 months, and (3) had access to a mobile phone that could play video on the internet. This study excluded patients who were medically unstable and reported having red flags for cervical and lumbar spine pathology. A list of red flag symptoms related to the cervical and lumber spine were displayed when the app was opened the first time and before the start of the questionnaire. The user would not be able to continue with the survey if more than two red flag symptoms were selected. These inclusion criteria were set to include a wide population because the app was designed to be accessed by members of the general public with experience with chronic neck and back pain.

Evaluation Questionnaire

The questionnaire included 14 questions that were intended to evaluate if using the AI-embedded mobile app may (1) increase adherence to therapeutic exercises, (2) affect pain level, and (3) reduce the need for other interventions. The World Health Organization defined the term "adherence" as "the extent to which patient behavior taking medication, following a diet, and/or executing lifestyle changes, corresponds with recommendations from a health care provider" in 2003 [16].

Textbox 1. The evaluation questions.

1. Have you experienced neck and low back pain within the past 3 months? *Responses: (1) yes; (2) no*

2. Are you age between: *Responses: (1) 18-25; (2) 26-30; (3) 31-40; (4) 41-50; (5) 51-60; (6) 60-65?*

3. What is your gender? *Responses: (1) male; (2) female*

4. How long have you used the AI-embedded mobile app? Responses: *(1) 1 day; (2) 1 week; (3) 1 month; (4) 3 months; (5) 6 months or over*

5. On a scale of 0-10 (0 being no pain, 10 being in extreme pain), how would you rate your pain level before using the rehabilitation program on the AI-embedded mobile app? *Responses: 0 to 10 Likert scale*

6. On a scale of 0-10, how would you rate your pain level after using the AI-embedded mobile app for rehabilitation program? *Responses: 0 to 10 Likert scale*

7. Prior to using the app, have you participated in any rehabilitation exercise program? *Responses: (1) yes; (2) no*

If "Yes," how much time did you spend on rehabilitation exercise a day on average? *Response: (1) 5 minutes or below; (2) 6 to 10 minutes; (3) between 10 and 30 minutes; (4) between 30 and 60 minutes; (5) 60 minutes or above*

8. How much time in a day have you followed the AI-embedded mobile app to perform the recommended exercise? *Responses: Amount of time entered by the user (value between 0 to 180 minutes)*

9. On a scale of 0-100 (0 being no improvement, 100 being completely resolved), how would you rate the overall improvement of your symptoms? *Responses: 0 to 10 Likert scale*

10. Have you received any of the following interventions prior to using the AI-embedded mobile app? *Responses: (1) acupuncture; (2) soft tissue therapy; (3) topical cream, (4) medication; (5) electrotherapy*

11. Have you received any of the following interventions prior to using the AI-embedded mobile app? *Responses: (1) acupuncture; (2) soft tissue therapy; (3) topical cream, (4) medication; (5) electrotherapy*

12. While using the AI-embedded mobile app, have you continued with any of the interventions mentioned? *Responses: (1) acupuncture; (2) soft tissue therapy; (3) topical cream, (4) medication; (5) electrotherapy; (6) no, I did not use any other intervention while using the AI-embedded mobile app*

13. Have you read the education material within the app? *Responses: (1) yes; (2) no.*

If "yes," how much time in a day have you spent on reading the educational material? *Responses: amount of time entered by the user (value between 0 and 180 minutes)*

14. Would those materials encourage you to adhere to the therapeutic exercise? *Responses: (1) yes; (2) no*

Based on this original definition, Donkin et al [17] provided a further definition of "the degree to which the user followed the program as it was designed" to incorporate e-therapies in 2011. A systematic review on the chronic low back pain population also found self-report diaries to be the most common measure of adherence [18]. A published study indicated a self-reported exercise log has acceptable agreement with objectively assessed exercise adherence for both exercise frequency and exercise duration [19]. Therefore, the operating definition of adherence in this study was the amount of "self-reported time spent on therapeutic exercises in a day" because the AI-embedded mobile app was designed to encourage daily therapeutic exercise.

The evaluation questions are presented in Textbox 1. The first two questions were used to assess if the responses fit the inclusion criteria.

In addition to clinical information, participants were asked to fill in the modified System Usability Scale (SUS). The scale had 10 questions that were validated to assess the generic usability of a product or a service [20]. The questions are as follows:

1. I think that I would like to use this product frequently.
2. I found the product unnecessarily complex.
3. I thought the product was easy to use.
4. I think that I would need the support of a technical person to able to use this product.
5. I found the various functions in this product were well integrated.
6. I thought there was too much inconsistency in this product.
7. I would imagine that most people would learn to use this product very quickly.
8. I found the system very awkward to use.
9. I felt very confident using the product.
10. I needed to learn a lot of things before I could get going with the product.

The SUS showed the domains as five scales numbered from 1 (strongly disagree) to 5 (strongly agree). Positively worded domains were equal to score–1, negatively worded domains were equal to 5–score, so the total SUS was calculated by summing the 10 domains and multiplying by 2.5.

Intervention: Well Health Mobile App

Credential and Affiliations of the Developers

The affiliations of the developers and the expert panel that were involved in developing the AI-embedded mobile app were clearly displayed within the "About us" section within the mobile app. The embedded AI system, including the assessment method, the exercise programs, and exercise instructions, were developed with an expert panel consisting of physiotherapists from Australia (Maxvale Physiotherapy Practice), the United Kingdom (University College London), and China (Sun Yat-sen University), including a clinical scientist and a rehabilitation

doctor from China (Sun Yat-sen University). Technical support in developing and maintaining the app and AI coding was done in collaboration with the information technology company "Well Health" (Ying Kang Wei Er Internet Service Co, Ltd, Guangzhou, China).

Development Process

The AI-embedded mobile app underwent 10 months of development. Members of the expert panel met in China for 3 weeks for residential development. During the residential period, the clinical content of the AI-embedded mobile app was determined. This included the development of the AI algorithm (initial training sample preparation, neuron processing methods), production of the exercise videos and their associated descriptions, and the education material. This was followed by 3 months of AI model validation. During this phase, experts went through computer-simulated data to ensure the AI-generated exercise programs were appropriate for the presented symptoms. The development of the mobile app interface took a further 6 months.

System usability and interface development of the app were conducted in the form of a semi-structured telephone interview (n=72) and focus group (n=8) prior to this study. The interview and focus groups were designed to gather information on (1) what users like to use the app for (eg, just browse for information or look for rehabilitation exercises for their back pain); (2) what additional functions they would like to have in the app; (3) how do they like the exercise content to be delivered (eg, exercises to be described by text or displayed as picture accompanied by text or in video with verbal guidance; (4) any in-app functions that users felt were missing; and (5) users' feelings on the app interface. No update was made to the content of the AI-embedded mobile app during the data collection phase.

The AI-embedded mobile app is available free of charge online to the general public in the US, UK, and China through the Apple app store and MI app store. The app is searchable using the keyword "Well Health."

Access

Access to the mobile app was completely free. The AI-embedded mobile app could be downloaded from any well-known app store. Participants were required to register an account using their mobile phone number or social media account. Technical support was provided by Well Health in registering for a user account, if needed.

Intervention Components

Physical Component

Well Health is a multiple-visit mobile app that provides adults with neck and back pain a tailored exercise rehabilitation program based on their presenting symptoms.

The AI-embedded mobile app follows the recommendations of the National Institute of Clinical Excellence (NICE) [21] in providing a combination of physical and self-management advice as well as the pathological causes of low back pain. It allows users to enter their symptoms using a self-report questionnaire. The questionnaire consists of the questions which physiotherapists routinely use to conduct a subjective assessment. The questionnaire asks for information on present condition, history of present condition, past investigations, 24-hour pattern, drug history, and social history. Once the subjective assessment is completed, the AI-embedded mobile app provides an exercise program based on the information provided. Exercise interventions provided by the AI-embedded mobile app were NICE guideline-recommended biomechanical exercises including stretching, (eg, lumbar flexion and neck flexion); motor control exercises, including Pilates exercises since their principles overlap with principles of motor control interventions [22] (eg, deep neck flexors exercise, transverse abdominis activation exercise); and strengthening exercises (eg, cervical side flexion exercise, side-lying trunk exercise). Figure 1 shows screenshots of these exercises. The suggested exercise duration was between 20 to 30 minutes a day.

Each exercise video was accompanied by detailed descriptions of the exercises and annotated diagrams were displayed to visually demonstrate the targeted muscles (see Multimedia Appendix 1). Users were expected to follow the video clip from start to finish once they started the set of training. The video clip could not be scrolled through, but could be paused/stopped. Points were awarded when the full length of the video was played.

The AI-embedded mobile app determined the exercise program based on the subjective information. The AI algorithm used in this study was based on the multilayered perceptron artificial neural network (MLP-ANN). It is the most commonly used AI algorithm in the medical field [23] and is capable of learning from historical examples, analyzing nonlinear data, and handling imprecise information [24]. The AI algorithm observed the data input from the subjective assessment (input layer), processed the information through the neurons to select the appropriate therapeutic exercises from the exercise bank (hidden layer), and then gave out the most appropriate therapeutic program (output layer). Figure 2 shows the architecture of the MLP-ANN of the AI used in this study. A total of 300 sets of training samples were initially provided by the experts during the expert meeting. Once the initial weighting of the MLP-ANN was trained by the samples provided by the experts, the back-propagation algorithm [25] was used to continue the training until an accuracy of at least 80% was achieved. Model validation was performed by comparing the AI-generated exercise program with the expert-generated exercise program.

Figure 1. Screenshots of stretching, motor control, and strengthening exercises: (a) assisted neck flexion stretch, (b) lumbar flexion stretch, (c) deep neck flexors exercise, (d) transverse abdominis activation exercise, (e) cervical side flexion exercise, and (f) side-lying trunk exercise.

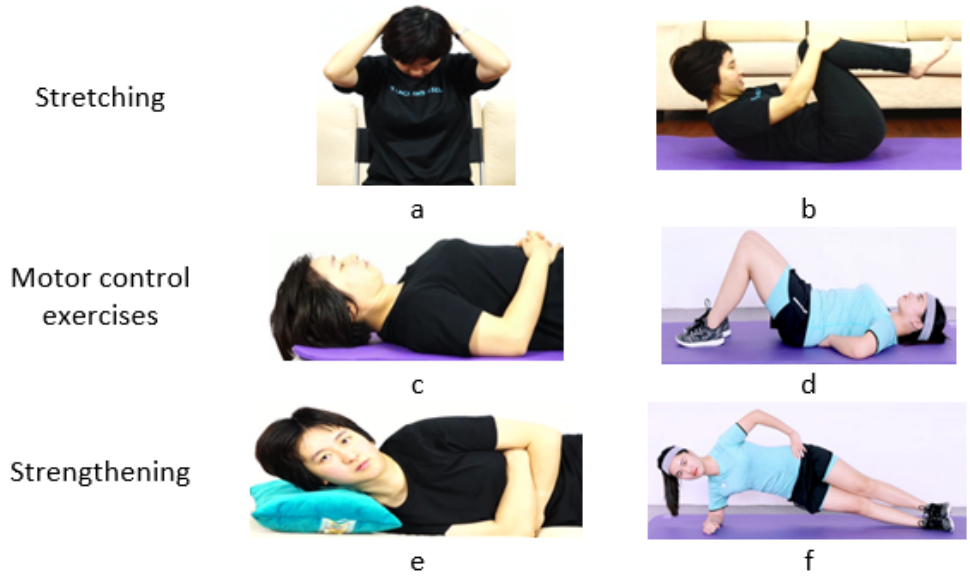
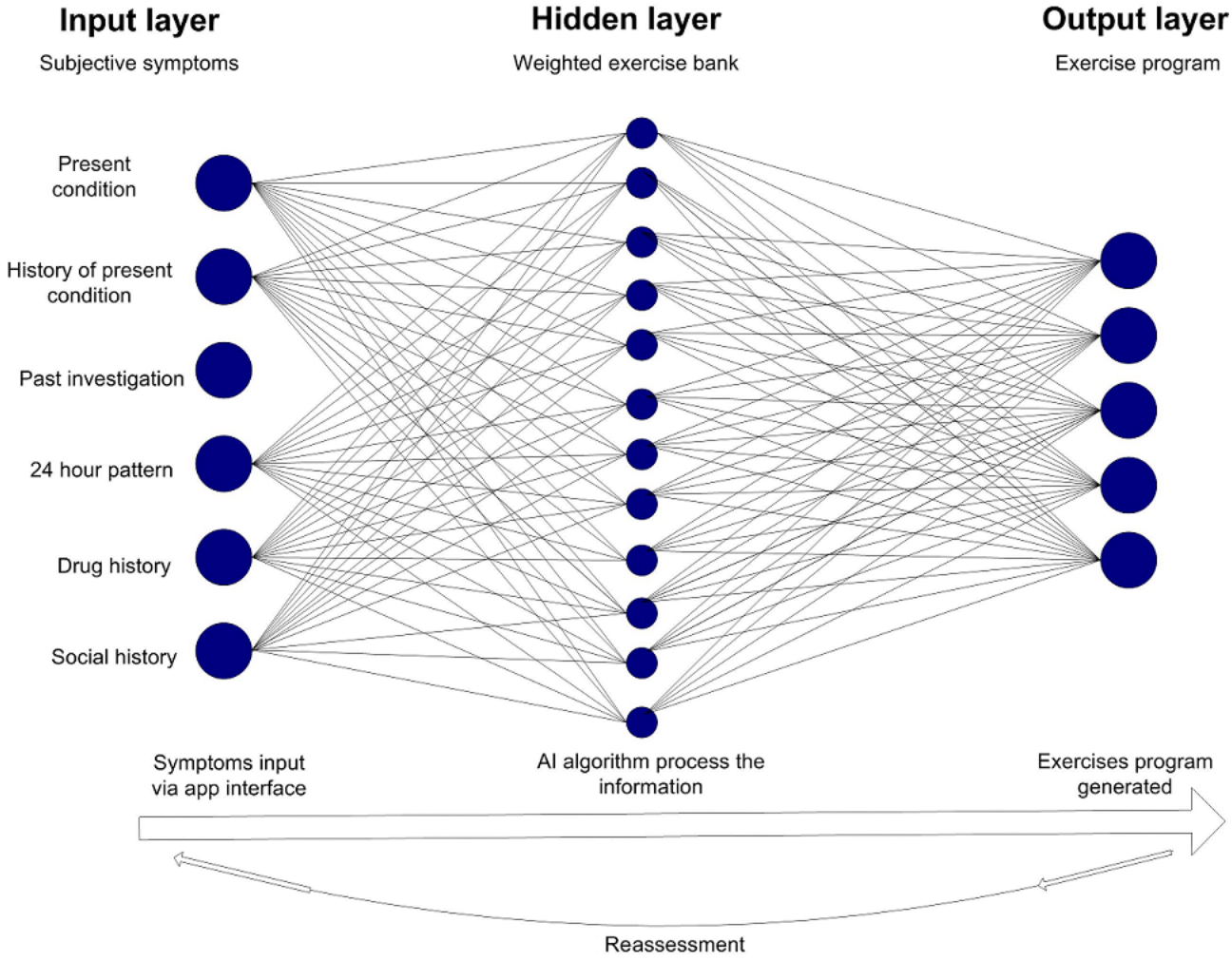


Figure 2. The architecture of the multilayered perceptron artificial neural network used in this study.



The basic principle of the AI algorithm was based on the following: (1) each symptom was assigned a weighting in accordance to the clinical importance as determined by the expert panel (ie, the pain score had a higher weighting than social history); (2) each therapeutic exercise was assigned a weighting in accordance to their effectiveness to each of the presented symptoms as deemed by the expert panel; (3) the AI algorithm then selected the five highest scoring exercises from

the database. Users were prompted to undertake reassessment every 14 days to report their progress and indicate the exercises that they found beneficial. The machine-learning algorithm would adjust the weighting of the exercise according to the individual's feedback and update the exercise program based on the input of the reassessment. In the case of acute symptoms worsening, the algorithm would adjust the exercise plan according to the new symptoms. All AI-embedded mobile app users could contact members of the medical team who were involved in developing the app via the in-app messaging function. As part of the disclaimer of using the AI-embedded mobile app, users were advised that it did not replace medical service and medical support should be sought at any time should they feel the need to do so.

A points-based reward was utilized to promote engagement with the AI-embedded mobile app. Incentivized conditions have been shown to be effective in engaging users in goal-directed behavior [26]. Points were awarded according to the daily task completed (see Multimedia Appendix 2). The task list included logging onto the app, completing the number of sets of exercise, sharing their exercise progress on the social media account, and posting a comment about the educational material they read. Users needed to collect sufficient points to "unlock" the next level. The use of a point-based system aimed to increase engagement with the AI-embedded mobile app. The AI-embedded mobile app automatically recorded the number of days and the amount of time users spent on playing the exercise videos. An exercise log was kept and users could track their progress at any time.

Educational Component

Educational material was "pushed" to users via a social media platform once every 3 days. This material included information about neck and back pain pathology, the pathological cause of back pain and the natural course of low back pain, pain physiology, principles to use exercise as an intervention to manage neck and back pain, and coping strategies. These materials were stored within the app and users could access these materials at any time. Users received one reminder daily via the app reminder function which the user could turn off.

Data Analysis

Data were analyzed in SPSS version 20.0 (IBM, Armonk, NY, USA) for Windows (Microsoft 10). Descriptive statistics were used to describe the sample population and gave a summary on all responses.

Changes in pain score and perceived self-improvement were analyzed as a group and as a subgroup based by the total duration of using the AI-embedded mobile app. Wilcoxon rank sum test was used to assess if the changes in Numerical Pain Rating Scale and self-perceived improvements were statistically significant.

Results

Overview

During the data collection period, the AI-embedded mobile app had 461 active users. A total of 161 users responded to the invitation. Three responses were excluded due to exclusion criteria (reported to be younger than 18 years of age). The sample population contained 119 males and 39 females. Of these, 30 and 31 responses reported to be in the age group of 18 to 25 years and 26 to 30 years, respectively. The age group between 31 and 40 years had the greatest number of responses (n=56). Nineteen and 21 responses reported to be in the age group of 41 to 50 years and 51 to 60 years, respectively. One response reported to be older than 60 years.

Time Spent on Therapeutic Exercises

Most respondents reported they used the AI-embedded mobile app for a month. In all, 14.6% (23/158) and 13.3% (21/158) of respondents indicated they used the AI-embedded mobile app for 3 months and 6 months or more, respectively. The responses from question 3 of the evaluation questionnaire were 1 day (n=35), 1 week (n=38), 1 month (n=41), 3 months (n=23), and 6 months or more (n=21).

A total of 60 users (37.98%) reported they had never participated in therapeutic exercises prior to using the AI-embedded mobile app. When using the AI-embedded mobile app, the mean time spent on rehabilitation exercises was 25 (SD 4) minutes per day. An increased number of responses was observed in the categories of 10 to 30 minutes and 30 to 60 minutes, but there was a decreased number of responses for the more than 60 minutes category. Figure 3 is a bar graph showing the self-reported differences in time spent on therapeutic exercise before and when using the AI-embedded mobile app.

Time Spent on Reading the Education Material

Overall, 142 users (89.9%) reported they had read the education material within the mobile app. Of these, 123 users (77.8%) indicated that the educational material had encouraged them to perform the AI-guided therapeutic exercise program. The mean time spent on reading the educational materials was 15 (SD 14) minutes a day.

Numerical Pain Rating Scale

An overall reduction in the Numerical Pain Rating Scale score was reported from users after using the AI-embedded mobile app for therapeutic exercises. The median Numerical Pain Rating Scale scores were 6 (interquartile range [IQR] 5-8) before and 4 (IQR 3-6) after using the AI-embedded mobile app (95% CI 1.18-1.81). A Wilcoxon rank sum test indicated the reduction in Numerical Pain Rating Scale score was statistically significant (P=.04).

The greatest reduction in Numerical Pain Rating Scale scores was reported from users who used the AI-embedded mobile app for 6 months. Figure 4 shows the changes in self-reported Numerical Pain Rating Scale scores reported by users with different tenures of using the AI-embedded mobile app.

Figure 3. Bar graph to illustrate the self-reported time spent on therapeutic exercises before and when using the AI-embedded mobile app.

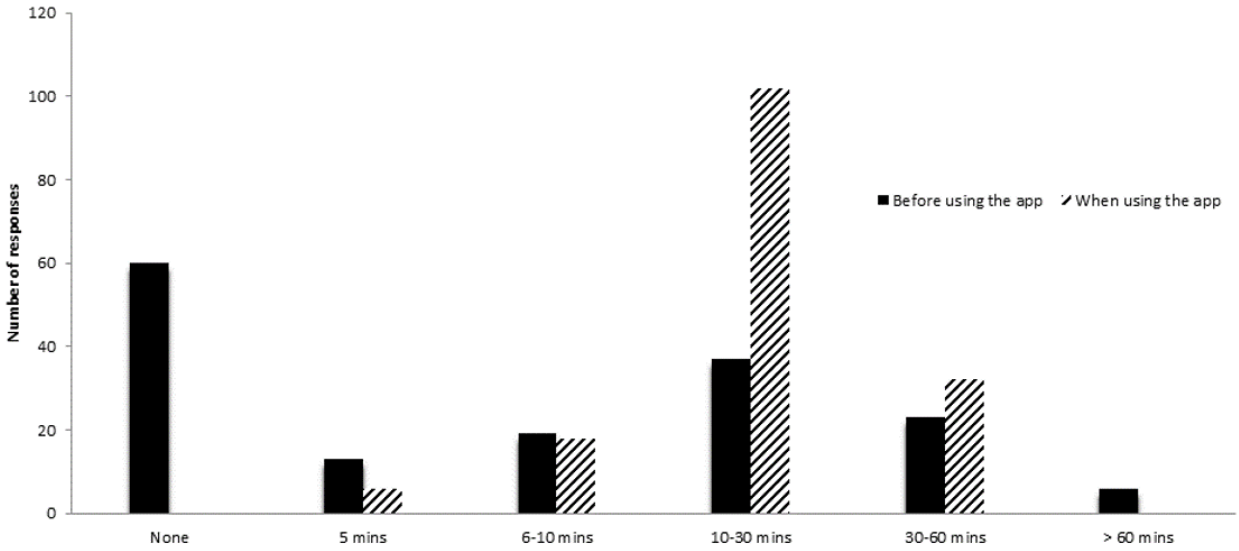
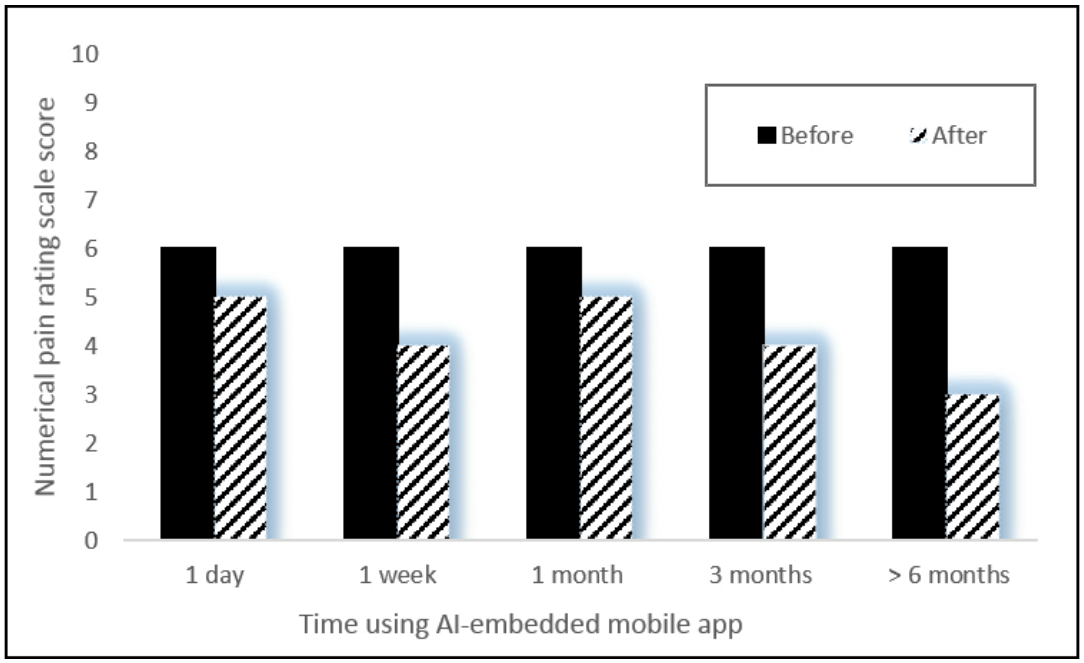


Figure 4. The reported difference in Numerical Pain Rating Scale changes reported by participants of different tenures.



Self-Perceived Improvement

The mean self-perceived improvement from all users was 65%. Users who utilized the AI-embedded mobile app for 1 day, 1 week, and 1 month reported improvements of 60.2% (SD 27.9%), 49.5% (SD 29.2%), and 47.2% (SD 26.4%), respectively. For users who used the app for over 3 and 6 months, improvements were reported to be 58.6% (SD 30.4%) and 71.1% (SD 25.8%), respectively.

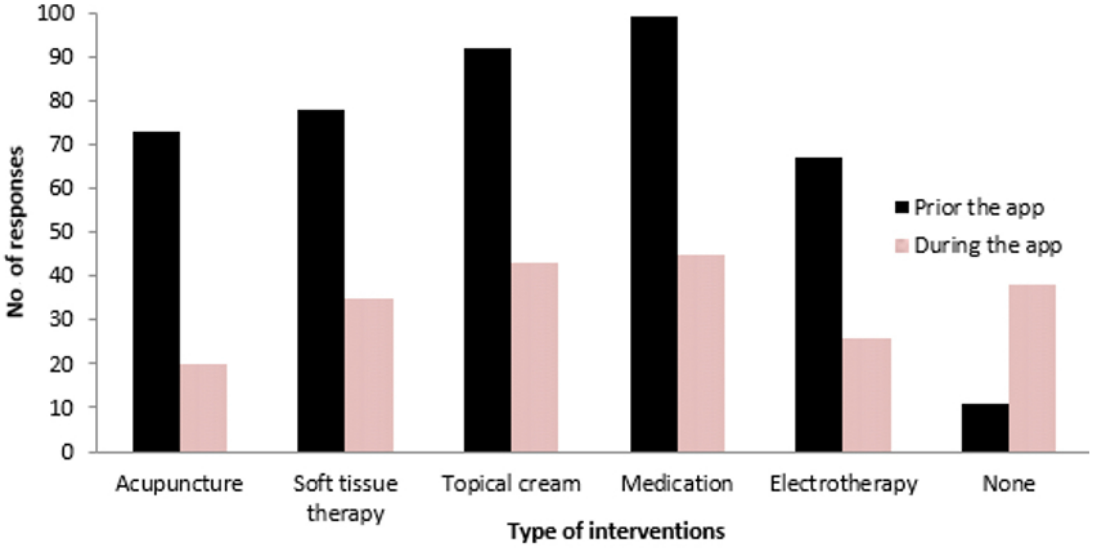
Usage of Other Interventions

A total of 11 respondents (7.0%) indicated that they had not received prior intervention from the listed modalities and 38 respondents (24.1%) indicated they did not receive any other intervention while using the AI-embedded mobile app. A reduction in responses was observed in all the listed interventions when using the AI-embedded mobile app. Figure 5 shows the number of responses of who had used other interventions prior to and when using the AI-embedded mobile app.

System Usability Scale

The median of the SUS from all responses was 73 (IQR 60-85), which would suggest an acceptable level of usability (score >68). All users, except for those who used the AI-embedded mobile app for a day (median 65, IQR 53-80), reported acceptable usability (score >68).

Figure 5. Bar graph to illustrate the number of responses of other interventions used prior to and when using the AI-embedded mobile app.



Discussion

Principal Findings

The main findings of this preliminary evaluation study were (1) users reported having spent more time on therapeutic exercises when using the AI-embedded mobile app, (2) an overall reduction in pain level was reported after using the AI-embedded mobile app, and (3) the usage of other interventions was reported to have been reduced with the use of the AI-embedded mobile app. This is among the first studies that investigated the feasibility and benefits of an AI-embedded mobile app for the management of chronic neck and low back pain.

Adherence to Therapeutic Exercise

Exercise therapy had been recommended as a core component in the management of chronic neck and back pain by several guidelines and reviews [27-29]. It was an unexpected finding that 38% of the users indicated that they had never participated in therapeutic exercises before. It was possible that some of the users might live in a remote part of China where rehabilitation service was not widely available. The use of the AI-embedded mobile app may therefore play a role in increasing access to exercise therapy intervention, as supported by the reported increase in time engaged in therapeutic exercises. Adherence to a therapeutic exercise program in the management of chronic neck and low back pain has been a long-standing issue. Early literature showed that between 50% [30] and 70% [31] of people with chronic low back pain were nonadherent to a home exercise program. The use of technologies has been suggested to be a way to increase adherence to a home exercise program [32]. Previous studies that investigated the effectiveness of using mobile phone text message reminders indicated continued adherence to exercises in adults with recurrent low back pain [7]. The AI-embedded mobile app also had a reminder function that reminded users to perform daily exercise. A survey conducted in 2016 on the Chinese population with chronic low back pain (N=113) reported that self-management behaviors were poor [15] and contributing factors included lack of disease knowledge and lack of understanding of the benefit of exercise. The educational material may therefore play a role in promoting regular therapeutic exercise because it provides information on these areas. The majority of users indicated the material motivated them to do therapeutic exercise. The increase in self-reported time spent on exercises observed in this study may suggest its potential to be a useful tool for patient adherence to therapeutic exercise programs, which may contribute to improving clinical outcomes.

Pain Level

A pain level reduction of 2 points (on a 0-10 Numerical Pain Rating Scale) is considered to be “much improved” [33] and a “meaningful” change [34] in people with chronic pain. The reduction in pain score was consistent with the finding of the Cochrane Review that indicated exercise, including motor control, flexibility, and strengthening exercises, may improve pain severity for people with chronic pain [35].

The difference in self-reported pain ratings was also consistent with a study that investigated the benefit of the combination of manual therapy and exercise therapy alone [36]. In this study, Alfuth and Welsink [36] investigated the pain level in people with chronic low back pain who attended physiotherapy practice for interventions. Interventions included a combination of manual therapy, stabilization exercises, and electrotherapy. A total of 85 participants received a mean number of treatment sessions of 42.6 (SD 2.3) within a year. A reduction of 3 points and 4 points on the Numerical Pain Rating Scale was observed after 3 and 6 months of treatment, respectively, when compared to baseline. Compared to the 2- and 3-point difference reported by users who used the AI-embedded mobile app for 3 and 6 months in this study, it would appear that using the AI-embedded mobile app to manage low back pain is feasible. An earlier study by Cambron et al [37] investigated the effect of a 4-week course of an active trunk extension exercise program at 3 months and 6 months postintervention. They reported a mean pain score reduction of 1.36 and 1.19 at 3 and 6 months postintervention, respectively. The difference in pain score

reduction may be related to the type of exercises that participants were asked to do. The AI-embedded mobile app selected from a range of progressive exercises, whereas participants in the Cambron et al study [37] were only asked to do trunk extension exercises.

Users' self-ratings of improvement of interventions is one of the core domains to be used in assessing impact of an intervention on chronic pain [38]. The self-rating of symptom improvement observed in this study corresponded to the reduction in pain score. The greatest symptom improvement was reported from users who had used the program for 6 months, followed by 3 months. A study published in 2010 reviewed 15 randomized controlled trials that were published between 2001 and 2007 [39]. The study summarized literature that reported long-term outcome data of physiotherapy exercise programs for adults with chronic low back pain. Trials were reviewed if they reported any outcome related to chronic low back pain measured at a follow-up period. Of the included trials, 13 reported follow-up data of up to 12 months and two trials reported follow-up data of up to 2 years. The results indicated consistent evidence for the effectiveness of an exercise program in pain reduction at the 6 months follow-up period when compared to a control group. Our study also observed the greatest self-reported pain score reduction and self-perceived improvement from users who used the AI-embedded mobile app for 6 months.

When compared to the only other trialed mobile app (FitBack) [40] that focused specifically on spinal pain management [10], the AI-embedded mobile app used in our study appeared to be at least as effective in reducing pain level. FitBack users reported a reduction in pain score from 2.86 at baseline to 2.64 and 2.16 at follow-up at 2 and 4 months, respectively. A Cochrane Review [35] and the NICE guideline [21] indicated that exercise programs should be tailored according to individual needs and capabilities. Therefore, the higher reduction of pain scores observed in this study may be related to the tailoring of the exercise content based on the presenting symptoms. But the higher magnitude of pain score reduction observed in our study should be interpreted with caution. The two studies have different designs, which may contribute to the differences in reported pain levels. It is typical for single-group data to show greater effects than randomized controlled trial data. The pain score data from our study may also contain recall bias [39], which further complicates the comparison.

Needs for Other Interventions

The results of our study indicate that using the app for the self-management of chronic neck and back pain might reduce the need for other interventions. The reported reduction in usage of other interventions was supported by an earlier study by Sculco et al [41]. Their study investigated the effect of 10 weeks of a prescribed home-based exercise program in people with chronic low back pain compared to a control group. The results suggested that participants in the exercise group had significantly lower numbers of physiotherapy referrals and medication use during the 2.5-year follow-up period. The reported reduction of other intervention usage suggests that the AI-embedded mobile app may at least reduce the ambulatory time to medical appointments, thus reducing the indirect health care costs that are associated with seeking intervention. Further study is recommended to understand its cost benefit within the local health care system.

Limitations

The results reported here must be viewed cautiously due to the limitations. This was a single-group study with no control which relies heavily on self-reported subjective data. This is among the first AI-embedded mobile apps developed for the management of low back pain. Thus, the primary goal of this preliminary study was to evaluate the perceived benefits of such a system by its users. With regard to the fidelity of the intervention, it is worth considering that the app was available for download by members of the general public. Users were not advised by anyone to download it; they downloaded it on their own initiative, assuming that they suffered from chronic neck and back pain and were looking for self-help material on the Web. Thus, it is reasonable to believe that users had followed the therapeutic program provided by the AI-embedded mobile app. This study defined the term "adherence" as the amount of self-reported time spent on the exercise generated by the AI-embedded mobile app. Therefore, the interpretation of adherence within this study cannot be generalized to the long-term adherence to therapeutic exercise. The "law of attrition" of eHealth trials [42] did not appear to be applicable in this study because this study only recorded response rate but not dropout rate. We could not verify whether participants' eligibility criteria information were accurate, which may affect their responses to the intervention. The amount of time participants viewed the video clips were not verified. However, the AI-embedded mobile app has a scoring system which the user had to satisfy by playing the video clip for the exercises. Further study that combines this with medical verification would provide greater confidence in the intervention effects. This study did not evaluate the functional aspect of the intervention. Thus, it was unclear if the changes in pain score may translate to improvement in function. We included active users at the time of the study, which many have affected the response rate. The study design also had sample selection bias because it relied on people who responded to the invitation, thus the sample population would be more likely to participate in the intervention. This was a retrospective evaluation study, which was likely to be associated with recall bias of preintervention symptoms and behaviors. The focus of this preliminary study was to assess the self-perceived benefits from users. Self-reported data on health care utilization are commonly used in health care research [43]. Self-reported data may be associated with recall bias; however, a study that investigated agreement between self-reported health care service usage and administrative records indicated good agreement between the two [44].

Conclusions

The key findings of this evaluation study support the perceived beneficiary effects of the AI-embedded mobile app to provide some personalized interventions that are tailored to individual needs for the self-management of chronic neck and back pain. The positive results of this study suggest that using the tool to

assist the self-management of chronic spinal pain may be feasible. The tool at least warrants further study to investigate the benefit of the AI-embedded mobile app and how it may compare with routine clinical care. Further study is also recommended to understand if the AI-embedded mobile app may induce long-term sustainable behavioral change.

Acknowledgments

This research project was supported by China Postdoctoral Science Foundation Grant (#2016M592581) and Guangzhou City Research & Technology Collaborative Innovation Projects (#2017010160387).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of a detailed description of the exercise (left) and of the annotated diagrams to show the targeted muscles (right).

[PNG File, 270KB - mhealth_v6i11e198_app1.png]

Multimedia Appendix 2

Screenshots of the daily task list (top) and the exercise log where users could track their progress (bottom).

[PNG File, 103KB - mhealth_v6i11e198_app2.png]

References

1. Leininger B, McDonough C, Evans R, Tosteson T, Tosteson AN, Bronfort G. Cost-effectiveness of spinal manipulative therapy, supervised exercise, and home exercise for older adults with chronic neck pain. Spine J 2016 Dec;16(11):1292-1304 [FREE Full text] [doi: 10.1016/j.spinee.2016.06.014] [Medline: 27345747]
2. Murray C, Vos T, Lozano R, Naghavi M, Flaxman A, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4]
3. GBD 2015 DALY and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Dec 08;388(10053):1603-1658 [FREE Full text] [doi: 10.1016/S0140-6736(16)31460-X] [Medline: 27733283]
4. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
5. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
6. Amorim A, Pappas E, Simic M, Ferreira M, Tiedemann A, Jennings M, et al. Integrating Mobile health and Physical Activity to reduce the burden of Chronic low back pain Trial (IMPACT): a pilot trial protocol. BMC Musculoskelet Disord 2016 Jan 19;17(1):a. [doi: 10.1186/s12891-015-0852-3]
7. Lilje S, Olander E, Berglund J, Skillgate E, Anderberg P. Experiences of older adults with mobile phone text messaging as reminders of home exercises after specialized manual therapy for recurrent low back pain: a qualitative study. JMIR Mhealth Uhealth 2017 Mar 30;5(3):e39 [FREE Full text] [doi: 10.2196/mhealth.7184] [Medline: 28360026]
8. Wallace LS, Dhingra LK. A systematic review of smartphone applications for chronic pain available for download in the United States. J Opioid Manag 2014;10(1):63-68. [doi: 10.5055/jom.2014.0193] [Medline: 24604571]
9. Miceli L, Bednarova R, Scarbolo M, Marzi R, Storelli E, Colonna U, et al. Development of an app helpful to manage patients with low back pain. Pain Pract 2014 Sep;14(7):e165-e166. [doi: 10.1111/papr.12235] [Medline: 25124448]
10. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
11. Chou R, Qaseem A, Snow V, Casey D, Cross JT, Shekelle P, Clinical Efficacy Assessment Subcommittee of the American College of Physicians, American College of Physicians, American Pain Society Low Back Pain Guidelines Panel. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 02;147(7):478-491. [Medline: 17909209]
12. Blödt S, Pach D, Roll S, Witt CM. Effectiveness of app-based relaxation for patients with chronic low back pain (Relaxback) and chronic neck pain (Relaxneck): study protocol for two randomized pragmatic trials. Trials 2014 Dec 15;15:490 [FREE Full text] [doi: 10.1186/1745-6215-15-490] [Medline: 25511185]

13. Machado G, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
14. World Health Organization. World Report on Disability. Malta: World Health Organization; 2011.
15. Xiang M, Zhang M, Zhang L. Relationship between self-management behaviors and disease knowledge in patients with chronic low back pain. J Nurs (China) 2016;23(15):42-45 [FREE Full text]
16. Sabate E. Adherence to Long-Term Therapies: Evidence for Action. Geneva: World Health Organization; 2003. URL: http://www.who.int/chp/knowledge/publications/adherence_report/en/ [accessed 2018-11-12] [WebCite Cache ID 73sN3jDzw]
17. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
18. Beinart N, Goodchild CE, Weinman JA, Ayis S, Godfrey EL. Individual and intervention-related factors associated with adherence to home exercise in chronic low back pain: a systematic review. Spine J 2013 Dec;13(12):1940-1950. [doi: 10.1016/j.spinee.2013.08.027] [Medline: 24169445]
19. Yuen HK, Wang E, Holthaus K, Vogtle LK, Sword D, Breland HL, et al. Self-reported versus objectively assessed exercise adherence. Am J Occup Ther 2013;67(4):484-489 [FREE Full text] [doi: 10.5014/ajot.2013.007575] [Medline: 23791324]
20. Bangor A, Kortum P, Miller J. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
21. National Institute for Health and Care Excellence. Low Back Pain and Sciatica in Over 16s: Assessment and Management. NICE Guideline. London: National Institute for Health and Care Excellence; 2016 Nov 30. URL: https://www.nice.org.uk/guidance/ng59/resources/low-back-pain-and-sciatica-in-over-16s-assessment-and-management-pdf-1837521693637 [accessed 2018-11-10] [WebCite Cache ID 73pzpVYKl]
22. Saragiotto B, Maher CG, Yamato TP, Costa LO, Menezes Costa LC, Ostelo RW, et al. Motor control exercise for chronic non-specific low-back pain. Cochrane Database Syst Rev 2016 Jan 08(1):CD012004. [doi: 10.1002/14651858.CD012004] [Medline: 26742533]
23. Steimann F. On the use and usefulness of fuzzy sets in medical AI. Artif Intell Med 2001;21(1-3):131-137. [Medline: 11154877]
24. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004 Sep;86(5):334-338 [FREE Full text] [doi: 10.1308/147870804290] [Medline: 15333167]
25. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986 Oct 9;323(6088):533-536. [doi: 10.1038/323533a0]
26. Przybylski A, Rigby CS, Ryan RM. A motivational model of video game engagement. Rev Gen Psychol 2010;14(2):154-166. [doi: 10.1037/a0019440]
27. Wise J. NICE recommends exercise and not acupuncture for low back pain. BMJ 2016 Mar 28;352:i1765. [doi: 10.1136/bmj.i1765] [Medline: 27025659]
28. Gross A, Paquin J, Dupont G, Blanchette S, Lalonde P, Cristie T, et al. Exercises for mechanical neck disorders: a Cochrane review update. Manual Ther 2016 Aug;24:25-45. [doi: 10.1016/j.math.2016.04.005]
29. Smith B, Littlewood C, May S. An update of stabilisation exercises for low back pain: a systematic review with meta-analysis. BMC Musculoskelet Disord 2014 Dec 09;15:416 [FREE Full text] [doi: 10.1186/1471-2474-15-416] [Medline: 25488399]
30. Friedrich M, Gittler G, Halberstadt Y, Cermak T, Heiller I. Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomized controlled trial. Arch Phys Med Rehabil 1998 May;79(5):475-487. [Medline: 9596385]
31. Reilly K, Lovejoy B, Williams R, Roth H. Differences between a supervised and independent strength and conditioning program with chronic low back syndromes. J Occup Med 1989 Jun;31(6):547-550. [Medline: 2525182]
32. Cunningham C, Malone S. Use of a smartphone app to improve adherence among low back pain group: a pilot RCT. Musculoskelet Sci Pr 2017 Apr;28:e18. [doi: 10.1016/j.math.2016.10.046]
33. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [Medline: 11690728]
34. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-291. [doi: 10.1016/j.ejpain.2003.09.004] [Medline: 15207508]
35. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Dec 24;4:CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub3] [Medline: 28436583]
36. Alfuth M, Welsink DW. Pain and functional outcomes after outpatient physiotherapy in patients with low back pain. Orthopäde 2017 Feb 15;46(6):522-529. [doi: 10.1007/s00132-017-3390-x]
37. Cambron JA, Gudavalli MR, Hedeker D, McGregor M, Jedlicka J, Keenum M, et al. One-year follow-up of a randomized clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. J Altern Complement Med 2006 Sep;12(7):659-668. [doi: 10.1089/acm.2006.12.659] [Medline: 16970536]

38. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [Medline: 14659516]
39. Smith C, Grimmer-Somers K. The treatment effect of exercise programmes for chronic low back pain. J Eval Clin Pract 2010 Jun;16(3):484-491. [doi: 10.1111/j.1365-2753.2009.01174.x] [Medline: 20438611]
40. Marceglia S, Conti C. A technology ecosystem for chronic pain: promises, challenges, and future research. Mhealth 2017;3:6 [FREE Full text] [doi: 10.21037/mhealth.2017.02.03] [Medline: 28300227]
41. Sculco AD, Paup DC, Fernhall B, Sculco MJ. Effects of aerobic exercise on low back pain patients in treatment. Spine J 2001;1(2):95-101. [Medline: 14588388]
42. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
43. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. Med Care Res Rev 2006 Apr;63(2):217-235. [doi: 10.1177/1077558705285298] [Medline: 16595412]
44. Carroll M, Sutherland G, Kemp-Casey A, Kinner SA. Agreement between self-reported healthcare service use and administrative records in a longitudinal study of adults recently released from prison. Health Justice 2016 Dec;4:11 [FREE Full text] [doi: 10.1186/s40352-016-0042-x] [Medline: 27942429]

Abbreviations

AI: artificial intelligence
eHealth: electronic health
ICT: information and communication technology
IQR: interquartile range
MLP-ANN: multilayered perceptron artificial neural network
NICE: National Institute of Clinical Excellence
NPRS: Numerical Pain Rating Scale
SUS: System Usability Scale

Edited by G Eysenbach; submitted 20.06.17; peer-reviewed by A Amorim, A Birney, M Luca, K Bosak, R Robinson; comments to author 12.07.17; revised version received 14.11.17; accepted 15.02.18; published 26.11.18.

Please cite as:
Lo WLA, Lei D, Li L, Huang DF, Tong KF
The Perceived Benefits of an Artificial Intelligence–Embedded Mobile App Implementing Evidence-Based Guidelines for the Self-Management of Chronic Neck and Back Pain: Observational Study
JMIR Mhealth Uhealth 2018;6(11):e198
URL: http://mhealth.jmir.org/2018/11/e198/
doi:10.2196/mhealth.8127
PMID:30478019

Original Paper

A Novel Approach to Evaluating Mobile Smartphone Screen Time for iPhones: Feasibility and Preliminary Findings

Aubrey D Gower[1], BS, BA; Megan A Moreno[1], MD, MPH, MEd

Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin - Madison, Madison, WI, United States

Corresponding Author:
Aubrey D Gower, BS, BA
Department of Pediatrics
School of Medicine and Public Health
University of Wisconsin - Madison
Suite 200
2870 University Avenue
Madison, WI, 53705
United States
Phone: 1 608 262 7103
Email: adgower@wisc.edu

Abstract

Background: Increasingly high levels of smartphone ownership and use pose the potential risk for addictive behaviors and negative health outcomes, particularly among younger populations. Previous methodologies to understand mobile screen time have relied on self-report surveys or ecological momentary assessments (EMAs). Self-report is subject to bias and unreliability, while EMA can be burdensome to participants. Thus, a new methodology is needed to advance the understanding of mobile screen time.

Objective: The objective of this study was to test the feasibility of a novel methodology to record and evaluate mobile smartphone screen time and use: battery use screenshot (BUS).

Methods: The BUS approach, defined for this study as uploading a mobile phone screenshot of a specific page within a smartphone, was utilized within a Web-based cross-sectional survey of adolescents aged 12-15 years through the survey platform Qualtrics. Participants were asked to provide a screenshot of their battery use page, a feature within an iPhone, to upload within the Web-based survey. Feasibility was assessed by smartphone ownership and response rate to the BUS upload request. Data availability was evaluated as apps per BUS, completeness of data within the screenshot, and five most used apps based on battery use percentage.

Results: Among those surveyed, 26.73% (309/1156) indicated ownership of a smartphone. A total of 105 screenshots were evaluated. For data availability, screenshots contained an average of 10.2 (SD 2.0) apps per screenshot and over half (58/105, 55.2%) had complete data available. The most common apps or functions included Safari and Home and Lock Screen.

Conclusions: Study findings describe the BUS as a novel approach for real-time data collection focused on iPhone screen time and use among young adolescents. Although feasibility showed some challenges in the upload capacity of young teens, data availability was generally strong across this large dataset. These data from screenshots have the potential to provide key insights into precise mobile smartphone screen use and time spent per mobile app. Future studies could explore the use of the BUS methodology on other mobile smartphones such as Android phones to correlate mobile smartphone screen time with health outcomes.

(JMIR Mhealth Uhealth 2018;6(11):e11012) doi:10.2196/11012

KEYWORDS

smartphone; youth; mobile apps; mobile phone; screenshot

Introduction

Smartphones and their vast functionalities have become an integral part of individuals' lives today [1]. In the United States alone, smartphone ownership has increased to 77%, with 95% of teens aged 13-17 years reporting access to or ownership of a smartphone (92%) [2,3]. Globally, similar trends are seen in smartphone ownership, including among emerging economies [4]. Among adolescent smartphone owners, 64% reported "everyday use" in a recent survey [5]. As ownership and accessibility to mobile smartphones have become ubiquitous, so has the need for research into the implications, such as positive and negative health consequences. Recent research suggests high screen media usage is associated with poor sleep and diminished academic performance among adolescents and adults [6-8]. Thus, accurate and feasible methodologies to study mobile screen time are necessary to further understand these relationships.

Previous methodologies to understand mobile screen time have typically relied on traditional self-report and cross-sectional research design [6-10]. However, self-report is vulnerable to systematic and confounding bias [11]. Previous work has shown conflicting results related to media use time with the use of self-report. Some studies have found self-report app usage to underestimate app and smartphone usage [12]. On the other hand, studies have also found self-report by participants to overestimate Web-based time [13,14]. These inconsistencies have led some researchers to suggest that self-reported smartphone use should be interpreted with caution [15].

Another approach to evaluate mobile media use is ecological momentary assessment (EMA). In this approach, participants are typically contacted multiple times per day to report their real-time screen use [16,17]. While this approach improves upon self-report biases, such as recall bias, it can be highly burdensome to participants [18]. In addition, these methodologies often fail to obtain large-scale and representative samples due to costs to researchers to provide the compensation necessary to attract participants [19]. Thus, a new methodology is needed to advance the assessment of mobile screen time that not only improves accuracy compared with self-report but is also not burdensome to participants.

One possible approach to understanding mobile smartphone screen time while limiting participant burden is to leverage passive tracking that smartphones are programmed for via battery use reporting. Most smartphones, such as iPhones, track battery use per app and phone function (eg, Home and Lock Screen). A majority of smartphones will then report this battery use by time, including both active onscreen and background use. The "battery use" function and display, thus, serve as an indicator for real-time app and smartphone activity.

In this research protocol, we present preliminary findings of the battery use screenshot (BUS) approach among young adolescents. The objective of this study was to test the feasibility of the BUS approach to obtain and evaluate iPhone screen time data.

Methods

Design

Data for this study were collected as a planned protocol evaluation as part of a larger study related to technology rules and health behaviors among adolescents aged 12-15 years. The BUS approach was defined for this study as the upload of a mobile phone screenshot of the "battery use" page within the participant's smartphone. To assess the feasibility of the BUS approach, this study was designed as a cross-sectional Web-based survey of 12- to 15-year-old adolescents. We collected data between June and July 2017 using the Qualtrics survey platform. The study was approved by the University of Washington Institutional Review Board.

Participants and Recruitment

Youth who were between the ages of 12 and 15 years and who could read English were eligible to complete the survey. As described in previous studies [20], Qualtrics draws upon previously established age panels within their Web-based database, allowing for targeted recruitment. Eligibility was assessed prior to beginning the Web-based informed consent process. Once informed consent from parents was obtained, participants could begin the survey. Compensation was provided to participants through Qualtrics for survey completion.

Data Collection Process

Variables

Phone Ownership

Participants were asked whether they currently had a smartphone capable of taking a screenshot. Participants were also asked to indicate whether the phone was their own, their parent's smartphone, or that they did not own a smartphone.

Demographics

Age, gender, parental education, and race or ethnicity were assessed.

Screenshot Upload

Participants were provided an example BUS within the survey (Figure 1) and then asked to create their own screenshot from the phone they currently used, if they had one. Instructions included first asking the participant to go to the "Settings" section of their smartphone and clicking the "Battery" option. Once in the "Battery" section, participants were asked to locate the "battery usage" data displayed. For iPhones, participants were asked to click both the "Last *x* days" and the clock symbol in the right-hand corner to display both onscreen and background activity for each app listed. Participants were asked to compare their display to the example provided to ensure the correct data were viewable. Participants were then asked to screenshot this display by holding down the home and power button at the same time to capture the image. Once the screenshot was captured, participants were asked to upload the image, as a .jpg or .png, to a file dropbox within the survey.

Figure 1. Example of an iPhone battery use screenshot from an Apple iPhone 6 with IOS 11; image was produced by first author.

Analysis

Descriptive statistics were used to analyze feasibility and data availability. The most commonly used apps were determined based on their frequency among the evaluated screenshots. Analysis focused on iPhone screenshots because of inconsistency among Android platforms in the display of battery usage. Screenshots were excluded if they contained anything other than the battery use page from a personal phone.

Outcomes

Feasibility

Measures to assess feasibility included a survey question asking about personal smartphone ownership and response rate to BUS upload request.

Data availability

To characterize data availability, content analysis assessed apps or functions per screenshot and completeness of the screenshot including display of "onscreen" and "background" time and displayed percentage of battery use per app. Functions were characterized as iPhone functions such as Home and Lock Screen that use battery life but are not considered typical downloadable apps. The five most commonly used apps or functions among evaluated screenshots were assessed. The most used apps or functions for each screenshot were defined based on displayed battery use percentage.

Results

Demographics

A total of 1156 adolescents with an average age of 13.6 (SD 1.09) years responded to the survey. The overall survey sample comprised 49.48% (572/1156) females and 72.84% (842/1156) Caucasian people. Among the sample of participants for which screenshots were evaluated, 52.3% (79/151) were females and 75.5% (114/151) were Caucasian people. Full demographics for both populations can be seen in Table 1.

Feasibility

Among the overall survey sample, 26.73% (309/1156) indicated they had their own phone. Among these adolescents, 48.9% (151/309) completed the BUS upload request (Figure 2).

Table 1. Demographic characteristics of participants.

Characteristics	Total survey sample (N=1156)	Sample screenshots evaluated (n=151)
Age in years, mean (SD)	13.6 (1.09)	13.7 (1.05)
Gender, n (%)		
Female	572 (49.48)	79 (52.32)
Male	573 (49.57)	69 (45.70)
Female to male transgender people	4 (0.35)	2 (1.32)
Male to female transgender people	0 (0.00)	0 (0.00)
Not sure	3 (0.26%)	1 (0.66)
Race or ethnicity, n (%)		
White or Caucasian	842 (72.84)	114 (75.50)
Black or African American	80 (6.92)	12 (7.95)
Hispanic or Latino	101 (8.74)	12 (7.95)
Asian	56 (4.84)	6 (3.97)
American Indian or Alaska Native	8 (0.69)	0 (0.00)
Native Hawaiian or Pacific Islander	4 (0.35)	1 (0.66)
More than one race	41 (3.55)	4 (2.65)
Other	10 (0.87)	2 (1.32)
Parental education level, n (%)		
High school Graduate	118 (10.21)	9 (5.96)
Tech school or associate degree	88 (7.61)	10 (6.62)
Some college	122 (10.55)	16 (10.60)
College degree	361 (31.23)	56 (37.09)
Some graduate school	39 (3.37)	11 (7.28)
Completed graduate degree	299 (25.87)	37 (24.50)
Other	116 (10.03)	12 (7.95)
Did not answer	13 (1.12)	0 (0.00)

Data Availability

A total of 105 iPhone screenshots, received as .jpg or .png images, were used for the evaluation of data availability. Each screenshot contained an average of 10.2 (SD 2.00) apps. More than half (58/105, 55.2%) had complete data availability, indicating successful implementation of instructions provided. Complete screenshots allowed for the ability to view the apps used, battery use percentage per app, and total minutes or hours of both “background” and “on screen” usage (Figure 1).

Most Commonly Used Apps

Figure 3 shows the frequency of the most common apps or functions to appear in the five most used apps or functions among the analyzed screenshots. Safari, an internet search engine, was the most common app among screenshots; 46.7% (49/105) of screenshots included Safari within the five most used apps.

Figure 2. Protocol results flowchart.

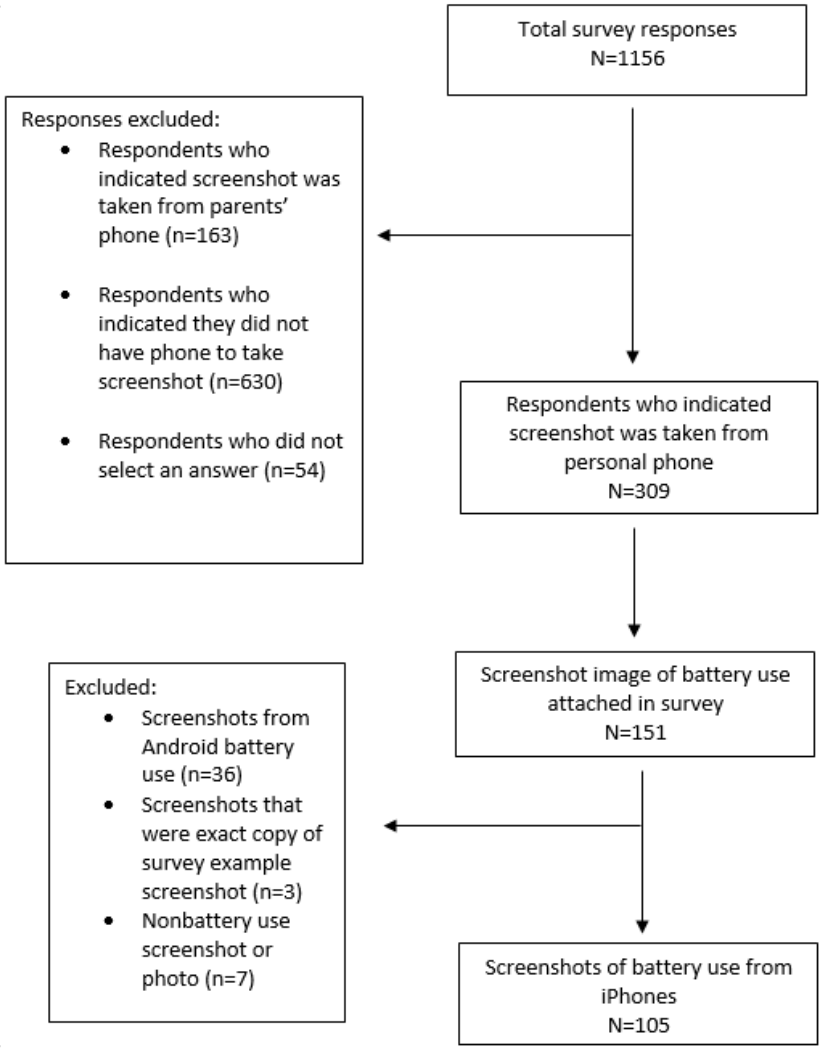
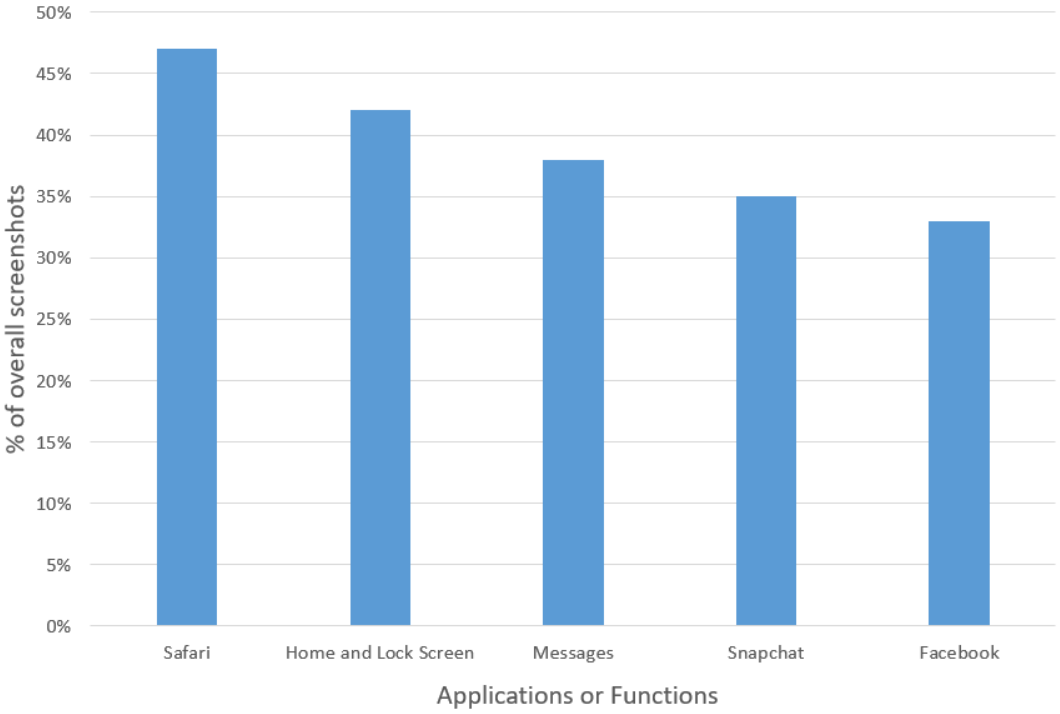


Figure 3. Most common apps or functions among analyzed screenshots.



Discussion

Principal Findings

This study describes a novel BUS approach to test feasibility, data availability, and most common apps among a young adolescent population. The inclusion of the BUS approach within a Web-based survey adds the capacity to assess mobile smartphone screen time that is accessible for both participants and researchers in a highly technological environment. Battery use is automatically monitored by the operating systems of most smartphones [21], reducing both the bias of self-report as well as burden on participants to record real-time smartphone and app use. Compared with typical EMA approaches, this methodology can improve the capacity to deliver real-time data as well as allow for data collection from larger, Web-based samples [22].

The feasibility of our approach was mixed with both strengths and challenges. Among the participants who did upload a screenshot, the images provided clear data that could be viewed, stored in a deidentified manner, and categorized for further analysis. Our challenges in feasibility may be explained by the rate of phone ownership in our young adolescent sample, which is lower than recent reports of 56% smartphone ownership among youth aged 8-18 years [23]. Among adolescents with their own phone, approximately half were willing or able to upload a screenshot. It is possible that a study population of older adolescents and adults may have higher phone ownership and better understanding and capacity to upload a screenshot. An additional explanation for this challenge may be the survey platform Qualtrics, which allows users to take surveys on both mobile phones and desktop computers. Participants who used a desktop computer to complete the survey may have experienced an additional burden in uploading a screenshot from a different device, which may have contributed to the overall feasibility of this study. Compensation of US $13 was determined by Qualtrics based on the initial suggested sample size for the larger study. Based on our results, low monetary compensation may be a factor in willingness to complete the screenshot task and should be considered when incorporating the BUS approach into future studies.

Over half of the screenshots analyzed contained complete data including the display of “onscreen” versus “background” time per app. One possible reason for cases of incomplete data may be that participants had an older iOS operating system, as the battery usage feature is only available on iOS 9 and newer operating systems [24]. This should be taken into consideration in future studies utilizing this methodology in iPhones. In addition, this study relied on a single screenshot, which may not have captured total app usage. To strengthen overall data availability, future studies might require as many screenshots as necessary to provide the full range of apps used by a participant.

While our study illustrates both strengths and challenges to the proposed research protocol, it serves as a valuable starting point for considering how to advance data collection methods to understand mobile smartphone screen time and media use in iPhones. Studies have concluded that smartphone apps can be beneficial in monitoring and evaluating patients [25-27] as well as increasing adherence to medical interventions [28,29]. The BUS approach offers the ability to take these studies further in understanding real-time use and overall time spent on apps as a factor that could contribute to health outcomes. This methodology offers a framework that could be adapted to Android phones, which offer varying displays of battery use per app and time spent per app. When data and display of battery use are available, a BUS approach may still be utilized as a data collection process. Instructions for the screenshot would need to be adjusted to reflect the varying battery use displays of Androids, and further exploration of these displays should be used to inform these instructions. By expanding the BUS approach to Android devices, future studies would also allow for greater sampling of socioeconomic status, with lower socioeconomic status individuals and families more likely to own Android devices.

The BUS approach may also be combined with a variety of research methodologies including Web-based cross-sectional surveys, as was done in this study, or used as a tool for monitoring smartphone app use longitudinally. In a previous pilot study conducted using BUS, older adolescent iPhone users were asked to submit weekly screenshots of battery use for 9 weeks, with a 94% retention rate over 5 weeks and 60% retention rate over 9 weeks [30]. A further advantage of the BUS approach is the ability to collect comprehensive data related to mobile smartphone screen time without the need for an additional app or programming. Thus, this methodology is accessible for researchers without the means for software or app development.

Limitations

There are limitations to this feasibility study. The sample size of screenshots analyzed is limited and is not representative of the larger adolescent population. The sample overall may not be representative of the US racial or ethnic makeup for youth aged under 18 years because current US census records differ from our study’s demographics and estimate this population to be 51% white, 15% black or African American, and 25% Latino people [31]. In addition, the battery use page only displays phone use while an iPhone is not charging and may not account for the full time spent using a smartphone device. Furthermore, adolescents may have access to other phones, including their parents’, which may underestimate overall screen time. Finally, in this feasibility study, only iPhone screenshots were evaluated. Results may not generalize to all smartphones or mobile devices with battery use tracking. We did not test the BUS tracking method against other methods of tracking app use to assess its accuracy.

Conclusion

Though feasibility with the BUS methodology showed challenges in phone ownership rates and upload capacity of young teens, availability of data was generally strong across this large dataset. The data available from screenshots have the potential to provide key insights into precise mobile smartphone screen use and amount of time spent per mobile app. The BUS approach may provide an innovative and complementary approach to understanding smartphone screen use without the

need for complex programming or mobile app development. Future studies could improve upon the BUS methodology to correlate mobile smartphone screen time with health outcomes.

Acknowledgments

This research was supported and funded by Seattle Children's Research Institute. We would also like to thank Esther Lam for her assistance on this project.

Conflicts of Interest

None declared.

References

1. Smith A. Pew Research Center: Internet & Technology. 2015 Apr 01. U.S. Smartphone Use in 2015 URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/03/PI_Smartphones_0401151.pdf [accessed 2018-03-31] [WebCite Cache ID 6yKuLZOqm]
2. Smith A. Pew Research Center. 2017 Jan 12. Record shares of Americans now own smartphones, have home broadband URL: http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/ [accessed 2018-03-30] [WebCite Cache ID 6yJ4vWBgw]
3. Anderson M, Jiang J. Pew Research Center: Internet & Technology. 2018 May 31. Teens, Social Media & Technology 2018 URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2018-08-04] [WebCite Cache ID 71QbOtxUb]
4. Poushter J. Pew Research Center. 2016. Smartphone ownership and internet usage continues to climb in emerging economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [WebCite Cache ID 6yI0eALhP]
5. Rideout V. Common Sense Media Incorporated. 2015. The common sense census: Media use by tweens and teens URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens [WebCite Cache ID 6zPgzRP48]
6. Hale L, Guan S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep Med Rev 2015 Jun;21:50-58. [doi: 10.1016/j.smrv.2014.07.007] [Medline: 25193149]
7. 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]
8. Lee KE, Kim S, Ha T, Yoo Y, Han J, Jung J, et al. Dependency on Smartphone Use and Its Association with Anxiety in Korea. Public Health Rep 2016;131(3):411-419 [FREE Full text] [doi: 10.1177/003335491613100307] [Medline: 27252561]
9. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
10. Lenhart A. Pew Research Center. 2015. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [WebCite Cache ID 6z6ArxOns]
11. Mehl M. Why researchers should think "real-time": A cognitive rationale. In: Handbook of Research Methods for Studying Daily Life. New York: Guilford Press; 2013.
12. Lin Y, Lin Y, Lee Y, Lin P, Lin S, Chang L, et al. Time distortion associated with smartphone addiction: Identifying smartphone addiction via a mobile application (App). J Psychiatr Res 2015 Jun;65:139-145. [doi: 10.1016/j.jpsychires.2015.04.003] [Medline: 25935253]
13. Moreno MA, Jelenchick L, Koff R, Eikoff J, Diermyer C, Christakis DA. Internet use and multitasking among older adolescents: An experience sampling approach. Computers in Human Behavior 2012 Jul;28(4):1097-1102. [doi: 10.1016/j.chb.2012.01.016]
14. Kobayashi T, Boase J. No Such Effect? The Implications of Measurement Error in Self-Report Measures of Mobile Communication Use. Communication Methods and Measures 2012 Apr;6(2):126-143. [doi: 10.1080/19312458.2012.679243]
15. Andrews S, Ellis DA, Shaw H, Piwek L. Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use. PLoS One 2015;10(10):e0139004 [FREE Full text] [doi: 10.1371/journal.pone.0139004] [Medline: 26509895]
16. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
17. Moskowitz DS, Young SN. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. J Psychiatry Neurosci 2006 Jan;31(1):13-20 [FREE Full text] [Medline: 16496031]
18. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological Momentary Assessment in Behavioral Research: Addressing Technological and Human Participant Challenges. J Med Internet Res 2017 Mar 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]

19. Stone AA, Mackie C, editors. Subjective well-being: Measuring happiness, suffering, and other dimensions of experience. In: Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press; 2013:978.
20. Len-Ríos ME, Hughes HE, McKee LG, Young HN. Early adolescents as publics: A national survey of teens with social media accounts, their media use preferences, parental mediation, and perceived Internet literacy. Public Relations Review 2016 Mar;42(1):101-108. [doi: 10.1016/j.pubrev.2015.10.003]
21. Falaki H, Mahajan R, Estrin D. SystemSens: a tool for monitoring usage in smartphone research deployments. Proceedings of the Sixth International Workshop on MobiArch - MobiArch 2011;11. [doi: 10.1145/1999916.1999923]
22. Soong A, Chen JC, Borzekowski DL. Using Ecological Momentary Assessment to Study Tobacco Behavior in Urban India: There's an App for That. JMIR Res Protoc 2015 Jun 24;4(2):e76 [FREE Full text] [doi: 10.2196/resprot.4408] [Medline: 26109369]
23. Lauricella AR, Cingel DP, Beaudoin-Ryan L, Robb MB, Saphir M, Wartella EA. Common Sense Media. 2016. The Common Sense census: Plugged-in parents of tweens and teens URL: https://www.commonsensemedia.org/research/the-common-sense-census-plugged-in-parents-of-tweens-and-teens-2016 [accessed 2018-10-10] [WebCite Cache ID 734FQR6zh]
24. Apple, Inc. Maximizing Battery Life and Lifespan URL: https://www.apple.com/batteries/maximizing-performance/ [accessed 2018-03-29] [WebCite Cache ID 6yHxSw7MP]
25. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
26. Tomlinson S, Behrmann S, Cranford J, Louie M, Hashikawa A. Accuracy of Smartphone-Based Pulse Oximetry Compared with Hospital-Grade Pulse Oximetry in Healthy Children. Telemed J E Health 2018 Jul;24(7):527-535. [doi: 10.1089/tmj.2017.0166] [Medline: 29215972]
27. Karduck J, Chapman-Novakofski K. Results of the Clinician Apps Survey, How Clinicians Working With Patients With Diabetes and Obesity Use Mobile Health Apps. J Nutr Educ Behav 2018 Jan;50(1):62-69.e1. [doi: 10.1016/j.jneb.2017.06.004] [Medline: 29325664]
28. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: Adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
29. Rathbone AL, Prescott J. The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
30. Gower A, Moreno M. New Approaches to Understanding App Use among Adolescents. Journal of Adolescent Health 2018 Feb;62(2):S85. [doi: 10.1016/j.jadohealth.2017.11.171]
31. United States Census Bureau. Washington; 2018. National Population by Characteristics 2010-2017 URL: https://www.census.gov/data/datasets/2017/demo/popest/nation-detail.html [accessed 2018-07-31] [WebCite Cache ID 71tZjJeIk]

Abbreviations

BUS: battery use screenshot
EMA: ecological momentary assessment

Edited by G Eysenbach; submitted 20.05.18; peer-reviewed by Z Ma, D Romer, M Lucassen; comments to author 28.06.18; revised version received 06.08.18; accepted 27.08.18; published 19.11.18.

Please cite as:
Gower AD, Moreno MA
A Novel Approach to Evaluating Mobile Smartphone Screen Time for iPhones: Feasibility and Preliminary Findings
JMIR Mhealth Uhealth 2018;6(11):e11012
URL: http://mhealth.jmir.org/2018/11/e11012/
doi:10.2196/11012
PMID:30455163

Original Paper

To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App

Niranjan Bidargaddi[1], BEng, PhD; Daniel Almirall[2], PhD; Susan Murphy[3], PhD; Inbal Nahum-Shani[2], PhD; Michael Kovalcik[2]; Timothy Pituch[4]; Haitham Maaieh[4]; Victor Strecher[4,5], PhD

[1]Personal Health Informatics, College of Medicine & Public Health, Adelaide, Australia

[2]Insitute for Social Research, Michigan University, Ann Arbor, MI, United States

[3]Department of Statistics, Harvard University, Boston, MA, United States

[4]Jool Health, Ann Arbor, MI, United States

[5]School of Public Health, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Niranjan Bidargaddi, BEng, PhD
Personal Health Informatics, College of Medicine & Public Health
Flinders University
GPO Box 2100 Adelaide SA 5001
Adelaide,
Australia
Phone: 61 72218840
Email: niranjan.bidargaddi@flinders.edu.au

Abstract

Background: Mobile health (mHealth) apps provide an opportunity for easy, just-in-time access to health promotion and self-management support. However, poor user engagement with these apps remains a significant unresolved challenge.

Objective: This study aimed to assess the effect of sending versus not sending a push notification containing a contextually tailored health message on proximal engagement, measured here as self-monitoring via the app. Secondary aims were to examine whether this effect varies by the number of weeks enrolled in the program or by weekday versus weekend. An exploratory aim was to describe how the effect on proximal engagement differs between weekday versus weekend by the time of day.

Methods: The study analyzes the causal effects of push notifications on proximal engagement in 1255 users of a commercial workplace well-being intervention app over 89 days. The app employs a microrandomized trial (MRT) design to send push notifications. At 1 of 6 times per day (8:30 am, 12:30 pm, 5:30 pm, 6:30 pm, 7:30 pm, and 8:30 pm; selected randomly), available users were randomized with equal probability to be sent or not sent a push notification containing a tailored health message. The primary outcome of interest was whether the user self-monitored behaviors and feelings at some time during the next 24 hours via the app. A generalization of log-linear regression analysis, adapted for use with data arising from an MRT, was used to examine the effect of sending a push notification versus not sending a push notification on the probability of engagement over the next 24 hours.

Results: Users were estimated to be 3.9% more likely to engage with the app in the next 24 hours when a tailored health message was sent versus when it was not sent (risk ratio 1.039; 95% CI 1.01 to 1.08; $P<.05$). The effect of sending the message attenuated over the course of the study, but this effect was not statistically significant ($P=.84$). The effect of sending the message was greater on weekends than on weekdays, but the difference between these effects was not statistically significant ($P=.18$). When sent a tailored health message on weekends, the users were 8.7% more likely to engage with the app (95% CI 1.01 to 1.17), whereas on weekdays, the users were 2.5% more likely to engage with the app (95% CI 0.98 to 1.07). The effect of sending a tailored health message was greatest at 12:30 pm on weekends, when the users were 11.8% more likely to engage (90% CI 1.02 to 1.13).

Conclusions: Sending a push notification containing a tailored health message was associated with greater engagement in an mHealth app. Results suggested that users are more likely to engage with the app within 24 hours when push notifications are sent at mid-day on weekends.

(JMIR Mhealth Uhealth 2018;6(11):e10123) doi:10.2196/10123

KEYWORDS

mobile applications; smartphone; self report; health promotion; lifestyle; ubiquitous computing; push notification

Introduction

Background

Evidence demonstrates that providing engaged individuals with a digital behavior change intervention improves health outcomes [1-3]. In recent years, owing to increasingly ubiquitous smartphone ownership at a population level and rapidly reducing digital divides [4], smartphone apps have emerged as an important channel for health behavior interventions. As individuals are in the vicinity of their smartphone most of the time, smartphone apps can be used to sense users' everyday context and offer behavioral interventions at the most suitable moments. Unlike the traditional health care system, most mobile apps are inexpensive and easy to access on demand. They overcome demographic, socioeconomic, and geographical barriers to access by being able to reach populations that were unreachable until recently [5]. Consequently, smartphone health apps are now widely recognized as a public health promotion resource. For example, the World Health Organization Mental Health Action Plan 2013-2020 recommends "the promotion of self-care, for instance, through the use of electronic and mobile health technologies" [6]. Websites of prominent and large public health organizations (such as the UK National Health Service [NHS] website NHS Choices or Reachout Australia [7]) have also begun formally endorsing and recommending effective apps.

A pressing concern for the field of mobile health (mHealth), however, is the high rate of disengagement among individuals who choose to install an app. After installing a mobile app, over 80% of app users use it only once and eventually delete it [8]. Only 5% of the apps continue to be used beyond a month [9]. Even among those who use the apps, the amount of use depends on an individual's health and behavioral characteristics [10-12].

The lack of participant engagement is not unique to mHealth apps. In fact, traditional Web-based interventions observe decreases in engagement over time, with large proportions of participants dropping out or discontinuing the use of the app completely [13]. Although when compared with traditional internet interventions, the more recent smartphone app interventions have the potential to reach participants at moments when they are most likely to engage. It is disconcerting that similar underlying disengagement patterns are emerging even with more advanced technology. Consequently, despite the potential of smartphone apps, lack of participant engagement with apps, which is critical to the success of interventions focused on health behavior change, is of major concern. Given the fast pace of smartphone health app development, there is a pressing need for novel research to enhance engagement with these apps.

Effective engagement with digital health apps is particularly critical when this engagement is part of the health intervention, as opposed to simply opening the app. For example, engagement might be classified as use of the app to self-monitor health behaviors and feelings; self-monitoring has been demonstrated to have a positive impact on well-being and improve health behaviors [14,15]. Engagement could also be the subjective quality of user experiences with the app, which can be influenced by design elements [16]. Individual characteristics such as age, education levels, and state of health are known to equally influence the adoption of mHealth apps, in addition to factors such as being female, younger, higher education levels, and lower body mass index increasing the odds [17,18]. Another factor affecting the dropout in mobile app use could be the timing of intervention offered. For example, participants in an app study reported they were likely to drop out if the intervention did not meet their expectations and needs at the right time [19]. On the other hand, some qualitative research suggests that individuals tend to use app-based interventions in brief bursts only, often when needed and in a fleeting manner [20,21]. In this context, it could be argued that even short-term or varying intensity patterns of use with app-based behavior change interventions could be beneficial to users. For example, in the case of a self-monitoring app intervention, some individuals might monitor more frequently than others, but irrespective of how often they engage, every act of self-monitoring enables users to learn some strategies that can be practiced to improve behaviors without additional guidance from the intervention. Thus, every time a user opens and interacts with the intervention activity of the app, they are effectively engaged.

As mobile phones are usually switched on and nearly always with users, the majority of app-based interventions adopt prompting as a strategy to encourage interaction and engagement. In apps, prompts are implemented as push notifications, which appear on the smartphone screen at a programmed time. Both the content and timing of a push notification are programmable. Consequently, smartphone users receive a deluge of push notifications daily, approximately 50 time-varying notifications, from a diverse set of apps [22]. This sheer volume of notifications throughout the day has the potential to further exacerbate disengagement with the apps. A few studies that observed how users responded to push notifications received on their smartphone over the course of a day suggest that users are most likely to ignore a vast majority of notifications even when the notifications come from apps of importance to them [23,24]. There are also concerns that receiving too many notifications might increase users' inattention and reduce well-being [25].

Due to advances in smartphone-sensing capabilities along with algorithmic advances, apps can now utilize users' context, for example, location, social setting, and activity level to determine the most opportune times to send push notifications. Studies suggest that users are receptive to notification interruptions at convenient times [26-28]. Within the human-computer interaction field, *interruptibility research* has emerged, with

many studies focused on understanding and anticipating suitable moments for interruption [29-32]. However, most studies concerning the effects of push notification interruptions on engagement are observational studies based on small convenience samples with neither randomization nor control conditions. Thus, although rapid advances are being made in devising algorithms to predict when to interrupt users, because of the lack of appropriately designed studies, there is a significant gap in knowledge concerning when particular types of interruptions are effective.

In contrast, the fields of public health and psychology contain several well-designed randomized controlled trial studies to evaluate text-messaging interventions that are similar to push notification interventions in mHealth apps [33,34]. Furthermore, a recent study randomized participants into 3 different conditions based on the different approaches used to determine the timing of push notification [35]. These studies, using baseline randomization of individuals into different conditions, were designed to compare between conditions in terms of a distal outcome (such as the percentage of notifications viewed or overall usage). However, baseline randomization is not suitable for addressing questions comparing prompts at different times or comparing a prompt (vs no prompt) in terms of their effect on near time, proximal engagement, and the conditions in which a prompt would be more or less beneficial. Indeed, on any given day, users may be concerned with the contingencies and demands of the day so that even a self-determined user may not think or remember to access support on the app.

To our knowledge, this is the first study focusing on the effects of push notifications containing contextually tailored health messages on near-time, proximal engagement with the app. Here, engagement in response to a prompt is operationalized as the user completing the self-monitoring intervention activity in the app within 24 hours of receiving the push notification (as opposed to just opening the app). This study examines data collected from a microrandomized trial (MRT) [36] implemented within the JOOL app, a commercial workplace well-being intervention product [37]. In an MRT, each user is randomized multiple times over a period of weeks and months. In the JOOL app, push notifications may be randomized at each of 6 time points per day. This repeat-randomization design adjusts for potential biases that might arise from within and between the individual factors. For instance, randomizing the decision to send a push notification at a time point ensures that within- and between-user factors contributing to day-to-day variations in engagement are balanced approximately evenly across both push notification and no push notification conditions. The specific primary and secondary aims and hypotheses of this study are presented in the next section.

Aims and Hypotheses

The specific primary and secondary aims and hypotheses are as follows:

Primary Aim

To test whether sending a push notification containing a contextually tailored health message versus not sending push notification (in moments of availability) results in an increased likelihood of proximal engagement with the app (ie, the user is self-monitored via the app at some point in the next 24 hours).

Hypothesis

We hypothesize that on an average, sending a push notification containing a contextually tailored message will lead to a greater likelihood of proximal engagement with the app as compared with not sending the push notification.

Secondary Aim 1

To test whether the effect of sending a push notification containing a tailored message (vs not sending a push notification) on the likelihood of proximal engagement with the app differs by week in the study (there are approximately 12 weeks in the study).

Hypothesis

We hypothesize that the effect of sending a push notification containing a tailored message will differ by week in the study. Specifically, we hypothesize that the effect will be greater earlier in the study than later in the study.

Secondary Aim 2

To test whether the effect of sending a push notification containing a tailored message (vs not sending a push notification) on the likelihood of proximal engagement with the app differs by weekdays (Monday to Friday) versus weekends (Saturday or Sunday).

Hypothesis

We hypothesize that the effect of sending a push notification containing a tailored message will differ by weekdays versus weekends. Specifically, we hypothesize that the effect will be greater when the push notification is on weekends than on weekdays.

The exploratory aim focused on whether the effect of sending a push notification containing a tailored message (vs not sending a push notification) on the likelihood of proximal engagement with the app differs by the time of day within weekday-versus-weekend.

For all aims, the effect of sending a push notification is defined more precisely as the effect of sending a push notification (and not sending a subsequent push notification over the next 24 hours) versus not sending a push notification now or over the next 24 hours.

Methods

Intervention and Push Notification

The JOOL app is a smartphone-based behavioral intervention using self-monitoring and feedback strategies to help users find their purpose in life and develop the energy and willpower they need to live in accordance with their purpose every day. To engage in self-monitoring activity, users open the app and record on a scale of 0 (worse) to 100 (best) values for daily energy, willpower, sleep, presence, physical activity, creativity, eating, and perceived alignment with the community, work, and personal purposes. Interventions that assist individuals to record and track these behaviors and feelings and assist with behavior

modification feedback are demonstrated to have a positive impact on well-being and improve health behaviors [14,15].

The app sends time-varying push notifications containing tailored health messages to provide feedback related to behavior modification and to encourage interaction with the app. The content of the feedback messages in the push notifications is drawn from a library of messages curated by JOOL to motivate, facilitate, and maintain behavior change. The content of messages is related to the purpose, energy, willpower, sleep, presence, activity, eating well, and creativity topics. The messages were created via tailoring strategies, are personalized, and the content was organized by types to both enhance message processing and interaction with the app [38] (see Figure 1). JOOL uses the answers from the self-monitoring along with environmental information (day of the week, temperature, and weather) to tailor the messages. Furthermore, to increase users' attention, interest, and motivation to process information, the feedback messages offered in the push notifications were sent in a context that is meaningful to the recipient. To contextually tailor, first, there was a determination of the user's context at the selected time point when the app is programmed to send a message. The context is determined by the user's current and past data from the self-monitoring, other app usage, and environmental data such as the time of day and day of week. Next, the tailoring algorithm identified a subset of messages from the library that are meaningful to the users' context at the decision time and randomly selects one of the messages to send to the user. For example, a user whose self-monitoring data indicate low energy is more likely to receive a message with a tip: "(...) Setting aside some time for meditation might give you more energy"; if the self-monitoring data includes reports of low willpower, the user might be sent a message such as: "(...) Little bursts of physical activity do bolster willpower."

When a user opens the app, either prompted or unprompted, they are always required to first complete the self-monitoring intervention activity in the app. Effective engagement in response to a push notification was operationalized as interacting with the self-monitoring intervention activity in the app within the next 24 hours.

Microrandomized Trial Design

The implementation of the MRT design for sending push notifications in the app is shown in Figure 2. Push notifications could be sent at 1 of 6 chosen time points throughout the day, and a user could either receive or not receive a push notification at a chosen time point. The insights JOOL Health has on its user group, who are primarily office workers in a 9 to 5 work environment, informed which time points are appropriate for sending prompts. During a typical day, these time points correspond to contexts such as before work (8:30 am), during lunch (12:30 pm), early evening at home (5:30 pm, 6:30 pm, and 7:30 pm), and just before bed (8:30 pm). These are times when an office worker is likely to be less busy and thus more receptive. As prior research suggests engagement typically occurs during nonworking hours, more time points outside work hours were included [39,40]. Multiple convenient time points were chosen to ensure the timing of prompts was uniformly distributed throughout the day. Specifically, at each time point on each day, a user is first classified as being available or unavailable to receive push notifications. Unavailable users advance to the next of the 6 time points on the same day at which time their availability will be assessed again. When a user was available at a selected time point, they were randomized to either consider that time point for a push notification randomization or to advance to the next time point. At each considered time point, users were randomized to either receive or not receive a push notification containing a tailored health message with a 50% probability. Once a time point is considered, the user is then considered unavailable for the remainder of the day.

Figure 1. Different types of contextually tailored messages.

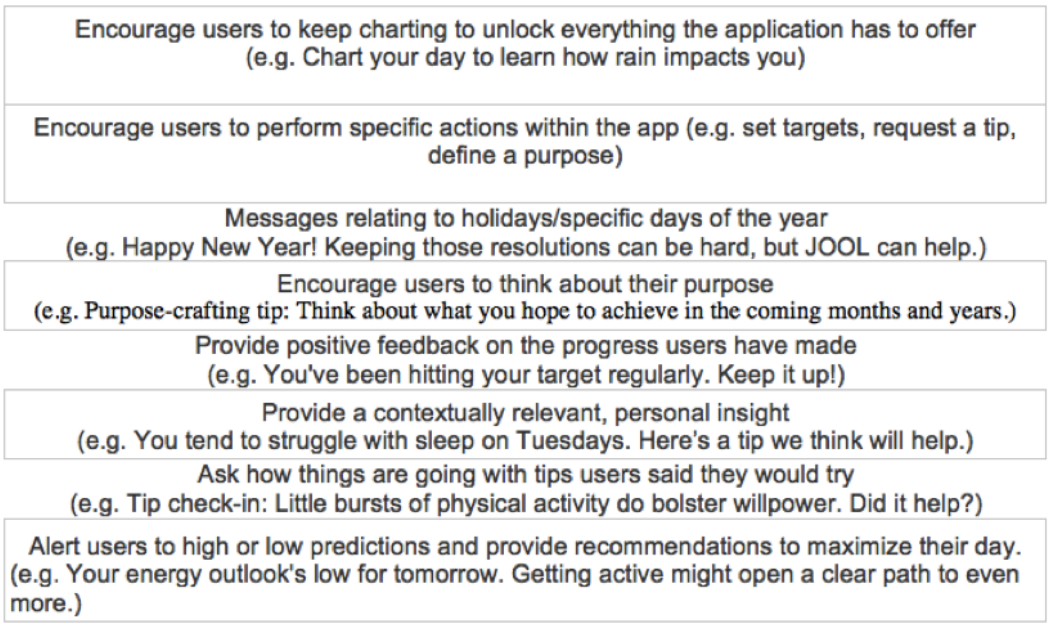
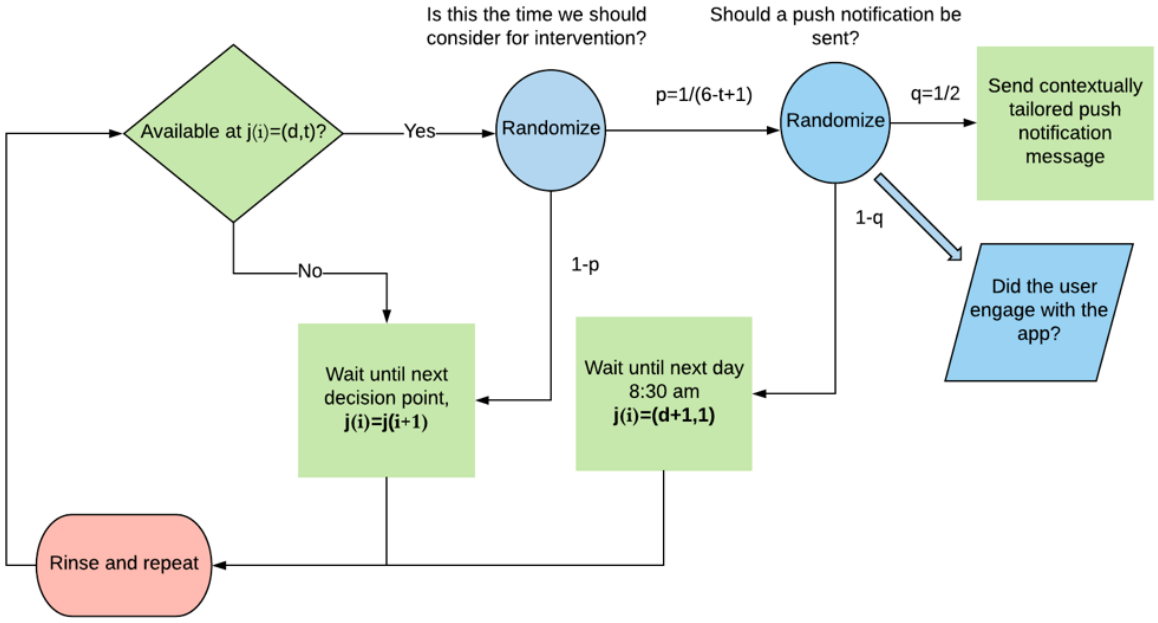


Figure 2. Micro-randomized trial design in JOOL app. Each decision point j (i)=(d, t) where i=1, 2,…,534, corresponding to a time of day t=1,2,…,6; 8:30 am, 12:30 pm, 5:30 pm, 6:30 pm, 7:30 pm, 8:30 pm with in a day d=1, 2, …, 89.



Data Gathering Process

Microrandomizations are a standard part of the JOOL app's quality improvement process. The MRT being reported in this paper, in particular, is part of the JOOL's effort to improve the quality of JOOL's "push notifications" feature. This quality improvement trial was rolled out to users in March 2017. All individuals who had the app installed on their phone and had push notifications enabled between March 2017, when the randomization software was rolled out, and August 2017 were included in this study. Users who did not use the app or used it just once after downloading the app were still included if they met the above eligibility criteria. Users who disabled push notification during the study were considered eligible until the next decision in the trial and then unavailable unless and until they re-enabled the push notifications.

A collaboration agreement was established between researchers and JOOL Health Inc. to undertake this study. The design of this study and analysis processes were carried out according to terms of service and privacy policy statements in the JOOL app, consented by the users when they created their account. For this study, a limited, nonidentifiable dataset was made available to fit the preplanned analytic models. The dataset included the MRT details and information about how and when users interacted with the app but did not include any of the details the users entered in the app. The data made available for analysis were anonymous, nonidentifiable, and housed only on JOOL Health's servers. Institutional review board letters from both the University of Michigan and Flinders University are available.

Outcome Measure

For all aims, the outcome is user engagement with the app, which we operationalize as whether or not the user charts in the app over the next 24 hours (as this is the first and most important interaction a user engages in after opening the app); this is a binary, longitudinal outcome. A total of 24 hours was allowed after a push notification for a user to respond because the self-monitoring is designed to be done daily; the self-monitoring questions concern behaviors and feelings over the prior 24-hour period. The primary aim tests whether there is an average effect of sending a push notification containing tailored health message versus not sending a push notification. This *main effect* is an average effect over time and over any other baseline or time-varying (eg, contextual) characteristics of the user. The secondary aims 1 and 2 focus on examining whether the effect of sending the push notification with a tailored health message differs by time (we examine week-of-study and weekdays vs weekends).

Availability

To mitigate the risk of users either turning off notifications or deleting the app due to receiving too many push notifications, users were classified as either "available" or "unavailable" at each time point, and only those time points when users were "available" were considered for push notification decision. Several rules were applied to determine availability. First, a user is "available" to receive only 1 push notification decision in a day, so after a push notification decision was made at a time point, users were classified as "not available" for consideration at subsequent time points in that day. Second, during weekends, users were considered as "unavailable" before noon. Finally, frequent prompts could potentially alienate inactive users, and result in them either disabling notifications or uninstalling the app [20,40]. To mitigate this risk, a user was also considered either "available" or "unavailable" on a day based on their longitudinal disengagement, classified as the number of days of inactivity with the app. Users who used the app less frequently were "unavailable" to receive push

notifications on a greater number of days compared with those who used the app more often. JOOL identified 4 distinct types of app users based on the clusters observed within the entire population of app users' inactivity patterns (see Table 1). The majority of individuals who stopped using the app were clustered within 0 to 9 days, followed by 10 to 29 days, and the remainder in 30+ days. As shown in Table 1, as the number of inactive days with the app increased, the number of "unavailable" days increased. This corresponds to the decrease in frequency of prompts received as users become inactive over longer periods. Users who stopped using the app recently (2 to 9 days), or in the early phases of disengagement, were "unavailable" for 2 days after a previous push notification decision. On the other hand, users who were inactive over 30 days were "unavailable" for 15 days after a previous notification decision. Similarly, during the first 2 days when the majority of users are most active with the app, users received prompts less frequently as they are already motivated and interacting with the app.

Data Analytic Plan

To analyze the data, we used a generalization of log-linear regression analysis specifically developed to ensure unbiased estimation of the causal effects of a time-varying treatment (ie, sending a push notification vs not) on a time-varying outcome (ie, charting over the next 24 hours) in mHealth settings. The method is a generalization of the approaches described in Boruvka et al (2016) and Dempsey et al (2017), with the use of a log-link function to accommodate the binary outcome [41,42]. These analyses pool time-varying, longitudinal data across all study participants. A separate analysis was conducted for each aim; each analysis involved prespecifying 2 sets of covariates before conducting the analyses. The first set of covariates, *X*, is used to examine moderation of the causal effect of sending a push notification with a tailored health message versus not sending a push notification (in moments of availability). The second set of control covariates (which can be time-varying) is a set of covariates that are expected to be highly correlated with app use in next 24 hours; these covariates are chosen to reduce the noise (ie, increase statistical efficiency) when assessing the effect of the push notification with a tailored health message versus none.

The causal effect is expressed on the "risk-ratio" scale, that is, on a scale that measures the probability ("risk") of completing the monitoring activity in the next 24 hours when a push notification is sent in a moment of availability, divided by the probability of completing the monitoring activity in the next 24 hours when a push is not sent in a moment of availability. If sending a push notification has a causal effect on the probability of completing the monitoring activity in the next 24 hours, the risk ratio will be different from 1. If sending a push increases the probability of completing the monitoring activity in the next 24 hours, the risk ratio will be greater than 1. Specifically, for each analysis, we modeled the log of the risk ratio linearly in *X*, using X^T beta, where the dimension of the unknown vector of parameters beta is the same as the number of covariates in *X*.

Table 2 provides the covariates in *X* corresponding to each aim and the hypothesis test corresponding to each aim. The covariate *week in study* is coded as 0 for the first week, 1 for the second week, and so on up to 12 for the final week of the study. The covariate *which day* is a binary variable. It has a value of 1 when the decision time is on a Monday to Friday, or 0 when the decision time is on a Saturday or Sunday. Each of the 3 preplanned hypothesis tests used a Wald statistic to test the null hypothesis that all the terms listed in column 3 of Table 1 are 0. For the primary and secondary aims, we set the type-1 error to 5%. All SEs were adjusted for within-person correlation in the binary outcome over time. We also reported estimates (and a 95% CI) of the average causal effect of sending a push notification (vs not) on the risk-ratio scale.

In the exploratory aim, we explored the effects of the push notification with a tailored health message by the time of day within weekday (vs weekend). Here, the time-of-day covariate has 6 possible levels corresponding to 6 possible decision points at which a push notification could be sent within a day: 8:30 am, 12:30 pm, 5:30 pm, 6:30 pm, 7:30 pm, and 8:30 pm. As these are exploratory analyses, we did not conduct hypothesis tests (ie, no *P* values will be reported for the exploratory aim). Instead, for these exploratory analyses, we provided plots of the model-based estimates of the effect of the push notification with a tailored health message versus no push notification (on the log of the risk-ratio scale) across different levels of the *X* covariates and report point-wise 90% CI around these estimates.

Table 1. Relationship between engagement patterns and push notification frequency.

Number of days since user last engaged with the app	Number of days to wait before sending a push notification (frequency of notification)
<2	3 (twice a week)
2-9	2 (2-3 times a week)
10-29	6 (weekly)
30+	15 (fortnightly)

Table 2. Covariates used to model the causal effect of sending a prompt versus not for each of the three aims.

Aim	Covariates, *X*	Hypothesis test
Primary aim	Intercept	Intercept
Secondary aim 1	Intercept (week in study)	Week in study
Secondary aim 2	Intercept, day (weekday or weekend)	Day (weekday or weekend)

For the primary and both the secondary aims, the following 5 control covariates were used: (1) *week in study*; (2) *time since last chart*; (3) *pushed indicator*, an indicator for whether a push has been sent in the past (once a push is sent, this indicator has a value of 1.0 for all remaining time points, otherwise it has a value of 0); (4) *push success ratio*, that is, total number of charts within 24 hours of receiving a push notification any time in the past divided by the total number of push notification sent any time in the past (note that *push success ratio* is nested within *pushed indicator*, ie, if no push notifications are sent in the past, the value is 0); and (5) *has charted 10 time* s, that is, whether the user has charted at least 10 times, which corresponds to the number of times a user must engage to unlock insights feature within the app. For the secondary aim 2, we additionally adjusted for *which day* (weekday vs weekend) as a control covariate.

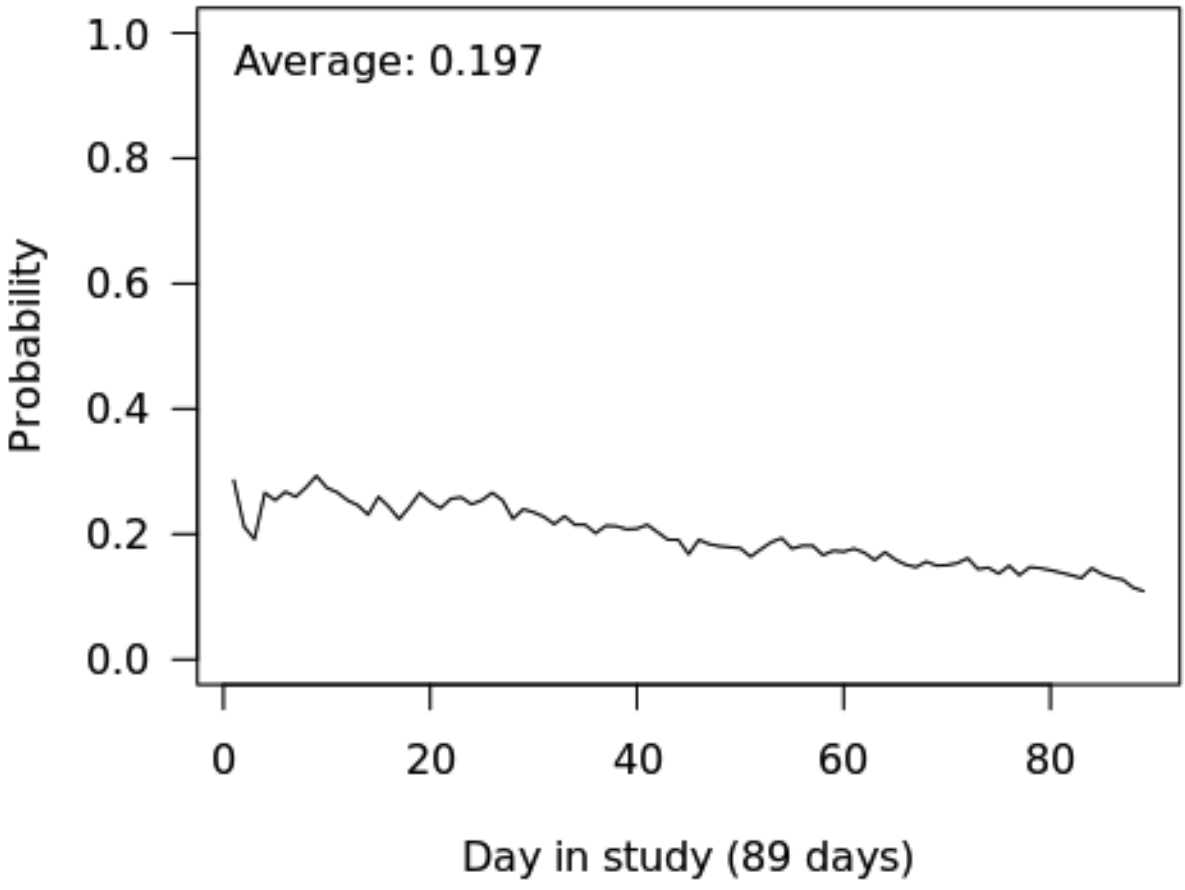
Results

Participant Characteristics

During the study period between March 2017 and August 2017, a total of 3300 users had the app installed on their phone, but 61.96% (2045/3300) of the users did not have push notifications enabled and were thus excluded from this study. The deidentified dataset analyzed in this study contained records from each of the 1255 eligible app users across 534 decision points (6 times per day over 89 days).

Among the study sample, 63.97% (790/1235) were females, 28.86% (357/1237) were aged under 30 years, 42.44% (525/1237) were aged between 30 and 50 years, and the remaining 28.70% (355/1237) were older than 50 years. Using a body mass index cut-off of 25 or higher, 52.88% (652/1233) of the participants were either overweight or obese. Figure 3 shows the percentage of individuals available by day in the study. On an average, over the duration of the study, JOOL app users were available approximately 20% of the time. The results for each of the aims are presented below.

Figure 3. Percentage of individuals available by day in study.



Primary Aim Analysis

There is sufficient evidence to reject the primary aim null hypothesis that states there is no effect of the push notification with a tailored health message ($P<.05$, Table 3). On the basis of the results of this analysis, it is estimated that on an average, individuals are 3.9% more likely to chart in the next 24 hours when a notification with a tailored health message is sent versus when a push notification is not sent (95% CI 1.01 to 1.08).

Secondary Aim 1 Analysis

On the basis of the results of this analysis (Table 4), there is insufficient evidence to reject the null hypothesis that the effect of a push notification containing a tailored health message versus not pushing a notification does not differ by week in study ($P=.84$). Figure 4 shows the estimated effects by week in study on the log risk-ratio scale (left) and risk-ratio scale (right) based on this model.

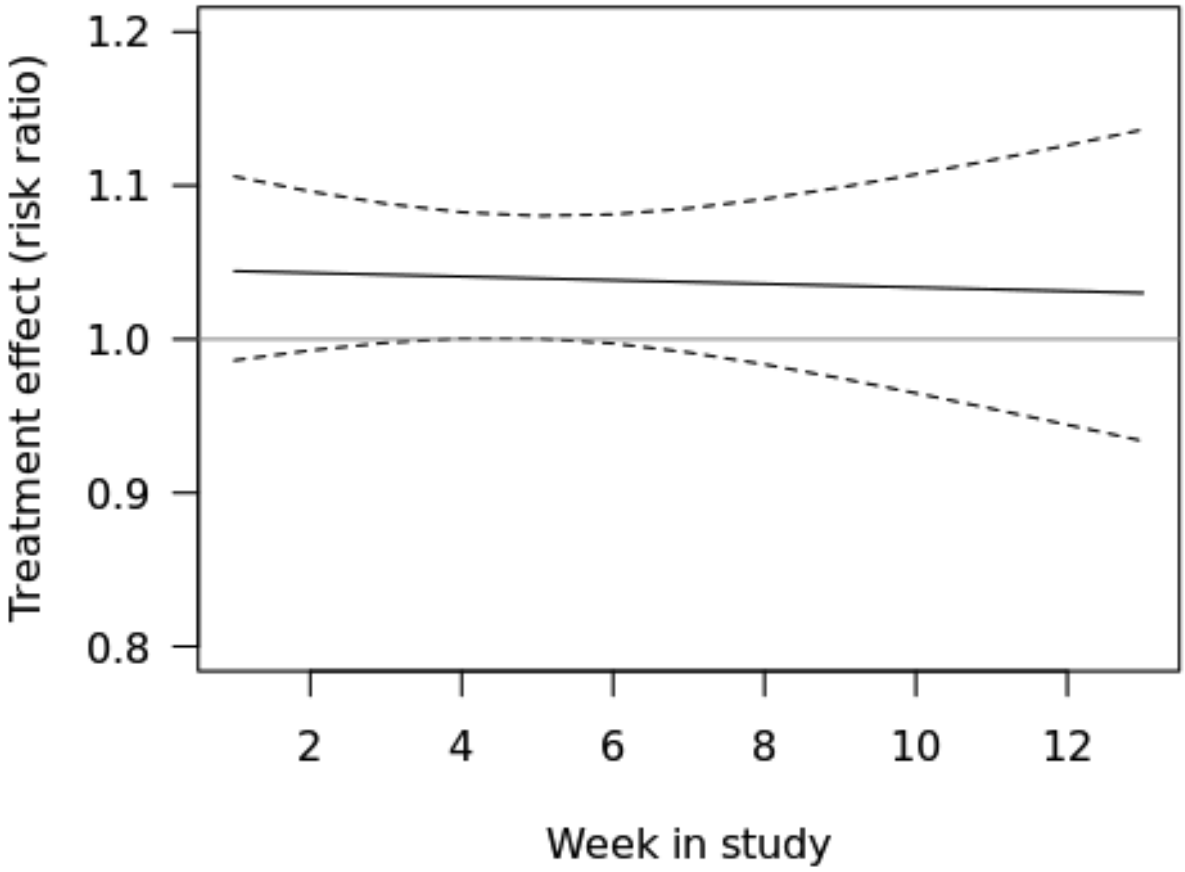
Table 3. Overall effects of the push notification with a tailored health message (primary aim).

Effect	Coefficient	SE	95% CI	*P* value
Causal				
Decision to push (=yes)	0.04	0.02	0 to 0.08	.047
Control covariates				
Intercept	−0.32	0.03	−0.38 to −0.26	
Week in study	0.00	0.01	−0.01 to 0.01	
Days since chart	−0.22	0.05	−0.31 to −0.13	
Pushed indicator	−0.91	0.12	−1.15 to −0.67	
Pushed indicator × push success ratio	0.75	0.11	0.54 to 0.97	
Has charted 10 times	0.16	0.05	0.07 to 0.26	

Table 4. Effects of push notification by week in the study (secondary aim 1).

Effect	Coefficient	SE	95% CI	*P* value
Causal				
Decision to push (=yes)	0.04	0.03	−0.01 to 0.10	.14
Decision to push (=yes) x week in study	0	0.01	−0.01 to 0.01	.84
Control covariates				
Intercept	−0.32	0.03	−0.38 to −0.26	
Week in study	0	0.01	−0.01 to 0.01	
Days since chart	−0.22	0.05	−0.31 to −0.13	
Pushed indicator	−0.91	0.12	−1.15 to −0.67	
Pushed indicator x success ratio	0.75	0.11	0.54 to 0.97	
Has charted 10 times	0.16	0.05	0.07 to 0.26	

Figure 4. Effects of push notification over course of the trial.



Secondary Aim 2 Analysis

There is insufficient evidence to reject the null hypothesis that the weekend and weekday effects are not different from each other (*P*=.18, Table 5). However, the effect of a push notification with a tailored health message versus no push notification is estimated to be somewhat larger on weekends than on weekdays. Specifically, it is estimated that, *on weekends*, individuals are 8.7% more likely to chart in the next 24 hours when pushed a tailored health message versus when not pushed (95% CI 1.01 to 1.17). Whereas, *on weekdays*, individuals are 2.5% more likely to chart in the next 24 hours when pushed a tailored health message versus when not pushed (95% CI 0.98 to 1.07).

Table 5. Effects of push notification by weekend versus weekday (secondary aim 2).

Effect	Coefficient	SE	95% CI	*P* value
Causal				
Decision to push (=yes)	0.084	0.04	0.007 to 0.156	.03
Decision to push (=yes) x which day	−0.059	0.044	−0.145 to 0.026	.18
Control covariates				
Intercept	−0.396	0.042	−0.477 to −0.313	
Which day	0.084	0.029	0.027 to 0.141	
Week in study	−0.002	0.006	−0.014 to 0.009	
Days since chart	−0.220	0.046	−0.311 to −0.129	
Pushed indicator	−0.898	0.123	−1.140 to −0.657	
Pushed indicator x push success ratio	0.757	0.111	−0.540 to 0.974	
Has charted 10 times	0.164	0.049	0.068 to 0.259	

Exploratory Aim Analysis

In this analysis, we examined the effect of a push notification containing a tailored health message versus no push notification by the time of day. There are 6 times of day: 8:30 am, 12:30 pm, 5:30 pm, 6:30 pm, 7:30 pm, and 8:30 pm; on weekends, no notifications are pushed at 8:30 am. There are 2 parts to this analysis. In part 1, we examined whether the effect of the push notification varies by the time of the day. In part 2, we examined whether the effect of the push notification varies by time of day within weekend versus weekday.

Part (1)

The data indicate that the largest effect of the push notification containing a tailored health message is at 12:30 pm (Figure 4). Specifically, it is estimated that when the notification is pushed at 12:30 pm (vs not pushed at 12:30 pm), individuals are 8.8% more likely to chart in the next 24 hours (90% CI 1.04 to 1.15).

Part (2)

During *weekdays*, the effect of the push notification was greatest at 12:30 pm (Figure 5). Specifically, it is estimated that when a tailored health message is pushed at 12:30 pm (vs not pushed at 12:30 pm), individuals are *exp* (0.071)=1.074 times as likely (ie, 7.4% as likely) to chart in the next 24 hours (90% CI 1.02 to 1.13). During *weekends*, the effects were greatest at 12:30 pm and 7:30 pm. Specifically at 12:30 pm on weekends, the risk ratio is 1.118 or 11.8% are more likely to chart in the next 24 hours (90% CI 1.02 to 1.23). The effect is similar at 7:30 pm on weekends. Figure 6 shows the effects by time of day and weekday versus weekends (with 90% confidence limits). These results are congenial with the results of the secondary aim 2 analyses, which suggested that the effects were somewhat stronger on weekends than on weekdays.

Figure 5. Effects of push notification over different times in a day.

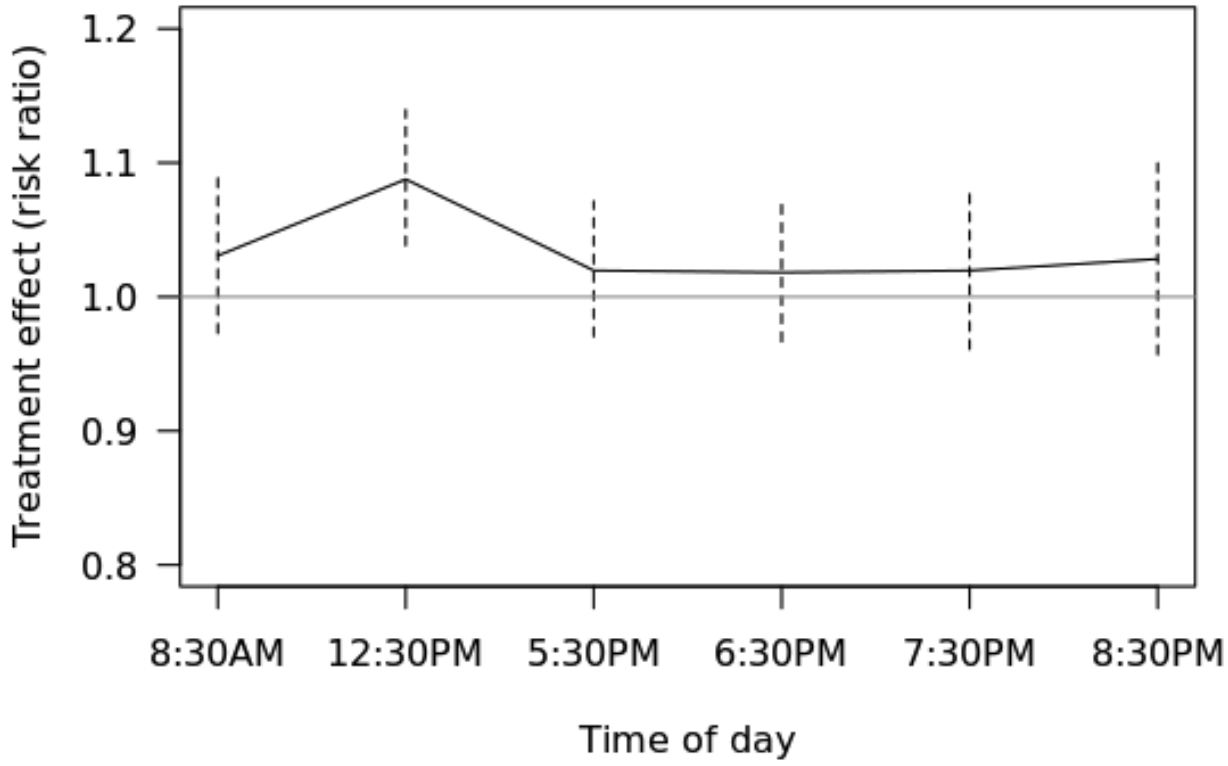
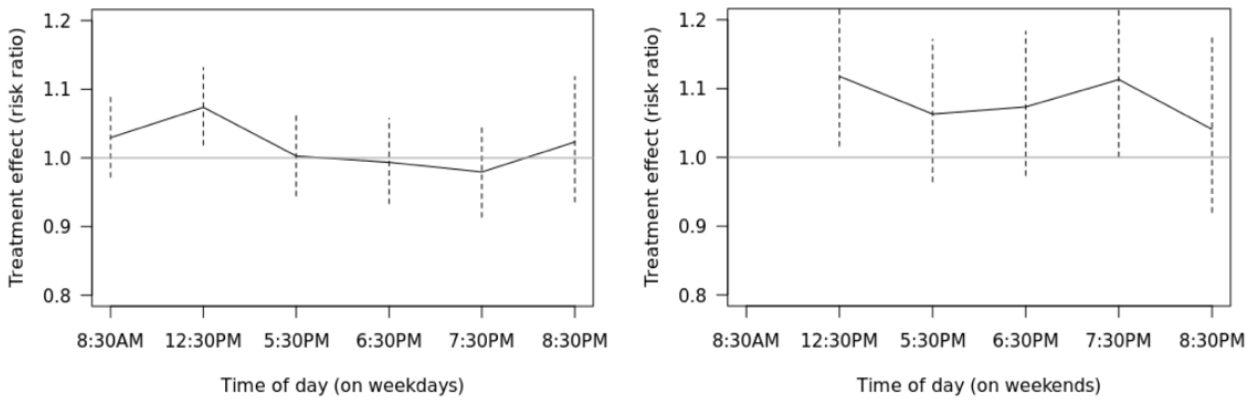


Figure 6. Effects of push notification over different times in a day (weekday and weekends).



Discussion

Principal Findings

The results indicate that pushing a notification with a tailored health message impacts near time, proximal engagement with the self-monitoring activity in the app. The effect of the pushed notifications is sustained over time, suggesting that push notifications containing tailored health messages can attenuate the rate at which users disengage. Both the positive effect of notifications [35,43], as well as the decline in effects over time on engagement have been noted in prior research [43]. Finally, the results suggest that push notifications with a tailored health message result in higher rates of engagement during weekends and according to time at mid-day. This is contrary to the lack of effect of the timing of notifications on engagement observed in prior research [35]. Substantial differences in trial design and relatively small sample sizes may explain why these effects were not observed in previous studies.

Implications

The findings suggest that users are more likely to engage with the app in the 24 hours following a push notification containing a tailored health message compared with no push notification. However, the effects are relatively small, with only a 3.9% greater likelihood of engaging with the app in the next 24 hours. The likelihood of users viewing the notifications within 24 hours is very high. Previously, it has been noted that the probability of users clicking a notification increases from 50% within 30 seconds to 83% in 5 min [24]. This suggests that prompts would have captured users' attention most of the time, but their attention would have translated to engagement with the self-monitoring activity in the app less frequently. On the other hand, even small effects on engagement can be of substantial benefit at the population level, given the scalability of app-based resources. Similarly, the effects may be small because not all push notifications, even when messages are contextually tailored and personalized, are likely to be persuasive to all users at all times. In the design of this study, push notifications were sent at different times and days of the week when users are likely to be in different contexts. As a result, over the course of the trial, some users could have responded fewer times to prompts if they had received them in certain contexts where receptiveness to interruptions is low. During such contexts, particularly if engaged in a cognitively demanding activity, they may be less likely to pay attention to notifications. In fact, the results from the exploratory aim suggest that effects vary by time of the day and day of the week. In general, the largest effect occurs when the notification is pushed at 12:30 pm—users are 8.8% more likely to chart in the next 24 hours. During the weekends, the largest effects occur at 2 time points (12:30 pm and 7:30 pm). Specifically at 12:30 pm on weekends, users are 11.8% more likely to chart in the next 24 hours. The effect is similar at 7:30 pm on weekends. These findings suggest that sending push notifications with contextually tailored messages over time, particularly if sent during those contextual moments when users are most receptive, can serve as an effective strategy to maximize engagement with mHealth app interventions.

As this study population involved office workers, users who are typically at work during weekdays, it is likely that they responded more to weekend prompts because they were less busy. This was suggested by the secondary analysis where the effect of push notification versus no notification was estimated to be somewhat larger on weekends than on weekdays. The exploratory analysis suggests that on weekdays, effects are the strongest at 12:30 pm, which coincides with the time office workers generally have a lunch break. This is consistent with previous research that suggests that engagement occurs more during nonworking hours [39,40]. In summary, the findings from this study suggest that push notification prompts hold promise as an effective engagement strategy for mHealth apps, and through contextual tailoring, further advancements can be achieved in reducing user disengagement.

Tackling the problem of poor engagement is further compounded by a lack of consensus on how to operationally define engagement with an app-based health intervention. Engagement has been typically operationalized in terms of usability or usage of the intervention, along with the factors that influence these [44]. Usage can refer to the frequency or the duration of either interaction within the app, or the practice of behavioral and cognitive strategies offered by the app in the real environment. Under both scenarios, more usage, either interacting with the app more often, or practicing behaviors and cognitive strategies learned through the app more often, is viewed as better engagement. This is based on the assumption that more usage is closely related to better outcomes. However, the relationship between usage and health outcomes is weak, as supportive evidence is mostly anecdotal or correlational [45]. Instead, shifting the focus on effective engagement with the digital intervention that may or may not require sustained usage but that mediates positive behaviors is an emerging alternative [46]. In published studies of different behavioral interventions, self-monitoring is encouraged at a variety of frequencies (up to multiple times per day) [2,12,47-49]. Evidence from these studies suggests effective self-monitoring can result in a profound positive impact on health outcomes. However, the frequency of monitoring varies between individuals and over time, and as a result, it is not evident how frequently users should be encouraged to self-monitor. In fact, encouraging high rates of tracking could potentially worsen stress and cause harm [50]. Further work evaluating different self-monitoring frequencies is needed.

There are several novel aspects and strengths to this study. First, this study is designed to investigate the effect of tailored push notifications (vs no notification) on immediate proximal engagement with an mHealth app. Second, this research is naturalistic, as the study did not pay users for study involvement nor did it employ clinical staff who text, telephone, or otherwise contact users to ensure that they stay engaged in the study. There is no concept of dropout in this study. When individuals stop using the app during the trial, their outcome is recorded as disengaged in the analysis as opposed to being a dropout and excluded from the analysis. Third, the use of the MRT design to repeatedly randomize users over the course of the trial allowed us to investigate the causal effects of push notifications on proximal engagement and the real-time, real-world conditions

that likely influence these proximal effects. In future, the MRT approach can also be used to adjust the probability of randomization to favor pushes in certain contexts. This experimental approach, therefore, provides the empirical evidence necessary to optimize engagement. Finally, investigating engagement in a large sample of users with an app that focuses on multiple health behaviors is another strength of this study.

Study Limitations

Individual characteristics such as personality traits and socioeconomic status (SES) are known to influence engagement with mHealth apps, and controlling for these variables could narrow the CIs to effects of time-varying push notifications. Qualitative assessment of users' experiences with push notifications would have offered greater insights into engagement, but as the study was constrained by data already collected within the app, such measures were not available for inclusion in our analysis. However, as noted earlier, randomizing the decision to send a push notification at a point in time ensures that within-user (eg, mood, location) and between-user (eg, personality traits, SES) factors contributing to the day-to-day variations in engagement are balanced across conditions.

Another limitation relates to the timescale used to define the proximal outcome in this study. Specifically, the proximal outcome was whether or not the individual engaged (ie, completed charting) with the app within 24 hours. A more sensitive proximal outcome measure might enable investigating the more immediate effects of prompts on engagement, potentially yielding larger effects than those observed in this study. Within these limitations, this study provides a first step to understanding whether and under what conditions push notifications promote proximal engagement in mHealth.

Conclusions

Research into approaches for optimizing engagement with mHealth interventions is warranted and potentially valuable to the public's health. Findings from this study suggest that push notifications can indeed influence engagement with a health app. Moreover, the results suggest that engagement effects are sustained over time but that the effect is different across contexts such as the time of day and day of week. On the basis of these results, mobile app developers are advised to incorporate push notifications as an engagement strategy and to pay attention to when prompts are sent and the types of prompts that are sent. Finally, the study offers an innovative trial design to optimize push notification delivery in mHealth apps. This approach can be incorporated into the structure of real-world apps.

Future research in this area should further investigate the contexts in which users respond to prompts and to use designs such as MRT to examine how various push-based interventions influence engagement within these contexts.

Acknowledgments

The authors thank the JOOL Health Engineering team, especially Lisa Schutte for feedback.

Conflicts of Interest

DA, SM, and NI declare no conflicts of interest. TP, HM, and VS are shareholders of JOOL Health. NB is a consultant of JOOL Health.

References

1. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
2. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Dec 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
3. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Dec;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
4. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
5. Martinez O, Wu E, Shultz AZ, Capote J, López RJ, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
6. Session S. Governing Body matters: Important issues arising out of the Sixty-sixth World Health Assembly and the 132nd and 133rd sessions of the WHO Executive Board. 2013 Sep. URL: http://apps.who.int/iris/handle/10665/128278?locale-attribute=de&locale=ar
7. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: randomized controlled trial. J Med Internet Res 2017 May 12;19(5):e141 [FREE Full text] [doi: 10.2196/jmir.6775] [Medline: 28500020]

8. Digitaltrends. Are you a rarity? Only 16 percent of people will try out an app more than twice URL: https://www.digitaltrends.com/mobile/16-percent-of-mobile-userstry-out-a-buggy-app-more-than-twice/ [accessed 2018-10-15] [WebCite Cache ID 73B9Gbm0V]
9. CHICStaff. Consumer-health. Motivating patients to use smartphone health apps URL: http://www.consumer-health.com/motivating-patients-to-use-smartphone-health-apps/ [accessed 2018-10-15] [WebCite Cache ID 73B9Vn62T]
10. Serrano K, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Dec;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
11. Torous J, Staples P, Slaters L, Adams J, Sandoval L, Onnela JP, et al. Characterizing smartphone engagement for schizophrenia: results of a naturalist mobile health study. Clin Schizophr Relat Psychoses 2017 Aug 04. [doi: 10.3371/CSRP.JTPS.071317] [Medline: 28777029]
12. Rahman Q, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, Manage My Pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
13. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
14. Strecher VJ. Life on Purpose. US: HarperCollins Publishers Inc; May 23, 2016.
15. Falk E, O'Donnell MB, Cascio CN, Tinney F, Kang Y, Lieberman MD, et al. Self-affirmation alters the brain's response to health messages and subsequent behavior change. Proc Natl Acad Sci U S A 2015 Feb 17;112(7):1977-1982 [FREE Full text] [doi: 10.1073/pnas.1500247112] [Medline: 25646442]
16. Kelders S, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
17. Carroll J, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Dec 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
18. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Dec 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
19. Zhang S, Elhadad N. Factors contributing to dropping-out in an online health community: static and longitudinal analyses. AMIA Annu Symp Proc 2016;2016:2090-2099 [FREE Full text] [Medline: 28269969]
20. Dennison L, Morrison L, Conway G, Yardley. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
21. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
22. Pielot M, Church K, de Oliveira R. An In-situ Study of Mobile Phone Notifications. 2014 Presented at: 16th International Conference on Human-computer Interaction with Mobile Devices & Services 233-242 (ACM); September 23-26, 2014; Toronto, ON, Canada.
23. Fischer JE, Greenhalgh C, Benford S. Investigating Episodes of Mobile Phone Activity As Indicators of Opportune Moments to Deliver Notifications. 2011 Presented at: 13th International Conference on Human Computer Interaction with Mobile Devices and Services 181-190 (ACM); August 30-September 02, 2011; Stockholm, Sweden. [doi: 10.1145/2037373.2037402]
24. Sahami Shirazi A. Large-scale Assessment of Mobile Notifications. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems 3055-3064 (ACM); April 26-May 01, 2014; Toronto, Ontario, Canada. [doi: 10.1145/2556288.2557189]
25. Kushlev K, Proulx J, Dunn EW. Silence Your Phones: Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. 2016 Presented at: CHI Conference on Human Factors in Computing Systems 1011-1020 (ACM); May 07-12, 2016; San Jose, California, USA. [doi: 10.1145/2858036.2858359]
26. Pejovic V, Musolesi M, Mehrotra A. Investigating The Role of Task Engagement in Mobile Interruptibility. 2015 Presented at: 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct 1100-1105 (ACM); August 24-27, 2015; Copenhagen, Denmark. [doi: 10.1145/2786567.2794336]
27. Exler A, Braith M, Schankin A, Beigl M. Preliminary Investigations About Interruptibility of Smartphone Users at Specific Place Types. 2016 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12-16, 2016; Heidelberg, Germany. [doi: 10.1145/2968219.2968554]
28. Mehrotra A, Müller SR, Harari GM, Gosling SD, Mascolo C, Musolesi M, et al. Understanding the role of places and activities on mobile phone interaction and usage patterns. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-22. [doi: 10.1145/3131901]

29. Pejovic V, Musolesi M. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. 2014 Presented at: 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, Washington p. 897-908. [doi: 10.1145/2632048.2632062]
30. Pejovic V, Musolesi M. Anticipatory mobile computing: a survey of the state of the art and research challenges. ACM Comput Surv 2015 Apr 16;47(3):1-29. [doi: 10.1145/2693843]
31. Mehrotra A, Pejovic V, Vermeulen J, Hendley R, Musolesi M. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. 2016 Presented at: 2016 CHI Conference on Human Factors in Computing Systems 1021-1032 (ACM Press); May 07-12, 2016; San Jose, California, USA. [doi: 10.1145/2858036.2858566]
32. Turner LD, Allen SM, Whitaker RM. Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. 2015 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing 801-812 (ACM); September 07-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2807514]
33. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
34. Spohr S, 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]
35. Morrison L, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2017;12(1):e0169162 [FREE Full text] [doi: 10.1371/journal.pone.0169162] [Medline: 28046034]
36. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
37. Joolhealth. URL: https://www.joolhealth.com/ [accessed 2018-10-18] [WebCite Cache ID 73FqY4qqE]
38. Hawkins R, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
39. Morrison L, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: a mixed-methods study. J Med Internet Res 2014 Oct 22;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
40. Bentley F, Tollmar K. The power of mobile notifications to increase wellbeing logging behavior. 2013 Presented at: SIGCHI International Conference on Human Factors in Computing Systems 1095-1098 (ACM); April 27-May 02, 2013; Paris, France. [doi: 10.1145/2470654.2466140]
41. Dempsey W, Liao P, Kumar S, Murphy SA. ArXiv. 2017. The stratified micro-randomized trial design: sample size considerations for testing nested causal effects of time-varying treatments URL: https://arxiv.org/pdf/1711.03587.pdf [accessed 2018-10-18] [WebCite Cache ID 73FpFMyfR]
42. Boruvka A, Almirall D, Witkiewitz K, Murphy S. Assessing time-varying causal effect moderation in mobile health. J AM Stat Assoc 2017 Mar 31;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274]
43. Freyne J, Yin J, Brindal E, Hendrie GA, Berkovsky S, Noakes M. Push Notifications in Diet Apps: Influencing Engagement Times and Tasks. Int J Human Comput Interact 2017 Jan 11. [doi: 10.1080/10447318.2017.1278896]
44. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
45. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
46. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Dec;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
47. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
48. Kanejima Y, Kitamura M, Izawa KP. Self-monitoring to increase physical activity in patients with cardiovascular disease: a systematic review and meta-analysis. Aging Clin Exp Res 2018 Apr 30. [doi: 10.1007/s40520-018-0960-7] [Medline: 29714027]
49. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Dec;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
50. Shieh C, Draucker CB. Self-monitoring lifestyle behavior in overweight and obese pregnant women: qualitative findings. Clin Nurse Spec 2018;32(2):81-89. [doi: 10.1097/NUR.0000000000000355] [Medline: 29419580]

Abbreviations

NHS: National Health Service
mHealth: mobile health
MRT: microrandomized trial

Edited by G Eysenbach; submitted 26.02.18; peer-reviewed by L Morrison, J Wang, H Singh; comments to author 12.04.18; revised version received 20.06.18; accepted 10.07.18; published 29.11.18.

Please cite as:
Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, Maaieh H, Strecher V
To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App
JMIR Mhealth Uhealth 2018;6(11):e10123
URL: http://mhealth.jmir.org/2018/11/e10123/
doi:10.2196/10123
PMID:30497999

Original Paper

Impact of a Wearable Device-Based Walking Programs in Rural Older Adults on Physical Activity and Health Outcomes: Cohort Study

Il-Young Jang[1], MD; Hae Reong Kim[2], MS; Eunju Lee[1], MD, PhD; Hee-Won Jung[3], MD, PhD; Hyelim Park[1,4], RN; Seon-Hee Cheon[4], RN; Young Soo Lee[1], MD, PhD; Yu Rang Park[2], PhD

[1]Division of Geriatric Medicine, Department of Internal Medicine, Asan Medical Center, Seoul, Republic of Korea

[2]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Graduate School of Medical Science And Engineering, Korea Advanced Institute Of Science And Technology, Daejeon, Republic of Korea

[4]Pyeongchang Health Center & Country Hospital, Gangwon-do, Republic of Korea

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 2493
Fax: 82 2 2228 2494
Email: yurangpark@yuhs.ac

Abstract

Background: Community-dwelling older adults living in rural areas are in a less favorable environment for health care compared with urban older adults. We believe that intermittent coaching through wearable devices can help optimize health care for older adults in medically limited environments.

Objective: We aimed to evaluate whether a wearable device and mobile-based intermittent coaching or self-management could increase physical activity and health outcomes of small groups of older adults in rural areas.

Methods: To address the above evaluation goal, we carried out the "Smart Walk" program, a health care model wherein a wearable device is used to promote self-exercise particularly among community-dwelling older adults managed by a community health center. We randomly selected older adults who had enrolled in a population-based, prospective cohort study of aging, the Aging Study of Pyeongchang Rural Area. The "Smart Walk" program was a 13-month program conducted from March 2017 to March 2018 and included 6 months of coaching, 1 month of rest, and 6 months of self-management. We evaluated differences in physical activity and health outcomes according to frailty status and conducted pre- and postanalyses of the Smart Walk program. We also performed intergroup analysis according to adherence of wearable devices.

Results: We recruited 22 participants (11 robust and 11 prefrail older adults). The two groups were similar in most of the variables, except for age, frailty index, and Short Physical Performance Battery score associated with frailty criteria. After a 6-month coaching program, the prefrail group showed significant improvement in usual gait speed (mean 0.73 [SD 0.11] vs mean 0.96 [SD 0.27], P=.02), International Physical Activity Questionnaire scores in kcal (mean 2790.36 [SD 2224.62] vs mean 7589.72 [SD 4452.52], P=.01), and European Quality of Life-5 Dimensions score (mean 0.84 [SD 0.07] vs mean 0.90 [SD 0.07], P=.02), although no significant improvement was found in the robust group. The average total step count was significantly different and was approximately four times higher in the coaching period than in the self-management period (5,584,295.83 vs 1,289,084.66, P<.001). We found that participants in the "long-self" group who used the wearable device for the longest time showed increased body weight and body mass index by mean 0.65 (SD 1.317) and mean 0.097 (SD 0.513), respectively, compared with the other groups.

Conclusions: Our "Smart Walk" program improved physical fitness, anthropometric measurements, and geriatric assessment categories in a small group of older adults in rural areas with limited resources for monitoring. Further validation through various rural public health centers and in a large number of rural older adults is required.

(JMIR Mhealth Uhealth 2018;6(11):e11335) doi:10.2196/11335

KEYWORDS

adherence; frailty; older adult; rural area; wearable device

Introduction

Unlike young adults, elderly individuals have different medical characteristics. Each older adult has a highly heterogeneous health status [1,2]. Even with definite illness, symptoms are frequently ambiguous and nonspecific in older adults [1]. Also, many frail older adults have diminished physiologic reservoirs and easily deteriorate regardless of their comorbidities [3]. Therefore, early detection and an appropriate preventive approach are crucial for older adults.

In recent decades, a frailty-based approach has been widely applied in communities having an older population [4]. Frailty is an age-related syndrome characterized by decreased physiologic reserve and increased vulnerability to stressors that lead to adverse health outcomes, such as disability, falls, institutionalization, and mortality [3]. Unlike the traditional comorbidity-based approach, a frailty-based strategy includes the concepts of an individualized approach, disability prevention, and enhancing quality of life regardless of age [5].

Many of the older adults living in the community show sedentary behavior. Reports have shown that >60% of community- dwelling older adults are sedentary [6]. Sedentary behavior is an important risk factor for cardiovascular disease [7], falls [8], and frailty [9] in older adults and is known to be associated with high mortality [10]. Therefore, preventing sedentary behavior and increasing physical activity among older adults are important in the public health care model.

However, community-dwelling older adults living in rural areas are in a less favorable environment for health care compared to those living in an urban environment. Healthcare facilities in rural areas are usually poor or have limited accessibility [11]. In addition, rural dwellers have a relatively lower socioeconomic status and education level and a higher prevalence of living alone, multimorbidity, frailty, and disability than urban older adults [12]. Thus, the role of a public health center is increasingly emphasized in rural areas with an optimized public health strategy. However, rural public health centers should also service relatively larger areas of dispersed older adults with insufficient resources, making it difficult to manage a cost-effective health care model. In addition, like several other Asian countries, South Korea has seen a more rapid older population growth in rural areas than in urban areas [13]. These greater burdens of aging-related health conditions and resource barriers for health care in rural areas facilitate the paradigm shift from disease management to health care and prevention.

The health benefits of wearable devices are known in the older population [14]. Accurately assessing the physical activity of older adults through interviews and examination requires considerable time and effort [15]. Wearable devices have been applied to monitor physical activity, falls, or behavior of community-dwelling older adults and have been applied to change the lifestyle and reduce metabolic risk of older adults with chronic diseases [16-19]. From the perspective of public health care, mobile health care services are in the limelight. Specifically, such mobile services have been shown to induce behavioral change by adding coaching or incentives to wearable devices [20]. Some wearable devices could be suitable upon adjusting for rural health resources. However, reports have shown that adherence is reduced to <10% when the benefit to the participants disappears [21]. Furthermore, these incentives can only be supported for a limited period. In rural areas, developing a wearable device-based health care program that can be operated with a small amount of resources and can encourage voluntary self-management after the end of the program is important.

Despite unfavorable conditions, public health centers in rural areas have several unique strengths for community studies [11]. Geographical isolation from private hospitals helps maximize the participation rate of the senior population within a short period, as well as the long-term retention rate [22,23]. We believed that a mobile health care service centered on a public health center is the most affordable, easy to access, and relatively costless method for health care of rural older adults. This study aimed to evaluate whether a simple mobile health care device and mobile-based intermittent coaching or self-management can increase the physical activity and health outcome of older adults in rural areas.

This study aimed to identify answers to the following three questions: (1) Can a wearable device improve physical activity and health outcomes of older adults in rural areas? (2) Are there differences in physical activity and health outcome improvement depending on frailty status? (3) Are there differences in wearable device adherence between coaching and self-management? To address the above questions, we carried out the “Smart Walk” program as a health care model using a wearable device to promote self-exercise among community-dwelling older adults managed by a community health center.

Methods

Study Design

We randomly selected older participants who had enrolled in the population-based, prospective cohort study of aging, the Aging Study of Pyeongchang Rural Area (ASPRA). The ASPRA cohort was established in October 2014 to determine the burden of frailty and geriatric syndromes in rural areas, understand the disparities between urban and rural older populations, and set the priority of public health interventions. The design and measurement protocol are described elsewhere [11,12]. Briefly, participants living in Pyeongchang rural area, located 180 kilometers east of Seoul, South Korea, were administered an annual comprehensive geriatric assessment including physical, mental, psychosocial, and frailty status, as well as medical conditions. The inclusion criteria were (1) being aged ≥65 years; (2) being registered in the National Healthcare Service; (3) being

ambulatory with or without an assistive device; (4) living at home; and (5) being able to provide informed consent. Those who were living in a nursing home, hospitalized, or bed-ridden and receiving nursing-home-level care at the time of enrollment were excluded. To conduct this project, an academic-public health collaborative team was organized, and about 95% of eligible older adults in the study area were enrolled. The characteristics of ASPRA participants were similar to those of the Korean rural population represented in the Korea National Health and Nutrition Examination Survey [11].

Frailty and Comprehensive Geriatric Assessment

In this study, we screened for potential participants using the Cardiovascular Health Study (CHS) frailty phenotype criteria, one of the most widely used assessment tools (Multimedia Appendix 1). For quantitative evaluation of frailty status, we used the frailty index suggested by Rockwood et al [24], which encompasses physical, cognitive, psychosocial, geriatric syndrome, disability, and underlying disease. Scores vary from 0 to 1; a higher score indicates a more severe frailty burden [23].

Trained nurses administered the comprehensive geriatric assessment using the following instruments: the International Physical Activity Questionnaire (IPAQ) short form, the Korean version of the Mini-Mental State Examination for cognitive function, the Korean version of the Center for Epidemiological Studies Depression scale for depressive mood, usual gait speed, the Mini Nutritional Assessment-Short Form (MNA-SF) score for malnutrition, multimorbidity, grip strength on dominant arm, the Short Physical Performance Battery score, and bioimpedance analysis using Inbody 620 (Inbody, Seoul, Korea). Detailed methods were described previously [11].

Smart Walk Program

The “Smart Walk” program was a 13-month program conducted from March 2017 to March 2018, consisting of 6 months of coaching management, 1 month of rest, and 6 months of self-management. Initially, the health center staff provided a wearable device to the participants, helping with the installation of apps on mobile phones and mobile phone pairing. Participants established a “buddy” relationship with health center staff through the mobile app, which allowed them to monitor the activities and walks of each participant. If a problem with the connection between the mobile phone and wearable device appeared, the participants were asked to visit the health center to solve the problem.

Coaching was performed by 8 health center staff through notification messages of the wearable device. If no record of device use existed, weekly follow-up was performed. During the first 2 months, health center staff set all participants’ wearable device to a goal of 5000 steps per day, and in the 6th month of coaching period, 7000 steps were finally set as the daily target. This goal was reset such that the daily target was increased by 1000 steps every 2 months, and 7000 steps per day were finally required via wearable device in the 6th month of the coaching period. If no device use is observed or if the target step number is not reached, a push alarm is sent first through the app to encourage device use and exercise. After that, the health center staff checked the health status of the participant by phone or visit within a few days. Regarding education, health center staff did not instruct the participants on how to exercise unless the participants so requested.

An incentive was provided (including two group picnics, US $50 worth of nutritional supplements, and a wearable device worn by the participant) to encourage participants during the 6-month coaching period only, followed by a 7-month follow-up with monthly questionnaires and data logs from the wearable device. Coaching was performed by 8 health center staff through notification messages of the wearable device. If no record of device use existed, weekly follow-up was performed.

As no existing criteria for adherence to a wearable device was available, this study defined adherence as continuous use if the device was used for at least 1 week per month. This criterion refers to the follow-up at weekly intervals during the coaching period.

This study was approved by the Institutional Review Board of the Asan Medical Center (Institutional review board no. 2015-0673). We obtained a research study personal information agreement and exercise commitment letter from the study participants.

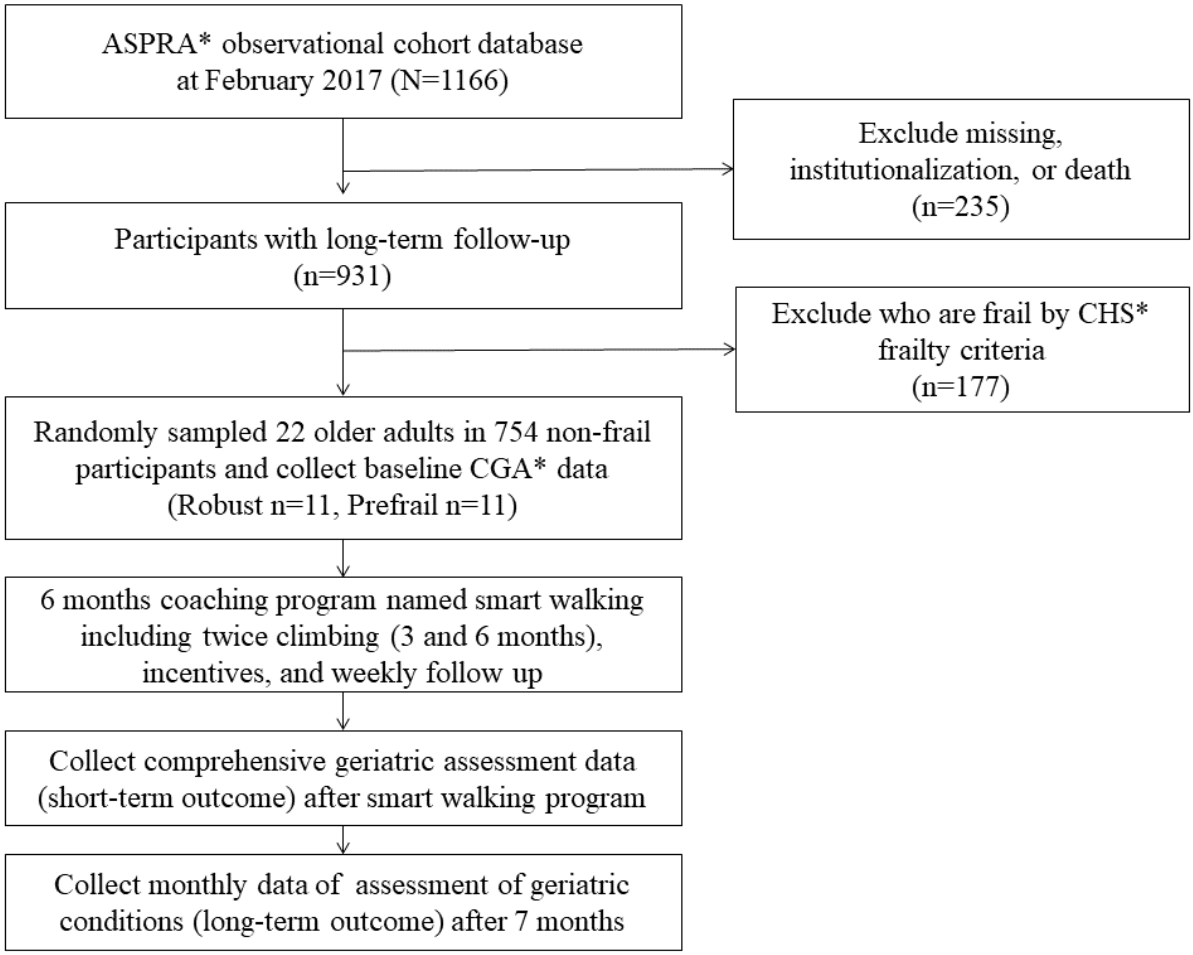
Data Description

In this study, we collected three types of data: (1) data from the whole period the device was worn; (2) Comprehensive geriatric assessment before and after the coaching program; and (3) monthly questionnaire data during the follow-up period. The wearable device used in this study was a Xiao Mi band 2. We chose this device because it was the model with the lowest battery consumption (up to 3 weeks on a single charge). Only step count was used for analysis. All participants underwent comprehensive geriatric assessment that encompassed the assessment of cognitive and physical function, depression, nutrition, and body composition using bioimpedance analysis identical with the protocol of the ASPRA cohort and an additional 3-minute walk test before and after the coaching program [25]. Detailed measurements and definitions are described elsewhere [11]. A monthly assessment of geriatric conditions, such as frailty screening, falls, number of hospital visits, and weight loss, was administered by an experienced nurse based on a one-on-one telephone or face-to-face interview. The nurse also obtained information on demographic characteristics, living status, occupation, income, education level, chronic conditions, etc (detailed variables can be found in a study by Hee-Won Jung et al [11]).

Data Analysis

Among a total of 1166 participants who participated in the ASPRA cohort, we excluded 177 participants who were graded within the frail state (based on the CHS frailty criteria [26]) given that their participation in the “Smart Walk” program would be difficult (Figure 1). Of the 754 participants who were classified as robust or prefrail, we randomly selected 22 older adults (robust group=11 and prefrail group=11) to participate in the “Smart Walk” program.

Figure 1. Recruitment of the Smart Walk program in the Aging Study of Pyeongchang Rural Area cohort. ASPRA: Aging Study of Pyeongchang Rural Area; CHS: Cardiovascular Health Study; CGA: Comprehensive Geriatric Assessment.



For random selection, we assigned random numbers to participants of the robust and prefrail groups except for frail participants using the CHS frailty criteria. The opportunity to participate in this study was given in the order of users with small random numbers. Random numbers were assigned using R software version 3.3.1 (R Foundation, Vienna, Austria). Those who could not walk 100 meters without an assistive device were excluded.

Initial data analysis compared the differences in physical activity and health outcomes according to frailty status. Since the total number did not exceed 30, we performed a nonparametric Mann-Whitney-Wilcoxon test and Fisher test for mean and ratio evaluation, respectively. To confirm the coaching effects before and after the "Smart Walk" program, we also performed a nonparametric Wilcoxon signed-rank test on the pairs. The last analysis analyzed health outcomes according to wearable device adherence. According to adherence, the participants of the program were divided into the following three groups: (1) coaching only; (2) short-term self; and (3) long-term self. The coaching only group comprised persons who used the band only during the 6-month coaching program. During the 6-month follow-up period, those who used the band for 3 months were included in the short-term self group and those who used the band for 6 months were included in the long-term self group. The Kruskal-Wallis test was performed to evaluate differences among the three groups, and all pairwise analyses were performed using the Mann-Whitney-Wilcoxon and Fisher tests for mean and ratio, respectively. All reported *P* values were two-sided, and *P* values <.05 were considered significant. Data analyses were conducted using R software, version 3.3.1 (R Foundation, Vienna, Austria).

Results

Overall Characteristics

From March 2017 to March 2018, 22 older adults participated in the Smart Walk program (robust group: n=11, male: 8/11; prefrail group: n=11, male: 6/11). The two groups were similar in most of the variables, except for age, frailty index, and Short Physical Performance Battery score associated with the CHS frailty criteria (Table 1). Although not statistically significant, the variables of living alone, including the risk of malnutrition and falls in the past year, were higher in the prefrail group. None of the participants had low income and received national medical aid.

Table 1. Overall characteristics of the Smart Walk program participants by the Cardiovascular Health Study frailty index.

Variables	Robust (n=11)	Prefrail (n=11)	Total (N=22)	*P* value
Age, mean (SD)	68.636 (1.85)	72.546 (4.298)	70.591 (3.801)	.03
Gender (male), n (%)	8 (73)	6 (55)	14 (64)	.66
Education level, mean (SD)	11.546 (4.22)	10.455 (5.3)	11 (4.711)	.66
Living alone, %	0 (0)	2 (18)	2 (9)	.48
Frailty index, mean (SD)	0.075 (0.04)	0.171 (0.085)	0.123 (0.083)	.005
Multimorbidity, %	0.099 (0.11)	0.141 (0.094)	0.12 (0.106)	.16
Cognition: MMSE[a] score, mean (SD)	28.818 (0.60)	27.909 (1.814)	28.364 (1.399)	.25
Mood: CES-D[b] score, mean (SD)	2.091 (2.73)	4.818 (4.262)	3.455 (3.764)	.09
Body mass index, mean (SD)	24.269 (2.74)	25.416 (3.478)	24.842 (3.115)	.95
At risk of malnutrition: MNA-SF[c], %	1 (9)	2 (18)	3 (14)	>.99
Short Physical Performance Battery score, mean (SD)	11.818 (0.40)	10.546 (0.934)	11.182 (0.958)	<.001
Dominant grip strength, mean (SD)	35.673 (9.01)	28.309 (7.99)	31.991 (9.125)	.13
Fall in the past year, %	0 (0)	3 (27)	3 (14)	.21

[a]MMSE: Mini-Mental State Examination.

[b]CES-D: Center for Epidemiological Studies Depression.

[c]MNA-SF: Mini Nutritional Assessment-Short Form.

Comparison of Health Improvement During the Coaching Program According to Frailty Status

We analyzed health improvement according to frailty status through coaching by managers of the public health center for the first 6 months of the Smart Walk program. No statistically significant difference was observed in the robust group, but the prefrail group showed significant improvement in usual gait speed (mean 0.73 [SD 0.11] vs mean 0.96 [SD 0.27], *P*=.02), IPAQ scores kcal (mean 2790.36 [SD 2224.62] vs mean 7589.72 [SD 4452.52], *P*=.01) [27], and European Quality of Life-5 Dimensions score (mean 0.84 [SD 0.07] vs mean 0.90 [SD 0.07], *P*=.02) [28]. In the total group, physical fitness, anthropometric measurements, and geriatric assessment categories, such as usual gait speed (mean 0.85 [SD 0.21] vs mean 1.02 [SD 0.27], *P*=.003), IPAQ score in kcal (mean 3013.63 [SD 2387.08] vs mean 7868.5 [SD 6250.56], *P*=.001), body mass index (BMI; mean 24.84 [SD 3.11] vs mean 24.52 [SD 3.36], *P*=.02), and total fat mass (mean 18.85 [SD 6.15] vs mean 17.82 [SD 6.41], *P*=.01) were significantly improved (Multimedia Appendix 2).

Comparison of Wearable Device Adherence During Coaching and Self-Management

A large difference in the proportion of continuous wearable device users was found between the coaching and self-management periods (average: 21.83 vs 8.16 persons, *P*<.001; see Figure 2). Figure 2 shows the average step count and number of continuous users in the robust and prefrail groups by month. The histogram plot indicates continuous users per month, and the line plot indicates the average total step count per month. The error bar of the line plot represents the SD of the average step count. In particular, the total step average was significantly different between the two periods; the total step average of the coaching period was about four times higher than that of the self-management period (5,584,295.83 vs 1,289,084.66, *P*<.001). The average monthly steps of the robust group and the prefrail group also differed markedly during the coaching period, while both were lower during the self-management period.

Of the 22 users, 5 (robust: 1 and prefrail: 4) people participated only in the coaching period (from March to September), and 11 (robust: 8 and prefrail: 3) people participated in the short self-management period (from March to December). The long self-management period (March 17 to March 18) included 6 (robust: 2 and prefrail: 4) people. An unusual finding was that prefrail persons used the wearable device longer than robust persons, but the average step number per month in the prefrail group was half that in the robust group.

Figure 3 compares the monthly average number of steps and SD between the three groups (coaching, short-term self, and long-term self). The long-term self group and the coaching group differed significantly (average: 106,309.15 vs 222,725.73 steps, *P*=.02). The short-term self group was similar to the coaching group at the beginning of the program, but it was more similar to the long-self group in the middle term.

Figure 2. Average step count and number of continuous users in the robust and prefrail groups by month.

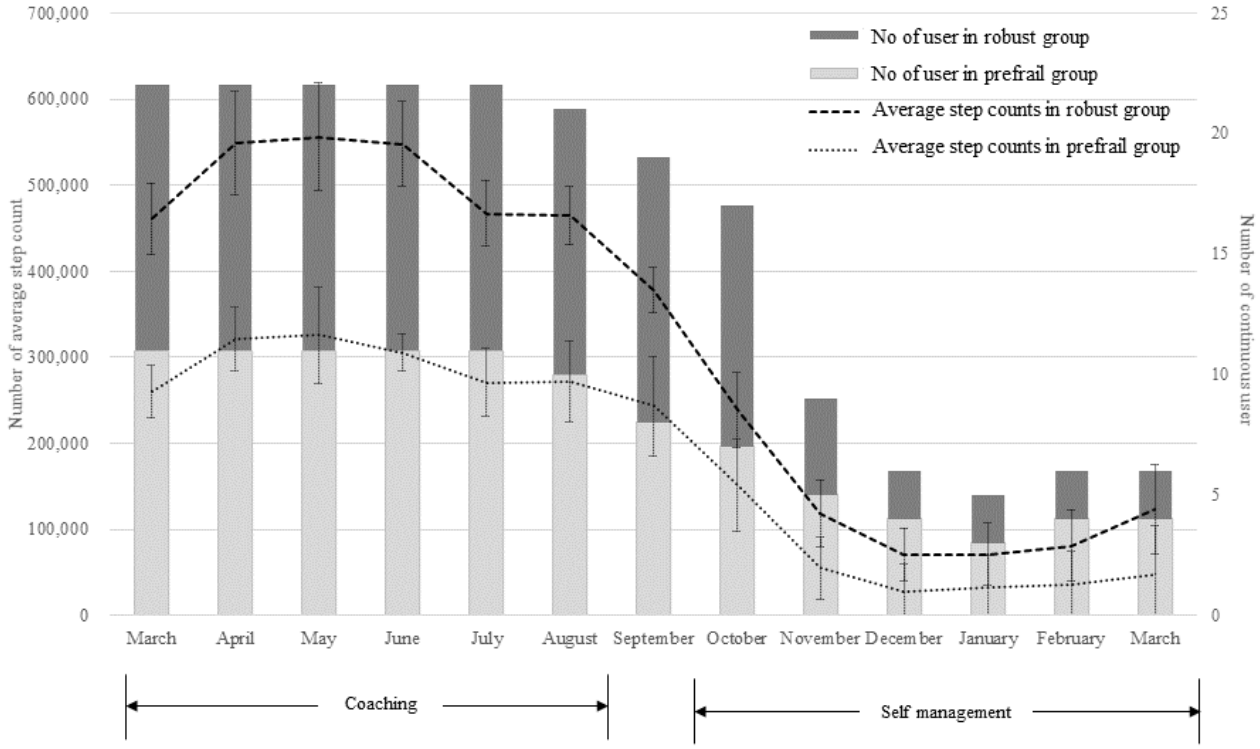


Figure 3. Monthly average step count by group of continuous use of the wearable device.

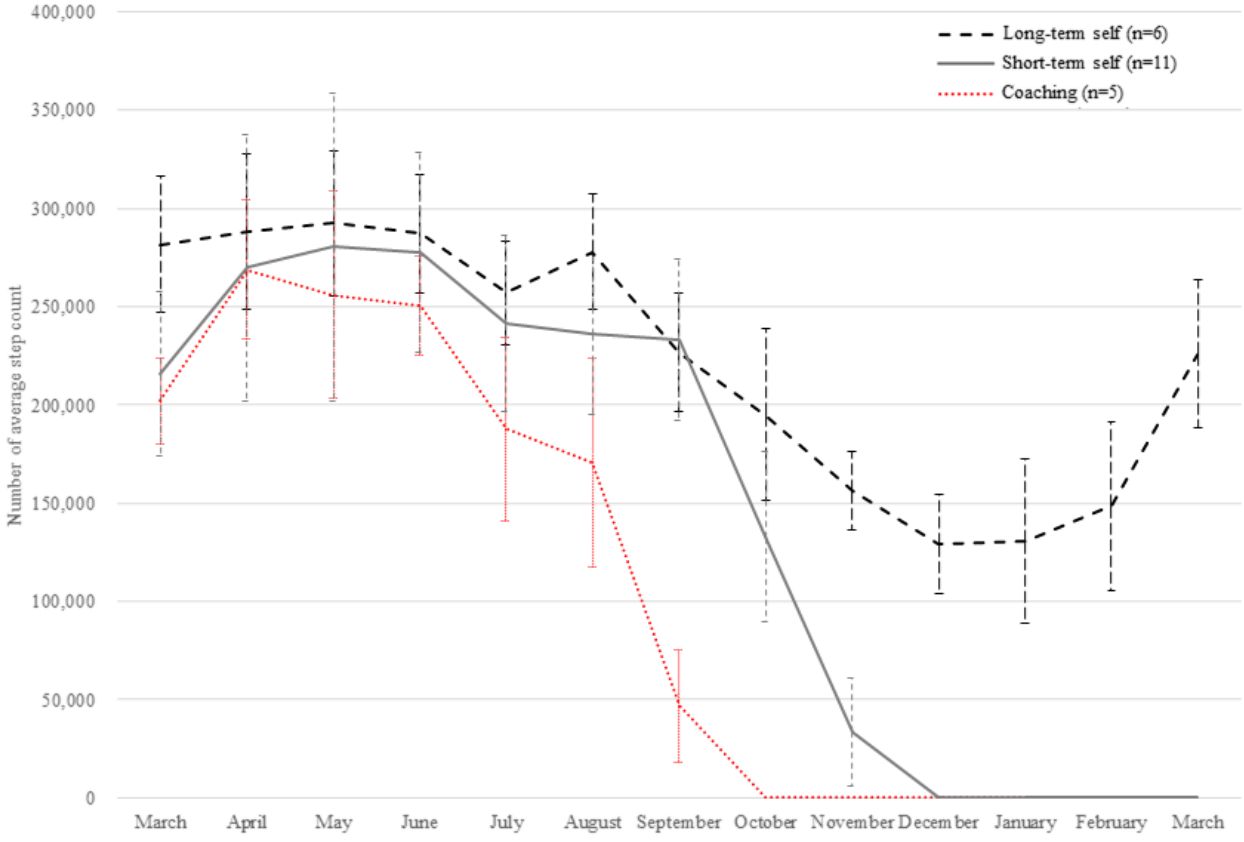


Table 2. Comparison of health improvement according to adherence type of the wearable device.

Characteristic	Coaching (n=4), mean (SD)	Short (n=10), mean (SD)	Long (n=6), mean (SD)	*P* value[a]
Number of falls	0.25 (0.5)	0.4 (1.265)	0 (0)	.49
Number of outpatient days	10.5 (10.472)	10 (6.864)	10.5 (4.722)	.10
Number of admission days	0 (0)	0 (0)	0 (0)	—[b]
Number of emergency room days	0.25 (0.5)	0 (0)	0 (0)	.14
Weight, kg	−1.325 (1.162)	−0.65 (1.824)	0.65 (1.317)	.07
Body mass index, kg/m^2	−0.416 (0.446)	−1.097 (3.373)	0.097 (0.513)	.30
Mini Nutritional Assessment score	0.25 (2.062)	−0.2 (1.687)	0.333 (0.516)	.66
K-frail score	0 (1.414)	0.3 (1.059)	0.167 (0.408)	.60

[a]The Krukal-Wallis test, a nonparametric test of ≥3 groups, was performed.

[b]Not applicable.

Health Improvement Comparison According to Wearable Device Adherence

During the research period, 2 participants were dropped. According to the CHS frailty criteria, both were in the robust group. The dropped participants refused geriatric assessment 4 months after the coaching period. However, they showed >90% adherence throughout the 6-month coaching period and were classified into the coaching group.

In the long-term self group, not a single fall was observed during the entire 13-month period, while an average of 0.25 and 0.4 falls occurred in the coaching and short-term self groups, respectively (Table 2). The average number of outpatient days was ≥10 in all three groups, with little difference between groups. No hospital admission occurred in all three groups. Emergency room visit was reported in only one case in the coaching group, wherein a food poisoning event occurred in a 72-year-old man at the end of the self-management period. Weight decreased by −1.325 (SD 1.824) and −0.65 (SD 1.317) in coaching and short-term self groups, respectively, and increased by 0.65 (SD 1.317) in the long-term self group. Similarly, BMI also increased in the long-term self group to 0.097 (SD 0.446) and decreased to −0.416 (SD 0.446) in the coaching group and to −1.097 (SD 3.373) in the short-term self group. MNA values decreased in the short-term self group but increased in the other two groups. No significant difference was found between all pairwise two-group comparisons (Multimedia Appendix 3).

Discussion

Principal Findings

This feasibility study showed that the wearable device-based intervention had a significant effect on physical performance (usual gait speed and IPAQ score) and anthropometric measurements (BMI and total fat mass) in rural older adults. In particular, usual gait speed (*P*=.02), IPAQ score (*P*=.01), and European Quality of Life-5 Dimensions-3L score (*P*=.02) were significantly improved in the prefrail group compared with the robust group. In addition, the compliance of wearable devices in this study confirmed a pattern of persistent use for ≥12 months by approximately 30% users (6/22 older adults) compared with the 10% 1-year continuous use rate in other studies [21]. The fact that 4 of the 6 users in the long-term self-management group belonged to the prefrail group supports the rationale that health concerns are among the factors that increase adherence to wearable devices. Also, we found that the long-term self group, which used the wearable device for the longest time, showed a 0.65 (SD 1.317) and 0.097 (SD 0.513) increase in body weight and BMI, respectively, compared with the other groups. Decreased body weight in older adults is not a positive indicator for health outcomes due to increasing loss of muscle mass [28]. In this respect, we improved the health outcomes of the older population through the long-term use of wearable devices. Our findings match those of previous studies reporting that individualized programs and self-management techniques could enhance physical activity adherence among older adults [29-31].

Our positive results may be explained by two factors. First, in order to maintain high adherence, we collaborated with public health centers to reduce the cost of managing and encouraging older adults and chose a wearable device with low management costs. Lewis et al reported that wearable-only interventions tend to produce only a modest effect on improving physical activity behavior [32]. Therefore, we decided to not only select a wearable device but also add human intervention via collaboration with the public health center. The burden of checking on the participants via smartphone every week and sending a message to those with poor exercise patterns did not exceed 5% of the weekly working time of a health center worker, even in a manpower-restricted rural public health center. When selecting the appropriate device for rural older adults, features like simplicity, ease of use, affordability, and long battery life (>3 weeks from a single charge) helped achieve a higher adherence.

Second, after comparing coaching and self-management components, older people demonstrated that the long-term use of wearable devices can be increased as needed. Many studies on the strengths of wearable devices are available, but several studies have still not maintained sustainable use and therefore have poor health outcomes [21,33]. However, the higher long-term use in the less healthy prefrail older adults in this study may lead to a new perspective on adherence to wearable

devices. In other words, if long-term management is performed for users who need health care, not the general public, the low adherence to wearable devices can be improved. In this study, the long-term self group using long-term wearable devices showed improved health outcomes in terms of nutrition (MNA score), physical activity (number of steps), and anthropometric measurements (weight and BMI).

Limitations

The main limitation of this study is that it did not have many participants. The first goal of this study was to examine the possibility of using wearable devices for older people. Hence, a feasibility study utilizing a small number of individuals with different frailty levels rather than a large number of users was required. As a result, we verified the high level of adherence to and health outcomes of wearable devices in this study. Based on the results of this study, further studies should plan to include many older adults with various frailty levels. The second limitation is that this study excluded frail older adults. The targeted daily step count of the wearable device was set by the researcher to increase by 1000 steps every 2 months and finally reach 7000 steps. Thus, frail older adults were not suitable for this study and were excluded, which could worsen the conditions of the frail adults. Further studies, however, are required to improve the health outcomes of these frail older adults.

Acknowledgments

We thank our study participants and their family members in Haanmi-ri and Gaesu-ri, Pyeongchang-gun, Gangwon-do, Republic of Korea. We are also grateful to the public health professionals at the Community Health Posts and County Hospital who provided administrative support.

Authors' Contributions

IYJ, HP, and SHC performed data acquisition. IYJ, HRK, HP, HWJ, YSL, and EL performed data preprocessing. IYJ, HRK, and YRP performed statistical analysis and interpretation and wrote the discussion. IYJ and YRP drafted the manuscript. IYJ, SHC, and EL supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The 34 components of frailty index.

[PDF File (Adobe PDF File), 22KB - mhealth_v6i11e11335_app1.pdf]

Multimedia Appendix 2

Comparison of health improvement according to the Cardiovascular Health Study frailty index with the coaching program.

[PDF File (Adobe PDF File), 41KB - mhealth_v6i11e11335_app2.pdf]

Multimedia Appendix 3

All pair wised comparison of health improvement according to the adherence type of the wearable device.

[PDF File (Adobe PDF File), 33KB - mhealth_v6i11e11335_app3.pdf]

References

1. Jung YJ, Yoon JL, Kim HS, Lee AY, Kim MY, Cho JJ. Atypical Clinical Presentation of Geriatric Syndrome in Elderly Patients With Pneumonia or Coronary Artery Disease. Ann Geriatr Med Res 2017 Dec 31;21(4):158-163. [doi: 10.4235/agmr.2017.21.4.158]
2. Limpawattana P, Phungoen P, Mitsungnern T, Laosuangkoon W, Tansangworn N. Atypical presentations of older adults at the emergency department and associated factors. Arch Gerontol Geriatr 2016;62:97-102. [doi: 10.1016/j.archger.2015.08.016] [Medline: 26323650]
3. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet 2013 Mar 02;381(9868):752-762 [FREE Full text] [doi: 10.1016/S0140-6736(12)62167-9] [Medline: 23395245]
4. Puts MTE, Toubasi S, Andrew MK, Ashe MC, Ploeg J, Atkinson E, et al. Interventions to prevent or reduce the level of frailty in community-dwelling older adults: a scoping review of the literature and international policies. Age Ageing 2017 May 01;46(3):383-392 [FREE Full text] [doi: 10.1093/ageing/afw247] [Medline: 28064173]

5. Shinkai S, Yoshida H, Taniguchi Y, Murayama H, Nishi M, Amano H, et al. Public health approach to preventing frailty in the community and its effect on healthy aging in Japan. Geriatr Gerontol Int 2016 Mar;16 Suppl 1:87-97. [doi: 10.1111/ggi.12726] [Medline: 27018287]
6. Collard RM, Boter H, Schoevers RA, Oude VRC. Prevalence of frailty in community-dwelling older persons: a systematic review. J Am Geriatr Soc 2012 Aug;60(8):1487-1492. [doi: 10.1111/j.1532-5415.2012.04054.x] [Medline: 22881367]
7. Hajduk AM, Chaudhry SI. Sedentary Behavior and Cardiovascular Risk in Older Adults: a Scoping Review. Curr Cardiovasc Risk Rep 2016 Jan;10(1) [FREE Full text] [doi: 10.1007/s12170-016-0485-6] [Medline: 27375828]
8. Thibaud M, Bloch F, Tournoux-Facon C, Brèque C, Rigaud AS, Dugué B, et al. Impact of physical activity and sedentary behaviour on fall risks in older people: a systematic review and meta-analysis of observational studies. Eur Rev Aging Phys Act 2011 Jun 30;9(1):5-15. [doi: 10.1007/s11556-011-0081-1]
9. Song J, Lindquist LA, Chang RW, Semanik PA, Ehrlich-Jones LS, Lee J, et al. Sedentary Behavior as a Risk Factor for Physical Frailty Independent of Moderate Activity: Results From the Osteoarthritis Initiative. Am J Public Health 2015 Jul;105(7):1439-1445. [doi: 10.2105/AJPH.2014.302540] [Medline: 25973826]
10. de Rezende LFM, Rey-López JP, Matsudo VKR, do Carmo Luiz O. Sedentary behavior and health outcomes among older adults: a systematic review. BMC Public Health 2014 Apr 09;14:333 [FREE Full text] [doi: 10.1186/1471-2458-14-333] [Medline: 24712381]
11. Jung H, Jang I, Lee YS, Lee CK, Cho E, Kang WY, et al. Prevalence of Frailty and Aging-Related Health Conditions in Older Koreans in Rural Communities: a Cross-Sectional Analysis of the Aging Study of Pyeongchang Rural Area. J Korean Med Sci 2016 Mar;31(3):345-352 [FREE Full text] [doi: 10.3346/jkms.2016.31.3.345] [Medline: 26952571]
12. Jang I, Jung H, Lee CK, Lee YS, Kim K, Kim KW, et al. Rural and Urban Disparities in Frailty and Aging-Related Health Conditions in Korea. J Am Geriatr Soc 2016 Apr;64(4):908-911 [FREE Full text] [doi: 10.1111/jgs.14074] [Medline: 27100601]
13. Kinsella K. Urban and rural dimensions of global population aging: an overview. J Rural Health 2001;17(4):314-322. [Medline: 12071553]
14. Kekade S, Hseieh C, Islam M, Atique S, Mohammed KA, Li YC, et al. The usefulness and actual use of wearable devices among the elderly population. Comput Methods Programs Biomed 2018 Jan;153:137-159. [doi: 10.1016/j.cmpb.2017.10.008] [Medline: 29157447]
15. Jang I, Jung H, Lee C, Lee Y, Lee E, Kim D. Comparison Between Korean Version of Physical Activity Scale for the Elderly and International Physical Activity Questionnaire-Short Form in Evaluation of Frailty Phenotype. Ann Geriatr Med Res 2017 Sep 30;21(3):101-107. [doi: 10.4235/agmr.2017.21.3.101]
16. Ehn M, Eriksson LC, Åkerberg N, Johansson A. Activity Monitors as Support for Older Persons' Physical Activity in Daily Life: Qualitative Study of the Users' Experiences. JMIR Mhealth Uhealth 2018 Feb 01;6(2):e34 [FREE Full text] [doi: 10.2196/mhealth.8345] [Medline: 29391342]
17. Li LC, Sayre EC, Xie H, Falck RS, Best JR, Liu-Ambrose T, et al. Efficacy of a Community-Based Technology-Enabled Physical Activity Counseling Program for People With Knee Osteoarthritis: Proof-of-Concept Study. J Med Internet Res 2018 Apr 30;20(4):e159 [FREE Full text] [doi: 10.2196/jmir.8514] [Medline: 29712630]
18. Kim BY, Lee J. Smart Devices for Older Adults Managing Chronic Disease: A Scoping Review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
19. Kim J, Kam HJ, Park YR, Yoo S, Oh JS, Kim Y, et al. Enchanted Life Space: Adding Value to Smart Health by Integrating Human Desires. Healthc Inform Res 2018 Jan;24(1):3-11 [FREE Full text] [doi: 10.4258/hir.2018.24.1.3] [Medline: 29503747]
20. Tedesco S, Barton J, O'Flynn B. A Review of Activity Trackers for Senior Citizens: Research Perspectives, Commercial Landscape and the Role of the Insurance Industry. Sensors (Basel) 2017 Jun 03;17(6) [FREE Full text] [doi: 10.3390/s17061277] [Medline: 28587188]
21. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
22. Jang I, Jung H, Lee CK, Yu SS, Lee YS, Lee E. Comparisons of predictive values of sarcopenia with different muscle mass indices in Korean rural older adults: a longitudinal analysis of the Aging Study of PyeongChang Rural Area. Clin Interv Aging 2018;13:91-99 [FREE Full text] [doi: 10.2147/CIA.S155619] [Medline: 29391783]
23. Jung HW, Jang IY, Lee CK, Yu SS, Hwang JK, Jeon C, et al. Usual gait speed is associated with frailty status, institutionalization, and mortality in community-dwelling rural older adults: a longitudinal analysis of the Aging Study of Pyeongchang Rural Area. Clin Interv Aging 2018;13:1079-1089 [FREE Full text] [doi: 10.2147/CIA.S166863] [Medline: 29922046]
24. Bergman H, Ferrucci L, Guralnik J, Hogan DB, Hummel S, Karunananthan S, et al. Frailty: an emerging research and clinical paradigm--issues and controversies. J Gerontol A Biol Sci Med Sci 2007 Jul;62(7):731-737 [FREE Full text] [Medline: 17634320]

25. Monfardini S, Ferrucci L, Fratino L, del Lungo I, Serraino D, Zagonel V. Validation of a multidimensional evaluation scale for use in elderly cancer patients. Cancer 1996 Jan 15;77(2):395-401. [doi: 10.1002/(SICI)1097-0142(19960115)77:2<395::AID-CNCR24>3.0.CO;2-W] [Medline: 8625250]
26. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [Medline: 11253156]
27. Rabin R, Gudex C, Selai C, Herdman M. From translation to version management: a history and review of methods for the cultural adaptation of the EuroQol five-dimensional questionnaire. Value Health 2014;17(1):70-76 [FREE Full text] [doi: 10.1016/j.jval.2013.10.006] [Medline: 24438719]
28. Aune D, Sen A, Prasad M, Norat T, Janszky I, Tonstad S, et al. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. BMJ 2016 Dec 04;353:i2156 [FREE Full text] [Medline: 27146380]
29. Cameron ID, Fairhall N, Langron C, Lockwood K, Monaghan N, Aggar C, et al. A multifactorial interdisciplinary intervention reduces frailty in older people: randomized trial. BMC Med 2013 Mar 11;11:65 [FREE Full text] [doi: 10.1186/1741-7015-11-65] [Medline: 23497404]
30. Tarazona-Santabalbina FJ, Gómez-Cabrera MC, Pérez-Ros P, Martínez-Arnau FM, Cabo H, Tsaparas K, et al. A Multicomponent Exercise Intervention that Reverses Frailty and Improves Cognition, Emotion, and Social Networking in the Community-Dwelling Frail Elderly: A Randomized Clinical Trial. J Am Med Dir Assoc 2016 Dec 01;17(5):426-433. [doi: 10.1016/j.jamda.2016.01.019] [Medline: 26947059]
31. Gill TM, Baker DI, Gottschalk M, Gahbauer EA, Charpentier PA, de Regt PT, et al. A prehabilitation program for physically frail community-living older persons. Arch Phys Med Rehabil 2003 Mar;84(3):394-404. [doi: 10.1053/apmr.2003.50020] [Medline: 12638108]
32. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: A systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
33. Monroe CM. Valuable steps ahead: promoting physical activity with wearables and incentives. Lancet Diabetes Endocrinol 2016 Dec;4(12):960-961. [doi: 10.1016/S2213-8587(16)30264-9] [Medline: 27717767]

Abbreviations

ASPRA: Aging Study of Pyeongchang Rural Area
BMI: body mass index
CHS: Cardiovascular Health Study
IPAQ: International Physical Activity Questionnaire
MNA-SF: Mini Nutritional Assessment-Short Form

Edited by G Eysenbach; submitted 20.06.18; peer-reviewed by R Sloan, JE Lee, JY Choi; comments to author 18.07.18; revised version received 01.08.18; accepted 01.08.18; published 21.11.18.

Please cite as:
Jang IY, Kim HR, Lee E, Jung HW, Park H, Cheon SH, Lee YS, Park YR
Impact of a Wearable Device-Based Walking Programs in Rural Older Adults on Physical Activity and Health Outcomes: Cohort Study
JMIR Mhealth Uhealth 2018;6(11):e11335
URL: http://mhealth.jmir.org/2018/11/e11335/
doi:10.2196/11335
PMID:30463838

Original Paper

Usability Study of Mainstream Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation

Jun Liang[1,2], MS; Deqiang Xian[3], MS; Xingyu Liu[4], MD; Jing Fu[5], MD; Xingting Zhang[6], MS, MD; Buzhou Tang[7], PhD; Jianbo Lei[8,9], MD, PhD

[1]IT Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China
[2]College of Information Engineering, China Jiliang University, Hangzhou, China
[3]Sichuan College of Traditional Chinese Medicine, Mianyang, China
[4]Affiliated Hospital of Stomatology, Southwest Medical University, Luzhou, China
[5]International School, Southwest Medical University, Luzhou, China
[6]Peking University Third Hospital, Beijing, China
[7]Shenzhen HIT Campus, Harbin Institute of Technology, Shenzhen, China
[8]Center for Medical Informatics, Peking University, Beijing, China
[9]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

Corresponding Author:
Jianbo Lei, MD, PhD
Center for Medical Informatics
Peking University
38 Xueyuan Road
Haidian District
Beijing, 100191
China
Phone: 86 8280 5901
Email: jblei@hsc.pku.edu.cn

Abstract

Background: Wearable devices have the potential to promote a healthy lifestyle because of their real-time data monitoring capabilities. However, device usability is a critical factor that determines whether they will be adopted on a large scale. Usability studies on wearable devices are still scarce.

Objective: This study aims to compare the functions and attributes of seven mainstream wearable devices and to evaluate their usability.

Methods: The wearable devices selected were the Apple Watch, Samsung Gear S, Fitbit Surge, Jawbone Up3, Mi Band, Huawei Honor B2, and Misfit Shine. A mixed method of feature comparison and a System Usability Scale (SUS) evaluation based on 388 participants was applied; the higher the SUS score, the better the usability of the product.

Results: For features, all devices had step counting, an activity timer, and distance recording functions. The Samsung Gear S had a unique sports track recording feature and the Huawei Honor B2 had a unique wireless earphone. The Apple Watch, Samsung Gear S, Jawbone Up3, and Fitbit Surge could measure heart rate. All the devices were able to monitor sleep, except the Apple Watch. For product characteristics, including attributes such as weight, battery life, price, and 22 functions such as step counting, activity time, activity type identification, sleep monitoring, and expandable new features, we found a very weak negative correlation between the SUS scores and price (r=−.10, P=.03) and devices that support expandable new features (r=−.11, P=.02), and a very weak positive correlation between the SUS scores and devices that support the activity type identification function (r=.11, P=.02). The Huawei Honor B2 received the highest score of mean 67.6 (SD 16.1); the lowest Apple Watch score was only 61.4 (SD 14.7). No significant difference was observed among brands. The SUS score had a moderate positive correlation with the user's experience (length of time the device was used) (r=.32, P<.001); participants in the medical and health care industries gave a significantly higher score (mean 61.1, SD 17.9 vs mean 68.7, SD 14.5, P=.03).

Conclusions: The functions of wearable devices tend to be homogeneous and usability is similar across various brands. Overall, Mi Band had the lowest price and the lightest weight. Misfit Shine had the longest battery life and most functions, and participants in the medical and health care industries had the best evaluation of wearable devices. The perceived usability of mainstream

wearable devices is unsatisfactory and customer loyalty is not high. A consumer's SUS rating for a wearable device is related to their personal situation instead of the device brand. Device manufacturers should put more effort into developing innovative functions and improving the usability of their products by integrating more cognitive behavior change techniques.

(JMIR Mhealth Uhealth 2018;6(11):e11066) doi:10.2196/11066

KEYWORDS

wearable devices; usability; System Usability Scale; function comparison; fitness

Introduction

Background

In recent years, wearable devices have gained tremendous momentum and have become part of people's daily lives. Wearable devices are portable mobile intelligent devices that can be worn directly or integrated into clothing [1]. The wearable devices in this study are wristbands and watches with step counting as their core function [2]. After combining IMS market research consulting with 2012 market research data from ABI Research, the US market research firm BI Intelligence believed that less than 40% of US consumers were familiar with wearable devices in 2012, whereas at least 54% of consumers developed a strong interest in them for health monitoring and fitness tracking in 2013. BI Intelligence expects the global wearable market to grow at a compound annual rate of 35% between 2015 and 2019 [3]. According to data from the International Data Corporation's Worldwide Quarterly Wearable Device Tracker, worldwide shipments of wearable devices grew by 1.2% in the first quarter of 2018 [4]. In addition, global smartwatch shipments in the second quarter of 2018 grew by 56% year on year [5]. Meanwhile, based on their original wearable market forecast from 2013 and interviews with industry experts, BI Intelligence concluded that, based on current trends, at least 33% of the US population will own wearable devices by the second quarter of 2017 [6]. The Tencent ISUX User Research Center released a report (smart wearable equipment market white paper) in 2015 [7] that contained a large-scale questionnaire survey with a total of 8083 domestic Chinese Tencent Instant Messenger user respondents. This stated that approximately 60% of internet users were familiar with wearable devices and that the number of Chinese internet users who owned a wearable device increased from 2.9% to 8.4% between November 2014 and May 2015. Among these users, wristband ownership was 4.6%, which was significantly higher than the 3.1% ownership of intelligent watches. Awareness of intelligent watches increased from 48% to 52%, and awareness of intelligent wristbands increased from 35% to 40% over the same period.

Wearable devices have significant potential for health monitoring and they are being widely investigated [8-13]. As a very popular tool, the use of wearable devices in medical and fitness applications is rapidly increasing. These devices can monitor vital signs such as motion [14], nutrition status [15], and heart rate [16], as well as a diabetic patient's blood sugar [17], cardiac disease status, apnea during sleep, and numerous other health parameters and statuses [18]. Luo and Fan [19] believed that the development of wearable devices may trigger innovation in medical treatment models. Kirk [20] believed that wearable medical technology may significantly improve the curative effect for patients, reduce physician time, and lower medical costs, representing future trends of precise personalized medical treatment. Appropriate technology, such as smart wearable devices, may create unprecedented opportunities for medical and health care service providers to obtain the accurate real-time data required to prevent, diagnose, and treat various chronic diseases in a more economical manner [21]. Recent studies have expressed concerns about the long-term use of wearable devices, emphasizing the need to combine behavioral change techniques such as target setting, feedback, and rewards with other evidence-based techniques in order to use them effectively to prompt behavior changes [22].

However, consumer acceptance of wearable devices has not been as positive as expected. Many important issues that affect the ultimate adoption of wearable devices by consumers have been investigated, including the reliability [23,24] and validity of the measures that they monitor [25-27]. In addition, studies have confirmed that the key to a product's success is the acceptance of wearable products by ordinary consumers and the comfort level of these products; product usability is an important aspect of this [28,29]. According to a definition by the International Organization for Standardization [30], usability refers to the effectiveness, efficiency, and user satisfaction rating of a product in a specific environment by a specific user for a specific purpose. It includes three aspects: effectiveness (ie, the accuracy and completeness of a goal that is achieved by a product); efficiency (ie, the effort required for a user to complete a task); and satisfaction (ie, the comfort and acceptability of a product). Usability tests and evaluations aim to make medical equipment easier, safer, and more effective and pleasant for users. A usability evaluation helps wearable devices to satisfy the requirements of the market and consumers [31].

In traditional usability studies, a thematic analysis based on heuristic evaluations, cognitive walkthroughs, think-alouds, and interviews is used; the results are then integrated together [32-34]. Although these methods are very popular, they have some limitations because results obtained in a laboratory environment can be difficult to interpret and translate into practice [35]. Questionnaires are another commonly employed method. They have been widely accepted by clinical medical informatics and consumer health informatics researchers and practitioners, and they are also widely applied in electronic health records, computerized physician order entry, and health applications [36-39]. At the same time, as a reliable, easy-to-operate, and low-cost method of usability evaluation, they have been recommended by the Agency for Healthcare Research and Quality [40]. Common questionnaire scales include the System Usability Scale (SUS) [41]; Ages and Stages

Questionnaire; Computer System Usability Questionnaire; Post-Study System Usability Questionnaire [42]; Use, Satisfaction, and Ease [43]; and Web Analysis and Measurement Inventory [44].

Research has continuously established that usability evaluation and the resulting improvements in wearable device usability are critical to the wide and rapid adoption of wearable devices [45]. Research on the usability of wearable devices for health monitoring is still in the early stages. Schenkenfelder et al [46] investigated the user experience during running for three types of wearable devices (sports wristband, intelligent watch, and intelligent glasses). By analyzing feedback from 18 participants they found that the wristband and watch received higher SUS scores than the glasses. However, no significant differences between the wristband and watch were observed. Kaewkannate and Kim [47] compared the functions and features of four wearable devices (Withings Pulse, Misfit Shine, Jawbone Up24, and Fitbit Flex) and recruited seven users for objective and subjective evaluations. The Withings Pulse received the highest user acceptance, followed by the Fitbit Flex, Jawbone Up24, and Misfit Shine.

Significance of This Study

This study has some advantages over other similar studies: it used a larger number of devices (seven different intelligent watches and wristbands), including mainstream international and domestic devices, and involved a significantly larger number of participants (388 persons, the largest sample size among similar studies). We compared the functions and features of various devices and evaluated the usability of wearable health tracking devices currently on the market using a mature and stable SUS scale. These results will help to assess the acceptance level of wearable devices and to identify the influencing factors. As there is a growing consensus that wearable devices should be developed for health applications, this study will help to examine the issues affecting their large-scale use and it will contribute to the research and development of better health applications.

Methods

Device Selection

The device selection was based on open market performance data from NPD (a leading global consumer and retail data supplier) and Canalys (a leading global technology market analyst with a distinct channel focus) [48-50]. Our inclusion criteria for the wearable activity trackers included (1) continuous monitoring of some kind of physical activity (eg, steps), and (2) provision of feedback via a separate mobile device or personal computer. We considered a device to be a wearable activity tracker if it contained an accelerometer and connected to a mobile platform. The device also had to be able to connect wirelessly with handheld or desktop computers, and be compatible with either Android 1.6+ or Apple's operating system iOS 6.0+. Therefore, current mainstream wearable devices with distinct market performance were selected as the research objects. These included the Apple Watch, Samsung Gear S, Fitbit Surge, Jawbone Up3, Mi Band, Huawei Honor Band B2, and Misfit Shine. Among these devices, the Apple Watch and Samsung Gear S are flagship products from leading mobile digital gadget vendors. The Fitbit Surge, Jawbone Up3, and Misfit Shine are well-known international sports wristband brands representative of different price ranges. Finally, the Mi Band and Huawei Honor Band B2 are leading brands of sports wristband in the Chinese market.

Questionnaire Selection

Usability refers to the effectiveness, efficiency, and user satisfaction rating of a product in a specific environment by a specific user for a specific purpose [51]. We chose to use the SUS, proposed by John Brooke in 1986, as the usability test tool. It is a simple scale based on a questionnaire and has been widely adopted in product usability evaluations. The SUS has the following advantages: (1) it is versatile and can be used to evaluate websites [52], software [53], mobile devices [54], and medical systems [55]; (2) it is a short questionnaire that is quick to answer; (3) a final score is provided with interpretation based on a well-established reference standard [56]; (4) it is free; (5) it is suitable even when applied to small samples (N<14); and (6) it has excellent reliability (0.85). Overall, the SUS is a quick and simple method for usability evaluation.

The SUS contains 10 questions based on the Likert five-point scale; questions 1, 3, 5, 7, and 9 are positive and questions 2, 4, 6, 8, and 10 are negative. The 10 questions are closely related and are employed for the comprehensive evaluation of a product. A higher SUS score indicates better product usability.

Details about the questionnaire's design (the 10 questions) and scoring method are given in Multimedia Appendix 1.

Subject Recruitment and Test Design

Recruitment was based on a convenience sampling method. It leveraged WeChat, the social media platform with the largest user population in China. Over a period from July 1, 2016 to October 31, 2016, the research team used their WeChat accounts to send the questionnaire link to 2180 friends (between 8 and 314 each) and various WeChat groups (approximately 20 groups) and invited their friends to forward the link. A total of 388 volunteers who had experience with the seven wearable devices were recruited.

Once the volunteers agreed to take part in the study, consent was obtained and they were invited to complete an evaluation questionnaire about their experiences with these products.

This research passed an audit of the Peking University Biomedical Ethics Committee #IRB00001052 - 16008 - Exampt, and consent was obtained from all participants.

Data Collection

The questionnaires were distributed and the results collected via "sojump.com," the largest free questionnaire survey platform in the world, which satisfied all the requirements for the questionnaires in this research [57]. The questionnaires were sent and completed via personal computers and mobile terminals. The final questionnaire "Comprehensive Evaluation Questionnaire for Intelligent Wearable Devices" is accessible online [58].

Statistical Analysis

For a descriptive statistical analysis, basic information about the respondents was collected, including gender, age, education, profession, and monthly income. Then, the product attributes (eg, price, weight, and battery life) and specific functions (whether 22 functions, such as step counting, activity time, activity type identification, sleep monitoring, and expandable new features, were supported) were summarized for the seven wearable devices. Finally, the SUS scores for the seven wearable devices used by the respondents were calculated.

For the inferential statistical analysis, the relationships between the SUS scores and product attributes (product price, weight, and battery life) were explored using Pearson product moment correlation (since price data are continuous and have a normal distribution) and Spearman rank correlation coefficient (since the weight and battery life do not have a normal distribution). The correlations between the SUS score and the functions of the 22 products (eg, step counting, activity time, activity type identification, sleep monitoring, and expandable new features) were analyzed using point-biserial correlation. The strength of the correlation was assessed based on Cohen criteria: correlations less than .30 are considered small, correlations between .30 and .50 are considered medium, and correlations greater than .50 are considered strong [59,60]. Finally, to analyze the participants' attitudes to the usability of different products, the analysis of variance (ANOVA) method was used with the brand and user demographic information. These analyses were all completed using IBM SPSS version 20 software (IBM Corp, Armonk, NY, USA).

Results

Comparison of Functions and Attributes of Devices

Based on the product specifications and official websites, the features and general attributes of the seven devices were summarized and compared (see Tables 1 and 2). Table 1 summarizes the functions of each device. All seven devices had very powerful functions in three major categories (activity, health, and miscellaneous) and 20 features. The Fitbit Surge had the most features, followed by the Apple Watch, Huawei Honor B2, Samsung Gear S, Jawbone UP3, Mi Band, and Misfit Shine. The Samsung Gear S had the most sports features, the Jawbone UP3 had the most health features, and the Apple Watch had the most other additional features. All the devices supported three basic features: step counting, activity timer, and distance record. The Fitbit Surge was the only device to record the number of floors climbed, the Samsung Gear S had a unique sports track recording feature, and the Huawei Honor B2 had a unique wireless earphone. The Apple Watch, Samsung Watch, Jawbone Up3, and Fitbit Surge could measure heart rate. All the devices could monitor sleep, with the exception of the Apple Watch. Table 2 lists some of the major attributes of each device. Among the seven devices, the Misfit Shine had the longest battery life of 3 months, followed by the Mi Band, Fitbit Surge, Jawbone Up3, Huawei Honor wristband, Samsung Gear S, and Apple Watch, which had a battery life of only 18 hours. The Fitbit Surge, which weighed 354 grams, was the heaviest device, followed by the Jawbone UP3, Misfit Shine, Apple Watch, Huawei Honor Band B2, Samsung Gear S, and Mi Band, which weighed only 5 grams.

System Usability Scale Questionnaire Respondents

A total of 388 volunteers completed the questionnaire for this research. The volunteers included 83 Apple Watch users, 36 Samsung Gear S users, 37 Fitbit Surge users, 32 Jawbone Up3 users, 122 Mi Band users, 47 Huawei Honor Band B2 users, and 31 Misfit Shine users. Of the volunteers, 257 (66.2%) were male and 131 (33.8%) were female. The demographic information of the participants is shown in Table 3.

System Usability Scale Scores for Each Device and Each Question

Based on the SUS calculation formula, the SUS scores for each brand are shown in Table 4. The mean SUS score for each question was calculated using SPSS, as shown in Table 5 (for full results, see Multimedia Appendix 2). For further visualization, please see corresponding boxplot diagrams in Multimedia Appendixes 3 and 4.

Correlation Between System Usability Scale Score and Product Characteristics

We first explored the relationship between the SUS score and product characteristics (see Table 2). Pearson product moment correlation was used to analyze the relationship between the SUS value and the price, and Spearman rank correlation coefficient was used to analyze the relationship between the SUS values and the weight and battery life. We found a very weak correlation between the SUS score and the price of the product (r=–.10, P=.03), and no significant correlation between the SUS scores and the weight (r=–.04, P=.39) or battery life (r=.09, P=.10).

We then explored the relationship between the SUS scores and the functions supported by the products (see Table 1) using point-biserial correlation. We found that the SUS score had a very weak positive correlation (r=.11, P=.02) with the devices that supported activity type identification, and a very weak negative correlation (r=–.11, P=.02) with devices that supported expandable new features, as shown in Table 6.

Correlation Between System Usability Scale Score, Device Brand, and Participants' Demographic Information

To investigate a participant's perception of the usability of each of the seven brands of wearable devices, seven ANOVA tests were performed. These were used to determine the correlation between the SUS scores, device brand, and participants' demographic information (length of time the device was used, gender, age, education, profession, and monthly income) (see Table 7).

Table 1. Function summary of the seven wearable devices.

Items	Apple Watch	Samsung Gear S	Jawbone Up3	Fitbit Surge	Misfit Shine	Huawei Honor B2	Mi Band
Activity							
Step counting	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Activity time	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Activity type identification	—[a]	—	Yes	Yes	Yes	Yes	Yes
Floor climbing	—	—	—	Yes	—	—	—
Energy consumption	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Running track	—	Yes	—	—	—	—	—
Activity goal	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Professional fitness record	Yes	Yes	—	Yes	—	—	Yes
Health							
Heart rate	Yes	Yes	Yes	Yes	—	—	—
Sleep monitoring	—	Yes	Yes	Yes	Yes	Yes	Yes
Diet record	—	—	Yes	Yes	Yes	—	—
Inactivity notification	Yes	—	Yes	—	Yes	Yes	Yes
Miscellaneous							
Incoming call alert	Yes	Yes	—	Yes	Yes	Yes	Yes
Vibrating alarm clock	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Text message display	Yes	Yes	—	Yes	Yes	—	—
Time display	Yes	Yes	—	Yes	Yes	—	—
Remote control photograph	Yes	—	—	—	Yes	—	—
Sports achievement sharing	Yes	—	Yes	Yes	Yes	Yes	Yes
Bluetooth synchronization	Yes	Yes	Yes	Yes	—	Yes	Yes
Expandable new features	Yes	Yes	—	—	—	—	—
Total	15	14	12	16	14	11	12

[a]Device does not support this feature.

Table 2. Major attributes of the seven wearable devices.

Device	Country	Date to market	Price (US$)	Weight (g)	Battery life (days)
Apple Watch	USA	September 10, 2014	$402	40.00	0.75
Samsung Gear S	Korea	September 25, 2013	$404	122.50	2.00
Fitbit Surge	USA	October 27, 2014	$279	354.00	7.00
Jawbone Up3	USA	November 5, 2014	$183	29.00	7.00
Mi Band	China	July 22, 2014	$12	5.00	30.00
Huawei Honor B2	China	March 1, 2015	$155	40.00	5.00
Misfit Shine	USA	September 14, 2013	$77	9.40	180.00

Table 3. Demographic information of the respondents (N=388).

Items	n (%)
Device to be evaluated	
Apple Watch	83 (21.4)
Samsung Gear S	36 (9.3)
Fitbit Surge	37 (9.5)
Jawbone Up3	32 (8.2)
Mi Band	122 (31.4)
Huawei Honor B2	47 (12.1)
Misfit Shine	31 (8.0)
Length of time the device was used	
<1 week	49 (12.6)
1 week-1 month	96 (24.7)
1-3 months	66 (17.0)
3-6 months	53 (13.7)
6 months-1 year	69 (17.8)
1-2 year	49 (12.6)
>2 years	6 (1.5)
Gender	
Male	257 (66.2)
Female	131 (33.8)
Age	
Under 18	10 (2.6)
18~25	107 (27.6)
26~30	88 (22.7)
31~40	113 (29.1)
41~50	56 (14.4)
51~60	13 (3.4)
>60	1 (0.3)
Education	
Under primary school	3 (0.8)
Primary school-high school	19 (4.9)
Junior college	46 (11.9)
Bachelor	185 (47.7)
Master of Science or Master of Arts	98 (25.3)
PhD	37 (9.5)
Profession	
Internet	114 (29.4)
Finance	28 (7.2)
Health care	92 (23.7)
Manufacturing	19 (4.9)
Fast consumer goods	4 (1.0)
Education	29 (7.5)
Law	6 (1.5)

Items	n (%)
Trade	6 (1.5)
Service	20 (5.2)
Advertising and media	6 (1.5)
Agriculture, forestry, fishery	2 (0.5)
Miscellaneous	62 (16.0)
Monthly income (US$)	
<145	27 (7.0)
145~290	34 (8.8)
291~580	45 (11.6)
581~1160	93 (24.0)
>1160	189 (48.7)

Table 4. Total System Usability Scale (SUS) score of each device (N=388).

Device	n	Mean (SD)	High score	Low score	% of total sum
Apple Watch	83	61.36 (14.69)	95.00	27.50	20.40
Samsung Gear S	36	62.08 (19.29)	100.00	5.00	9.00
Fitbit Surge	37	63.85 (21.97)	100.00	0.00	9.50
Jawbone Up3	32	65.94 (17.53)	100.00	27.50	8.50
Mi Band	122	65.12 (14.73)	100.00	35.50	31.80
Huawei Honor B2	47	67.61 (16.12)	100.00	22.50	12.70
Misfit Shine	31	65.97 (20.21)	100.00	10.00	8.20

Table 5. Mean System Usability Scale (SUS) score for each question and device.

Device	SUS score, mean (SD)									
	1	2	3	4	5	6	7	8	9	10
Apple Watch (n=83)	2.86 (1.00)	1.71 (1.16)	2.90 (0.89)	2.30 (1.26)	2.60 (0.95)	2.11 (1.14)	2.84 (0.89)	2.20 (1.23)	2.78 (0.96)	2.23 (1.26)
Samsung Gear S (n=36)	2.72 (1.26)	1.61 (1.29)	3.08 (1.02)	2.31 (1.37)	2.78 (0.96)	1.92 (1.27)	2.86 (1.02)	2.14 (1.44)	3.06 (0.95)	2.36 (1.31)
Fitbit Surge (n=37)	2.64 (1.38)	2.00 (1.41)	3.00 (1.07)	2.56 (1.52)	2.53 (1.06)	2.06 (1.26)	2.81 (1.17)	2.44 (1.38)	2.92 (1.08)	2.67 (1.22)
Jawbone Up3 (n=32)	2.90 (1.14)	1.90 (1.40)	2.97 (0.91)	2.58 (1.41)	2.71 (1.01)	2.16 (1.21)	2.94 (1.06)	2.55 (1.29)	3.10 (0.91)	2.42 (1.52)
Mi Band (n=122)	2.65 (1.02)	1.87 (1.30)	2.94 (1.04)	2.80 (1.33)	2.43 (1.01)	2.27 (1.13)	2.93 (1.04)	2.82 (1.11)	2.50 (1.13)	2.84 (1.19)
Huawei Honor B2 (n=47)	2.70 (1.07)	1.72 (1.19)	2.96 (1.07)	2.85 (1.28)	2.78 (0.94)	2.46 (1.21)	2.83 (0.97)	2.76 (1.21)	3.13 (0.88)	2.80 (1.13)
Misfit Shine (n=31)	2.87 (1.01)	2.33 (1.45)	3.10 (1.09)	2.87 (1.36)	2.40 (1.13)	2.27 (1.28)	2.83 (1.21)	2.60 (1.25)	2.77 (1.14)	2.50 (1.36)
Total (N=388)	2.74 (1.09)	1.84 (1.09)	2.97 (1.00)	2.61 (1.35)	2.57 (1.00)	2.20 (1.19)	2.87 (1.02)	2.54 (1.25)	2.80 (1.05)	2.58 (1.27)

Table 6. Point-biserial correlation analysis between the System Usability Scale (SUS) scores and product functions (N=388).

Product function	Point-biserial correlation	
	r	*P* value
Activity type identification	.11	.02
Expandable new features	−.11	.02
Floor climbing	.03	.58
Running track	−.04	.40
Professional fitness record	−.09	.08
Heart rate	−.09	.07
Sleep monitoring	.09	.07
Diet record	.04	.45
Inactivity notification	.01	.81
Incoming call alert	.01	.85
Text message display	−.06	.22
Time display	−.06	.22
Remote control photograph	−.05	.35
Sports achievement sharing	.04	.40
Bluetooth synchronization	.08	.13

Table 7. ANOVA tests of the System Usability Scale (SUS) scores of the seven devices and of the volunteer's demographic information (profession, length of time the device was used).

Measures	Sum of squares	Degrees of freedom	Mean square	*F* (df1,df2)	*P* value
SUS scores of devices					
Between groups	1671.52	6	278.59	0.98 (6,381)	.44
Within groups	108,057.91	381	283.62	—[a]	—
Total	109,729.43	387	—	—	—
Profession					
Between groups	5618.23	11	510.75	1.84 (11,376)	.04
Within groups	104,111.19	376	276.89	—	—
Total	109,729.43	387	—	—	—
Length of time device used					
Between groups	12,581.19	6	2096.86	8.22 (6,381)	.001
Within groups	97,148.24	381	254.98	—	—
Total	109,729.43	387	—	—	—

[a]Not applicable.

First, we found that there was no statistically significant difference between the SUS scores of the different brands (see Table 7), but there were significant differences in the SUS scores according to occupation and user experience (length of time the device was used). No significant difference was found for other features such as gender, age, education, and monthly income, as detailed in Multimedia Appendix 5. Second, we further used independent sample *t* tests to compare the mean SUS scores of respondents in the health care and internet industries; this indicated that the former scores were significantly higher than the latter. The SUS scores for the health care industry (n=92) were mean 68.7 (SD 14.5, range 35-100); for the internet industry (n=114) they were mean 61.1 (SD 17.9, range 5-100; t_{204}=−3.24, *P*=.001). Third, we used Pearson product moment correlation method to explore the relationship between the SUS scores and the length of time the device was used. The SUS score had a moderate positive correlation (*r*=.32, *P*<.001) with user experience. In other words, the SUS score was related to a volunteer's user experience (how long they had used the product). Other factors had no significant influence on the SUS score.

Discussion

Low System Usability Scale Scores for Seven Wearable Devices

The mean SUS score for the seven devices was 64.3 (SD 16.8). Although the Huawei Honor Band B2 received the highest score (mean 67.6, SD 16.1), based on the ANOVA analysis, there was no statistically significant difference in the mean SUS score of the seven products. Based on the SUS score, we believe that these products have little difference in terms of the ease of use. Aaron et al from AT&T Labs [61] added one question to the end of the SUS questionnaire, "Overall, I would rate the user friendliness of this product as…" This allowed users to rate the user interface using adjectives such as poor, OK, and good. The purpose of this was to associate the SUS with these adjectives. According to the data in this research, the user interfaces rated as "good" by users received a mean SUS score of 71.4. Therefore, the usability of all seven wearable devices was "good" or "OK." The Huawei Honor Band B2, which received the highest SUS score of 67.6 (SD 16.1), was also in the "good" range, whereas the Apple Watch received the lowest score of 61.3 (SD 14.6). Although the seven brands received different SUS scores, the ANOVA test of the SUS scores versus device brands revealed no significant correlation between them. We believe that the usability of a wearable health tracking device is moderate with no significant differences between brands. These products are immature and additional breakthroughs in the core technology are required. In this sector, there is no leading brand with an absolute competitive edge, and there is no significant difference between domestic and foreign brands.

Correlation Between System Usability Scale Score and Demographic Characteristics

The SPSS analysis indicated that the SUS scores for wearable devices are related to demographic characteristics. Factors such as user experience and profession may affect the scores. The results also revealed that user experience (the time length the device was used) could be an important factor affecting the SUS score. We found that there was a moderate positive correlation between the SUS scores and user experience (r=.32, P<.001); this matches the conclusions of MacDorman et al [62] and Kortum et al [63]. The participants in medical and health care industries gave a significantly higher score (mean 68.7, SD 14.5) than the participants in other industries, particularly those in the internet industry (mean 61.1, SD 17.9), with the former giving scores that approached the "good" level (71.4) in the acceptance diagram [61]. We conducted independent sample t tests to compare the mean SUS scores of participants in the health care and internet industries, which indicated that the former score was significantly better than the latter (P<.001). This suggests that in our sample set medical and health care employees evaluated the devices more highly and that their acceptance level of wearable device usability was higher than internet employees, indirectly showing that the demand for wearable devices is more urgent in health-related industries, a potential reason or phenomenon worthy of further research and analysis in the future. This may be due to the serious shortage of medical resources in China, especially in remote areas. The gap between supply and demand provides opportunities for mobile medical-based wearable equipment, whereas the rapid development of mobile internet and big data technologies provide the support required for the development of mobile medical treatments. In the future, patients with chronic diseases such as coronary heart disease, hypertension, and diabetes will not only receive drug treatment, but integrated disease management programs including remote monitoring, tele-treatment plan adjustments, and lifestyle management through wearable technology. Previous SUS research [61] indicated that no control was needed for gender or that gender was not a key influencing factor for consumer-grade products. Our result has validated this conclusion.

Correlation Between System Usability Scale Score and Product Characteristics

Some characteristics of wearable devices are related to the SUS scores. Our study found that the intelligent recognition of activity type had a very weak positive correlation with the SUS score, and that expandable new functions and price have a very weak negative correlation with the SUS score. We believe that these findings, to some extent, show a "user portrait" of Chinese wearable device consumers who are likely to prefer devices that are functional and easy to operate (a weak positive correlation between the SUS score and the intelligent recognition function of the device). The price was a favorable factor, although it was not strong (the SUS score had a weak negative correlation with the price of device). We found that this is in line with the current market strategy of the Mi Band bracelet which applies the lowest price (only US $12) and the most core functions (such as steps, exercise time, sleep time, and resting time) to maximize the market share. Using this strategy, Mi Technologies Inc sold more than 3.6 million Mi Band bracelets in China in the first quarter of 2017, surpassing the giant manufacturers Fitbit, Apple, and Samsung to become the world's largest wearable device maker [64]. In addition, we found that Q3 and Q7 (see Multimedia Appendix 1) had the highest scores: 2.97 and 2.87, respectively (see Table 5 and Multimedia Appendix 2). These scores were related to the ease of use (accessibility) of the product and the fact that people are generally more inclined to use products that do not require experience or focus and do not produce, or produce less, cognitive stress [65]. We think that this may be one reason for the high short-term usage rate of wearable devices. Although the use of wearable devices can enhance the ability to monitor the user's behavior, if they perform similar activities and movements every day then, over time, it becomes easier to stop using the device as the user's perception of the link between the results of the behavior intervention and the device monitoring is lost [66]. In order to prolong the use-life cycle and long-term wear rate of wearable devices, we need to integrate more cognitive behavior change strategies and functions to promote and consolidate the changes in users' habits. This includes the identification of behavioral disorders and adjustment of cognitive attitudes [67]. At the same time, we believe that the low SUS scores of the current fitness wearable devices are due to the users' dissatisfaction with the function and effect of the equipment, or their lack of cognitive motivation to change their behavior for a healthy lifestyle [22]. Therefore, we suggest that developers focus on the core

requirements of the user, combined with social support, environmental support factors, customizable self-monitoring schemes, and personalized feedback, which should be integrated into the wearable device's regular monitoring functions. In this way, we can help to improve the user's overall usability evaluation of such products, prolong their use time, increase product adhesion, and cultivate user loyalty.

Implications of Feature Analysis: Function Homogeneity, Awaiting Killer Function

Wearable devices have great potential in personal health management and clinical care. In future, the cost of wearable devices will decrease, such as the Mi Band which is sold for only US $12. Wearable devices can record sports data, help set exercise goals, and avoid excessive movement, making it easy to obtain sports performance data and maintaining and promoting the passion for sport. In addition, wearable devices can play a role in clinical practice. An obvious characteristic of wearable devices is the continuous monitoring of health data, which can provide clues for disease diagnosis, identify disease symptoms in a timely manner, and help doctors to make more comprehensive and accurate judgments.

Current mainstream wearable devices, including the seven brands in this study, are powerful; they not only track and record health data, such as sports data, sleep, and heart rate, but also provide numerous additional functions (including call notification, social media sharing, and Bluetooth). However, the function and feature analyses show that their functions are homogeneous and most concentrated on three functions: activity, reminders, and health monitoring. From this point of view, the homogenization of the product functions is very serious, and the functions of each product are lacking. There are no killer functions that solve major health management or clinical questions. Accordingly, the perceived usability of mainstream wearable devices is unsatisfactory and customer loyalty is not high. In fact, some recent studies have shown that many users use activity trackers for only a short time [68].

Characteristics and Limitation of This Research

In this research, the features of different devices were compared and their usability evaluated using the SUS. Compared to similar studies, we investigated the most extensive range of devices and used the largest group of participants. However, this research has its limitations. First, we adopted the snowball convenient randomization method; the participants were recruited based on the research teams' circle of WeChat friends. Although the final respondents recruited were distributed across the country, this sample is not sufficient to represent consumers of wearable devices nationwide. Nonetheless, since our sample size was large and the responses were autonomous, we believe that our group still has a certain degree of representativeness. This is reinforced by the fact that the proportion of each brand in our study (Mi 31.4%, Huawei 12.1%, and Samsung 9.3%) is similar to the market share of intelligent wristbands in the 2014 report by Tencent ISUX [69] (Mi 34.3%, Huawei 12.5%, and Samsung 10.5%). Second, the survey was conducted using online questionnaires. Although this method is more convenient, faster, and more efficient than a conventional paper questionnaire in terms of distribution, collection, and analysis, many factors are uncontrollable and some factors, such as behavioral cognitive factors [70], social network factors [71], and environmental support factors [72], may affect the results.

Conclusions

The homogenization of wearable device functions is obvious, the usability of popular wearable devices on the market is unsatisfactory, and these devices have not been completely accepted by consumers. Usability is similar among the different brands with no absolute leader. No significant differences between domestic and international brands were observed. A consumer's SUS score for wearable devices is related to their personal situation rather than the device brand. Device manufacturers should put more effort into developing innovative functions and improving the usability of their products.

Acknowledgments

We thank all volunteers who participated in this research. This study was sponsored by the National Natural Science Foundation of China (grants #81471756, #81771937, and #81871455) and the Medical and Health Planning Project of Zhejiang Province of China (grant #2017KY386).

Authors' Contributions

XZ and DX were responsible for the questionnaire preparation and data collection; XL was responsible for literature review; JL was responsible for the data analysis and manuscript draft; JF and BT supervised the data analysis and interpretation; and JL was responsible for study design and revision of the manuscript. All authors reviewed this final manuscript and expressed agreement regarding its content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

System Usability Scale (SUS) questionnaire design and scoring method.

[PDF File (Adobe PDF File), 33KB - mhealth_v6i11e11066_app1.pdf]

Multimedia Appendix 2

System Usability Scale (SUS) score of each question for each device.

[PDF File (Adobe PDF File), 30KB - mhealth_v6i11e11066_app2.pdf]

Multimedia Appendix 3

System Usability Scale (SUS) scores for all devices.

[PDF File (Adobe PDF File), 511KB - mhealth_v6i11e11066_app3.pdf]

Multimedia Appendix 4

System Usability Scale (SUS) scores for all questions.

[PDF File (Adobe PDF File), 436KB - mhealth_v6i11e11066_app4.pdf]

Multimedia Appendix 5

ANOVA tests of the System Usability Scale (SUS) scores and the volunteer's demographic information.

[PDF File (Adobe PDF File), 32KB - mhealth_v6i11e11066_app5.pdf]

References

1. Sungmee P, Jayaraman S. A transdisciplinary approach to wearables, big data and quality of life. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4155-4158. [doi: 10.1109/EMBC.2014.6944539] [Medline: 25570907]
2. Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, et al. Categorizing health outcomes and efficacy of mHealth apps for persons with cognitive impairment: a systematic review. J Med Internet Res 2017 Aug 30;19(8):e301. [doi: 10.2196/jmir.7814] [Medline: 28855146]
3. Danova T. Business Insider. 2015 Jan 07. Apple Watch will control nearly half of the entire smartwatch market by 2017 URL: http://www.businessinsider.com/wearable-computing-market-statistics-2015-1 [accessed 2018-03-05] [WebCite Cache ID 6zSSbEfOI]
4. Shirer M, Liamas R, Ubrani J. IDC. 2018 Jun 04. Wearable device shipments slow in Q1 2018 as consumers shift from basic wearables to smarter devices, according to IDC URL: http://www.idc.com/getdoc.jsp?containerId=prUS43900918 [accessed 2018-08-21] [WebCite Cache ID 73BeXyGOr]
5. Waltzer S. Strategy Analytics. 2018 Aug 02. Global smartwatch vendor market share by region: Q2 2018 URL: https://tinyurl.com/y72zuydk [accessed 2018-08-21] [WebCite Cache ID 73BgvvmQS]
6. Danova T. Yahoo! Finance. 2014 Oct 29. The Wearable Computing Market Report: growth trends, consumer attitudes, and why smartwatches will dominate URL: https://tinyurl.com/y86hey85 [accessed 2018-03-05] [WebCite Cache ID 6zST0P4Ln]
7. 36kr. 2015 Jul 03. Tencent ISUX released domestic wearable devices report, the user retention rate is low, professional is the main demand URL: http://36kr.com/p/5035222.html [accessed 2018-03-08] [WebCite Cache ID 6zSVCYZRl]
8. Brakenridge CL, Healy GN, Winkler EA, Fjeldsoe BS. Usage, acceptability, and effectiveness of an activity tracker in a randomized trial of a workplace sitting intervention: mixed-methods evaluation. Interact J Med Res 2018 Mar 02;7(1):e5 [FREE Full text] [doi: 10.2196/ijmr.9001] [Medline: 29500158]
9. Koydemir HC, Ozcan A. Wearable and implantable sensors for biomedical applications. Annual Rev Anal Chem 2018 Jun 12;11(1):127-146. [doi: 10.1146/annurev-anchem-061417-125956]
10. Niwano S, Sekiguchi Y, Ishii Y, Iwasaki Y, Kato R, Okamura H, et al. Clinical usefulness of wearable cardioverter defibrillator (WCD) and current understanding of its clinical indication in Japan. Circ J 2018;82(6):1481-1486. [doi: 10.1253/circj.CJ-17-1336]
11. Shen J, Naeim A. Telehealth in older adults with cancer in the United States: The emerging use of wearable sensors. J Geriatr Oncol 2017 Nov;8(6):437-442. [doi: 10.1016/j.jgo.2017.08.008] [Medline: 28888556]
12. Liu Y, Wang H, Zhao W, Zhang M, Qin H, Xie Y. Flexible, stretchable sensors for wearable health monitoring: sensing mechanisms, materials, fabrication strategies and features. Sensors (Basel) 2018 Feb 22;18(2) [FREE Full text] [doi: 10.3390/s18020645] [Medline: 29470408]
13. Nenggan Z, Zhaohui W, Man L, Yang L, Gang P. Infrastructure and reliability analysis of electric networks for e-Textiles. IEEE Trans Syst Man Cybern C 2010 Jan;40(1):36-51. [doi: 10.1109/TSMCC.2009.2031497]
14. Chen J, Lum PS. Pilot testing of the spring operated wearable enhancer for arm rehabilitation (SpringWear). J Neuroeng Rehabil 2018 Mar 02;15(1):13 [FREE Full text] [doi: 10.1186/s12984-018-0352-4] [Medline: 29499712]

15. Bell W, Colaiezzi BA, Prata CS, Coates JC. Scaling up dietary data for decision-making in low-income countries: new technological frontiers. Adv Nutr 2017 Nov;8(6):916-932 [FREE Full text] [doi: 10.3945/an.116.014308] [Medline: 29141974]
16. Lee H, Lee H, Whang M. An enhanced method to estimate heart rate from seismocardiography via ensemble averaging of body movements at six degrees of freedom. Sensors (Basel) 2018 Jan 15;18(1) [FREE Full text] [doi: 10.3390/s18010238] [Medline: 29342958]
17. Lee H, Hong YJ, Baik S, Hyeon T, Kim D. Enzyme-based glucose sensor: from invasive to wearable device. Adv Healthc Mater 2018 Apr;7(8):e1701150. [doi: 10.1002/adhm.201701150] [Medline: 29334198]
18. Lee J, Lee DH, Oh TJ, Kim KM, Choi SH, Lim S, et al. Clinical feasibility of continuously monitored data for heart rate, physical activity, and sleeping by wearable activity trackers in patients with thyrotoxicosis: protocol for a prospective longitudinal observational study. JMIR Res Protoc 2018 Feb 21;7(2):e49 [FREE Full text] [doi: 10.2196/resprot.8119] [Medline: 29467121]
19. Luo X, Fan W. Medical wearable devices and the health care model innovation. Chinese J Health Manage 2016;2016(1):70-75. [doi: 10.3760/cma.j.issn.1674-0815.2016.01.016]
20. Kirk SR. Wearable medical technology-the future of personalized medicine is here. Chinese J Med Manage Sci 2016;6(1):34-37. [doi: 10.4072/j.issn.2095-7432.2016.01.006]
21. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018 Jan;54(1):144-150. [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
22. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
23. Wen D, Zhang X, Liu X, Lei J. Evaluating the consistency of current mainstream wearable devices in health monitoring: a comparison under free-living conditions. J Med Internet Res 2017 Dec 07;19(3):e68 [FREE Full text] [doi: 10.2196/jmir.6874] [Medline: 28270382]
24. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform 2017 Dec;102:87-92. [doi: 10.1016/j.ijmedinf.2017.03.008] [Medline: 28495352]
25. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. Ann Intern Med 2017 Apr 18;166(8):610-612 [FREE Full text] [doi: 10.7326/L16-0353] [Medline: 28395305]
26. Salisbury JP, Keshav NU, Sossong AD, Sahin NT. Concussion assessment with smartglasses: validation study of balance measurement toward a lightweight, multimodal, field-ready platform. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e15 [FREE Full text] [doi: 10.2196/mhealth.8478] [Medline: 29362210]
27. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
28. Wen D, Zhang X, Lei J. Consumers' perceived attitudes to wearable devices in health monitoring in China: a survey study. Comput Methods Programs Biomed 2017 Mar;140:131-137. [doi: 10.1016/j.cmpb.2016.12.009] [Medline: 28254069]
29. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
30. Usability Partners. 1998. ISO Standards in usability and user-centred design—ISO 9241-11 (1998) guidance in usability URL: http://www.usabilitypartners.se/about-usability/iso-standards [accessed 2018-03-08] [WebCite Cache ID 6zSTOOYVm]
31. Yan Y, Wang G, Liu S, Wu H, Zhang Q. Usability evaluation of infusion pump based on system usability scale. China Medical Devices 2012;10(1):25-27. [doi: 10.3969/j.issn.1674-1633.2012.10.005]
32. Khajouei R, Hajesmaeel Gohari S, Mirzaee M. Comparison of two heuristic evaluation methods for evaluating the usability of health information systems. J Biomed Inform 2018 Apr;80:37-42. [doi: 10.1016/j.jbi.2018.02.016] [Medline: 29499315]
33. Khajouei R, Zahiri Esfahani M, Jahani Y. Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems. J Am Med Inform Assoc 2017 Apr 01;24(e1):e55-e60. [doi: 10.1093/jamia/ocw100] [Medline: 27497799]
34. Hudson D, Kushniruk A, Borycki E, Zuege DJ. Physician satisfaction with a critical care clinical information system using a multimethod evaluation of usability. Int J Med Inform 2018 Apr;112:131-136. [doi: 10.1016/j.ijmedinf.2018.01.010] [Medline: 29500010]
35. Ho GW. Examining perceptions and attitudes. West J Nurs Res 2017 Dec;39(5):674-689. [doi: 10.1177/0193945916661302] [Medline: 27456460]
36. Hoonakker PL, Carayon P, Walker JM. Measurement of CPOE end-user satisfaction among ICU physicians and nurses. Appl Clin Inform 2010 Jan 01;1(3):268-285 [FREE Full text] [doi: 10.4338/ACI-2010-03-RA-0020] [Medline: 21423791]

37. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. J Med Internet Res 2014 Dec 03;16(12):e268 [FREE Full text] [doi: 10.2196/jmir.3395] [Medline: 25472463]
38. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013 Apr 16;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
39. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. J Med Internet Res 2004 Sep 15;6(3):e29 [FREE Full text] [doi: 10.2196/jmir.6.3.e29] [Medline: 15471755]
40. Johnson C, Johnston D, Crowley P, Culbertson H, Rippen H, Damico D. EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records. Rockville, MD: Agency for Healthcare Research and Quality; 2011. URL: https://healthit.ahrq.gov/sites/default/files/docs/citation/EHR_Usability_Toolkit_Background_Report.pdf [accessed 2018-08-01] [WebCite Cache ID 73TXMQxJ7]
41. Jordan P, Jordan JW, McClelland I. Usability Evaluation in Industry. London: Taylor & Francis; 1996.
42. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Int 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
43. Lund A. ResearchGate. 2001. Measuring usability with the USE questionnaire URL: https://www.researchgate.net/profile/Arnold_Lund/publication/230786746_Measuring_Usability_with_the_USE_Questionnaire/links/56e5a90e08ae98445c21561c/Measuring-Usability-with-the-USE-Questionnaire.pdf [accessed 2018-08-01] [WebCite Cache ID 73TXz4NQU]
44. Kirakowski J, Cierlik B. Measuring the usability of web sites. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 1998 Oct 02 Presented at: 42nd Human Factors and Ergonomics Society Annual Meeting; Oct 5-9, 1998; Chicago, IL p. 424-428. [doi: 10.1177/154193129804200405]
45. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
46. Schenkenfelder R, Selinger S. A Comparison of Multiple Wearable Devices Regarding their User Experience During Running. URL: http://ffhoarep.fh-ooe.at/bitstream/123456789/686/1/125_218_Schenkenfelder_FullPaper_en_Final.pdf [accessed 2018-03-08] [WebCite Cache ID 6zSTYWYt5]
47. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 Dec 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
48. Canalys. 2015 May 20. Media alert: Fitbit maintains leadership share of wearable band market before Apple Watch entrance URL: https://tinyurl.com/y9bus445 [accessed 2018-03-08] [WebCite Cache ID 6zSTny381]
49. Canalys. 2014 May 21. Fitbit accounted for nearly half of global wearable band shipments in Q1 2014 URL: http://www.canalys.com/newsroom/fitbit-accounted-nearly-half-global-wearable-band-shipments-q1-2014 [accessed 2018-03-08] [WebCite Cache ID 6zSTyFV3T]
50. Dolan B. MobiHealthNews. 2014 Jan 15. Fitbit, Jawbone, Nike had 97 percent of fitness tracker retail sales in 2013 URL: http://www.mobihealthnews.com/28825/fitbit-jawbone-nike-had-97-percent-of-fitness-tracker-retail-sales-in-2013 [accessed 2018-03-08] [WebCite Cache ID 6zSUC0gI]
51. Bevan N, Carter J, Harker S. ISO 9241-11 Revised: what have we learnt about usability since 1998? Lect Notes Comput Sci 2015;9169(1):143-151. [doi: 10.1007/978-3-319-20901-2_13]
52. Flavián C, Guinalíu M, Gurrea R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Inform Manage 2006 Jan;43(1):1-14. [doi: 10.1016/j.im.2005.01.002]
53. Greene K, Byrne M, Everett S. A comparison of usability between voting methods. In: EVT. 2006 Presented at: USENIX/Accurate Electronic Voting Technology Workshop 2006 on Electronic Voting Technology Workshop; July 31-Aug 4, 2006; Vancouver, BC.
54. Campbell BA, Tossell CC, Byrne MD, Kortum P. Voting on a smartphone: evaluating the usability of an optimized voting system for handheld mobile devices. In: Proc Hum Factors Ergon Soc. 2011 Sep 01 Presented at: Human Factors and Ergonomics Society Annual Meeting; Sep 19-23, 2011; Las Vegas p. 1100-1104. [doi: 10.1177/1071181311551230]
55. Trafton J, Martins S, Michel M, Lewis E, Wang D, Combs A, et al. Evaluation of the acceptability and usability of a decision support system to encourage safe and effective use of opioid therapy for chronic, noncancer pain by primary care providers. Pain Med 2010 Apr;11(4):575-585. [doi: 10.1111/j.1526-4637.2010.00818.x] [Medline: 20202142]
56. Lewis J, Sauro J. The factor structure of the system usability scale. Lect Notes Comput Sc 2009;5619(1):94-103. [doi: 10.1007/978-3-642-02806-9_12]
57. WJX. Introduction of Sojump URL: http://www.wjx.cn/html/aboutus.aspx [accessed 2018-03-08] [WebCite Cache ID 6zSUEYFip]
58. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Comprehensive evaluation questionnaire for intelligent wearable devices. 2018. URL: https://www.wjx.cn/jq/5733344.aspx [accessed 2018-11-02] [WebCite Cache ID 73dF5aIKk]
59. Wong WS, Chen PP, Chow YF, Wong S, Fielding R. A study of the reliability and concurrent validity of the Chinese version of the Pain Medication Attitude Questionnaire (ChPMAQ) in a sample of Chinese patients with chronic pain. Pain Med 2016 Jun;17(6):1137-1144. [doi: 10.1093/pm/pnv086] [Medline: 26843593]
60. Cohen J. Statistical Power Analysis for the Behavioral Sciences. London: Taylor & Francis; 1988.
61. Aaron B, Philip K, James M. Journal of Usability Studies. 2009. Determining what individual SUS scores mean: adding an adjective rating scale URL: https://tinyurl.com/yax6fx5n [accessed 2018-08-01] [WebCite Cache ID 73TZmtS9R]

62. MacDorman KF, Whalen TJ, Ho C, Patel H. An improved usability measure based on novice and expert performance. Int J Hum-Comput Int 2011 Feb 08;27(3):280-302. [doi: 10.1080/10447318.2011.540472]
63. Kortum PT, Bangor A. Usability ratings for everyday products measured with the system usability scale. Int J Hum-Comput Int 2013 Jan;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
64. IDC. 2017 Oct. Worldwide wearables market shares, 1H17: Xiaomi leads the market by volume, Apple leads by value URL: http://www.idc.com/getdoc.jsp?containerId=US43141317 [accessed 2018-06-01] [WebCite Cache ID 73BiT0BMV]
65. Lidwell W, Holden K, Butler J. Universal Principles of Design, Revised and Updated 125 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions. Beverly, MA: Rockport Publisher; 2010.
66. Karapanos E, Gouveia R, Hassenzahl M, Forlizzi J. Wellbeing in the making: peoples' experiences with wearable activity trackers. Psychol Well Being 2016;6:4 [FREE Full text] [doi: 10.1186/s13612-016-0042-6] [Medline: 27376017]
67. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
68. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
69. Tencent ISUX. 2014. 2014 smart wearable market white paper URL: http://digi.tech.qq.com/a/20141231/032756.html [accessed 2018-03-08] [WebCite Cache ID 6zSVCYZRl]
70. Herning MM, Cook JH, Schneider JK. Cognitive behavioral therapy to promote exercise behavior in older adults: implications for physical therapists. J Geriatr Phys Ther 2005;28(2):34-38. [Medline: 16236226]
71. Rovniak LS, Kong L, Hovell MF, Ding D, Sallis JF, Ray CA, et al. Engineering online and in-person social networks for physical activity: a randomized trial. Ann Behav Med 2016 Dec;50(6):885-897 [FREE Full text] [doi: 10.1007/s12160-016-9814-8] [Medline: 27405724]
72. Sallis JF, Cerin E, Conway TL, Adams MA, Frank LD, Pratt M, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. Lancet 2016 May 28;387(10034):2207-2217. [doi: 10.1016/S0140-6736(15)01284-2] [Medline: 27045735]

Abbreviations

ANOVA: analysis of variance
SUS: System Usability Scale

Edited by G Eysenbach; submitted 17.05.18; peer-reviewed by A Sullivan, G Signorelli, J Quiroz; comments to author 13.06.18; revised version received 03.08.18; accepted 20.10.18; published 08.11.18.

Please cite as:
Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, Lei J
Usability Study of Mainstream Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation
JMIR Mhealth Uhealth 2018;6(11):e11066
URL: http://mhealth.jmir.org/2018/11/e11066/
doi:10.2196/11066
PMID:30409767

Original Paper

Feasibility of Using a Commercial Fitness Tracker as an Adjunct to Family-Based Weight Management Treatment: Pilot Randomized Trial

Thao-Ly Tam Phan[1,2,3], MD, MPH; Nadia Barnini[4], MS, RCEP, CSCS; Sherlly Xie[5], ScM; Angelica Martinez[6], MPH; Lauren Falini[3], MSc; Atiera Abatemarco[6], MPH, CHES; Maura Waldron[7], MS, CSCS, CPT; Jane M Benton[6], MD, MPH; Steve Frankenberry[6], LMHC; Cassandra Coleman[2], MD; Linhda Nguyen[3], PA-C; Cindy Bo[8], MBA; George A Datto[2,3], MD; Lloyd N Werk[6,9], MD, MPH

[1]Center for Healthcare Delivery Science, Nemours Children's Health System, Wilmington, DE, United States
[2]Department of Pediatrics, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, United States
[3]Division of Weight Management, Nemours Children's Health System, Wilmington, DE, United States
[4]Cardiac Center, Nemours Children's Hospital, Orlando, FL, United States
[5]Biostatistics Core, Department of Biomedical Research, Nemours Children's Health System, Wilmington, DE, United States
[6]Division of General Academic Pediatrics, Nemours Children's Hospital, Orlando, FL, United States
[7]Department of Kinesiology, West Chester University, West Chester, PA, United States
[8]Strategy and Business Development, Nemours Children's Health System, Wilmington, DE, United States
[9]College of Medicine, University of Central Florida, Orlando, FL, United States

Corresponding Author:
Thao-Ly Tam Phan, MD, MPH
Center for Healthcare Delivery Science
Nemours Children's Health System
1600 Rockland Road
Wilmington, DE, 19803
United States
Phone: 1 3023317342
Email: tphan@nemours.org

Abstract

Background: Fitness trackers can engage users through automated self-monitoring of physical activity. Studies evaluating the utility of fitness trackers are limited among adolescents, who are often difficult to engage in weight management treatment and are heavy technology users.

Objective: We conducted a pilot randomized trial to describe the impact of providing adolescents and caregivers with fitness trackers as an adjunct to treatment in a tertiary care weight management clinic on adolescent fitness tracker satisfaction, fitness tracker utilization patterns, and physical activity levels.

Methods: Adolescents were randomized to 1 of 2 groups (adolescent or dyad) at their initial weight management clinic visit. Adolescents received a fitness tracker and counseling around activity data in addition to standard treatment. A caregiver of adolescents in the dyad group also received a fitness tracker. Satisfaction with the fitness tracker, fitness tracker utilization patterns, and physical activity patterns were evaluated over 3 months.

Results: A total of 88 adolescents were enrolled, with 69% (61/88) being female, 36% (32/88) black, 23% (20/88) Hispanic, and 63% (55/88) with severe obesity. Most adolescents reported that the fitness tracker was helping them meet their healthy lifestyle goals (69%) and be more motivated to achieve a healthy weight (66%). Despite this, 68% discontinued use of the fitness tracker by the end of the study. There were no significant differences between the adolescent and the dyad group in outcomes, but adolescents in the dyad group were 12.2 times more likely to discontinue using their fitness tracker if their caregiver also discontinued use of their fitness tracker (95% CI 2.4-61.6). Compared with adolescents who discontinued use of the fitness tracker during the study, adolescents who continued to use the fitness tracker recorded a higher number of daily steps in months 2 and 3 of the study (mean 5760 vs 4148 in month 2, *P*=.005, and mean 5942 vs 3487 in month 3, *P*=.002).

Conclusions: Despite high levels of satisfaction with the fitness trackers, fitness tracker discontinuation rates were high, especially among adolescents whose caregivers also discontinued use of their fitness tracker. More studies are needed to determine how to sustain the use of fitness trackers among adolescents with obesity and engage caregivers in adolescent weight management interventions.

(JMIR Mhealth Uhealth 2018;6(11):e10523) doi:10.2196/10523

KEYWORDS

fitness trackers; pediatric obesity; health behavior; accelerometry

Introduction

Prevalence rates of obesity continue to increase among adolescents [1,2], translating into significant health consequences [3,4]. Expert guidelines recommend that pediatricians refer patients with obesity to tertiary care weight management clinics if they are unable to achieve a healthy weight in the primary care setting [5]. Patients who are adherent to interdisciplinary treatment in tertiary care weight management clinics demonstrate an improvement in weight status [6,7]. However, adherence to treatment is low, and attrition rates are high in tertiary care weight management clinics [8-10], especially among adolescents [11,12], thus limiting the number of patients who actually benefit from treatment. Indeed, tertiary care weight management clinics consistently report attrition rates >50% [8,9]. This underscores the need for innovative approaches to engage adolescents in weight management treatment.

One of the ways to engage patients in weight management treatment is to promote self-monitoring of lifestyle behaviors such as physical activity, which are associated with improved weight and health outcomes [13,14]. Self-monitoring is based on self-regulation theory, which proposes that monitoring one's own behaviors leads to self-evaluation of how these behaviors impact progress toward a goal, which then positively reinforces behaviors that positively impact and negatively reinforces behaviors that negatively impact progress toward that goal [13,14]. Despite self-monitoring being a cornerstone of weight management treatment [15,16], it can be an arduous process [13,15,17]. Commercial wearable devices that track physical activity (fitness trackers) automate the self-monitoring process [18] and may be useful for engaging adolescents who are heavy technology users [19]. Fitness trackers have shown moderate success in assisting with weight management efforts in adults [20-23], but these studies have not extensively evaluated newer fitness trackers that are more accurate [24-27] and able to automate cognitive behavioral techniques such as goal setting and reinforcement [21].

Although there are many studies using research accelerometers with pediatric populations, commercial fitness trackers have not been extensively studied in the pediatric population. A recent systematic review found only 4 studies evaluating fitness trackers as an intervention among children, with only 1 of these studies evaluating fitness trackers as an intervention among adolescents [28]. These 4 pediatric studies reported positive effects of fitness trackers on physical activity but were limited by their use of older fitness tracker models and small sample sizes. In addition, none of these studies evaluated the use of fitness trackers in a clinical setting where physical activity data can be used to tailor counseling about physical activity for patients with obesity. Finally, none of these studies has evaluated the impact of providing caregivers with fitness trackers despite evidence that adolescent physical activity level is associated with the physical activity level of their caregivers [29-32]. Therefore, the purpose of this study was to test the use of a fitness tracker as an adjunct to family-based weight management treatment for adolescents. Our aim was to explore whether providing caregivers with a fitness tracker, in addition to providing adolescents with a fitness tracker, would improve an adolescent's fitness tracker satisfaction, fitness tracker utilization, and physical activity levels.

Methods

Study Design

A randomized trial of a fitness tracker was conducted over 3 months. Participants were randomized (1:1) using computer block randomization techniques into 1 of 2 groups: (1) adolescent group with the adolescent receiving a fitness tracker or (2) dyad group with the adolescent and their caregiver receiving a fitness tracker (Figure 1). Adolescents in both groups also received counseling regarding physical activity data as part of standard tertiary care weight management treatment, as described in detail below. At the time of trial implementation, the study did not meet requirements for registration on clinicaltrial.gov because of the sample size, lack of a control group naïve to an intervention, and intent to test the feasibility of the study procedures.

Participants

Participants recruited for the trial were adolescents (aged 13-17 years) who were new patients to 1 of 2 tertiary care weight management clinics in a children's health system. Patients were referred by their primary care provider to the clinic for having a body mass index percentile ≥85th percentile for age and having failed attempts at weight management in the primary care setting. One of the adolescent's caregivers was also enrolled as a participant in the study. Participants were excluded if the adolescent had a primary genetic or endocrine syndrome associated with obesity or was taking a medication that would predispose them to weight gain; if the caregiver was not the adolescent's legal guardian, did not reside in the same household, or was of limited English proficiency; or if either the adolescent or their caregiver was unable to understand how to use the fitness tracker, had used a fitness tracker before, did not have a smartphone or tablet computer, or had a condition limiting physical activity. Participants were also excluded from

the study if they were unable to attend monthly weight management clinic visits for logistical or insurance reasons. This study was approved by the institutional review board, with both adolescents and their caregivers signing consent forms at the time of enrollment.

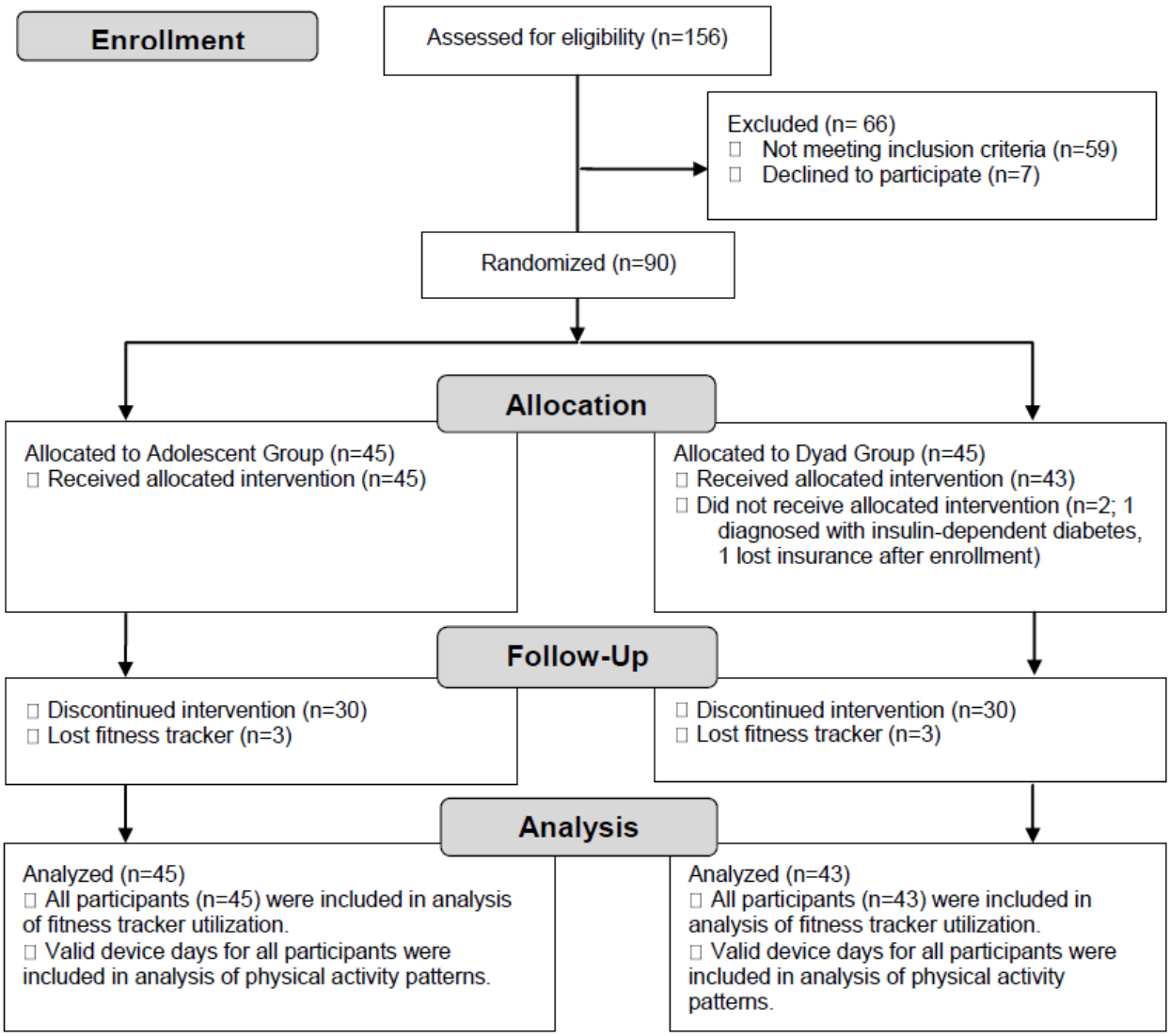
Sample Size Derivation

The primary aim of this study was to assess the differential impact of an adolescent using the fitness tracker alone or in conjunction with their caregiver on physical activity patterns. As this is a novel comparison of the intervention, the effects are unknown. To detect a clinically meaningful difference in mean daily steps of 2000 (equating to approximately 1 mile of additional walking) between groups with a power of 80% and a level of significance of .05, assuming a 30% attrition rate for this 3-month study, we estimated that a sample size of 90 participants (45 in each experimental group) would be needed.

Setting

Each of the 2 tertiary care weight management clinics was located inside a tertiary care children's hospital. One clinic was located in the mid-Atlantic, and 1 clinic was located in the South Atlantic. Each clinic provides individualized, interdisciplinary weight management treatment to patients aged younger than 18 years (please see below section on Standard Weight Management Treatment for more details). Participants in the South Atlantic clinic were more likely to be of Hispanic background (48% [14/29] vs 14% [8/59] in the mid-Atlantic, $P<.001$), less likely to be of non-Hispanic white background (21% [6/29] vs 44% [26/59], $P=.03$), and more likely to have Medicaid insurance (55% [16/29] vs 29% [17/59], $P=.02$). There were no significant differences between participants at the 2 clinical sites in terms of gender, age, and weight status.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram for pilot randomized trials.



Intervention

Standard Weight Management Treatment

All participants received the same standard weight management treatment, which was the typical routine care provided to all patients in the tertiary care weight management clinics. At their initial visit to the clinic, adolescents and their families underwent an assessment by a medical provider and received an individualized care plan. A typical plan included monthly clinic visits with an interdisciplinary team, which could include a medical provider, health educator, exercise physiologist, dietitian, or psychologist depending on identified needs. Care provided during each visit included management of medical comorbidities, nutrition education, physical activity counseling, and discussion of behavioral strategies such as goal setting and self-monitoring, with elements of motivational interviewing used to encourage behavior change. Care was targeted toward the entire family and stressed the importance of adoption of healthy lifestyle behaviors by the entire family. There was opportunity for the adolescent to participate in weekly personal training sessions at the mid-Atlantic site, which only 6 adolescents participated in.

Fitness Tracker Intervention

Adolescents in both the adolescent and dyad groups received a free fitness tracker. Caregivers of adolescents assigned to the dyad group also received a free fitness tracker. Participants were allowed to keep their fitness tracker as an incentive but were not provided additional incentives during the study to prevent influence on fitness tracker utilization patterns.

The fitness tracker used in this study was a slim, adjustable, water-resistant device worn on the wrist with a rubber wristband with a battery life of 5 days. The fitness tracker houses an accelerometer that tracks physical activity and sleep. This particular fitness tracker was found to be the second most accurate fitness tracker among 7 fitness trackers tested, with a reported error rate of 10% to 18% in measuring energy expenditure [25,33,34]. Each of 5 LED lights on the wristband illuminates with every 20% progress toward a daily activity goal set by the participants and vibrates when the daily goal is reached.

Data from the accelerometer wirelessly sync to computers and smartphones, with data viewable through a Web-based dashboard or mobile device app operated by the device manufacturer. These apps include features that allow participants to track physical activity (including steps taken, calories burned, and intensity of activity) and sleep (including duration of sleep and periods of restlessness) over different time intervals (daily, weekly, and monthly). Users are also able to earn virtual badges for reaching goals and compete in challenges with other fitness tracker users in real time through the Web-based app. Finally, participants can track food and water intake with the apps. Weekly reports are automatically emailed to users by the device manufacturer, containing a summary of their physical activity statistics, reminders about charging the fitness tracker, and links to healthy lifestyle resources.

At their initial visit to the weight management clinic, research staff assisted participants with installing the mobile app on their smartphone or tablet computer, setting up a fitness tracker account, and properly fitting the fitness tracker to each participant's wrist. Research staff instructed participants on how to use the fitness tracker, including its different apps and functions, and how to sync the device. In addition, participants in the dyad group were encouraged to participate in fitness tracker challenges with one another. To account for variability in baseline physical activity, a phone call from research staff was made to adolescents 1 week after their initial visit to increase their step goal by 1000 steps above the average number of steps the adolescent had taken over the 3 days before the call. Research staff also provided assistance with fitness tracker issues during this call.

A secure research database was used to automatically and continuously collect coded data from the fitness tracker every time it was synced. Research staff texted generic reminders to participants to use, charge, and sync their fitness tracker on a weekly basis if they were noted to have zero steps recorded that week on the research database dashboard. During each weight management clinic visit, health care providers viewed the participant's research database dashboard and reviewed their physical activity and progress for the past month, including amount of moderate to vigorous physical activity (MVPA) and step counts. Participants were encouraged to increase their MVPA and advised to strive for a new step goal that was 1000 steps above the average daily number of steps taken during their most consistent week.

Measures

Adolescent Demographics

Adolescent age, sex, and insurance status were collected in the electronic health record as part of routine clinical care and entered into a REDCap database [35] by research staff at the initial visit. Adolescent race and ethnicity were reported via a REDCap questionnaire at the initial visit. Date of enrollment was also captured in REDCap and categorized according to season (summer being June to August, fall being September to November, winter being December to February, and spring being March to May).

Fitness Tracker Utility

A questionnaire was administered via REDCap survey to adolescents at each follow-up clinic visit (1, 2, and 3 months). Three items were asked to assess the adolescent's perception about whether the fitness tracker was helping them meet their healthy lifestyle goals, be more physically active, and be more motivated to achieve a healthy weight. The adolescent was asked to rate how strongly they agreed with each statement on a 5-point Likert scale. Items were considered positively endorsed if they received a 4 or 5 on the Likert scale. Five additional items were asked to assess how frequently the adolescent used the functions of the fitness tracker (mobile app, Web-based dashboard, virtual rewards, challenges with their caregiver, and food and water log). Items responses were on a 5-point Likert scale. Items were considered frequently used (ie, used at least weekly) if they scored a 3, 4, or 5 on the Likert Scale.

Fitness Tracker Utilization and Physical Activity Patterns

Data on number of steps taken, number of calories burned, and number of minutes spent in MVPA were automatically and continuously collected by the fitness tracker and synced to the research database. Only physical activity on valid device days, which were defined as a day where 16 or more hours of nonzero steps were tracked, similar to criteria used in other studies [36,37], was used to the describe physical activity. For the purpose of analysis, physical activity data were described in terms of daily steps, daily minutes of MVPA, and daily calories burned. A mean daily value for each physical activity metric was calculated for each individual for each month of the study. Mean daily step data were further categorized by time of day (during school hours from 8 am to 4 pm or during nonschool hours from 4 pm to 11 pm) for some analyses. Adolescents were considered to have discontinued use of the fitness tracker if there was a time point after which they had no more valid device days, for at least 7 consecutive days.

Data Analysis

Adolescent demographics were summarized and compared between the adolescent and dyad groups. To describe fitness tracker satisfaction, responses to the Fitness Tracker Utility questionnaire were summarized. Data collected from the fitness tracker were summarized to describe fitness tracker utilization and physical activity patterns. To explore the impact of providing caregivers with a fitness tracker as part of the intervention, intent-to-treat analysis was used to compare adolescent fitness tracker satisfaction, fitness tracker utilization, and physical activity patterns between the adolescent and dyad groups. Due to differences in season of enrollment and differences in demographics between the 2 clinical sites, we also tested to see if there were any differences in adolescent fitness tracker utility responses, fitness tracker utilization, and physical activity patterns between adolescents based on season of enrollment and clinical site. Due to high fitness tracker discontinuation rates, we also conducted exploratory analyses comparing demographic and physical activity patterns between adolescents who discontinued use of the fitness tracker during the study and those who did not. Chi-square or Fisher exact analyses were used to compare categorical variables, and 2-sample *t* tests were used to compare continuous variables. All tests were 2 tailed, with an overall level of significance of .05. Statistical software R was used.

Results

Adolescent Demographics

A total of 90 adolescents were recruited to receive a fitness tracker, with random assignment of 45 to the adolescent group and 45 to the dyad group (Figure 1). Two adolescents were excluded from the dyad group in the first month of the study because of insurance problems and diagnosis of insulin-dependent type 2 diabetes. The 88 remaining adolescents were representative of the patient population with the majority being female (61/88, 69%), having severe obesity (55/88, 63%), and representing a diversity of racial and ethnic backgrounds (non-Hispanic black [32/88, 36%], non-Hispanic white [26/88, 29.5%], and Hispanic [20/88, 23%]). There were no significant differences in baseline demographics between the adolescent and dyad groups (Table 1).

Fitness Tracker Utility

Of the 72 adolescents who filled out the questionnaire at the 1-month follow-up visit, the majority reported that the fitness tracker was helping them meet their healthy lifestyle goals (50/72, 69%), be more physically active (49/72, 68%), and be more motivated to achieve a healthy weight (48/72, 66%). Of the 41 adolescents who completed the questionnaire at the 3-month follow-up visit, an even greater majority reported that the fitness tracker was helping them meet their healthy lifestyle goals (32/41, 78%), be more physically active (34/41, 82%), and be more motivated to achieve a healthy weight (33/41, 80%). Every month, more than 75% of adolescents reported accessing their mobile app weekly, but less than 45% of adolescents reported accessing their dashboard on the consumer website weekly. One-third of all adolescents used virtual rewards, and one-third of adolescents in the dyad group engaged in challenges with their caregiver. Food and activity logs were used rarely. There were no significant differences on any Fitness Tracker Utility questionnaire responses between the adolescent and dyad groups, between participants based on clinical site, or between participants based on the season of enrollment.

Fitness Tracker Utilization and Physical Activity Patterns

Table 2 describes fitness tracker utilization and physical activity patterns only on valid device days. Sixty-six (66/88, 68%) adolescents discontinued use of the fitness tracker before the end of the study, with 25 adolescents (25/88, 28%) discontinuing use of the fitness tracker in the first month. Notably, adolescents in the dyad group were 12.2 times more likely to discontinue using their fitness tracker if their caregiver also discontinued use of their fitness tracker (95% CI 2.4-61.6). Even among adolescents who continued to use the fitness tracker, the number of valid device days declined over time, from 29 days in the first month to 18 days in the third month of the study. Mean daily steps, minutes of MVPA, and calories burned on valid device days were consistently less than 8000 steps and 20 min of MVPA during each month. Number of steps per hour was significantly higher during school hours than during nonschool hours (484 vs 362 in the adolescent group and 473 vs 340 in the dyad group, P=.01).

There were no significant differences between the adolescent and dyad groups or between participants based on clinical site on any measures of fitness tracker utilization or physical activity patterns. However, participants who were enrolled in the summer burned a significantly higher number of calories in month 2 (mean 2549 vs 2241 among those enrolled in the fall and 2222 among those enrolled in the winter, P=.04) and month 3 (mean 2602 vs. 2191 among those enrolled in the fall and 2257 among those enrolled in the winter, P<.001).

Table 1. Adolescent demographics.

Adolescent demographics	Adolescent group[a] (n=45)	Dyad group[b] (n=43)	*P* value
Female, n (%)	31 (69)	30 (70)	1.00
Race/ethnicity, n (%)			.75
Non-Hispanic black	18 (40)	14 (33)	
Non-Hispanic white	12 (27)	14 (33)	
Hispanic	11 (24)	9 (21)	
Other	4 (9)	6 (14)	
Medicaid, n (%)	15 (33)	18 (42)	.55
Age (years), mean (SD)	15 (1.4)	14.7 (1.2)	.25
Baseline weight category, n (%)			.11
Overweight or obese (BMI[c] <99% for age)	21 (47)	12 (28)	
Severe obesity (BMI ≥99% for age)	24 (53)	31 (72)	
Clinic location, n (%)			1.00
Mid-Atlantic region	30 (67)	29 (67)	
South Atlantic region	15 (33)	14 (33)	
Season of enrollment, n (%)			.62
Summer (June to August)	17 (39)	20 (47)	
Fall (September to November)	22 (50)	17 (40)	
Winter (December to February)	5 (11)	6 (14)	

[a]Participants randomized to receive fitness tracker and counseling about physical activity data during standard weight management treatment.

[b]Participants randomized to receive fitness tracker and counseling about physical activity data during standard weight management treatment and for caregiver to also receive fitness tracker.

[c]BMI: body mass index.

Table 2. Adolescent fitness tracker utilization and physical activity patterns.

Fitness tracker utilization and physical activity patterns	Adolescent group[a] (n=45)	Dyad group[b] (n=43)	*P* value
Discontinued using the fitness tracker,[c] n (%)			
By 3 months	30 (67)	30 (70)	.93
Month 1	14 (31)	11 (26)	.57
Month 2	10 (22)	12 (28)	.53
Month 3	6 (13)	7 (16)	.70
Valid device days[d] before device discontinuation, mean number of days (% per month)			
Month 1	29 (97)	28 (92)	.44
Month 2	26 (86)	24 (81)	.43
Month 3	19 (62)	17 (58)	.41
Daily steps on valid device days[d], mean (SD)			
Month 1	7541.1 (2891.2)	7356.3 (2611.4)	.75
Month 2	7717.9 (3264.1)	7287.2 (2332.8)	.51
Month 3	7873.7 (3035.3)	7150.4 (2543.4)	.31
Weekdays	7859.1 (3002.2)	7676.2 (2431.1)	.75
Weekends	6865.2 (3099.5)	6126.1 (2675.0)	.24
Daily minutes of moderate to vigorous physical activity on valid device days[d], mean (SD)			
Month 1	20.7 (34.0)	17.5 (29.8)	.64
Month 2	15.3 (29.9)	13.6 (26.1)	.78
Month 3	17.7 (33.1)	12.2 (24.7)	.38
Daily calories burned on valid device days[d], mean (SD)			
Month 1	2665.9 (500.5)	2670.3 (477.4)	.97
Month 2	2700.7 (553.0)	2666.6 (514.8)	.78
Month 3	2706.0 (554.2)	2653.4 (529.3)	.70
Hourly steps on valid device days[d], mean (SD)			
School hours (8 am to 4 pm)	483.5 (187.4)	473.9 (145.2)	.79
Nonschool hours (4 pm to 11 pm)	361.6 (174.7)	340.1 (162.3)	.55

[a]Participants randomized to receive fitness tracker and counseling about physical activity data during standard weight management treatment.

[b]Participants randomized to receive fitness tracker and counseling about physical activity data during standard weight management treatment and for caregiver to also receive fitness tracker.

[c]Fitness tracker considered to be discontinued if there was a time point before the last week of the study after which there were 0 valid device days.

[d]A day was considered to be nonvalid if there were 16 or more hours of 0 steps.

Comparison of Adolescents Based on Fitness Tracker Discontinuation

Table 3 compares demographic characteristics and physical activity patterns between adolescents who discontinued use of the fitness tracker during the study and adolescents who continued to use the fitness tracker for the duration of the study. Adolescents who continued to use the fitness tracker were older (mean age 15.3 vs 14.6 years, *P*=.02) and were more likely to have been enrolled in the study in the summer (*P*=.05). Adolescents who continued to use the fitness tracker recorded a significantly higher number of mean daily steps in months 2 and 3 of the study (5760 vs 4148 in month 2, *P*=.002, and 5942 vs 3487 in month 3, *P*=.002) and calories burned in month 3 of the study (2539 vs 2262, *P*=.03) on valid device days than adolescents who discontinued use of the fitness tracker during the study.

Table 3. Comparison of adolescents based on fitness tracker discontinuation.

Adolescent demographics	Discontinued fitness tracker[a] (n=60)	Continued fitness tracker (n=28)	*P* value
Male, n (%)	20 (33)	7 (25)	.47
Race/ethnicity, n (%)			.39
Non-Hispanic black	19 (32)	13 (46)	
Hispanic	16 (27)	4 (14)	
Non-Hispanic white	19 (32)	7 (25)	
Other	6 (10)	4 (14)	
Medicaid, n (%)	26 (43)	7 (25)	.16
Age (years), mean (SD)	14.6 (1.3)	15.3 (1.2)	.02
Baseline weight category, n (%)			1.00
Overweight/obese (BMI[b] <99% for age)	23 (38)	10 (36)	
Severe obesity (BMI ≥99% for age)	37 (62)	18 (64)	
Clinic location, n (%)			.40
Mid-Atlantic region	38 (63)	21 (75)	
South Atlantic region	22 (37)	7 (25)	
Season of enrollment, n (%)			.05
Summer (June to August)	23 (38)	14 (50)	
Fall (September to November)	32 (53)	8 (29)	
Winter (December to February)	5 (8)	6 (21)	
Daily steps on valid device days[c], mean (SD)			
Month 1	5768.0 (3027.4)	6096.0 (2421.3)	.62
Month 2	4147.6 (3133.3)	5760.4 (1902.1)	.005
Month 3	3486.5 (3111.3)	5942.0 (2360.1)	.002
Daily minutes of moderate to vigorous physical activity on valid device days[c], mean (SD)			
Month 1	20.6 (33.3)	16.4 (29.4)	.56
Month 2	14.7 (30.8)	14.2 (23.5)	.93
Month 3	15.8 (32.4)	13.2 (25.9)	.71
Daily calories burned on valid device days[c], mean (SD)			
Month 1	2469.8 (544.7)	2568.5 (472.4)	.41
Month 2	2279.2 (606.6)	2525.0 (404.5)	.16
Month 3	2261.8 (545.5)	2538.9 (428.7)	.03

[a]Fitness tracker considered to be discontinued if there was a time point before the last week of the study after which there were 0 valid device days.

[b]BMI: body mass index.

[c]A day was considered to be nonvalid if there were 16 or more hours of 0 steps.

Discussion

Principal Findings

Our study is the first to describe the use of a fitness tracker as an adjunct to family-based weight management treatment among a cohort of adolescents with obesity, a historically difficult to engage patient population with lower levels of physical activity and worse health consequences, and to describe the impact of providing fitness trackers to caregivers in support of their adolescents. Although the majority of adolescents in this study reported that the fitness tracker was helpful in their weight management efforts, discontinuation rates were high. In addition, there was no impact of providing caregivers with fitness trackers on adolescent physical activity levels; caregiver fitness tracker discontinuation rates were also high and associated with adolescent fitness tracker discontinuation. Adolescents who did not discontinue use of the fitness tracker by the end of the study were able to maintain physical activity levels, whereas those who discontinued use of their fitness tracker during the study had a decline in physical activity levels over time.

The majority of adolescents in this study reported that the fitness tracker was helpful in their weight management efforts and reported accessing their physical activity data weekly through a mobile app, supporting the utility of physical activity feedback especially when easily accessible via mobile technology. This is consistent with other studies that have demonstrated high levels of acceptability of fitness trackers among children and adolescents [38-40]. For example, the majority of adolescents in a study by Slootmaker et al found value in the physical activity assessments provided by fitness trackers [40]. In addition, qualitative studies by Schaefer et al found that pre-adolescents enjoyed the feedback feature of fitness trackers and that these features promoted self-reflection about their physical activity [38,39].

Despite high acceptability rates, the majority of adolescents discontinued use of the fitness tracker by the end of the study. Even among adolescents who continued to use the device for the entire study, the number of days the fitness tracker was used declined each month. Adherence to fitness trackers and discontinuation of fitness trackers have also been a major challenge in other studies [28,38,40,41]. A systematic review of randomized trials using accelerometers to measure physical activity in children found that one-third of children were not adherent to using accelerometers. In trials of commercial fitness trackers as the actual intervention, adherence is even worse. For example, only 24% of adolescents in Slootmaker et al's study provided fitness tracker data for the entire study [40], and among pre-adolescents in Schaefer et al's study, there was a 50% decrease in the use of the fitness tracker in the first month [38].

There may be several reasons for fitness tracker discontinuation over time. Seasonality appears to play a role, with participants who enrolled in the summer being more likely to continue use of their fitness tracker. This is likely because adolescents have more time when not in school, and the weather is more conducive in the summer for engaging in physical activity and use of the fitness tracker. There may also be other reasons that are more psychological in nature. For example, a recent mixed-methods study by Kerner and Goodyear found that adolescents lost motivation to use a fitness tracker over time because of feelings of incompetence when not achieving their goal [42], and a review by Sullivan and Lachman noted that inactive individuals struggle to come up with a plan for how to be physically active and overcome obstacles to being physically active [43], barriers that are not addressed with current fitness trackers.

In yet another study, Schaefer et al noted that utilization of the fitness tracker was highest during times where there was contact with research staff, suggesting that in-person support may still be a necessary component to engagement despite the automatic support inherent in the fitness tracker itself [38]. Similarly, utilization of the fitness tracker was highest in our study during the first month when contact with research staff was more frequent. At the same time, only about one-third of adolescents in our study used the automated functions of the fitness tracker that provide positive reinforcement such as rewards and challenges, which are important for behavior change, especially among inactive individuals [43], suggesting that more directive guidance on how to use these functions may also help to increase utilization of the fitness tracker and promote physical activity.

Fitness tracker discontinuation rates of adolescents in the dyad group were strongly associated with those of caregivers in the dyad group, suggesting that there is a strong influence of caregiver actions on adolescents. This is not surprising, given the importance of caregivers as role models to their children and the importance of the home environment to a child's lifestyle habits [29-32]. However, caregivers in the dyad group also demonstrated low fitness tracker utilization rates, which may explain why adolescents in the dyad group did not demonstrate better fitness tracker utilization and physical activity patterns than adolescents in the adolescent group. Indeed, these findings suggest that more attention to increasing the physical activity of caregivers may be necessary to indirectly promoting physical activity in their adolescents.

At the same time, the lack of difference in outcomes between the adolescents in the adolescent and dyad groups may also speak to the need to focus efforts on peer-based interventions, especially among this age group that is more responsive to peer influence and support [44-46]. For example, Schaefer et al found that pre-adolescents in their study engaged with their fitness tracker most when competing with peers [38]. This concept is supported by studies demonstrating the positive effect of exergaming on physical activity among children, especially when peer socialization elements are incorporated [47-49]. Social media trends such as Pokemon GO further exemplify the role of social gaming in motivating individuals to be physically active [50,51]. Incorporating peer socialization as a component of fitness tracker interventions may also enhance physical activity outside of the school setting, which is important given that our study and others have observed significantly lower physical activity levels after school hours than during school hours [52].

Physical activity levels did not improve over the course of the study and were lower than evidence-based recommendations of 10,000 steps or 60 min of MVPA every day [53]. This lack of improvement in physical activity is in contrast to other studies reporting positive effects of fitness trackers on physical activity levels [20,28] but is consistent with studies reporting generally low levels of MVPA among children in this country, especially those with obesity and of adolescent age, who are a very difficult group to motivate to be more active [52,54,55] for many of the reasons that Sullivan and Lachman note in their study on inactive adults [43]. Notably, adolescents who continued to use the fitness tracker for the entire study period maintained their physical activity levels, whereas adolescents who discontinued use of the fitness tracker by the end of the study demonstrated a decline in their physical activity levels over the course of the study. This suggests that fitness tracker use may at least promote maintenance of physical activity levels among adolescents, even if it does not encourage an increase in physical activity levels. Because higher physical activity levels are associated with long-term weight loss and health benefits [12,13,56], this finding supports the importance of using fitness trackers with adolescents who are most likely to benefit from them or finding ways to increase adolescent engagement with fitness trackers.

Limitations

There were several limitations to this study. Although our study was larger than other fitness tracker trials in pediatric populations, the sample size was still small, and the duration of our study was short. As this was a pilot study, we also did not include a control group of adolescents receiving standard treatment in the weight management clinic. Because we wanted to evaluate a commercial fitness tracker as an adjunct to care, leveraging the automated functions of the fitness tracker, we provided minimal intervention or incentives beyond counseling around the fitness tracker data at routine clinic visits, and use of the intervention may have been limited by functions specific to the fitness tracker, including the need to sync and charge the device. Also, the majority of adolescents in our study reported that the fitness tracker was helpful, but this may have been subject to self-report bias. Collecting qualitative data about why adolescents discontinued use of their fitness tracker may have provided meaningful information to inform future interventions. In addition, fitness tracker discontinuation rates were high, limiting the data that were available for analysis of physical activity, potentially leading to under- or overestimation of physical activity among participants.

Conclusions

Despite high levels of satisfaction with the fitness trackers among adolescents enrolled in a weight management clinic, fitness tracker discontinuation rates were high, suggesting that more guidance and support may be needed beyond what is provided by the fitness tracker alone. This includes the support of caregivers, among whom fitness tracker discontinuation rates were also high and associated with adolescent fitness tracker discontinuation. Adolescents who continued use of the fitness tracker for the entire study demonstrated sustained physical activity levels compared with the decline in physical activity levels seen among adolescents who discontinued use of the fitness tracker during the study, suggesting that fitness tracker use may promote maintenance of physical activity levels over time. We believe the findings of our study provide valid and valuable insight into how adolescents with obesity use fitness trackers and underscores the importance of continued research to identify which subgroups will benefit most from this type of intervention while identifying innovative ways to engage adolescents in use of fitness trackers over time, including enhanced caregiver or peer involvement and features that address barriers to physical activity specific to obese patients.

Acknowledgments

This research was supported by the Nemours Children's Health System. TLTP receives support from the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (award #K23HD083439). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Ogden CL, Carroll MD, Lawman HG, Fryar CD, Kruszon-Moran D, Kit BK, et al. Trends in Obesity Prevalence Among Children and Adolescents in the United States, 1988-1994 Through 2013-2014. JAMA 2016 Jun 07;315(21):2292-2299. [doi: 10.1001/jama.2016.6361] [Medline: 27272581]
2. Skinner AC, Ravanbakht SN, Skelton JA, Perrin EM, Armstrong SC. Prevalence of Obesity and Severe Obesity in US Children, 1999-2016. Pediatrics 2018 Feb 26. [doi: 10.1542/peds.2017-3459] [Medline: 29483202]
3. Trasande L, Liu Y, Fryer G, Weitzman M. Effects of childhood obesity on hospital care and costs, 1999-2005. Health Aff (Millwood) 2009;28(4):w751-w760. [doi: 10.1377/hlthaff.28.4.w751] [Medline: 19589800]
4. Bass R, Eneli I. Severe childhood obesity: an under-recognised and growing health problem. Postgrad Med J 2015 Nov;91(1081):639-645. [doi: 10.1136/postgradmedj-2014-133033] [Medline: 26338983]
5. Barlow SE, Expert C. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. Pediatrics 2007 Dec;120 Suppl 4:S164-S192. [doi: 10.1542/peds.2007-2329C] [Medline: 18055651]
6. Hampl SE, Borner KB, Dean KM, Papa AE, Cordts KP, Smith TR, et al. Patient Attendance and Outcomes in a Structured Weight Management Program. J Pediatr 2016 Dec;176:30-35. [doi: 10.1016/j.jpeds.2016.05.032] [Medline: 27289497]
7. Skelton JA, DeMattia LG, Flores G. A pediatric weight management program for high-risk populations: a preliminary analysis. Obesity (Silver Spring) 2008 Jul;16(7):1698-1701 [FREE Full text] [doi: 10.1038/oby.2008.243] [Medline: 18451781]
8. Halvorson EE, Skelton JA. Appointment attendance in a pediatric weight management clinic. Clin Pediatr (Phila) 2012 Sep;51(9):888-891 [FREE Full text] [doi: 10.1177/0009922811410876] [Medline: 21622690]
9. Skelton JA, Beech BM. Attrition in paediatric weight management: a review of the literature and new directions. Obes Rev 2011 May;12(5):e273-e281 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00803.x] [Medline: 20880126]

10. Hampl S, Demeule M, Eneli I, Frank M, Hawkins MJ, Kirk S, et al. Parent perspectives on attrition from tertiary care pediatric weight management programs. Clin Pediatr (Phila) 2013 Jun;52(6):513-519. [doi: 10.1177/0009922813482515] [Medline: 23539682]
11. Skelton JA, Irby MB, Geiger AM. A systematic review of satisfaction and pediatric obesity treatment: new avenues for addressing attrition. J Healthc Qual 2014;36(4):5-22 [FREE Full text] [doi: 10.1111/jhq.12003] [Medline: 23414547]
12. Hampl S, Paves H, Laubscher K, Eneli I. Patient engagement and attrition in pediatric obesity clinics and programs: results and recommendations. Pediatrics 2011 Sep;128 Suppl 2:S59-S64 [FREE Full text] [doi: 10.1542/peds.2011-0480E] [Medline: 21885646]
13. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
14. Hallal PC, Victora CG, Azevedo MR, Wells JCK. Adolescent physical activity and health: a systematic review. Sports Med 2006;36(12):1019-1030. [Medline: 17123326]
15. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
16. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behavior Therapy 1993 Jun;24(3):377-394. [doi: 10.1016/S0005-7894(05)80212-6]
17. Burke LE, Swigart V, Warziski TM, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
18. Helsel DL, Jakicic JM, Otto AD. Comparison of techniques for self-monitoring eating and exercise behaviors on weight loss in a correspondence-based intervention. J Am Diet Assoc 2007 Oct;107(10):1807-1810. [doi: 10.1016/j.jada.2007.07.014] [Medline: 17904942]
19. Lenhart A. Pew Research Center. 2015 Apr 09. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2018-10-30] [WebCite Cache ID 6xm4w55rk]
20. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
21. de Vries HJ, Kooiman TJM, van Ittersum MW, van Brussel M, de Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring) 2016 Oct;24(10):2078-2091 [FREE Full text] [doi: 10.1002/oby.21619] [Medline: 27670401]
22. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
23. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: A systematic review. BMC Public Health 2015;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
24. Adam NJ, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. J Med Eng Technol 2013 Oct;37(7):456-462. [doi: 10.3109/03091902.2013.831135] [Medline: 24007317]
25. Iowa State University Physical Activity and Health Promotion Lab. Results of new consumer based physical activity monitor study! URL: http://www.physicalactivitylab.org/consumer-monitor-study-new.html [accessed 2018-10-30] [WebCite Cache ID 6xm5t52CY]
26. 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;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
27. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
28. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
29. Craig CL, Cameron C, Tudor-Locke C. Relationship between parent and child pedometer-determined physical activity: a sub-study of the CANPLAY surveillance study. Int J Behav Nutr Phys Act 2013 Jan 18;10:8 [FREE Full text] [doi: 10.1186/1479-5868-10-8] [Medline: 23331386]
30. Fuemmeler BF, Anderson CB, Mâsse LC. Parent-child relationship of directly measured physical activity. Int J Behav Nutr Phys Act 2011 Mar 08;8:17 [FREE Full text] [doi: 10.1186/1479-5868-8-17] [Medline: 21385455]
31. Tu AW, Watts AW, Masse LC. Parent-Adolescent Patterns of Physical Activity, Sedentary Behaviors and Sleep Among a Sample of Overweight and Obese Adolescents. J Phys Act Health 2015 Nov;12(11):1469-1476. [doi: 10.1123/jpah.2014-0270] [Medline: 25621567]

32. Verloigne M, Veitch J, Carver A, Salmon J, Cardon G, De Bourdeaudhuij I, et al. Exploring associations between parental and peer variables, personal variables and physical activity among adolescents: a mediation analysis. BMC Public Health 2014 Sep 18;14:966 [FREE Full text] [doi: 10.1186/1471-2458-14-966] [Medline: 25231062]
33. Top Ten Reviews. Fitbit Flex URL: http://www.toptenreviews.com/electronics/health/pg-3/ [accessed 2018-10-30] [WebCite Cache ID 6xm7WJBgN]
34. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent Validity of Wearable Activity Trackers Under Free-Living Conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
35. 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]
36. Schneider M, Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. BMC Res Notes 2016 Sep 21;9(1):448 [FREE Full text] [doi: 10.1186/s13104-016-2253-6] [Medline: 27655477]
37. Romashkin A. McMaster University. 2014 Aug. Sleep, physical activity, and cardiometabolic health in overweight/obese adolescents URL: https://macsphere.mcmaster.ca/bitstream/11375/16359/2/Romashkin_Anna_2014_September_Master%20of%20Science.pdf [accessed 2018-10-30] [WebCite Cache ID 6xm7joRBd]
38. Schaefer SE, Ching CC, Breen H, German JB. Wearing, Thinking, and Moving: Testing the Feasibility of Fitness Tracking with Urban Youth. American Journal of Health Education 2016 Jan 08;47(1):8-16. [doi: 10.1080/19325037.2015.1111174]
39. Schaefer SE, Van Loan M, German JB. A feasibility study of wearable activity monitors for pre-adolescent school-age children. Prev Chronic Dis 2014;11:E85 [FREE Full text] [doi: 10.5888/pcd11.130262] [Medline: 24854236]
40. Slootmaker SM, Chinapaw MJM, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and Internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]
41. Howie EK, Straker LM. Rates of attrition, non-compliance and missingness in randomized controlled trials of child physical activity interventions using accelerometers: A brief methodological review. J Sci Med Sport 2016 Oct;19(10):830-836. [doi: 10.1016/j.jsams.2015.12.520] [Medline: 26874648]
42. Kerner C, Goodyear V. The Motivational Impact of Wearable Healthy Lifestyle Technologies: A Self-determination Perspective on Fitbits With Adolescents. American Journal of Health Education 2017 Jul 28;48(5):287-297 [FREE Full text] [doi: 10.1080/19325037.2017.1343161]
43. Sullivan AN, Lachman ME. Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
44. Ali MM, Amialchuk A, Heiland FW. Weight-related behavior among adolescents: the role of peer effects. PLoS One 2011;6(6):e21179 [FREE Full text] [doi: 10.1371/journal.pone.0021179] [Medline: 21731665]
45. Salvy S, Roemmich JN, Bowker JC, Romero ND, Stadler PJ, Epstein LH. Effect of peers and friends on youth physical activity and motivation to be physically active. J Pediatr Psychol 2009 Mar;34(2):217-225 [FREE Full text] [doi: 10.1093/jpepsy/jsn071] [Medline: 18617572]
46. Camacho-Miñano MJ, LaVoi NM, Barr-Anderson DJ. Interventions to promote physical activity among young and adolescent girls: a systematic review. Health Educ Res 2011 Dec;26(6):1025-1049. [doi: 10.1093/her/cyr040] [Medline: 21680763]
47. Staiano AE, Calvert SL. Exergames for Physical Education Courses: Physical, Social, and Cognitive Benefits. Child Dev Perspect 2011 Jun;5(2):93-98 [FREE Full text] [doi: 10.1111/j.1750-8606.2011.00162.x] [Medline: 22563349]
48. Chin AP, Jacobs WM, Vaessen EP, Titze S, van Mechelen W. The motivation of children to play an active video game. J Sci Med Sport 2008 Apr;11(2):163-166. [doi: 10.1016/j.jsams.2007.06.001] [Medline: 17706461]
49. Staiano AE, Abraham AA, Calvert SL. Motivating effects of cooperative exergame play for overweight and obese adolescents. J Diabetes Sci Technol 2012 Jul 01;6(4):812-819 [FREE Full text] [doi: 10.1177/193229681200600412] [Medline: 22920807]
50. Howe K, Suharlim C, Ueda P, Howe D, Kawachi I, Rimm EB. Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study. BMJ 2016 Dec 13;355:i6270 [FREE Full text] [doi: 10.1136/bmj.i6270] [Medline: 27965211]
51. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on Physical Activity: Study and Implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
52. Belton S, O'Brien W, Issartel J, McGrane B, Powell D. Where does the time go? Patterns of physical activity in adolescent youth. J Sci Med Sport 2016 Nov;19(11):921-925. [doi: 10.1016/j.jsams.2016.01.008] [Medline: 26897391]
53. U.S. Department of Health and Human Services. 2008 Physical Activity Guidelines for Americans. Hyattsville, MD: Department of Health and Human Services; 2008.
54. Howie EK, Olds T, McVeigh JA, Abbott RA, Straker L. It's A-bout Time: Detailed Patterns of Physical Activity in Obese Adolescents Participating in a Lifestyle Intervention. J Phys Act Health 2015 Nov;12(11):1453-1460. [doi: 10.1123/jpah.2014-0480] [Medline: 25621487]
55. Treuth MS, Catellier DJ, Schmitz KH, Pate RR, Elder JP, McMurray RG, et al. Weekend and weekday patterns of physical activity in overweight and normal-weight adolescent girls. Obesity (Silver Spring) 2007 Jul;15(7):1782-1788 [FREE Full text] [doi: 10.1038/oby.2007.212] [Medline: 17636097]

56. Dankel SJ, Loenneke JP, Loprinzi PD. Does the fat-but-fit paradigm hold true for all-cause mortality when considering the duration of overweight/obesity? Analyzing the WATCH (Weight, Activity and Time Contributes to Health) paradigm. Prev Med 2016 Feb;83:37-40. [doi: 10.1016/j.ypmed.2015.12.002] [Medline: 26687100]

Abbreviations

BMI: body mass index
MVPA: moderate to vigorous physical activity

Edited by G Eysenbach; submitted 28.03.18; peer-reviewed by W Jayawardene, A Sullivan, J Demongeot; comments to author 28.06.18; revised version received 14.08.18; accepted 10.09.18; published 27.11.18.

Please cite as:
Phan TLT, Barnini N, Xie S, Martinez A, Falini L, Abatemarco A, Waldron M, Benton JM, Frankenberry S, Coleman C, Nguyen L, Bo C, Datto GA, Werk LN
Feasibility of Using a Commercial Fitness Tracker as an Adjunct to Family-Based Weight Management Treatment: Pilot Randomized Trial
JMIR Mhealth Uhealth 2018;6(11):e10523
URL: http://mhealth.jmir.org/2018/11/e10523/
doi:10.2196/10523
PMID:30482743

Corrigenda and Addenda

Corrigendum and Editorial Warning Regarding Use of the MMAS-8 Scale (The Health Buddies App as a Novel Tool to Improve Adherence and Knowledge in Atrial Fibrillation Patients: A Pilot Study)

Lien Desteghe[1,2], MSc; Kiki Kluts[1], MSc; Johan Vijgen[2], MD, FESC; Pieter Koopman[2], MD; Dagmara Dilling-Boer[2], MD, FESC; Joris Schurmans[2], MD; Paul Dendale[1,2], MD, PhD, FESC; Hein Heidbuchel[1,3], MD, PhD, FESC, FEHRA

[1]Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

[2]Heart Center Hasselt, Jessa Hospital, Hasselt, Belgium

[3]University of Antwerp and Antwerp University Hospital, Antwerp, Belgium

Corresponding Author:
Lien Desteghe, MSc
Faculty of Medicine and Life Sciences
Hasselt University
Martelarenlaan 42
Hasselt, 3500
Belgium
Phone: 32 11 30 82 26
Fax: 32 11 33 70 38
Email: lien.desteghe@uhasselt.be

Related Article:

Correction of: http://mhealth.jmir.org/2017/7/e98/

(JMIR Mhealth Uhealth 2018;6(11):e12202) doi:10.2196/12202

Authors' Corrigendum

The authors of "The Health Buddies App as a Novel Tool to Improve Adherence and Knowledge in Atrial Fibrillation Patients: A Pilot Study" (JMIR Mhealth Uhealth. 2017;5(7):e98) require corrections in relationship to the use of the Morisky Medication Adherence Scale (MMAS-8).

The following text was omitted from the Acknowledgements and should be added:

> *Use of the MMAS is protected by US copyright laws. Permission for use is required. A license agreement is available from: Donald E Morisky, ScD, ScM, MSPH, Professor, Department of Community Health Sciences, UCLA School of Public Health, 650 Charles E Young Drive South, Los Angeles, CA 90095-1772.*

This statement under the section "Feasibility, Data Collection, and Outcome Measures" requires an advisory warning:

> *A MMAS-8 score of 0-5 indicates a low adherence, 6-7 is medium adherence, and a score of 8 represents a highly adherent patient.*

The new text, which includes the warning, reads as follows:

> *A MMAS-8 score of 0-5 indicates a low adherence, 6-7 is medium adherence, and a score of 8 represents a highly adherent patient. This MMAS-8 scoring and coding criteria is incorrect, and, if used, would invalidate the MMAS-8 results and potentially put patients at risk of harm.*

Two additional references should be cited after reference 17 in the following sentence:

> *First, the self-reported 8-item Morisky medication adherence scale (MMAS-8) was used to get an idea about the adherence level from the viewpoint of the patient [17].*

- Krousel-Wood M, Islam T, Webber LS, Re RN, Morisky DE, Muntner P. New medication adherence scale versus pharmacy fill rates in seniors with hypertension. Am J Manag Care 2009 Jan;15(1):59-66.
- Morisky DE, DiMatteo MR. Improving the measurement of self-reported medication nonadherence: response to authors. J Clin Epidemiol 2011 Mar;64(3):255-7; discussion 258.

These will become references 18 and 19, respectively, and all subsequent references will be renumbered accordingly.

The correction will appear in the online version of the paper on the JMIR website on November 30, 2018, 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 also has been resubmitted to those repositories.

Editorial Notice

Authors and journal were forced to publish this correction due to legal threats by Steven Trubow and Donald Morisky from the company MMAS Research LLC, the copyright holder of the instrument. This is unfortunately not an isolated case, as the developers of this scale are known to comb the literature and ask those who used the scale for research to pay for a retroactive license which may cost thousands or tens of thousands of dollars, and to add references to their work [1].

The Committee on Publication Ethics (COPE) has recently discussed the ethics of this type of behavior by copyright holders of scales ("holding authors to ransom in this way") and recommends to emphasize "the fact that this is not good for the advancement of scientific knowledge or in the public interest" [2]. As open access and open science publisher, JMIR Publications couldn't agree more and we remind our authors of our policies and preference for public and free availability of research tools, including questionnaires [3]. We actively discourage use of MMAS and other instruments which are not available under a Creative Commons Attribution license, and encourage our authors to use or develop/validate new instruments. We are also hereby issuing a special call for papers for short paper instruments or electronic tools licensed under Creative Commons or available under an Open Source license that can be used instead of MMAS to measure medication adherence, and will waive the article submission fee for such development and validation papers describing new instruments that can be used as a free alternative to MMAS.

References

1. Marcus A. Pay up or retract? Survey creator's demands for money rile some health researchers. Science 2017 Sep 13;15(3):169-182. [doi: 10.1126/science.aap9445]
2. Committee on Publication Ethics (COPE). License for a Published Scale. 2018. URL: https://publicationethics.org/case/licence-published-scale [accessed 2018-11-29]
3. JMIR Publications. What is your policy regarding access to critical research tools and instruments (eg, questionnaires)?. 2018. URL: https://jmir.zendesk.com/hc/en-us/articles/360000547811 [accessed 2018-11-29]

Edited by G Eysenbach; submitted 13.09.18; this is a non–peer-reviewed article;accepted 28.09.18; published 30.11.18.

Please cite as:
Desteghe L, Kluts K, Vijgen J, Koopman P, Dilling-Boer D, Schurmans J, Dendale P, Heidbuchel H
Corrigendum and Editorial Warning Regarding Use of the MMAS-8 Scale (The Health Buddies App as a Novel Tool to Improve Adherence and Knowledge in Atrial Fibrillation Patients: A Pilot Study)
JMIR Mhealth Uhealth 2018;6(11):e12202
URL: https://mhealth.jmir.org/2018/11/e12202/
doi:10.2196/12202
PMID:30578241

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org