JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 8 (2020), Issue 7 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Original Papers

Reviews

Original Paper

Validation of a Mobile Version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form: An Observational Randomized Crossover Trial

Jingyi Hou[1*], MD; Qingyue Li[1*], MD; Menglei Yu[1], MD; Fangqi Li[1], MD; Yiyong Tang[1], MD; Yi Long[1], MD; Yamuhanmode Alike[1], MD; Yuanhao Zhang[1], MD; Maslah Idiris Ali[1], MD; Congda Zhang[1], MD; Weiping Li[1], MD; Rui Yang[1], MD

Department of Orthopedics, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Rui Yang, MD
Department of Orthopedics
Sun Yat-sen Memorial Hospital
Sun Yat-sen University
107 Yan Jiang Rd W
Guangzhou, 510120
China
Phone: 86 13694200667
Email: yangr@mail.sysu.edu.cn

Abstract

Background: The American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES) questionnaire is an effective tool for evaluating shoulder joint function. The development and usage of a mobile version of the ASES questionnaire has the potential to save time, money, and effort.

Objective: The aim of this study is to assess the equivalence between the paper and mobile versions of the ASES questionnaire and their acceptability among patients.

Methods: The paper and mobile versions of the ASES questionnaire were used to evaluate the shoulder joint function of 50 patients with shoulder pain. This study included patients from the shoulder clinic of Sun Yat-sen Memorial Hospital. The intraclass correlation coefficient (ICC) and Bland-Altman method were used to evaluate the agreement (reliability) of the scores obtained by the two methods (paper versus mobile).

Results: Of the 50 patients recruited from March 2018 to May 2019, 46 (92%) completed the study. There was a high agreement between the paper and mobile versions of the ASES questionnaire (ICC=0.979, 95% CI 0.943-0.987; P<.001). The mean difference between the scores of the mobile and paper versions was 1.0, and only 1/46 (2%) had a difference greater than the minimal clinically important difference of 12 points. About 75% of patients preferred the mobile version to the paper version.

Conclusions: Our study shows that the mobile version of the ASES questionnaire is comparable to the paper version, and has a higher patient preference. This could prove to be a useful tool for epidemiological studies and patient follow-up over longer periods of time.

(JMIR Mhealth Uhealth 2020;8(7):e16758) doi:10.2196/16758

KEYWORDS

ASES; ePROM; smartphone

Introduction

Shoulder pain is the third most common musculoskeletal problem and it can result in an inability to work and high medical costs [1]. It affects up to one-third of the general population, and is especially prevalent among the elderly [2]. When treating shoulder pain, we need to get information from the patients' perspective to assess the severity of symptoms and the level of disability. The American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form (ASES) is

a patient-reported outcome measure (PROM); it is widely used in sports medicine to evaluate shoulder function, and applied both in clinical research and clinical practice [3-8]. It includes two parts: the patient self-report section and the doctor's evaluation section. The patient self-report section has been validated in many languages and is considered to be a reliable and valid evaluation tool [3,6,7,9-14].

With the popularity of smartphones and the development of patient-focused software, the standard PROM is increasingly shifting from conventional paper and pen toward electronic administration of PROMs (ePROMs). Advantages of ePROMs include ease of use, reduced time of filling in the questionnaire, ease of data collection, high-quality data, reduced data attrition, reduced missing items, and improved patient compliance [15-19].

Some studies have compared the mobile and paper versions of PROM. Although most studies mention equivalence in scores, some studies have shown nonequivalence in scores [20], leading to the conclusion that simply digitizing existing PROMs without reliability testing cannot assure the reliability of ePROMs.

Our research group developed a mobile version of the ASES questionnaire. The aim of this study is to test the equivalence between the paper and mobile versions of the ASES questionnaire and their acceptability among patients.

Methods

Design

This was an observational randomized crossover trial. Participants completed both versions on the same day.

Study Participants

Participants were chosen among patients of the shoulder clinic of Sun Yat-sen Memorial Hospital, Sun Yat-sen University. Inclusion criteria consisted of patients with shoulder pain, aged 18 years and above, and good written communication skills in Chinese. Exclusion criteria included limitations in understanding the Chinese language, difficulty in operating a touch screen device, a mental status that prevented the completion of the survey, or an unwillingness to participate. Ethical approval for this research was provided by the Ethics Committees of the Sun Yat-sen Memorial Hospital (SYSEC-KY-KS-2019-059).

ASES Questionnaire

The self-report section of the ASES questionnaire is divided into two parts: (1) pain score and (2) daily activities. The total ASES score is derived from both parts. The pain score was obtained using the Visual Analogue Scale (VAS), which ranges from 0 (“No pain”) to 10 (“Worst pain”). For assessing the activities of daily living (ADL), 10 items are presented (Table 1) and graded on a 4-point ordinal scale. Scores ranged from 0 (“Unable to do”) to 3 (“Not difficult”). A weighted average was taken of the cumulative ADL score and the pain score, and this was merged into a total score. The formula is the following: ASES score = 5 × ([10 - ASES pain VAS] + ASES cumulative ADL score/3) [3].

Table 1. Patient self-evaluation: activities of daily living questionnaire[a].

Activity	Right arm	Left arm
1. Put on a coat	0 1 2 3	0 1 2 3
2. Sleep on your painful or affected side	0 1 2 3	0 1 2 3
3. Wash back/do up bra in the back	0 1 2 3	0 1 2 3
4. Manage toileting	0 1 2 3	0 1 2 3
5. Comb hair	0 1 2 3	0 1 2 3
6. Reach a high shelf	0 1 2 3	0 1 2 3
7. Lift 10 pounds above shoulder	0 1 2 3	0 1 2 3
8. Throw a ball overhand	0 1 2 3	0 1 2 3
9. Do usual work. List:	0 1 2 3	0 1 2 3
10. Do usual sport. List:	0 1 2 3	0 1 2 3

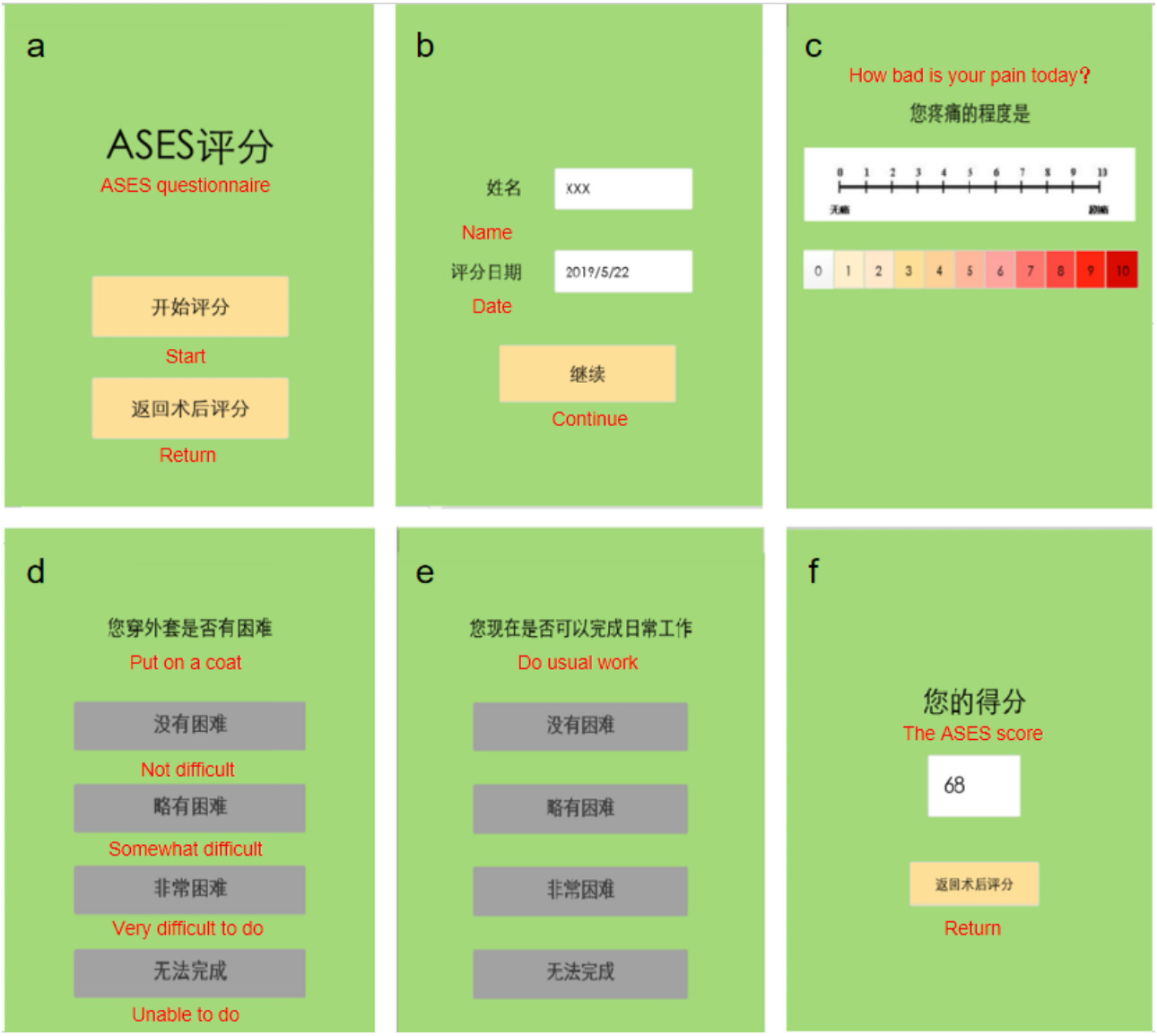
[a]Patients circle the number in the box that indicates their ability to do the activity listed: 0=Unable to do; 1=Very difficult to do; 2=Somewhat difficult; 3=Not difficult.

Mobile Version

The software of the mobile version of the ASES questionnaire was developed by our group. It can run on various operating systems including Android, iOS, and Windows. The major change to the mobile version of the ASES was that only one question was shown per page. Below each question, several options are placed and the screen changes to the next question when an option is selected. There are 11 options on a scale from 0 to 10 on the VAS page. As the patients may remember their answers to the first questionnaire and this could affect the results of the second questionnaire, the ADL items in the mobile version are set to appear randomly, with the options given in reverse order (ie, from 3 to 0). Once the patient clicked on an answer, the questionnaire automatically jumps to the next question, removing the possibility of unintentionally overlooking

questions. For incorrectly filled questions, the mobile version allows the patient to go back and modify the answer. If the completion of the questionnaire is interrupted, the patient can retrieve it and continue without losing the previously entered information. Patients could log in to the software and fill out the questionnaire using an account and password given by the researchers. Once finished, the score would be displayed on the screen and stored on the server. Date and completion time are automatically recorded for each questionnaire. The score obtained using the paper version of the ASES questionnaire is input manually into the device. In this study, the test was administered on an iPad Mini (Apple Inc) with an A5 processor and a 20.1 cm screen (1024×768 screen resolution). Figure 1 depicts a screenshot of the mobile version of the ASES questionnaire.

Figure 1. Screenshots of the iPad Mini screen showing (A) the login screen, (B) the home screen, (C) VAS scores, (D) and (E) ADL scores, and (F) the result screen. ADL: activities of daily living; ASES: American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form; VAS: Visual Analogue Scale.



Study Procedures

With informed consent, patients were asked basic questions about demographics, smartphone familiarity, location, and clinical history; their answers were recorded by the researcher. Based on a computer-generated random number, patients either complete the mobile or the paper version of the questionnaire first. The aim of randomization is to prevent the sequence of trials from influencing the results. Patients were given the second questionnaire 2 hours after completing the first one. Each time a patient filled out the questionnaire, a stopwatch was used to measure the task completion time (accurate to seconds). Finally, the time required to complete the questionnaire was recorded. Patients were asked which questionnaire they preferred and the reason for their choice.

Data Analysis

The same Excel sheet was used to summarize the data from the mobile and paper versions. Data from the mobile version could be automatically transferred to the Excel sheet with our software. For the paper version, two researchers would separately calculate each patient's score and enter it into the clinical data entry form

established with EpiData software (EpiData Association). The data was then transferred to the same Excel sheet. As described above, a stopwatch was used to record the time required to sort the data. The data sorting time for the paper version is an average of the time taken by the two researchers. Descriptive statistics included the mean of aggregate scores, the SD, the mean score difference, and the SD of difference. Correlations between the scores obtained from the paper and mobile versions were assessed using the Pearson correlation coefficient, *r*. The intraclass correlation coefficient (ICC) and Bland-Altman method were used to evaluate the agreement (reliability) between the two methods (paper versus mobile) [21]. To determine whether the score differences between the two versions were clinically significant, we compared the difference with the minimal clinically important difference (MCID), which is the minimum change in score for a patient to notice differences in functional outcomes for the ASES questionnaire. As described in previous studies, 12 points was identified as the MCID for the ASES scores [22,23]. The *t* test was adopted for comparison of the time taken to fill out the form and sort data. Patient preference was studied by a simple content analysis.

Results

Participant Characteristics

From March 2018 to May 2019, 50 patients were enrolled in the study. Of these, 4 patients were excluded as they did not complete the second questionnaire. In total, 46 participants completed the study and were included in the final analysis. Details of the patients are shown in Table 2. The data of the patients in this study were consistent with the baseline population of our clinic, as shown in Multimedia Appendix 1.

Table 2. Details of patients included in the study.

Characteristics	Values
Gender, n (%)	
Male	27 (59)
Female	19 (41)
Age (years), mean (range)	43.87 (18-68)
Diseases, n (%)	
Rotator cuff tear	20 (44)
Frozen shoulder	6 (13)
Impingement syndrome	3 (7)
Instability of shoulder	5 (11)
AC joint arthritis	5 (11)
Superior labrum anterior and posterior (SLAP)	3 (7)
Biceps tendonitis	4 (9)
Prior use of smartphones, n (%)	
Yes	41 (89)
No	5 (11)

Consistency

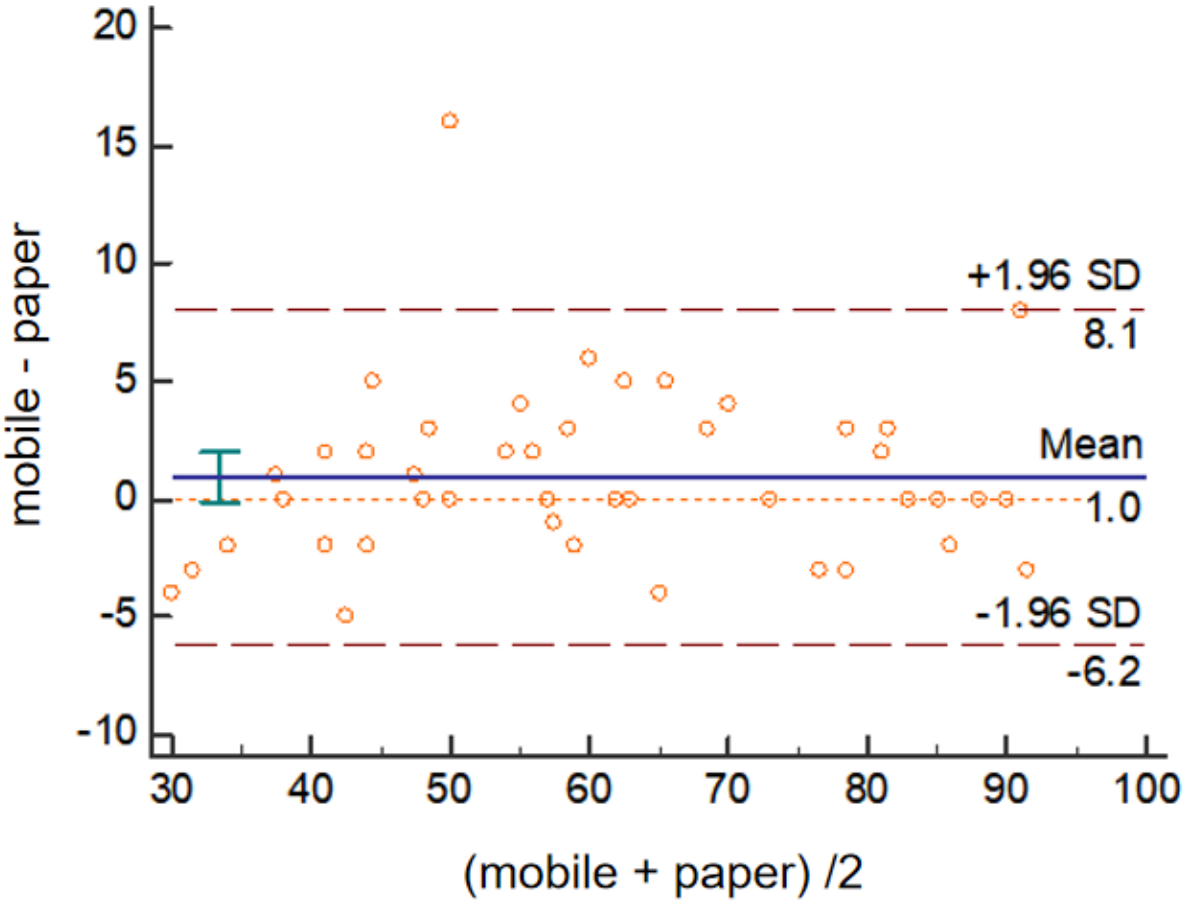
The mean score of the paper version was 60.50 (SD 17.93) and the mean score of the mobile version was 61.46 (SD 18.17). The mean score difference was 0.96 (SD 0.24). The scores of the mobile version were strongly correlated with the scores of the paper version (Pearson $r>0.98$; $P<.001$).

As shown in Table 3, there was minimal variation between the mobile and paper versions. The ICC was 0.979 (95% CI 0.943-0.987; $P<.001$), confirming very good agreement between the versions. A Bland-Altman analysis for the ASES questionnaire showed that the mean difference between scores of the mobile and paper versions was 1.0 of a maximum of 100, and the 95% limits of agreement of the two methods was –6.2 to 8.1 (Figure 2). Only 1 patient of 46 (2%) had a score difference greater than the ASES MCID of 12 points (16 points), and the score difference was within 5 points for 93% (43/46) of cases. These results indicated excellent consistency between these two methods.

Table 3. Consistency analysis of paper and mobile versions.

Comparison	Intraclass correlation coefficient	*P* value	Bland-Altman analysis, mean difference
Paper versus mobile	0.979	<.001	1.0 (-6.2 to 8.1)

Figure 2. Bland-Altman plots between the paper and mobile versions of the ASES questionnaire. ASES: American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form.



Patient Acceptance

Patients were asked for their preference regarding the questionnaire version. Overall, 9 of 46 (19.6%) respondents expressed no preference as they found no obvious difference. Of those who expressed a preference, 28 of 37 (75.7%) preferred using the mobile version over the paper version, while 9 preferred the paper version. Among those who preferred the mobile version, 54% (15/28) found it easy to use and 29% (8/28) thought it was eco-friendly. Although 15.2% (7/46) of the patients had not used a smartphone or tablet before, more than half of that group (4/7) liked using the mobile version. Among patients who preferred the paper version, 8 of 9 thought the mobile version was not intuitive enough to use, as it only showed one question per page.

Time

The paper version of the ASES questionnaire took an average of 1.76 minutes to fill in, while the mobile version took 1.01 minutes. The time taken for data sorting was 3.52 minutes for the paper version and 0.46 minutes for the mobile version (P<.001; Table 4).

Table 4. Time taken to fill out the form and sort data.

Group	Paper version	Mobile version	*P* value
Time taken to fill out the form	1.76 (0.48-3.35) minutes	1.01 (0.42-2.88) minutes	<.001
Time taken to sort data	3.52 minutes	0.46 minutes	<.001

Discussion

PROMs have been widely used in sports medicine, both in clinical research and practice. Compared to the paper version of PROM, ePROM has many advantages. The objective of this study was to confirm the reliability of a mobile version of the ASES questionnaire and its acceptability among patients. The results of this study showed that there is a strong correlation in the ASES score between the mobile and paper versions. The high ICC of 0.979 indicates that the mobile and paper versions of the ASES questionnaire have excellent consistency. The results are in line with previous review articles comparing ePROM validation outcomes [24-28]. In most cases, the difference between the ASES scores of the mobile and paper versions was lower than 5 points. Very few cases (1/46, 2%) had a difference greater than the ASES MCID of 12 points,

indicating that use of the ASES mobile version in place of the paper version would not affect the clinical interpretation of outcomes.

Overall, most participants preferred the mobile version of the ASES questionnaire, because it is eco-friendly and has a user-friendly interface. Even among patients lacking experience using smartphones, more than half preferred the mobile version. This preference may improve cooperation with PROMs.

The mobile version of the ASES questionnaire brings about many advantages. In clinical practice, several PROMs are typically administered simultaneously for shoulder pain [29-32]. A longer amount of time taken to fill in the PROMs may reduce the patient's interest, which could lead to a drop in the quality of responses. In our study, the time taken to fill in the mobile version was significantly less than the paper version (1.01 versus 1.76 minutes; $P<.001$).

One major difference between the two versions is that the mobile version shows the next question only once the current question has been answered, ensuring that no items are missing from the testing process. This improved the reliability of outcomes. This also helps the patient focus only on one question without being distracted. The true benefit of the mobile version can be seen with data processing. The calculation process to obtain the total score of the ASES questionnaire is prone to error and not intuitive. Giving it an advantage over the paper version, the mobile version can automatically record test times, calculate total scores, and export the results to an Excel spreadsheet; it also provides an automated data retrieval system. This entire process is almost instantaneous. Conversely, for the paper version, it takes significantly more time to fill in the questionnaire and to process data. Our results show that handling the data takes 7 times longer in the paper version. The use of the mobile version is clinically significant as it can greatly reduce the workload of medical staff and save time, making clinical data collection easier. Considering the need to use several questionnaires during the same visit when evaluating shoulder function, the advantage of digitalization will be more pronounced. When sorting data for the paper version, the whole process was completed by two researchers separately to avoid calculation errors and transcription errors. We used the verification feature of EpiData to compare the two sets of data, and found no calculation or transcription errors.

Another advantage of the mobile version is that it can be used on mobile devices and be loaded on patients' mobile phones directly, which means it can be completed at home and patients do not need to return to the hospital during long-term follow-up. In addition, the mobile version could automatically remind patients to take the test, thereby reducing dropout rates. In addition, patients can take the self-assessment test when they are feeling their worst to more accurately estimate the ASES score. This can help doctors develop better treatment or recovery plans.

We do acknowledge this study has some limitations. Due to the crossover design of this study, the impact of memory recall cannot be ignored. As the patient's pain score can change rapidly, the washout period could not be too long. In this study, the washout period was 2 hours. In addition, some researchers have tried new methods to overcome these challenges, such as creating two functionally equivalent halves of the item bank [33]. However, for the ASES questionnaire, there are relatively few items, and each item evaluates a specific functional direction, so this method cannot be adopted. Therefore, we explored a new approach. To further reduce the impact of memory recall, we set the ADL items to appear randomly in the mobile version, with the options given in reverse order from the paper version. Since these items are independent of each other and have no logical progression, such a change should be feasible in this situation. In the digitalization of other questionnaires, this method may be used as a reference when the questions are independent of each other. At the same time, the sequence by which the patients were given the questionnaire may affect the time taken to complete the questionnaire, although we found that among the first-filled questionnaires, the mobile version took less time than the paper version. Another limitation is that, since this software interface is in the Chinese language, it is currently only available to Chinese-literate countries and patients. Although we have verified the efficacy of this mobile version, further validation is still required upon the translation of this mobile version to different languages. Furthermore, people in some less-developed areas may not be able to use it due to a lack of advanced technologies.

ePROMs are becoming important in daily practice, and more of them will be used in clinical and research environments in the near future. The outcomes of our study show that the digitalization of the ASES questionnaire is feasible and useful. It reduces the workload of medical workers in collecting and processing data. Additionally, it saves patients' time and allows them to evaluate their condition at any time. This can improve patient compliance and the accuracy of disease assessment, facilitating the implementation of personalized medicine. At the same time, as the difficulty of data collection is reduced, it will be beneficial for the development of real-world studies and predictive medicine.

In conclusion, our study shows that the mobile version of the ASES questionnaire is comparable to the paper version. More patients indicated a preference toward the mobile version of the ASES questionnaire as it is user-friendly and eco-friendly. The mobile version can save time for patients and doctors, and its automated data retrieval system allows for more efficient data collection and analysis. Therefore, this mobile version might prove to be useful in other epidemiological studies and long-term patient follow-up.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (81972067), Guangzhou Science and Technology Program key projects (number 201704020132), and Guangdong Natural Science Foundation (number 2018A030313780).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the patients included in the study and all clinic patients in 2018.
[DOCX File , 17 KB - mhealth_v8i7e16758_app1.docx]

References

1. Boykin RE, Heuer HJ, Vaishnav S, Millett PJ. Rotator cuff disease – basics of diagnosis and treatment. Rheumatol Rep 2010 Mar 08;2(1):1. [doi: 10.4081/rr.2010.e1]
2. Dinnes J, Loveman E, McIntyre L, Waugh N. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. Health Technol Assess 2003;7(29):iii, 1-iii166 [FREE Full text] [doi: 10.3310/hta7290] [Medline: 14567906]
3. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994 Nov;3(6):347-352. [doi: 10.1016/S1058-2746(09)80019-0] [Medline: 22958838]
4. Angst F, Schwyzer H, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S174-S188 [FREE Full text] [doi: 10.1002/acr.20630] [Medline: 22588743]
5. Gagnier JJ, Robbins C, Bedi A, Carpenter JE, Miller BS. Establishing minimally important differences for the American Shoulder and Elbow Surgeons score and the Western Ontario Rotator Cuff Index in patients with full-thickness rotator cuff tears. J Shoulder Elbow Surg 2018 May;27(5):e160-e166. [doi: 10.1016/j.jse.2017.10.042] [Medline: 29307675]
6. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg Am 2005 Sep;87(9):2006-2011. [doi: 10.2106/JBJS.C.01624] [Medline: 16140816]
7. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg 2002;11(6):587-594. [doi: 10.1067/mse.2002.127096] [Medline: 12469084]
8. Unger RZ, Burnham JM, Gammon L, Malempati CS, Jacobs CA, Makhni EC. The Responsiveness of Patient-Reported Outcome Tools in Shoulder Surgery Is Dependent on the Underlying Pathological Condition. Am J Sports Med 2019 Jan;47(1):241-247. [doi: 10.1177/0363546517749213] [Medline: 29323925]
9. Celik D, Atalar AC, Demirhan M, Dirican A. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg Sports Traumatol Arthrosc 2013 Sep;21(9):2184-2189. [doi: 10.1007/s00167-012-2183-3] [Medline: 22932692]
10. Goldhahn J, Angst F, Drerup S, Pap G, Simmen BR, Mannion AF. Lessons learned during the cross-cultural adaptation of the American Shoulder and Elbow Surgeons shoulder form into German. J Shoulder Elbow Surg 2008;17(2):248-254. [doi: 10.1016/j.jse.2007.06.027] [Medline: 18207432]
11. Knaut LA, Moser ADL, Melo SDA, Richards RR. Translation and cultural adaptation to the portuguese language of the American Shoulder and Elbow Surgeons Standardized Shoulder assessment form (ASES) for evaluation of shoulder function. Rev Bras Reumatol 2010;50(2):176-189. [doi: 10.1590/s0482-50042010000200007] [Medline: 21125153]
12. Padua R, Padua L, Ceccarelli E, Bondi R, Alviti F, Castagna A. Italian version of ASES questionnaire for shoulder assessment: cross-cultural adaptation and validation. Musculoskelet Surg 2010 May;94 Suppl 1:S85-S90. [doi: 10.1007/s12306-010-0064-9] [Medline: 20383686]
13. Yahia A, Guermazi M, Khmekhem M, Ghroubi S, Ayedi K, Elleuch MH. Translation into Arabic and validation of the ASES index in assessment of shoulder disabilities. Ann Phys Rehabil Med 2011 Mar;54(2):59-72 [FREE Full text] [doi: 10.1016/j.rehab.2010.12.002] [Medline: 21354384]

14. Piitulainen K, Paloneva J, Ylinen J, Kautiainen H, Häkkinen A. Reliability and validity of the Finnish version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section. BMC Musculoskelet Disord 2014 Aug 11;15:272 [FREE Full text] [doi: 10.1186/1471-2474-15-272] [Medline: 25112412]
15. Sindhu BS, Shechtman O, Tuckey L. Validity, reliability, and responsiveness of a digital version of the visual analog scale. J Hand Ther 2011;24(4):356-363. [doi: 10.1016/j.jht.2011.06.003] [Medline: 21820864]
16. Hauth F, Bizu V, App R, Lautenbacher H, Tenev A, Bitzer M, et al. Electronic Patient-Reported Outcome Measures in Radiation Oncology: Initial Experience After Workflow Implementation. JMIR mHealth uHealth 2019 Jul 24;7(7):e12345 [FREE Full text] [doi: 10.2196/12345] [Medline: 31342906]
17. Leece P, Bhandari M, Sprague S, Swiontkowski MF, Schemitsch EH, Tornetta P, et al. Internet versus mailed questionnaires: a controlled comparison (2). J Med Internet Res 2004 Oct 29;6(4):e39 [FREE Full text] [doi: 10.2196/jmir.6.4.e39] [Medline: 15631963]
18. Mangunkusumo RT, Moorman PW, Van Den Berg-de Ruiter AE, Van Der Lei J, De Koning HJ, Raat H. Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. J Adolesc Health 2005 Jan;36(1):70.e1-70.e6. [doi: 10.1016/j.jadohealth.2004.02.020] [Medline: 15661601]
19. Bälter KA, Bälter O, Fondell E, Lagerros YT. Web-based and mailed questionnaires: a comparison of response rates and compliance. Epidemiology 2005 Jul;16(4):577-579. [doi: 10.1097/01.ede.0000164553.16591.4b] [Medline: 15951679]
20. Buchanan T, Ali T, Heffernan TM, Ling J, Parrott AC, Rodgers J, et al. Nonequivalence of on-line and paper-and-pencil psychological tests: the case of the prospective memory questionnaire. Behav Res Methods 2005 Feb;37(1):148-154. [doi: 10.3758/bf03206409] [Medline: 16097355]
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
22. Werner BC, Wong AC, Mahony GT, Craig EV, Dines DM, Warren RF, et al. Causes of poor postoperative improvement after reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2016 Aug;25(8):e217-e222. [doi: 10.1016/j.jse.2016.01.002] [Medline: 27068387]
23. Wong SE, Zhang AL, Berliner JL, Ma CB, Feeley BT. Preoperative patient-reported scores can predict postoperative outcomes after shoulder arthroplasty. J Shoulder Elbow Surg 2016 Jun;25(6):913-919. [doi: 10.1016/j.jse.2016.01.029] [Medline: 27039669]
24. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 07;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]
25. van Kollenburg RAA, de Bruin DM, Wijkstra H. Validation of the Electronic Version of the International Index of Erectile Function (IIEF-5 and IIEF-15): A Crossover Study. J Med Internet Res 2019 Jul 02;21(7):e13490 [FREE Full text] [doi: 10.2196/13490] [Medline: 31267983]
26. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
27. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429. [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
28. Geerards D, Pusic A, Hoogbergen M, van der Hulst R, Sidey-Gibbons C. Computerized Quality of Life Assessment: A Randomized Experiment to Determine the Impact of Individualized Feedback on Assessment Experience. J Med Internet Res 2019 Jul 11;21(7):e12212 [FREE Full text] [doi: 10.2196/12212] [Medline: 31298217]
29. Kim Y, Jin H, Lee H, Cho H, Lee W, Jang H. Is It Safe to Inject Corticosteroids Into the Glenohumeral Joint After Arthroscopic Rotator Cuff Repair? Am J Sports Med 2019 Jun;47(7):1694-1700. [doi: 10.1177/0363546519843910] [Medline: 31084488]
30. Ranalletta M, Rossi LA, Bongiovanni SL, Tanoira I, Elizondo CM, Maignon GD. Corticosteroid Injections Accelerate Pain Relief and Recovery of Function Compared With Oral NSAIDs in Patients With Adhesive Capsulitis: A Randomized Controlled Trial. Am J Sports Med 2016 Feb;44(2):474-481. [doi: 10.1177/0363546515616238] [Medline: 26657263]
31. Privitera DM, Sinz NJ, Miller LR, Siegel EJ, Solberg MJ, Daniels SD, et al. Clinical Outcomes Following the Latarjet Procedure in Contact and Collision Athletes. J Bone Joint Surg Am 2018 Mar 21;100(6):459-465. [doi: 10.2106/JBJS.17.00566] [Medline: 29557861]
32. Pogorzelski J, Horan MP, Hussain ZB, Vap A, Fritz EM, Millett PJ. Subpectoral Biceps Tenodesis for Treatment of Isolated Type II SLAP Lesions in a Young and Active Population. Arthroscopy 2018 Feb;34(2):371-376. [doi: 10.1016/j.arthro.2017.07.021] [Medline: 28899638]
33. Brodey BB, Gonzalez NL, Elkin KA, Sasiela WJ, Brodey IS. Assessing the Equivalence of Paper, Mobile Phone, and Tablet Survey Responses at a Community Mental Health Center Using Equivalent Halves of a 'Gold-Standard' Depression Item Bank. JMIR Ment Health 2017 Sep 06;4(3):e36 [FREE Full text] [doi: 10.2196/mental.6805] [Medline: 28877861]

Abbreviations

ASES: American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form
ePROM: electronic patient-reported outcome measure
ICC: intraclass correlation coefficient
PROM: patient-reported outcome measure
MCID: minimal clinically important difference
VAS: Visual Analogue Scoring

Edited by G Eysenbach; submitted 26.10.19; peer-reviewed by T Jamieson, A Korchi; comments to author 28.11.19; revised version received 26.01.20; accepted 03.06.20; published 17.07.20.

Please cite as:
Hou J, Li Q, Yu M, Li F, Tang Y, Long Y, Alike Y, Zhang Y, Ali MI, Zhang C, Li W, Yang R
Validation of a Mobile Version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form: An Observational Randomized Crossover Trial
JMIR Mhealth Uhealth 2020;8(7):e16758
URL: http://mhealth.jmir.org/2020/7/e16758/
doi:10.2196/16758
PMID:32706731

Review

Mobile Phone Technologies in the Management of Ischemic Heart Disease, Heart Failure, and Hypertension: Systematic Review and Meta-Analysis

Praveen Indraratna[1,2], MBBS, FRACP; Daniel Tardo[3,4], MBBS, BAsc; Jennifer Yu[1,2], MBBS, BSci, FRACP; Kim Delbaere[5,6], BA, MPT, MEd, PhD; Matthew Brodie[5,7], BE, PhD; Nigel Lovell[7], BE (Hons), PhD; Sze-Yuan Ooi[1,2], MD, MBBS, FRACP

[1]Department of Cardiology, Prince of Wales Hospital, Sydney, Australia

[2]Prince of Wales Clinical School, The University of New South Wales, Sydney, Australia

[3]Department of Medicine, St Vincent's Hospital, Sydney, Australia

[4]Faculty of Medicine, The University of Notre Dame, Sydney, Australia

[5]Falls, Balance and Injury Research Centre, Neuroscience Research Australia, Sydney, Australia

[6]School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia

[7]Graduate School of Biomedical Engineering, The University of New South Wales, Sydney, Australia

Corresponding Author:
Praveen Indraratna, MBBS, FRACP
Department of Cardiology
Prince of Wales Hospital
Barker Street, Randwick
Sydney
Australia
Phone: 61 293822222
Email: praveen@unsw.edu.au

Abstract

Background: Cardiovascular disease (CVD) remains the leading cause of death worldwide. Mobile phones have become ubiquitous in most developed societies. Smartphone apps, telemonitoring, and clinician-driven SMS allow for novel opportunities and methods in managing chronic CVD, such as ischemic heart disease, heart failure, and hypertension, and in the conduct and support of cardiac rehabilitation.

Objective: A systematic review was conducted using seven electronic databases, identifying all relevant randomized control trials (RCTs) featuring a mobile phone intervention (MPI) used in the management of chronic CVD. Outcomes assessed included mortality, hospitalizations, blood pressure (BP), and BMI.

Methods: Electronic data searches were performed using seven databases from January 2000 to June 2019. Relevant articles were reviewed and analyzed. Meta-analysis was performed using standard techniques. The odds ratio (OR) was used as a summary statistic for dichotomous variables. A random effect model was used.

Results: A total of 26 RCTs including 6713 patients were identified and are described in this review, and 12 RCTs were included in the meta-analysis. In patients with heart failure, MPIs were associated with a significantly lower rate of hospitalizations (244/792, 30.8% vs 287/803, 35.7%; n=1595; OR 0.77, 95% CI 0.62 to 0.97; P=.03; I^2=0%). In patients with hypertension, patients exposed to MPIs had a significantly lower systolic BP (mean difference 4.3 mm Hg; 95% CI −7.8 to −0.78 mm Hg; n=2023; P=.02).

Conclusions: The available data suggest that MPIs may have a role as a valuable adjunct in the management of chronic CVD.

(JMIR Mhealth Uhealth 2020;8(7):e16695) doi:10.2196/16695

KEYWORDS

mobile phone; text messaging; telemedicine; myocardial ischemia; heart failure; hypertension

Introduction

Background

Cardiovascular disease (CVD) is the leading cause of death worldwide [1] and is a leading cause of hospital readmission. With an aging population and a rising prevalence of obesity in developed countries, the physical and economic burden of CVD will only increase. Repeat cardiac events contribute significantly to the burden of disease [2], and reductions of more than 80% of such events may be achieved through secondary prevention [3]. Secondary prevention relies upon monitoring and control of modifiable risk factors such as hypertension, dyslipidemia, and obesity. This is achieved through (1) patient education and empowerment, (2) the prescription of and compliance with optimal pharmacotherapy, (3) lifestyle modification, and (4) the early identification of clinical deterioration.

Traditional cardiac rehabilitation (CR) programs are the embodiment of these principles, offering support, tailored education, and supervised exercise [4]. They are proven to reduce repeat events and mortality, and their adoption is strongly advocated in the current guidelines for the management of acute coronary syndromes [5]. However, globally, CR participation rates are, at best, 30% [6]. The reasons for this are varied, but importantly in the context of this review, they include accessibility, convenience, patient availability, and a fear of group-based settings [7]. Therefore, novel strategies are required.

The outpatient management of heart failure is ideally performed using a multidisciplinary approach, including nurse-led medication titration services, which reduce readmission rates and mortality. However, these services are heavily dependent upon the limited resources of skilled nursing staff. Again, a novel approach to outpatient management is required.

The telemedicine care process involves using communication networks to deliver health care services and move patient information between locations. Literature reviews have underlined several advantages of using telemedicine to reduce inequalities in cardiovascular outcomes [8] and provide improved care for patients with heart failure [9]. Systematic review evidence has found telemedicine as a cost-effective option for treating many chronic conditions [10]. However, not all remote monitoring apps have equal efficacy, and the reported benefits depend on the type of technology used, the presence of organizational support, and the level of care provided to control groups [11].

Mobile phones (particularly smartphones) provide new opportunities to remotely care for patients with cardiac conditions. Traditionally, telemedicine required the provision of home-based specialized monitoring equipment to patients. Smartphones, mobile phones, and wearable technology, however, offer tremendous potential for monitoring health through phone calls, text messages, data recording, highly portable peripheral devices, and activity monitoring, which may find utility for novel models of health care delivery that are cost-effective, accessible, and patient-centric. Mobile phones are ubiquitous, and the recent landmark Nature publication [12] highlights the staggering potential for big data to encourage people to be physically active and to influence health policy. However, concerns exist regarding the use of and inconsistencies in unvalidated smartphone-based interventions for health research [13]. Mobile phone apps have been used to target individuals with ischemic heart disease, cardiac failure, and hypertension and in those undergoing CR; however, gaps exist between the reported potential for mobile phone technology to transform health care services and current clinical practice. The evidence for improved health outcomes now needs to be systematically assessed, and consequently, no clear guidelines exist on the use of these new technologies in clinical practice.

Objectives

The purpose of this study was to systematically review and meta-analyze the evidence for mobile phone technology in the management of cardiac conditions according to the following questions: (1) What are the specific interventions available and do they involve an interface whereby the clinician can use the data to intervene (henceforth referred to as a *back-end*)? (2) Can mobile phone technologies improve patient outcomes with respect to mortality and hospitalizations? and (3) Can mobile phone technologies reduce risk factors for cardiac events, specifically medication compliance and hypertension?

Methods

Data Search Strategy

Electronic data searches were performed using Ovid MEDLINE, PubMed, EMBASE, Database of Abstracts of Review of Effects, American College of Physicians Journal Club, National Health Service Economic Evaluation Database, and Cochrane Database of Systematic Reviews from January 2000 to June 2019. A combination of search terms were used to maximize sensitivity: *mobile applications*, *cell phones*, *smart phones*, *mobile phones*, and *text messaging* were combined with *cardiovascular disease*, *heart failure*, *ischemic heart disease*, *acute coronary syndrome*, *myocardial infarction*, *cardiac rehabilitation*, and *hypertension* as either Medical Subject Headings (MeSH) terms or key words. Full texts of selected publications were retrieved following review of all abstracts. A manual review of the reference lists of each relevant manuscript was performed to identify further results. Recording of results followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. Bias assessments were based on the Cochrane Risk of Bias assessment tool (RoB 2).

Selection Criteria

Studies were considered eligible for this systematic review if a mobile phone app or text messaging (used interchangeably henceforth with SMS) were used in a randomized control trial (RCT) in the management of ischemic heart disease, cardiac failure, or hypertension in adult patients. Studies limited to the management of obesity, dyslipidemia, diabetes, sedentary lifestyle, and smoking cessation in patients without CVD were not included. Studies examining a combined population such as patients with either dyslipidemia or hypertension were not included unless the groups were analyzed separately. Where telephone calls were the primary intervention, the study was

excluded as such an intervention could occur using landline telephones. Similarly, web-based interventions were not included as they could occur using a computer or tablet. Studies that recruited less than 10 subjects in each arm were excluded. Qualitative studies or those with no clinical endpoints were not included. Non-English language results were not included. Abstracts, case reports, editorials, and conference presentations were excluded.

Data Extraction and Critical Appraisal

Article screening was performed by reviewing abstracts (by PI). Clinical outcome data were extracted from article text, tables, and figures independently by 2 researchers (PI and DT) from articles where it was available in the text, tables, figures, or supplementary material. Any discrepancies were resolved after the collaborative review. The final results were reviewed by all authors.

Meta-Analysis

Meta-analysis was performed by combining event rates of dichotomous variables and using the supplied means and standard deviations for continuous variables. The odds ratio (OR) was used as a summary statistic for dichotomous variables. A random effect model was used. Chi-square tests were used to study heterogeneity between trials. The I^2 statistic was used to estimate the percentage of total variation across studies due to heterogeneity rather than chance. An I^2 value of greater than 50% was considered to represent substantial heterogeneity. Subgroup analyses were not possible due to the lack of patient-level data. All *P* values were two-sided. All statistical analyses were conducted with Review Manager Version 5.3 (Cochrane Collaboration, Software Update).

Results

Quantity and Quality of Studies

Using the search strategy described earlier, 306 unique references were retrieved (465 before deduplication). The screening process is summarized in the PRISMA chart in Figure 1. A total of 26 references were included in the final systematic review, comprising a total of 6713 patients. These are summarized in Tables 1 and 2 and in Multimedia Appendices 1 and 2.

Of the 26 RCTs, the target population was ischemic heart disease in 6 studies [14-19], heart failure in 6 [20-25], hypertension in 6 [26-31], and CR in 8 [32-38]. Of these 8, 1 paper included 2 separate trials [35].

The studies were performed in 17 different countries, including 9 studies from Europe, 6 from North America, 5 from Australia/New Zealand, 4 from Asia, and 1 each from Africa and South America. In total, 10 studies examined a 1-way SMS intervention, 3 examined an interactive SMS intervention—where participants could reply, 4 examined an automatic telemonitoring system in which metrics were transmitted to the research team without the need for manual entry, 6 examined a manual telemonitoring system, and 3 studies examined a smartphone app that did not fit the previous criteria. Moreover, 10 studies included a *back-end*, whereby participants were able to transmit data that were viewable by the researchers or clinicians.

Blinding of the participants was not possible in any of the 26 identified studies. This was expected given the nature of the interventions. Only 8 studies featured blinding of the researchers or outcome assessors, and only 7 studies were adjudicated as having a low risk of bias (Multimedia Appendix 3).

Figure 1. Search strategy and results for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method.

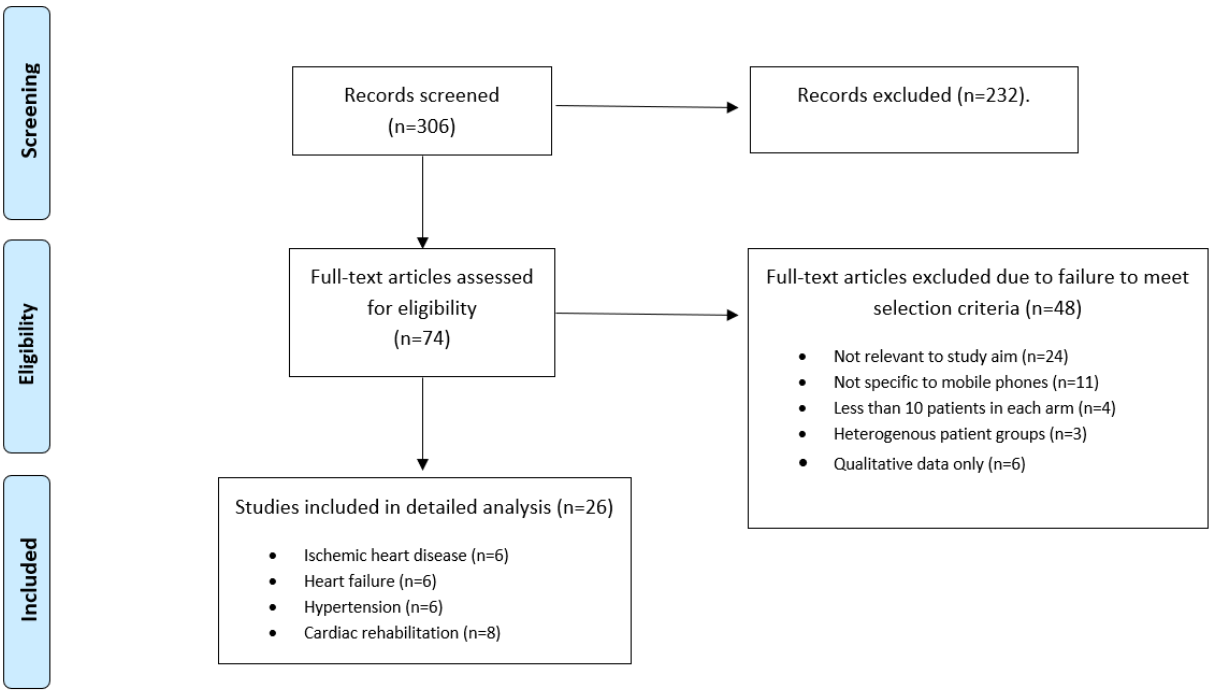


Table 1. Characteristics of the studies included in the systematic review.

Reference	Country	Multicenter	Population	Number of patients (N)	Follow-up (weeks)	Intervention	Back-end
Blasco 2012 [14]	Spain	No	IHD[a]-ACS[b]	203	52	Manual telemonitoring	Yes
Chow 2015 [15]	Australia	No	IHD-ACS and stable CAD[c]	710	26	One-way SMS	No
Fang 2016 [16]	China	No	IHD-stable CAD	271	26	One-way SMS	No
Khonsari 2015 [17]	Malaysia	No	IHD-ACS	62	8	One-way SMS	No
Park 2015 [18]	United States	No	IHD-ACS, stable	90	4.5	Interactive SMS	No
Quilici 2013 [19]	France	No	IHD-ACS	499	4.5	One-way SMS	No
Chen 2019 [20]	China	No	HF[d]	512	26	One-way SMS	No
Dendale 2012 [21]	Belgium	Yes	HF	160	26	Auto telemonitoring	Yes
Koehler 2011 [22]	Germany	Yes	HF	710	52	Auto telemonitoring	Yes
Scherr 2009 [23]	Austria	No	HF	120	26	Manual telemonitoring	Yes
Seto 2012 [24]	Canada	No	HF	100	26	Auto telemonitoring	Yes
Vuorinen 2014 [25]	Finland	No	HF	94	26	Manual telemonitoring	Yes
Bobrow 2016 [26]	South Africa	No	HTN[e]	1372	52	Interactive SMS	No
Kiselev 2012 [27]	Russia	No	HTN	199	52	One-way SMS	No
Logan 2012 [28]	Canada	Yes	HTN	110	52	Auto telemonitoring	Yes
Morawski 2018 [29]	United States	Yes	HTN	411	12	Manual telemonitoring	No
Morikawa 2011 [30]	Japan	No	HTN	41	4	One-way SMS	No
Varleta 2017 [31]	Chile	Yes	HTN	314	26	One-way SMS	No
Bravo-Escobar 2017 [32]	Spain	No	CR[f]-ischemic CM[g]	28	8	App (other)	Yes
Del Rosario 2018 [33]	Australia	No	CR (mixed)	66	12	Manual telemonitoring	No
Maddison 2019 [34]	New Zealand	Yes	IHD-ACS	162	24	App (other)	Yes
Pandey 2017 [35]	Canada	No	CR-ACS	34	52	One-way SMS	No
Pandey 2017 [35]	Canada	No	CR-ACS	50	52	One-way SMS	No
Pfaeffli Dale 2015 [36]	New Zealand	Yes	CR-ACS and stable CAD	123	26	Interactive SMS	No
Piotrowicz 2010 [37]	Poland	No	CR (HF)	152	8	Manual telemonitoring	No
Varnfield 2014 [38]	Australia	No	CR-ACS	120	6	App (other)	Yes

[a]IHD: ischemic heart disease.

[b]ACS: acute coronary syndrome.

[c]CAD: coronary artery disease.

[d]HF: heart failure.

[e]HTN: hypertension.

[f]CR: cardiac rehabilitation.

[g]CM: cardiomyopathy.

Table 2. Endpoints examined in randomized trials of mobile phone technology in cardiovascular disease.

Author	Population	Primary endpoint	Primary result	Major secondary endpoints				
				Mortality	Hospitalization	Blood pressure	BMI	Medication adherence
Blasco 2012 [14]	IHD[a]	Cardiovascular risk	NS[b]	NR[c]	NR	NS	NS	NR
Chow 2015 [15]	IHD	Lipid profile	P[d]	NR	NR	P	P	NR
Fang 2016 [16]	IHD	Medication adherence	P	NR	NR	NR	NR	P1[e]
Khonsari 2015 [17]	IHD	Medication adherence	P	NS	NS	NR	NR	P1
Park 2015 [18]	IHD	Medication adherence	P	NR	NR	NR	NR	P1
Quilici 2013 [19]	IHD	Medication adherence	P	NR	NR	NR	NR	P1
Chen 2019 [20]	HF[f]	Mortality	NS	NS1[g]	P	NR	NR	P
Dendale 2012 [21]	HF	Mortality	P	P1	NS	NR	NR	NR
Koehler 2011 [22]	HF	Mortality	NS	NS1	NS	NR	NR	NR
Scherr 2009 [23]	HF	Mortality	NS	NS1	NS	NR	NR	NR
Seto 2012 [24]	HF	BNP[h]	NS	NR	NR	NR	NR	NR
Vuorinen 2014 [25]	HF	Readmissions	NS	NS	NS1	NR	NR	NR
Bobrow 2016 [26]	HTN[i]	Blood pressure	P	NR	NS	P1	NR	P
Kiselev 2012 [27]	HTN	Blood pressure	P	NR	NR	P1	NS	NR
Logan 2012 [28]	HTN	Blood pressure	P	NR	NR	P1	NR	NR
Morawski 2018 [29]	HTN	Blood pressure	NS	NR	NR	NS1	NR	P
Morikawa 2011 [30]	HTN	Blood pressure	P	NR	NR	P1	NS	NR
Varleta 2017 [31]	HTN	Medication adherence	NS	NR	NR	NS	NR	NS1
Bravo-Escobar 2017 [32]	CR[j]	Physical fitness	NS	NR	NR	NS	NS	NR
Del Rosario 2018 [33]	CR	CR completion rate	P	NR	NR	NS	NS	NR
Maddison 2019 [34]	CR	Physical fitness	NS	NR	NR	NR	NR	NR
Pandey 2017 [35]	CR	Medication adherence	NS	NR	NR	NR	NR	NS1
Pandey 2017 [35]	CR	Lifestyle adherence	NS	NR	NR	NR	NR	NR
Pfaeffli Dale 2015 [36]	CR	Lifestyle adherence	NS	NR	NR	NS	NS	P
Piotrowicz 2010 [37]	CR	Functional status	P	NR	NR	NR	NR	NR

Author	Population	Primary endpoint	Primary result	Major secondary endpoints				
				Mortality	Hospitalization	Blood pressure	BMI	Medication adherence
Varnfield 2014 [38]	CR	CR completion rate	NS	NR	NR	NS	P	NR

[a]IHD: ischemic heart disease.

[b]NS: not significant.

[c]NR: not reported.

[d]P: positive.

[e]P1: positive primary endpoint.

[f]HF: heart failure.

[g]NS1: not significant primary endpoint.

[h]BNP: brain natriuretic peptide.

[i]HTN: hypertension.

[j]CR: cardiac rehabilitation.

Description of Studies

Ischemic Heart Disease

A total of 6 interventions were identified that were targeted at patients with ischemic heart disease, excluding patients who were exclusively recruited from CR. The most commonly assessed primary endpoint was medication adherence, and the results are tabulated in Multimedia Appendix 4.

The largest and most comprehensive study was the pivotal Tobacco, Exercise and Diet Messages (TEXT ME) trial [15]. A total of 352 patients received 4 motivational text messages a week, randomly during daylight hours over a 6-month period. The messages focused on secondary prevention strategies. The primary endpoint was low-density lipoprotein cholesterol levels, which were lower in the intervention group (79 mg/dL [2.04 mmol/L] vs 84 mg/dL [2.17 mmol/L]; *P*=.04). Several other endpoints were examined, and significant improvements in systolic blood pressure (BP; −7.6 mm Hg; *P*<.001), physical activity, and smoking rates (88/339, 25.9% vs 152/354, 42.9%; *P*<.001) were noted. The majority of patients found the intervention motivational, educational, and useful. Medication adherence was not assessed.

A Spanish RCT investigated a 2-way messaging service in patients who had experienced an acute coronary syndrome [14]. Participants were provided with a mobile phone and a sphygmomanometer, a glucose meter, and a lipid meter. After measuring their results, they were entered into an app that transmitted the data to a cardiologist, who remotely monitored the data and responded with recommendations to the participant via text message. The content and nature of these messages were not described. Mean BMI improved in the intervention arm but did not change in the control arm (−0.37 [95% CI –0.08 to 0.04] vs 0.38 kg/m^2 [95% CI –0.11 to 0.85]; *P*=.02). The proportion of patients meeting the BP target was higher in the intervention arm (63/101, 62.3% vs 44/102, 43.1%; *P*=.01), although there was no significant difference in BP between the 2 groups. Medication adherence was not assessed.

Fang et al [16] examined the use of SMS via both standard delivery and in combination with a messaging app called *Micro Letter*, and compared these results with standard care, in a 3-arm RCT [16]. Compliance was assessed using the Morisky Medication Adherence Scale (MMAS-4), and results were presented in a logistic regression analysis. Patients receiving messages via 2 platforms were more compliant than those just receiving standard SMS. Both intervention groups demonstrated superior compliance with the control group.

A small study performed in Malaysia randomized 62 patients with a recent acute coronary syndrome to receive an SMS reminder before scheduled medication times or usual care only, for an 8-week period post discharge [17]. Patients were more likely to have *high compliance*, as measured by a score of 8 on the MMAS-8 scoring scale (20/31, 65% vs 4/31, 13%; *P*<.001).

Park et al [18] randomized patients to receive SMS reminders and education, SMS education only, and usual care. Electronic pill bottles, along with the MMAS score, were used to evaluate compliance. The percentage of correct doses taken for antiplatelets over 30 days was higher in the intervention groups than in the control groups (88% and 87% vs 72%; *P*=.04). However, no significant difference was observed for compliance with statin therapy.

Quilici et al [19] performed a randomized trial of 499 patients to receive daily text messages or standard care and assessed aspirin adherence based on patient interview data and also platelet function. The self-reported rates of nonadherence were significantly different between the intervention group and the control group in both the self-reported (7/250, 2.8% vs 18/249, 7.2%; *P*=.02) and the platelet function testing endpoints (13/250, 5.2% vs 28/249, 11.2%; *P*=.01).

Cardiac Failure

A total of 6 RCTs assessed the efficacy of telemedicine-based interventions in the management of heart failure; 5 out of the 6 studies demonstrated at least one clinical benefit for the intervention, although there was no endpoint that was shown to be consistently improved across all studies. One study demonstrated an improvement in mortality [21]. The results are summarized in Multimedia Appendix 5.

A recent large RCT from China randomized patients with chronic heart failure into 3 arms: structured telephone support

(a single phone call within 30 days of discharge with the opportunity to speak to a nurse during work hours), an SMS-based support system, or a control group [20]. The SMS system consisted of daily educational messages for 10 days and weekly reminder messages thereafter. These messages were automated, not personalized, and could not be replied to. A comparison of the SMS group with the control group demonstrated significantly lower readmission rates as well as higher rates of medication compliance. There was no significant difference in mortality or quality of life. It was not specified how rates of medication noncompliance were measured.

Dendale et al [21] conducted a multicenter study in Belgium that compared patients who were established on an automated monitoring platform including BP, heart rate, and weight with those receiving standard care [21]. Alerts were received by the patient's general practitioner when abnormal parameters were encountered on 2 consecutive days. Despite its small sample size, a statistically significant difference in all-cause mortality was observed (4/80, 5% vs 14/80, 17%; *P*=.01).

The study by Koehler et al [22] comprised 710 patients with severely reduced left ventricular ejection fraction (LVEF) who were randomized 1:1 to receive 24-hour home telemonitoring or usual care. Devices for electrocardiogram (ECG), BP, and body weight were paired via Bluetooth to a personal digital assistant, which transmitted the data via a mobile phone service to a central data monitoring unit. Patients were followed up for a minimum of 12 months (mean 26 months). There was no significant difference in mortality (15% in both groups, 54/354 intervention vs 55/356 in the control group) or hospitalizations, although patients in the intervention arm may have had better quality of life as evidenced by higher SF-36 scores.

The MOBIle TELemonitoring in Heart Failure Patients study compared the use of a manual monitoring system with standard care [23]. Patients were given a mobile phone and asked to measure their weight and BP daily. The results were manually entered into the mobile phone app. The physicians received alerts for measurements that were abnormal. Patients in the intervention group had lower rates of hospitalization, and functional status improved by one New York Heart Association (NYHA) score. When patients were hospitalized, the median length of stay in the intervention group was also shorter than that of the control group (6.5 vs 10 days; *P*=.04). Of note, this trial was prematurely terminated due to a high rate of technical difficulties; 12 out of 54 patients (22%) in the intervention arm were unable to operate the app.

Seto et al [24] recruited 100 patients who were randomized into telemonitoring and control groups. The patients in the telemonitoring group performed BP and weight measurements, which were transmitted via Bluetooth and a smartphone to a data monitoring unit. Patients were required to answer a daily symptom-based questionnaire. Alerts of varying priorities were sent to the treating clinician based on the results. Quality of life scores were improved in the telemonitoring group (Minnesota Living With Heart Failure Questionnaire [MLHFQ]), but there was no difference in overall mortality, hospitalization, or ejection fraction.

Vuorinen et al [25] studied a Finnish cohort of patients with severe heart failure with reduced ejection fraction (<35%), who were symptomatic with a NYHA score of 2 or more, and known to the established outpatient heart failure service. Patients were provided with a mobile phone as well as a sphygmomanometer and a weight scale, and they entered these results via an app, which also contained a symptomatology questionnaire. Results were entered once a week. The authors noted no difference in days spent in hospital for admission and no difference in N-terminal prohormone of brain natriuretic peptide concentrations or LVEF. There was a significantly higher use of resources for patients in the intervention arm, specifically with regard to the number of visits to the clinic.

Hypertension

In total, 6 RCTs examined smartphone apps in the management of patients previously diagnosed with hypertension; 4 of these demonstrated a statistically significant reduction in systolic BP. The results are summarized in Multimedia Appendix 6.

The SMS Text Adherence Support trial randomized 1372 patients to receive SMS information, interactive SMS, or usual care [26]. The follow-up period was 12 months. Patients receiving the information only experienced a mean systolic BP reduction of 2.2 mm Hg (95% CI −4.4 to −0.04, *P*=.04). There was no significant difference in the group that received interactive SMS. Patients in the SMS group were also more likely to be adherent to their medications, with adherence defined as medications correctly taken on over 80% of days (62.1% vs 49%; *P*<.001 for both groups).

Kiselev et al [27] performed a randomized study of 199 patients. The intervention group received SMS reminders to promote medication compliance as well as healthy behaviors, and the control group received standard care. BP fell significantly in both groups; however, the difference was more marked in the intervention group (23.7 mm Hg vs 6.9 mm Hg; *P*=.04). Of note, the baseline BP was significantly higher in the control group, suggesting problems with randomization. Nearly 77% (47/62) of the patients in the intervention group achieved their goal BP. There was a significant dropout rate in the intervention arm, with only 62 of 97 (63.9%) participants attending the 12-month follow-up.

A multicenter Canadian study of 110 diabetic patients examined the impact of an automatic telemonitoring system using a smartphone app and Bluetooth-enabled BP machine [28]. Abnormal readings were responded to by an automated system, which transmitted self-care messages to the patient. Clinicians were alerted only to highly abnormal values. The control group received the same BP machine, but without any intervention from the automated system. The mean daytime ambulatory systolic BP decreased significantly in the intervention group by 9.1 (SD 5.6) mm Hg (*P*<.001) and did not significantly change in the control group.

The Medication Adherence Improvement Support App For Engagement—Blood Pressure app used a web-based recruitment platform to randomize 411 patients with essential hypertension to receive either an automatic, Bluetooth-enabled sphygmomanometer or the companion Medisafe smartphone

app [29]. The app incorporated medication reminders after either the user had input their medication list or it was auto-populated via an electronic medical record. Participants were followed up for 12 weeks. Medication adherence, as determined by the MMAS-8 questionnaire score, increased by 0.4 points in the intervention group compared with the control group. This result was statistically significant, but not thought to be clinically significant, as the authors targeted an MMAS-8 score improvement of 2.0 to be a meaningful improvement. There was no significant difference in BP.

Morikawa et al [30] studied the effect of minimizing salt intake on BP in a group of hypertensive railroad employees. A total of 41 patients were allocated, via a quasi-randomized process, to an intervention or control group. Those in the intervention group were asked to measure urine sodium concentration via a salt sensor and container provided by the investigators. The results were then used to trigger a personalized email, delivered via mobile phone, to the participants. BP and salt intake were assessed over 4 weeks. The intervention group had a larger mean reduction in systolic BP (5.4 mm Hg vs 2.2 mm Hg), although this did not reach statistical significance. Diastolic BP, however, fell significantly (6.2 mm Hg vs 1.6 mm Hg; *P*=.01). The intervention was shown to increase the number of patients modifying their dietary salt intake.

A multicenter study from Chile examined the effect of 1 SMS every 2 weeks on BP and medication adherence [31]. There was no significant difference for either of these endpoints, although there was a very high dropout rate, meaning over 43.3% (136/314) of participants did not have their BP measured at the completion of the study.

Cardiac Rehabilitation

Eight randomized controlled trials studied the addition of mobile phone interventions (MPIs) to standard CR. All trials demonstrated at least one benefit in the intervention group, although specific positive results varied significantly between trials.

Del Rosario et al [33] randomized patients undergoing home-based or hospital-based CR for the first time to receive a sphygmomanometer and weighing scale that could transmit data via a mobile phone using near field communication technology as an adjunct to standard CR or standard CR alone. The authors noted an improvement in CR completion rates within the intervention arm (27/33, 88% vs 20/33, 67%; *P*=.04).

The Text4Heart trial, undertaken in New Zealand, examined the addition of a personalized 24-week program of educational and motivational text messages, delivered daily, to standard CR for patients with postmyocardial infarction [36]. Apart from self-reported medication compliance at 6 months, there was no significant difference between other self-reported endpoints at the completion of 6-month follow-up.

Piotrowicz et al [37] examined the utility of a mobile phone paired with a 3-lead ECG monitor in patients with cardiac failure. Patients were randomized to receive home telemonitoring for CR or standard CR and were followed up over an 8-week period. Patients in the home telemonitoring group answered questions regarding symptomatic status before a session of CR and transmitted their ECG to a remote center. Once reviewed, the patients were given approval to commence the session. ECGs were automatically transmitted during the session and at the conclusion of the session, with the patient able to transmit an ECG if they experienced chest pain or any other concerning symptoms. Although clinical endpoints were similar, there was a 20% dropout rate in the standard CR group, predominantly driven by patients being unable to afford the costs of attending CR and due to difficulties attending CR due to time constraints.

Varnfield et al [38] compared a home-based CR service that was delivered by a smartphone with standard CR in patients following an acute coronary syndrome. The components of the program included a step-counter, a *wellness diary*, weekly teleconferences with a mentor, motivational text messages, and videos. Uptake rates in the intervention group were significantly better (48/60, 80% vs 37/60, 62%; *P*=.04); however, there were no significant differences in clinical outcomes over 6 weeks.

Pandey et al [35] published an RCT of 50 postmyocardial infarction patients who received 4 SMS messages daily reminding them to exercise. The patients kept a logbook of the days they exercised and the duration. According to this self-reported endpoint, patients in the intervention arm exercised more frequently (17 days per 12.5 hours per month vs 13 days per 8.5 hours per month). There was, however, no difference in cardiorespiratory fitness as measured by metabolic equivalents during a Bruce protocol exercise stress test. In the same paper was a separate RCT of 34 patients with recent myocardial infarction attending CR receiving daily SMS medication reminders [35]. Only patients taking once-daily medications were included in the trial, which introduces selection bias, as patients taking once-daily medications may have less difficulty adhering to their prescribed regimen compared with those requiring multiple daily dosing. The messages were generic, such as *please remember to take your morning medications now* and did not specify any information about the patient or the medications themselves. The patient was not able to reply to the message. Medication compliance was assessed using logbooks kept by the patient. The compliance rate, as measured by days where medications were all taken, was significantly higher in the intervention group (94% vs 80%; *P*<.001).

Two studies focused on the use of a wearable ECG monitoring system using smartphone technology. An RCT from New Zealand compared home-based CR using a smartphone-based platform including an ECG monitoring vest and web-based education with center-based CR. This noninferiority trial demonstrated comparable average physical fitness (as measured by maximal oxygen consumption: VO_2 max) between the 2 groups [34]. Preliminary results from a single center in a Spanish study revealed no significant differences in exercise-related outcomes between a group of patients with ischemic cardiomyopathy who utilized a wearable ECG vest with a smartphone connection for home-based CR and those undergoing a traditional hospital-based program [32].

Meta-Analysis

For the heart failure cohort, all 6 studies were included in the meta-analysis. There was no significant difference in mortality

(measured at 6 months in all studies, with the exception of Koehler et al [22], who reported mortality at 12 months). The mortality rate in the intervention group was 10.4% compared with 11.6% in the control group (87/836 vs 98/847; *P*=.45; I^2=45%; Figure 2).

Readmissions due to heart failure over 6 months were less common in the intervention group than in the control group for the 3 studies that reported this endpoint (96/686, 14.0% vs 129/696, 18.5%; OR 0.69, 95% CI 0.48 to 0.98; *P*=.04; I^2=26%; Figure 3).

The rate of hospitalization for any reason over 6 months was significantly lower in the intervention group (244/792, 30.8% vs 287/803, 35.7%; OR 0.77, 95% CI 0.62 to 0.97; *P*=.03; I^2=0%; Figure 4).

The difference in systolic BP was analyzed from 5 studies that reported the endpoint at 6 or 12 months. The mean systolic BP was 4.3 mm Hg less in the intervention group than in the control group (95% CI −7.8 to −0.78 mm Hg; *P*=.02; Figure 5). Substantial heterogeneity was identified (I^2=78%). Two studies that reported the endpoint were excluded from analysis, one due to a significantly different BP at baseline between the 2 groups [27] and one due to a very high withdrawal rate of over 40% [31].

Four studies reported the percentage of patients who reached the target BP, defined as 140/90 mm Hg in 3 studies [14,25,26] and 130/80 mm Hg in the other [28]. In the meta-analysis, patients in the intervention arm were more likely to achieve their target BP than those in the control group (596/865, 68.9% vs 472/885, 53.3%; OR 2.07, 95% CI 1.29 to 3.32; *P*=.002; I^2=78%; Figure 6).

There was no significant difference in the change in BMI between the 4 studies that reported the endpoint after 6 or more months (mean difference −0.46; 95% CI −1.44 to 0.52; *P*=.36; I^2=82%; Multimedia Appendix 7).

A meta-analysis of medication adherence could not be performed, as there was no uniform measurement for assessing the outcome.

Figure 2. Forest plot of the odds ratio (OR) of mortality in patients with heart failure who were involved in randomized controlled trials comparing a mobile phone intervention versus control. The estimate of the OR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. The summary OR is represented by the middle of the solid diamond. A test of heterogeneity is given below the summary statistics. MPI: mobile phone intervention, df: degrees of freedom.

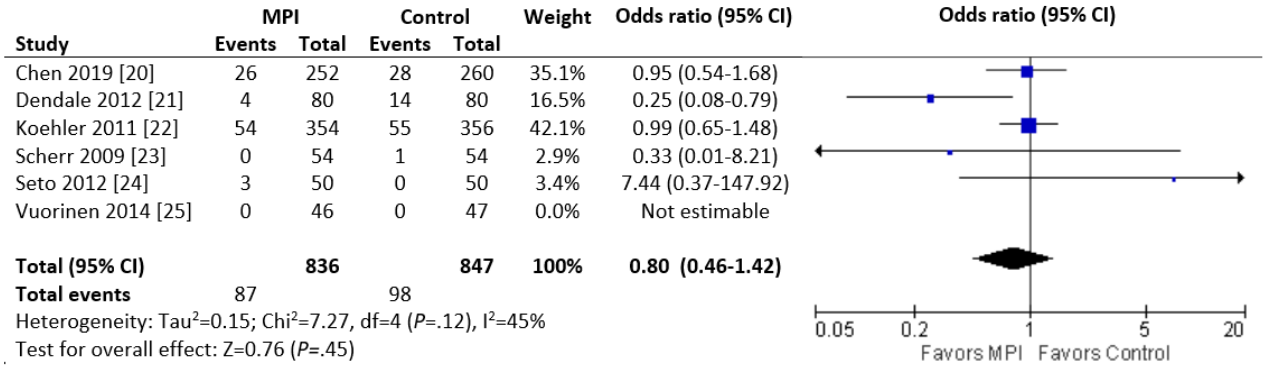


Figure 3. Forest plot of the odds ratio (OR) of heart failure readmissions in patients with heart failure who were involved in randomized controlled trials comparing a mobile phone intervention versus control. The estimate of the OR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. The summary OR is represented by the middle of the solid diamond. A test of heterogeneity is given below the summary statistics. MPI: mobile phone intervention, df: degrees of freedom.

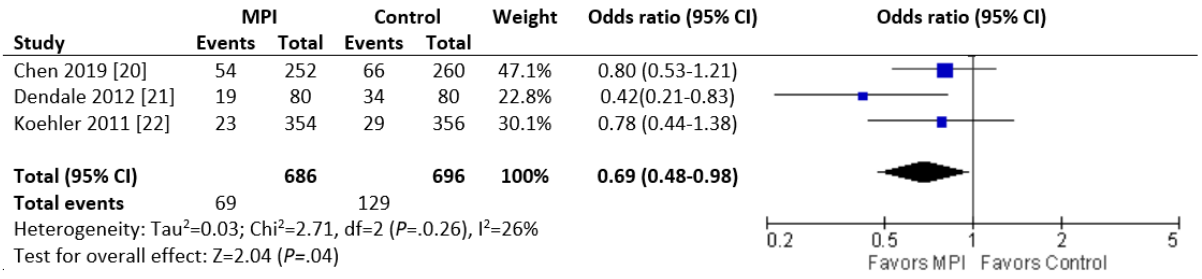


Figure 4. Forest plot of the odds ratio (OR) of all-cause readmissions in patients with heart failure who were involved in randomized controlled trials comparing a mobile phone intervention versus control. The estimate of the OR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. The summary OR is represented by the middle of the solid diamond. A test of heterogeneity is given below the summary statistics. MPI: mobile phone intervention, df: degrees of freedom.

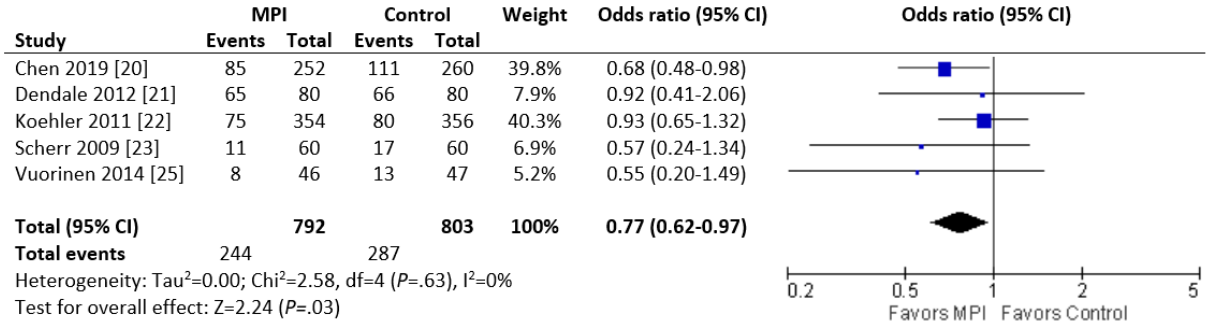


Figure 5. Forest plot of the mean difference in blood pressure in patients with hypertension who were involved in randomized controlled trials comparing a mobile phone intervention versus control. The mean difference of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. The summary mean difference is represented by the middle of the solid diamond. A test of heterogeneity is given below the summary statistics. MPI: mobile phone intervention, df: degrees of freedom.

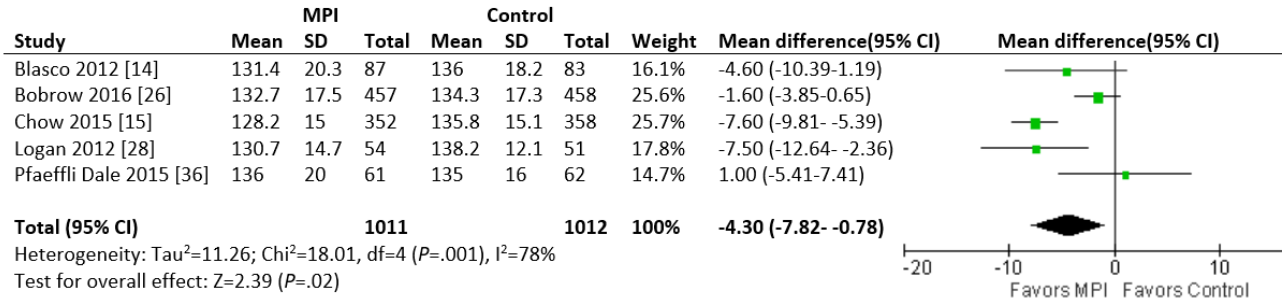
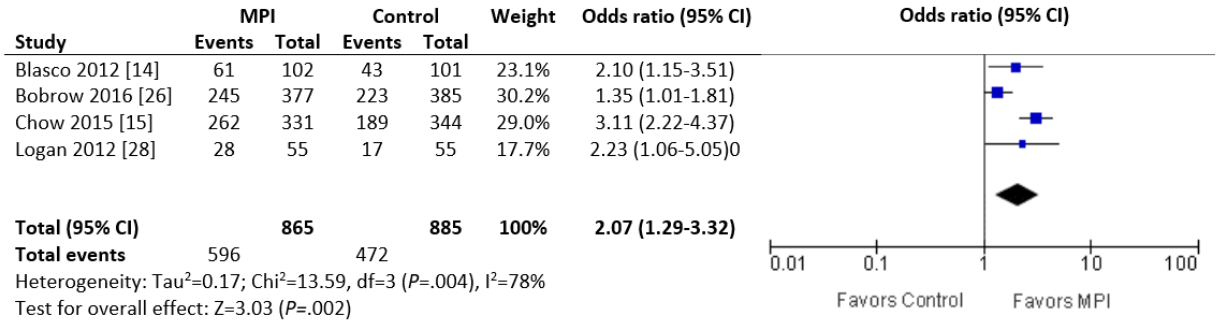


Figure 6. Forest plot of the odds ratio (OR) in patients with hypertension who achieved the prespecified target blood pressure and who were involved in randomized controlled trials comparing a mobile phone intervention versus control. The estimate of the OR of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% CI. The summary OR is represented by the middle of the solid diamond. A test of heterogeneity is given below the summary statistics. MPI: mobile phone intervention, df: degrees of freedom.



Discussion

Principal Findings

This meta-analysis demonstrated that in patients with heart failure, the use of MPIs reduced the rate of hospital admission, both in relation to total admissions and heart failure admissions. There was no significant difference in mortality rates between the groups. In patients with hypertension, those who used MPIs had a significantly lower systolic BP and were more likely to reach the target BP. There was no significant difference in BMI.

Mobile phone and smartphone technology represent a significant opportunity for health care providers to improve outcomes for large populations of patients with CVD. Although no single holistic cardiac care app has been rigorously trialed, a multitude of small, targeted apps have been studied. In general, these heterogeneous and underpowered data do not allow for clear conclusions to be made on the overall benefit of MPIs; however, several important observations are made.

In patients with ischemic heart disease, MPIs universally improved medication compliance. It was not possible to perform meta-analysis of this endpoint due to the variation in reporting systems used between studies, but this is an important finding. Reasons for noncompliance are multiple; however, by providing physical reminders as well as motivational support, text messaging appears to be an effective method of reinforcing adherence, particularly in the context of asymptomatic disease. Given the relatively low cost and negligible risk of text messaging, it could be considered a mainstream management strategy for patients possessing a mobile phone. It remains to be seen, however, whether improved compliance leads to a clinically significant benefit, as cardiovascular event rates were generally not examined in these trials. The successful use of mobile phone technology to promote medication compliance has also been demonstrated in other fields of medicine [24], with one recent meta-analysis estimating an improvement in patient compliance from 50% to 67.8% with consistent use of text message reminders for patients with chronic disease [39,40].

In hypertensive patients, reductions in BP are also likely a reflection of improved compliance [26]. In addition, apps that emphasize positive lifestyle modifications such as dietary improvement and increased exercise regularity are likely to provide an impetus for nonpharmacological reduction in BP, as was seen in trials aimed at the general population [41,42]. Reduction of BP, and improvement in compliance, particularly in relation to antiplatelet agents following revascularization, is likely to reduce long-term recurrence rates of ischemic heart disease; however, none of the studies examined a follow-up period beyond 6 months. Furthermore, all studies examining SMS compared SMS with a control arm. Different types of SMS (personalized vs generic and interactive vs nonreply) were not compared with each other and neither were variations in SMS frequency (eg, daily vs multiple times daily). Although the body of evidence showing the benefits of SMS interventions for compliance against a control group is growing, the optimal nature and frequency of SMS has not been established and should be the focus of future studies. The majority of studies reviewed here relied on self-reported measures of compliance or compliance questionnaires, both of which are subjective. A reduction in systolic BP and an increased likelihood of reaching target BP was observed in the meta-analysis, although substantial heterogeneity contributed to this result, due to the variety of patient populations and interventions considered.

In the cardiac failure cohort, five studies used home monitoring of BP and weight. Although benefits were shown with regard to quality of life and functional status, only one study demonstrated a difference in mortality, and the results of the meta-analysis were negative. The negative result for mortality was driven largely by the study by Koehler et al [22], which used remote monitoring of ECG, weight, and BP in combination with a 24-hour physician monitoring service. The authors recruited a cohort of patients with severe cardiomyopathy and ejection fractions of <30% [22]. These patients had a poor prognosis, as evidenced by the 15% mortality rate at the end of the 26-month follow-up period. It is possible that the severity of cardiomyopathy in these patients meant that despite remote monitoring, readmissions and mortality were unavoidable in such a cohort. In addition, it is not clarified what the triggers and frequency for contacting patients were. Seto et al [24] utilized a similar model using a mobile phone and alerts to clinicians when the results were abnormal. This study of 100 patients was underpowered to detect any statistically significant reductions in mortality or hospitalization rates. Vuorinen et al [25] also studied a cohort of patients with severely impaired left ventricular function. The authors concluded that the intervention had a net negative benefit, as there were no improvements in clinical outcomes, and an increase in clinic visits for those patients. Patients in this study, however, only had measurements performed weekly. Heart failure with preserved ejection fraction was largely underrepresented in all these studies.

From these data, it is clear that not all interventions are equal. Dendale et al [21] demonstrated an intervention that improved both mortality and the rate of heart failure readmissions. This intervention used daily automatic data transmission, an alert-based system to *flag* patients with abnormal parameters, and involved the care of the patient's general practitioner. All these components appear to be beneficial. Automatic data transmission, rather than manual data entry, eases the work burden on the patient and may improve compliance with the system. Although weekly data entry, as used in the study by Vuorinen et al [25], eases the compliance burden on the patient, it may not be sensitive enough to detect early decompensation. Using an alert-based system helps to identify patients who need closest monitoring and avoid data saturation of clinicians. Utilizing the expertise of a general practitioner who is familiar with the patient is also likely to be beneficial in optimizing therapy. Similarly, the study by Logan et al [28] in hypertensive patients used an automatic BP transmission process, which alerted the patient's usual primary care clinician to abnormal values. This particular intervention proved 1 of 2 interventions that demonstrated a statistically significant BP reduction over at least 6 months. Therefore, it would appear that an optimal telemonitoring MPI for cardiovascular patients should include (1) collection of patient data using automatic methods rather than manual entry; (2) an automated back-end that will filter and identify abnormal data; (3) the input of the patient's regular clinician; (4) an educational component, perhaps by text messaging; and (5) ease of use and limited technical issues.

The results for mobile phone technology as an adjunct to CR suggested potential improvements in medication adherence [36], participation rates [37-39], and quality of life scores [43]. Smartphone-based CR allows patients to participate from home, which is desirable, particularly for participants who may find it inconvenient or costly to attend hospital-based programs.

Other studies of telemonitoring in heart failure, using technologies other than mobile phones, have shown mixed

results. Multiple meta-analyses have shown superior outcomes for telemonitoring in heart failure patients compared with standard care [44-46], although 2 large individual trials have been negative [47,48]. A recent meta-analysis included 37 RCTs with 9582 heart failure patients [46]. The authors identified a reduction in all-cause mortality for telemedical interventions compared with usual care (relative risk 0.81; 95% CI 0.70 to 0.94; I^2=0.16). The methodologies and delivery of these interventions were highly variable, and the vast majority (32/37) did not use mobile phone–based systems. Examples of delivery methods include videoconference, telephone calls, websites, and purpose-built telemedical units capable of transmitting data. A head-to-head comparison of telemedical delivery methods has not been performed.

The cost-effectiveness of mobile phone technology for any of the aforementioned indications in CVD has not been conclusively studied. It is believed that a reduction of adverse clinical outcomes and an associated reduction in costs of hospitalization would likely offset costs of implementing the software and monitoring data, although there was a notable paucity of cost-effectiveness data. There remains only a single published cost-effectiveness analysis of an MPI based on randomized trial data [49]. The authors extrapolated the results of the clinical endpoints in the TEXT ME trial to a hypothetical cohort of 50,000 patients in a Markov model. They estimated cost savings of over Aus $10 million (US $6.53 million) in such a model.

Assuming that the mobile phone is not operated during driving or other dangerous tasks, there are no significant risks to the patient in any of the described interventions. There are several limitations of the available data contained within this systematic review and meta-analysis, and the results should be interpreted with caution. Although all studies were RCTs, they were generally small, with varying methodologies, and prone to bias. No follow-up period was longer than 26 months; thus, data on recurrent clinical events in the medium to long term were lacking. In addition, patient compliance with the mobile phone technology itself over longer periods is unknown. The utility of such interventions in the older population is uncertain, given that advanced age is one of the most significant risk factors for all forms of CVD. One study demonstrated a high rate of dropout due to inability of patients to operate the app (22%) [23], whereas another study reported a lower rate of unsuccessful app use (10%) recruited patients from online communities [29], which introduces a degree of selection bias as patients accessing these communities are likely to be more familiar with mobile phone technologies. Compliance data are typically measured using questionnaires and are thus prone to recall bias.

Several gaps remain in this literature. There is significant potential of this technology to gather data that can be reviewed in real-time and subsequently allow for rapid modifications in patient therapy in response, although this was only examined in a small number of trials. When patients are reviewed routinely by a clinician in the community, the clinician only sees a *snapshot* of the patient's current health status. Trends in data such as weight and BP can only be compared between visits, rather than on a daily or regular basis. The connectivity of smartphone devices allows for data to be transmitted to a clinician who may be able to interpret them and provide management advice even from remote locations, thus allowing appropriate disease management to be instigated before the patient's presentation with an acute medical event. It would appear that there is a large scope for mobile phone technology in this setting, and further study is needed to identify the optimal characteristics of such an app or program. Second, the impact of usability on the success of the MPIs was not adequately addressed in trials, with minimal reporting of this endpoint. We would suggest that any publication of an RCT describing an MPI have some evaluation of usability, as this will almost certainly influence compliance and hence the efficacy of the intervention, perhaps even more so than the capabilities of the MPI itself. Finally, the longest minimum follow-up period for any heart failure study was 12 months. A longer duration of monitoring in the studies may have allowed for a difference in readmission rates and, in particular, mortality to have been identified. Given that heart failure is a chronic condition, the benefits of an MPI may not be observed until several years of follow-up have occurred. As telehealth is a rapidly evolving field, it is surmised that studies are published with a short follow-up period to avoid publication at a time where the intervention is outdated or obsolete.

The majority of smartphone apps examined here had a single aim or function, and the management of CVD entails the optimization of multiple factors. Therefore, there remains a need for an adequately powered RCT examining the effect of a holistic smartphone intervention with multiple features that possess the ability to react to collected data and improve therapy and therefore clinical endpoints, and there is a need to identify which patients would benefit the most.

Conclusions

MPIs have been applied to a variety of target groups in CVD. These fall into several categories including SMS apps, automatic and manual monitoring, and purpose-built apps. A number of RCTs have been published. The results suggest that mobile phone technology may improve medication adherence in patients with ischemic heart disease, BP in individuals with hypertension, and hospitalization rates in patients with heart failure. Further large RCTs with longer follow-up periods and a greater focus on clinical endpoints are required. However, given the relatively low risk and cost of such interventions, they should be considered as an adjunctive therapy in the management of patients with CVD or at risk of CVD.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study interventions and target populations.
[DOCX File , 14 KB - mhealth_v8i7e16695_app1.docx]

Multimedia Appendix 2
Other secondary endpoints of included studies.
[DOCX File , 16 KB - mhealth_v8i7e16695_app2.docx]

Multimedia Appendix 3
Bias assessment according to the Cochrane Risk of Bias Tool.
[DOCX File , 17 KB - mhealth_v8i7e16695_app3.docx]

Multimedia Appendix 4
Summary of mobile phone interventions for medication adherence in patients with ischaemic heart disease.
[DOCX File , 13 KB - mhealth_v8i7e16695_app4.docx]

Multimedia Appendix 5
Mobile phone interventions for treatment of heart failure.
[DOCX File , 14 KB - mhealth_v8i7e16695_app5.docx]

Multimedia Appendix 6
Mobile phone interventions for treatment of hypertension.
[DOCX File , 13 KB - mhealth_v8i7e16695_app6.docx]

Multimedia Appendix 7
Forest plot of the mean difference in body mass index (BMI) in patients who were involved in randomised controlled trials comparing a mobile phone intervention (MPI) versus control. The mean difference of each trial corresponds to the middle of the squares, and the horizontal line shows the 95% confidence interval (CI). The summary mean difference, is represented by the middle of the solid diamond. A test of heterogeneity is given below the summary statistics where df refers to degrees of freedom.
[PNG File , 12 KB - mhealth_v8i7e16695_app7.png]

References

1. World Health Organization. 2019. The Top 10 Causes of Death URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2019-06-29]
2. Australian Institute of Health and Welfare. Cardiovascular disease series no 35. Canberra: AIHW; 2017. Cardiovascular disease: Australian facts 2011 URL: https://www.aihw.gov.au/getmedia/13cd081b-8123-4660-ad68-3d780c12ffeb/12116-20111005.pdf.aspx [accessed 2020-05-22]
3. Smith SC, Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, et al. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation 2011 Nov 29;124(22):2458-2473. [doi: 10.1161/CIR.0b013e318235eb4d] [Medline: 22052934]
4. Anderson L, Thompson DR, Oldridge N, Zwisler AD, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016 Jan 5(1):CD001800 [FREE Full text] [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]
5. NHFA CSANZ Heart Failure Guidelines Working Group, Atherton JJ, Sindone A, de Pasquale CG, Driscoll A, MacDonald PS, et al. National heart foundation of Australia and cardiac society of Australia and New Zealand: guidelines for the prevention, detection, and management of heart failure in Australia 2018. Heart Lung Circ 2018 Oct;27(10):1123-1208 [FREE Full text] [doi: 10.1016/j.hlc.2018.06.1042] [Medline: 30077227]
6. Hoo SY, Gallagher R, Elliott D. Predictors of cardiac rehabilitation attendance following primary percutaneous coronary intervention for ST-elevation myocardial infarction in Australia. Nurs Health Sci 2016 Jun;18(2):230-237. [doi: 10.1111/nhs.12258] [Medline: 26858136]
7. Dalal HM, Wingham J, Palmer J, Taylor R, Petre C, Lewin R, REACH-HF Investigators. Why do so few patients with heart failure participate in cardiac rehabilitation? A cross-sectional survey from England, Wales and Northern Ireland. BMJ Open 2012;2(2):e000787 [FREE Full text] [doi: 10.1136/bmjopen-2011-000787] [Medline: 22454188]
8. Wade V, Stocks N. The use of telehealth to reduce inequalities in cardiovascular outcomes in australia and new zealand: a critical review. Heart Lung Circ 2017 Apr;26(4):331-337. [doi: 10.1016/j.hlc.2016.10.013] [Medline: 27993487]

9. Kotb A, Cameron C, Hsieh S, Wells G. Comparative effectiveness of different forms of telemedicine for individuals with heart failure (HF): a systematic review and network meta-analysis. PLoS One 2015;10(2):e0118681 [FREE Full text] [doi: 10.1371/journal.pone.0118681] [Medline: 25714962]
10. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
11. Baroi S, McNamara RJ, McKenzie DK, Gandevia S, Brodie MA. Advances in remote respiratory assessments for people with chronic obstructive pulmonary disease: a systematic review. Telemed J E Health 2018 Jun;24(6):415-424. [doi: 10.1089/tmj.2017.0160] [Medline: 29083268]
12. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
13. Brodie MA, Pliner EM, Ho A, Li K, Chen Z, Gandevia SC, et al. Big data vs accurate data in health research: large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. Med Hypotheses 2018 Oct;119:32-36. [doi: 10.1016/j.mehy.2018.07.015] [Medline: 30122488]
14. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
15. 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. J Am Med Assoc 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
16. Fang R, Li X. Electronic messaging support service programs improve adherence to lipid-lowering therapy among outpatients with coronary artery disease: an exploratory randomised control study. J Clin Nurs 2016 Mar;25(5-6):664-671. [doi: 10.1111/jocn.12988] [Medline: 26522838]
17. Khonsari S, Subramanian P, Chinna K, Latif LA, Ling LW, Gholami O. Effect of a reminder system using an automated short message service on medication adherence following acute coronary syndrome. Eur J Cardiovasc Nurs 2015 Apr;14(2):170-179. [doi: 10.1177/1474515114521910] [Medline: 24491349]
18. Park LG, Howie-Esquivel J, Whooley MA, Dracup K. Psychosocial factors and medication adherence among patients with coronary heart disease: a text messaging intervention. Eur J Cardiovasc Nurs 2015 Jun;14(3):264-273. [doi: 10.1177/1474515114537024] [Medline: 24853566]
19. Quilici J, Fugon L, Beguin S, Morange PE, Bonnet J, Alessi M, et al. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. Int J Cardiol 2013 Sep 20;168(1):568-569. [doi: 10.1016/j.ijcard.2013.01.252] [Medline: 23462636]
20. Chen C, Li X, Sun L, Cao S, Kang Y, Hong L, et al. Post-discharge short message service improves short-term clinical outcome and self-care behaviour in chronic heart failure. ESC Heart Fail 2019 Feb;6(1):164-173 [FREE Full text] [doi: 10.1002/ehf2.12380] [Medline: 30478888]
21. Dendale P, de Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340 [FREE Full text] [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
22. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 3;123(17):1873-1880. [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
23. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
24. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
25. Vuorinen A, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014 Dec 11;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
26. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-text adherence support [STAR]): a single-blind, randomized trial. Circulation 2016 Feb 9;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]

27. Kiselev AR, Gridnev VI, Shvartz VA, Posnenkova OM, Dovgalevsky PY. Active ambulatory care management supported by short message services and mobile phone technology in patients with arterial hypertension. J Am Soc Hypertens 2012;6(5):346-355. [doi: 10.1016/j.jash.2012.08.001] [Medline: 22995803]
28. Logan AG, Irvine MJ, McIsaac WJ, Tisler A, Rossos PG, Easty A, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. Hypertension 2012 Jul;60(1):51-57. [doi: 10.1161/HYPERTENSIONAHA.111.188409] [Medline: 22615116]
29. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MediSAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 1;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
30. Morikawa N, Yamasue K, Tochikubo O, Mizushima S. Effect of salt reduction intervention program using an electronic salt sensor and cellular phone on blood pressure among hypertensive workers. Clin Exp Hypertens 2011;33(4):216-222. [doi: 10.3109/10641963.2011.583966] [Medline: 21699447]
31. Varleta P, Acevedo M, Akel C, Salinas C, Navarrete C, García A, et al. Mobile phone text messaging improves antihypertensive drug adherence in the community. J Clin Hypertens (Greenwich) 2017 Dec;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
32. Bravo-Escobar R, González-Represas A, Gómez-González AM, Montiel-Trujillo A, Aguilar-Jimenez R, Carrasco-Ruíz R, et al. Effectiveness and safety of a home-based cardiac rehabilitation programme of mixed surveillance in patients with ischemic heart disease at moderate cardiovascular risk: a randomised, controlled clinical trial. BMC Cardiovasc Disord 2017 Feb 20;17(1):66 [FREE Full text] [doi: 10.1186/s12872-017-0499-0] [Medline: 28219338]
33. Rosario MB, Lovell NH, Fildes J, Holgate K, Yu J, Ferry C, et al. Evaluation of an mHealth-based adjunct to outpatient cardiac rehabilitation. IEEE J Biomed Health Inform 2018 Nov;22(6):1938-1948. [doi: 10.1109/JBHI.2017.2782209] [Medline: 29990228]
34. Maddison R, Rawstorn JC, Stewart RA, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129 [FREE Full text] [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
35. Pandey A, Krumme A, Patel T, Choudhry N. The impact of text messaging on medication adherence and exercise among postmyocardial infarction patients: randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Aug 3;5(8):e110 [FREE Full text] [doi: 10.2196/mhealth.7144] [Medline: 28778843]
36. Dale LP, 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 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
37. Piotrowicz E, Baranowski R, Bilinska M, Stepnowska M, Piotrowska M, Wójcik A, et al. A new model of home-based telemonitored cardiac rehabilitation in patients with heart failure: effectiveness, quality of life, and adherence. Eur J Heart Fail 2010 Feb;12(2):164-171 [FREE Full text] [doi: 10.1093/eurjhf/hfp181] [Medline: 20042423]
38. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
39. Sarabi RE, Sadoughi F, Orak RJ, Bahaadinbeigy K. The effectiveness of mobile phone text messaging in improving medication adherence for patients with chronic diseases: a systematic review. Iran Red Crescent Med J 2016 May;18(5):e25183 [FREE Full text] [doi: 10.5812/ircmj.25183] [Medline: 27437126]
40. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
41. Liu Z, Chen S, Zhang G, Lin A. Mobile phone-based lifestyle intervention for reducing overall cardiovascular disease risk in Guangzhou, China: a pilot study. Int J Environ Res Public Health 2015 Dec 17;12(12):15993-16004 [FREE Full text] [doi: 10.3390/ijerph121215037] [Medline: 26694436]
42. 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 9;4(11) [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
43. 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 Jul 23;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
44. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010 Aug 4(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
45. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]

46. Yun JE, Park J, Park H, Lee H, Park D. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. J Card Fail 2018 Jan;24(1):19-28. [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]
47. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
48. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
49. Burn E, Nghiem S, Jan S, Redfern J, Rodgers A, Thiagalingam A, et al. Cost-effectiveness of a text message programme for the prevention of recurrent cardiovascular events. Heart 2017 Jun;103(12):893-894. [doi: 10.1136/heartjnl-2016-310195] [Medline: 28235776]

Abbreviations

BP: blood pressure
CR: cardiac rehabilitation
CVD: cardiovascular disease
ECG: electrocardiogram
LVEF: left ventricular ejection fraction
MMAS: Morisky Medication Adherence Scale
MPI: mobile phone intervention
NYHA: New York Heart Association
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCTs: randomized control trials

Edited by G Eysenbach; submitted 15.10.19; peer-reviewed by S Omboni, B Visser; comments to author 01.03.20; revised version received 30.03.20; accepted 08.04.20; published 06.07.20.

Please cite as:
Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, Ooi SY
Mobile Phone Technologies in the Management of Ischemic Heart Disease, Heart Failure, and Hypertension: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2020;8(7):e16695
URL: https://mhealth.jmir.org/2020/7/e16695
doi:10.2196/16695
PMID:32628615

Original Paper

Understanding Clinicians' Adoption of Mobile Health Tools: A Qualitative Review of the Most Used Frameworks

Christine Jacob[1,2], BSc, MSc; Antonio Sanchez-Vazquez[3], PhD; Chris Ivory[3], PhD

[1]Anglia Ruskin University, Cambridge, United Kingdom

[2]University of Applied Sciences Northwestern Switzerland, Brugg, Switzerland

[3]Innovation and Management Practice Research Centre, Anglia Ruskin University, Cambridge, United Kingdom

Corresponding Author:
Christine Jacob, BSc, MSc
Anglia Ruskin University
East Rd
Cambridge, CB1 1PT
United Kingdom
Phone: 41 798769376
Email: christine.k.jacob@gmail.com

Abstract

Background: Although there is a push toward encouraging mobile health (mHealth) adoption to harness its potential, there are many challenges that sometimes go beyond the technology to involve other elements such as social, cultural, and organizational factors.

Objective: This review aimed to explore which frameworks are used the most, to understand clinicians' adoption of mHealth as well as to identify potential shortcomings in these frameworks. Highlighting these gaps and the main factors that were not specifically covered in the most frequently used frameworks will assist future researchers to include all relevant key factors.

Methods: This review was an in-depth subanalysis of a larger systematic review that included research papers published between 2008 and 2018 and focused on the social, organizational, and technical factors impacting clinicians' adoption of mHealth. The initial systematic review included 171 studies, of which 50 studies used a theoretical framework. These 50 studies are the subject of this qualitative review, reflecting further on the frameworks used and how these can help future researchers design studies that investigate the topic of mHealth adoption more robustly.

Results: The most commonly used frameworks were different forms of extensions of the Technology Acceptance Model (TAM; 17/50, 34%), the diffusion of innovation theory (DOI; 8/50, 16%), and different forms of extensions of the unified theory of acceptance and use of technology (6/50, 12%). Some studies used a combination of the TAM and DOI frameworks (3/50, 6%), whereas others used the consolidated framework for implementation research (3/50, 6%) and sociotechnical systems (STS) theory (2/50, 4%). The factors cited by more than 20% of the studies were usefulness, output quality, ease of use, technical support, data privacy, self-efficacy, attitude, organizational inner setting, training, leadership engagement, workload, and workflow fit. Most factors could be linked to one framework or another, but there was no single framework that could adequately cover all relevant and specific factors without some expansion.

Conclusions: Health care technologies are generally more complex than tools that address individual user needs as they usually support patients with comorbidities who are typically treated by multidisciplinary teams who might even work in different health care organizations. This special nature of how the health care sector operates and its highly regulated nature, the usual budget deficits, and the interdependence between health care organizations necessitate some crucial expansions to existing theoretical frameworks usually used when studying adoption. We propose a shift toward theoretical frameworks that take into account implementation challenges that factor in the complexity of the sociotechnical structure of health care organizations and the interplay between the technical, social, and organizational aspects. Our consolidated framework offers recommendations on which factors to include when investigating clinicians' adoption of mHealth, taking into account all three aspects.

(JMIR Mhealth Uhealth 2020;8(7):e18072) doi:10.2196/18072

KEYWORDS

telemedicine; smartphone; electronic health record; workflow; workload; workplace; public health practice; technology; perception; health education; mHealth; mobile health; telehealth; eHealth

Introduction

Background

Health care sectors worldwide are witnessing the rapid emergence of new tools using mobile technology to support medical practice; new apps are created and launched daily. The World Health Organization global observatory of eHealth defines mobile health (mHealth) as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, Personal Digital Assistants (PDAs), and other wireless devices." Telemedicine itself is a growing and well-established subcategory of mHealth and is defined as "the communication or consultation between health professionals about patients using the voice, text, data, imaging, or video functions of a mobile device. But it can be applied to other situations; the management of chronic diseases of patients living at home is one example" [1].

Previous research has shown that such tools have promising potential. They can save time and improve efficiency [2], increase patient access and decrease workload [3], empower patients and address their information needs [4], enhance patient-clinician communication by prompting some underreported and sensitive discussions [5,6], and make patients feel more taken care of [5,6]. The patient data generated through some of these tools can also help care teams to prioritize and tailor treatment plans accordingly [5]. Research has also shown that the potential benefits encourage even older populations to adopt such tools [7] despite the widespread notion of a *digital divide* and boost patients' quality of life through tailored and personalized treatment plans [4,5].

However, it remains that the dynamic and liberal way in which the mHealth market is organized makes it difficult to assess the quality of the tools on offer, making adoption decisions more difficult, especially from the perspective of a clinician [8]. In addition, the amount of health data generated by such apps creates a need for more comprehensive privacy regulations, in particular, around the use of personal data that are clinically and scientifically meaningful [9,10]. Furthermore, high workload, lack of resources, and necessary workflow adjustments were frequently raised as barriers to adoption [11-16].

Objective

Although there is a drive toward encouraging mHealth adoption to harness its potential benefits, the reality is that there are many challenges that go beyond the technology itself but stem instead from the complexity of the health care ecosystem itself. Such challenges have slowed down the acceptance and adoption of mHealth in specific health care settings [17,18], with some failing to progress beyond the pilot stage [19] or failing to become a part of the regular care landscape [20,21]. To fully account for the adoption and nonadoption of technology, it is thus essential to shift our focus beyond the technology itself to tackle clinicians' differing concerns, improving clinical workflow, and workload issues [18,22,23]. Therefore, this review aimed to identify the frameworks most used to successfully understand clinicians' adoption of mHealth to enable future researchers to design studies that take all the important factors into account, leading to more successful adoption and implementation.

We were guided in our review by the field of social studies of technology that views individuals and technological artifacts as entangled and interacting elements in any organizational or social setting [24-27], bearing in mind that such interactions may trigger or enable new forms of organizing work, new roles, or new hierarchies [28,29]. This ontological approach enabled us to widen our scope and identify potential shortcomings in the most frequently used frameworks. We found that the most frequently applied models tend to focus on the technical or individual perspectives of technological adoption.

Methods

Search Strategy and Study Selection

This review was an in-depth subanalysis of a larger systematic review focusing on the social, organizational, and technical factors impacting clinicians' adoption of mHealth [30]. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [31] and the Cochrane handbook [32] were followed to ensure a systematic and rigorous approach. A structured search was carried out using Medical Literature Analysis and Retrieval System Online, or MEDLARS Online (MEDLINE), PubMed, the Cochrane Library, and Serial Analysis of Gene Expression (SAGE) databases for studies published in the English language between January 2008 and July 2018, yielding 4993 results. Of these 4993 studies, 171 met the inclusion criteria, of which only 50 studies used a theoretical framework. These latter 50 studies were the subject of this qualitative review. The review presented here reflects more deeply on the most common frameworks used and on how researchers can be supported to design more robust and reliable studies that properly investigate and reflect the inherent complexity of the issue of mHealth adoption.

Data Collection and Synthesis

The diversity of measures and results that were identified in the studied sample was not homogenous enough to enable a quantitative synthesis of the data. Therefore, a narrative synthesis was used and structured around the organizational, social, and technological factors impacting the clinician's adoption of mHealth solutions. NVivo (QSR International), a computer-assisted qualitative data analysis software, was used to assist in this task.

Data coding began with an initial data extraction grid that included themes based on the most used technology acceptance frameworks in the studied sample; more themes were added as they emerged during the review process. A thematic analysis

as per Braun and Clarke [33] was used to identify and extract themes that addressed the review's research question; the phases of the thematic analysis are explained in detail in Multimedia Appendix 1. The first author (CJ) conducted the initial analysis and coding; she is a digital strategist with more than 18 years of experience and has contributed to the creation and realization of several digital solutions in health care. Then, the second author (ASV) reviewed the coding; any cases of disagreement about coding were discussed in conjunction with the last author (CI) and mutually agreed.

The research themes were guided by the most used theoretical frameworks in the studied sample, in addition to Leonardi's *Methodological Guidelines for the Study of Materiality and Affordances* [27]; hence, they were split into three key groups: material and technological factors, social and personal factors, and organizational and policy factors. This process lasted from August to December 2019.

Results

A summary of each of the included studies, the theoretical framework(s) they used, sample size, sample composition, data collection method, and key findings are detailed in Multimedia Appendix 2.

Characteristics of the Included Studies

The sample characteristics of the included papers are detailed in Multimedia Appendix 3. In terms of data collection methods, 56% (28/50) of the studies used quantitative methods, 26% (13/50) of the studies used qualitative methods, and 18% (9/50) of the studies used a mixed approach. Altogether, 42% (21/50) of the studies focused on clinicians (these were studies that included multiple clinician groups, with a clinician defined as "a person qualified in the clinical practice of medicine, psychiatry, or psychology as distinguished from one specializing in laboratory or research techniques or in theory" [34]), 30% (15/50) of the studies focused on physicians, 10% (5/50) of the studies focused on nurses, 4% (2/50) of the studies focused on medical students, and 14% (7/50) of the studies focused included multiple clinician groups and other populations such as patients or caregivers. Sample sizes varied, that is, 18% (9/50) of the studies had a sample size of 1 to 20 participants, 18% (9/50) of the studies had a sample size of 21 to 50 participants, 20% (10/50) of the studies had a sample size of 51 to 100 participants, 28% (14/50) of the studies had a sample size of 101 to 200 participants, and 16% (8/50) of the studies had a sample size of more than 200 participants.

Geographically, 24% (12/50) of the studies were based in the United States [35-46], 10% (5/50) of the studies were based in the United Kingdom [47-51], and the rest were spread across many countries: Australia [52,53], Australia and the United Kingdom [15], Austria [54], Canada [55-57], Germany [58-60], Iran [61], Iraq [62], Japan [63], the Netherlands [64], the Netherlands, Spain, and the United Kingdom [65], Nigeria [66], Norway [67,68], South Korea [69,70], Spain [71-74], Spain Colombia and Bologna [75], Sri Lanka [76], Sweden [77,78], Switzerland [79], Taiwan [80,81], and Turkey [82].

The studies also cover diverse disease areas and medical specialties, including ambulatory care [58], cognitive behavioral therapy [40], cardiovascular disease [56,83], dermatology [73], diabetes [50,63,65], general practice [35,52,59], intensive care [36], neurology [48], pediatric [45], primary and acute care [46,49,55,71,74], psychiatry and mental health [41,53,68], residential aged care [57], reproductive health [37], and substance use recovery [39,44].

Frameworks Mostly Used in Studying Clinicians' Adoption of Mobile Health

The most commonly used frameworks were different forms of extensions of the Technology Acceptance Model (TAM), the diffusion of innovation theory (DOI), and different forms of extensions of the unified theory of acceptance and use of technology (UTAUT). An overview of the most used frameworks is shown in Table 1 and visualized in Multimedia Appendix 4.

Table 1. Overview of the most used frameworks.

Framework(s)	Prevalence in the studied sample (N=50), n (%)
TAM[a] extensions	17 (34)
Others	11 (22)
DOI[b]	8 (16)
UTAUT[c] extensions	6 (12)
DOI and TAM	3 (6)
CFIR[d]	3 (6)
Sociotechnical theory	2 (4)

[a]TAM: Technology Acceptance Model.

[b]DOI: diffusion of innovation theory.

[c]UTAT: unified theory of acceptance and use of technology.

[d]CFIR: consolidated framework for implementation research.

TAM extensions were used in 34% (17/50) of the studies and varied from additions stemming from the literature and the research-specific context [35,36,38,57,60,61,68,70,71,73,74] to extensions using other frameworks such as Chau and Hu's

model of telemedicine acceptance and theory of interpersonal behavior (TIB) [83], the organizational readiness for change model [45], theory of reasoned action (TRA), and theory of planned behavior [74], and in combination with the UTAUT [58,69].

DOI was used in 16% (8/50) of the studies [37,40,50,53,56,62,76,78], and UTAUT extensions were used in 12% (6/50) of the studies; some of these UTAUT extensions were based on the research context and previous research [64,66,84], and other included extensions from other frameworks such as the De Lone and McLean information success model [54], use of technology [65], and a combination of TAM, TPB, and DOI [82]. Furthermore, certain studies used a combination of the TAM and DOI frameworks (3/50, 6%) [42,43,59]; others used the consolidated framework for implementation research (CFIR; 3/50, 6%) [39,41,67] and the STS (2/50, 4%) [49,55].

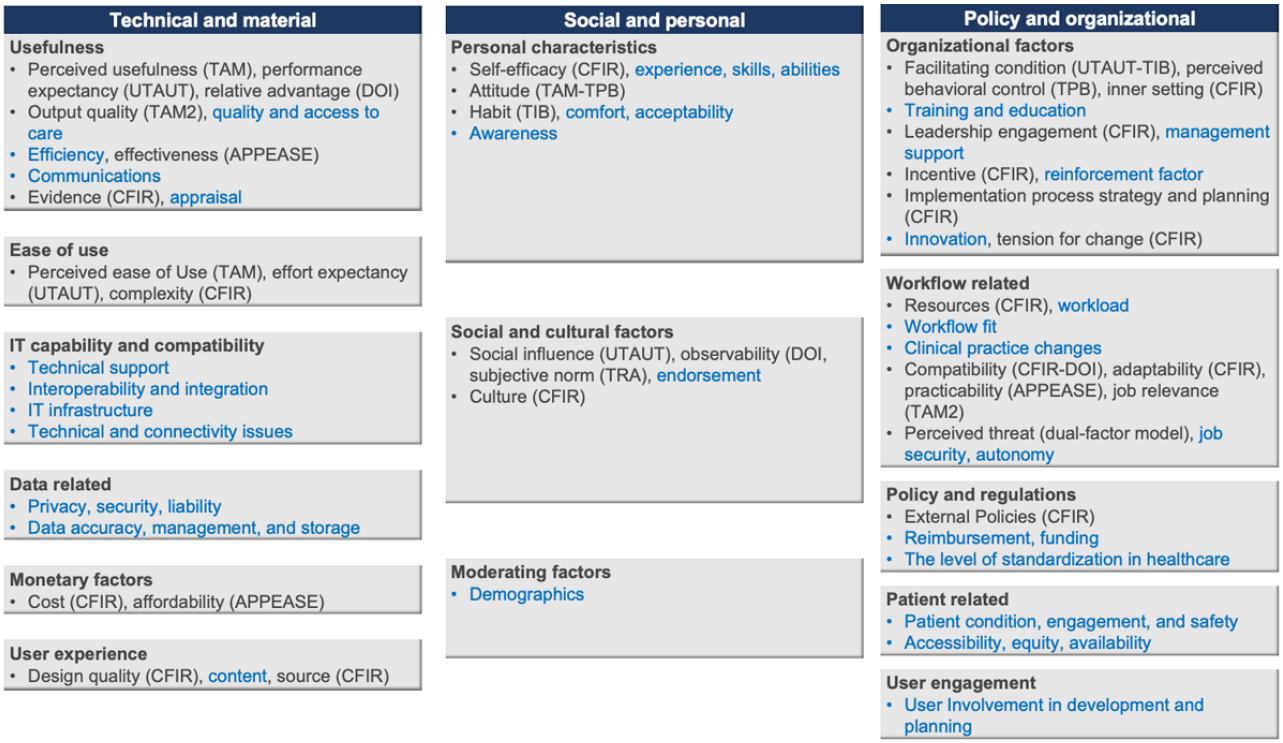
Other frameworks were used in 22% (11/50) of the studies, including the affordability, practicability, effectiveness, acceptability, side effects/safety and equity (APEASE) framework [52]; an extended De Lone and McLean information system success model [63]; Giddens structuration theory [51]; normalization process theory (NPT) [15]; organizational theory of implementation effectiveness [46]; reach, effectiveness, adoption, implementation, and maintenance framework [44]; technological frames (TF) [77]; the design science research methodology (DSRM) [85]; theory of change [48]; and TPB [80].

Framework-Based Versus Additional Factors

We will not present an analysis of each factor in detail, as this was extensively covered in the initial systematic review [30]; we will instead focus on the framework-based versus additional factors to better identify possible shortcomings in the most used frameworks and provide recommendations based on the specificities of the health care setting, as reported in the studied sample.

Factors cited in more than 20% of the included studies were usefulness, output quality, ease of use, technical support, data privacy, self-efficacy, attitude, organizational inner setting, training, leadership engagement, workload, and workflow fit. Most factors could be linked to one framework or another, but there were no frameworks that could cover all factors in our findings unless modified or used in combination with others. An overview of the emerging framework-based and additional factors is presented in Figure 1. Factors defined as specific constructs in one of the used frameworks include the name of the framework following the factor name between brackets, whereas the factors that emerged directly from the data but were not identified as a specific construct in any of the used frameworks are not followed by brackets and are marked in blue font for clarity. These additional factors stemmed mostly from qualitative studies or were predefined by the researchers' review of previous literature findings rather than established frameworks.

Figure 1. Gap analysis of emerging framework-based versus additional factors. APEASE: affordability, practicability, effectiveness, acceptability, safety/side effects, and equity; CFIR: consolidated framework for implementation research; DOI: diffusion of innovation theory; IT: information technology; TAM: Technology Acceptance Model; TAM2: an expanded version of the original TAM; TPB: theory of planned behavior; TRA: theory of reasoned action; UTAUT: unified theory of acceptance and use of technology.



Some factors with more or less the same meaning were named differently in different frameworks. For example, when we talk about the tool's usefulness, it can be referred to as perceived usefulness if the study uses TAM, as performance expectancy if the UTAUT is used, or as relative advantage if DOI or CFIR is used. Similarly, when we talk about the tool's ease of use, it

can be referred to as perceived ease of use if the study uses TAM, as effort expectancy if the UTAUT is used, or as complexity if DOI or CFIR is used. We grouped these similar factors together in our analysis to avoid an overly congested schema of factors. Multimedia Appendix 5 summarizes the definitions of the most used framework-based factors grouped by similarity to emphasize the meaning and overlaps between the different theories.

Technical and Material Factors

When we look into the most reported technical and material factors, as visualized in Figure 2, we can see that the clusters of ease of use and usefulness were the most reported in the included papers; however, other important acceptance elements such as health data–related factors and system interoperability may be strong influencing aspects, but these are not specifically considered in most of the original models.

Figure 2. The most cited technical and material factors. APEASE: affordability, practicability, effectiveness, acceptability, safety/side effects, and equity; CFIR: consolidated framework for implementation research; DOI: diffusion of innovation theory; IT, information technology; TAM: Technology Acceptance Model; TAM2: an expanded version of the original TAM; TPB: theory of planned behavior; TRA: theory of reasoned action; UTAUT: unified theory of acceptance and use of technology.

The cluster of usefulness factors tops the list of technical and material factors. Most studies (n=31) reported usefulness, in general, using theory-based constructs such as perceived usefulness (TAM), performance expectancy (UTAUT), or relative advantage (DOI). Many of these studies, especially those that mainly relied on TAM, did not necessarily dig into the hidden elements that might influence the views of users on usefulness; yet, some still attempted to dig deeper to get at the underlying factors that influence perceived usefulness. For example, Gagnon et al [83] and Orruno et al [73] assumed that compatibility with one's work would impact the perception of a tool's usefulness, whereas Liu and Cheng [81] anticipated

that a user's perception that a tool is a threat to their job would also impact their perceived usefulness, and Ray et al [45] proposed that there is a whole cluster of contextual elements such as organizational and individual factors that may impact perceived usefulness. Most of the studies found that usefulness is one of the most if not the most influential factor when it comes to adoption decisions [36,66,69,71,83], whereas other studies reported that a lack of perceived advantage of the system would be considered a barrier to adoption [50].

Output quality (an additional factor in the expanded version of the original TAM [TAM2]) or quality and access to care were more specific expressions of usefulness, as reported in 12 studies. This is a health care specific factor, as mHealth was seen as the most useful and adding value when it enhances patient care [42,46,67] and facilitates access to care for people who would not otherwise access it easily [45,53]. Moreover, eight studies named increased efficiency or effectiveness (APEASE) as a facilitating factor, as it allows optimal use of staff time [47,50], but some also reported mixed views on whether such tools can really improve efficiency and the necessity of streamlining processes to enable such gains [35,52]. Similarly, eight studies specified communication; however, some saw it as a facilitator [42,47,50], and others reported the risk of miscommunication as a barrier [51,53,76], whereas six studies specified evidence (CFIR) or appraisal, showing that growing the evidence base and the *positive stories* about such tools can help increase adoption [39,48,53].

The ease of use cluster (n=23) was named differently in different theories: perceived ease of use in TAM, effort expectancy in UTAUT, and complexity in CFIR. Similar to the general factor usefulness, *ease of use* lacks the specification of what makes a tool easy to use, making it difficult to understand the hidden factors that can influence a user's perception of ease of use. This pushed several researchers to operationalize this generic factor by breaking it down into more explicit elements; for example, Gagnon et al [83] and Asua et al [71] assumed that factors such as users' habits (ie, behaviors that became automatized because of a person's habit of doing it repetitively) could influence their perception of a tool's ease of use; whereas Liu and Cheng [81] anticipated that mobility (location flexibility) could impact perceived ease of use. Furthermore, Ray et al [45] proposed that it is influenced by a whole cluster of contextual factors as well as elements such as resource availability, whereas Rho et al [70] proposed that it is impacted by self-efficacy and accessibility of medical records, and Sezgin et al [82] suggested that elements such as demonstrability of results, personal innovativeness, mobile anxiety (fear of telephones), mobile self-efficacy, and habit can have an impact on users' effort expectancy.

The cluster information technology (IT) capability and compatibility includes factors such as technical support (n=12), interoperability and integration (n=8), IT infrastructure (8), and technical and connectivity issues (n=5). These factors can clearly influence adoption; however, they were not specifically spelled out in the frameworks used. The availability of good technical support and collaboration with the IT department may enable better adoption [44,46,47,71,73,79,82,83], whereas the lack of technical support was reported as a barrier [51,68]. Similarly, if the tool is not integrated into or interoperable with the hospital or clinical systems, it becomes harder to adopt it as it usually creates an increased workload (eg, inputting data), and therefore, the potential for information to be lost [44,47,53,78]. Several studies also reported that an insufficient IT infrastructure could be a barrier [42,53,62,73,76,83] and some specified difficulties such as technical and connectivity issues that seem to be a common problem in some hospitals [50,62,67,68,79].

Similarly, the data-related factors cluster such as data privacy and security (n=17), and elements such as data accuracy, management, or storage (n=7) can clearly impact adoption in the health care setting but were not specifically defined as distinct constructs in the frameworks used. Such factors are quite specific to the health care ecosystem and are crucial for successful implementation. This is mainly driven by the highly regulated nature of the health care sector and the sensitivity of health data that pushes clinicians to be extra cautious when dealing with any system that captures or stores patient data, highlighting the importance of data privacy and security concerns [37,47,53,56,58,62,75], patient confidentiality [48,55,65,68], and the related medicolegal liability [35,42,44].

Monetary factors such as costs (CFIR) and affordability (APEASE) were reported in eight studies; cost, financial constraints, and uncertainty of future funding were usually perceived as barriers to sustained implementation [39,53,64]. User experience factors such as design quality (CFIR), content, and source (CFIR) were mentioned in three studies; they show that aspects such as the trustworthiness of the tool's content and where it is coming from may influence adoption [54,55,67].

Social and Personal Factors

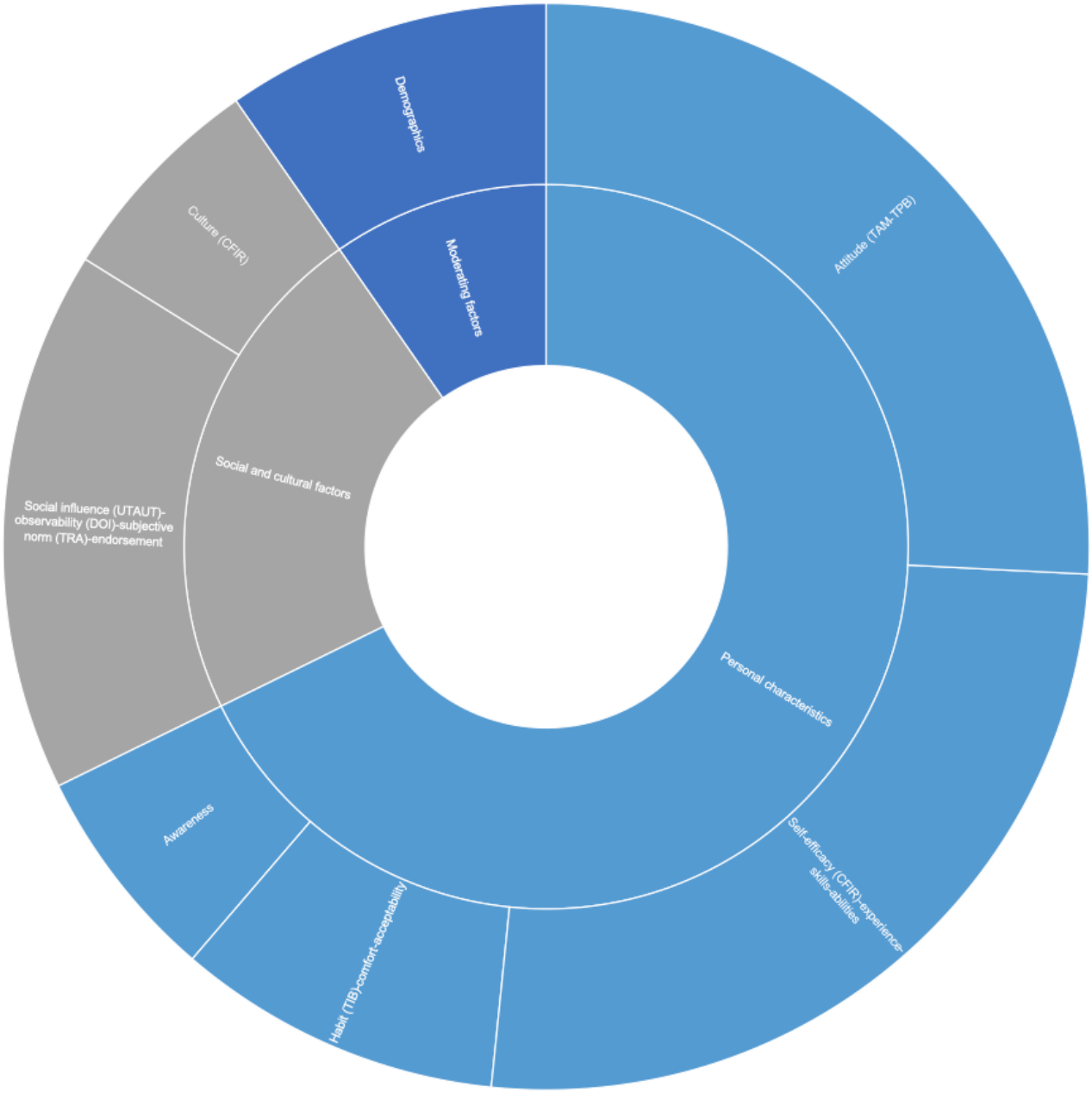
The personal characteristics of adopters were reported in more studies compared with social and cultural factors. Aspects such as self-efficacy (CFIR) were reported in 16 studies, highlighting the importance of the clinicians' previous experiences, technical skills, knowledge, and abilities in general [36,39,42,53,58,59,64,70,85]. These factors are often linked to the users' perceived ease of use [61,82]. Similarly, attitude (TAM-TPB) was reported in 16 studies, demonstrating the significance of how the positive or negative feelings users have about using such tools may influence their decision to adopt [36,41,42,53,59,62,69,75,80]. Nevertheless, it was also reported that a positive attitude is not enough to ensure adoption [73]. The importance of factors such as habit (TIB), comfort, and acceptability was reported in six studies and can be influenced by other elements such as trust and mobile use in personal life [15,52,55]. It was also reported that anxiety in using mHealth tools might influence adoption negatively [82].

Social factors (n=10) such as the influence of others on a user's decision to adopt or the extent to which users believe that people who are important to them will approve their adoption of a particular behavior were taken into account in some of the frameworks used under different names: social influence (UTAUT), observability (DOI), subjective norm (TRA), and endorsement. This shows that factors such as linking the app to a reputable organization, coworkers, opinion leaders, image within the organization, and observing other clinicians using such tools may impact adoption [43,57,63-65,74,78,79].

Conversely, some studies found the influence of such social factors to be insignificant [66]. At the same time, culture (CFIR) came up in four studies, explaining how a conservative or risk-averse culture could be a barrier to adoption [48,53,62,67].

Although demographics are not an initial construct in any of the frameworks used, there were mixed results in the studies that used them as moderating factors (n=6). There were studies that found that elements such as age or gender may have an impact on the decision to adopt [37,59,60], whereas some studies did not find personal demographics to be a significant factor [42,43,64]. An overview of the social and personal factors is shown in Figure 3.

Figure 3. The most cited social and personal factors. CFIR: consolidated framework for implementation Research; DOI: diffusion of innovation theory; TAM: Technology Acceptance Model; TIB: theory of interpersonal behavior; TPB: theory of planned behavior; TRA: theory of reasoned action; UTAUT: unified theory of acceptance and use of technology.



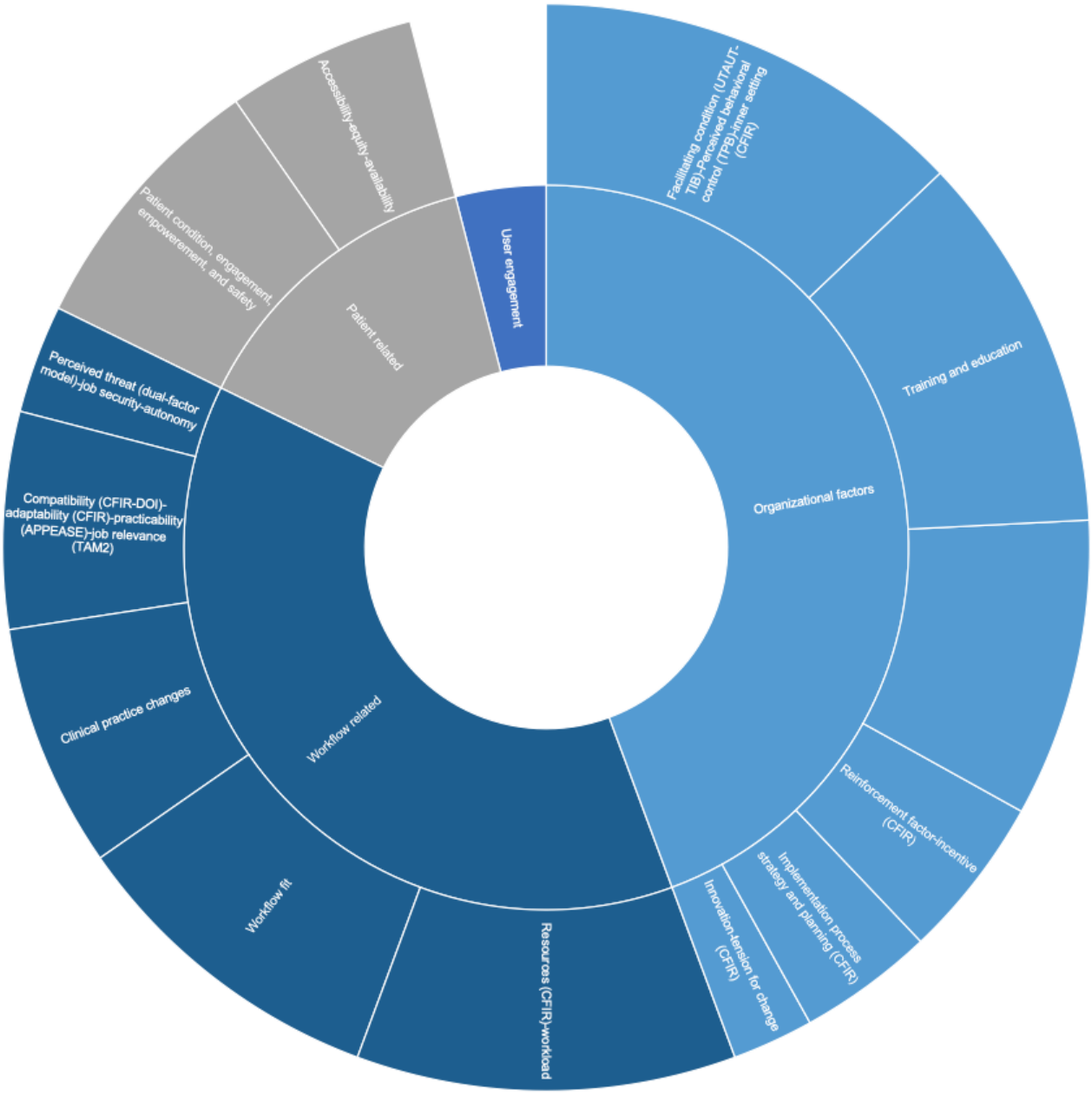
Organizational and Policy Factors

As shown in Figure 4, the organizational and policy factors were divided into 5 key clusters: organizational, workflow-related, policy and regulations, patient-related, and user engagement factors.

Organizational factors (n=16) were taken into account in some of the frameworks used under different names: facilitating conditions (UTAUT-TIB), which is the users' perception that organizational support and technical facilitating conditions are in place to aid technology use; perceived behavioral control (TPB) designating the perceptions of the availability of skills, opportunities, and resources required for using technology; and inner setting (CFIR) defining the qualities of the organization in which the intervention is implemented. Many studies reported that such organizational facilitators are among the most important factors influencing adoption [41-43,53,65,66,71,73,79,80,83] and explained how the lack of

such conditions can hinder adoption [48,67]. More specific factors such as training and education were also explicitly cited in 14 studies [15,39,41,64,68,71,73,82,83], showing their importance, but were not spelled out as a specific construct in most of the frameworks used.

Figure 4. The most cited organizational and policy factors. APEASE: affordability, practicability, effectiveness, acceptability, safety/side effects, and equity; CFIR: consolidated framework for implementation research; DOI: diffusion of innovation theory; TAM2: an expanded version of the original Technology Acceptance Model; TIB: theory of interpersonal behavior; TPB: theory of planned behavior; TRA: theory of reasoned action; UTAUT: unified theory of acceptance and use of technology.



Leadership engagement (CFIR) and management support were other important factors identified in 11 studies. The findings show that top management approval, engagement, and support [15,39,41,47,53,62,67], as well as professional organizations and societies [85] are vital for implementation success and adoption. Reinforcement factors and incentives (CFIR) were reported in six studies as elements that may positively impact adoption [45,48,66,70]. Implementation process, strategy, and planning (CFIR) was reported in five studies, including ideas such as a step-wise approach or piloting the tools before a complete rollout [53,56,58,67]. Factors such as innovation and tension for change (CFIR), which indicate the degree to which stakeholders perceive the current situation as intolerable or needing change, were mentioned in three studies [49,67,75].

The workflow-related cluster was separated from the other organizational factors because of the many factors relating to workflow, which necessitated distinguishing them and highlighting their specificity. Although resource, in general, is a construct in the CFIR, workload was more specifically mentioned in many of the included studies (n=14), showing that an elevated workload or the perception that such tools would increase workload may hinder adoption

[46,47,49-51,55,67,68,77,79,85]. On the other hand, the users' perception that the tool may ease the existing workload was also found to encourage adoption [52,53].

Workflow fit, which indicates how the tool would fit into and improve the day-to-day work of clinical staff, is another specific factor that was not defined as a distinct construct in most frameworks used, although it was mentioned in 12 studies [39,44,46,49,52,60,77,84,86]. Similarly, changes to clinical practices were specifically reported in nine studies, indicating that mHealth adoption sometimes required some process changes and adaptation to clinical practice, and this may hinder adoption if not addressed properly [15,41,45,49,53,71,77,87].

Compatibility (CFIR-DOI), adaptability (CFIR), practicability (APEASE), and job relevance (TAM2) were mentioned in eight studies and mostly related to the degree to which mHealth fits into, can be adapted, tailored, or refined to meet local needs [39,40,57,61,62,67,71,82]. Furthermore, four studies discussed a factor that was not specified as a distinct construct in most of the frameworks used, which is the impact of the perception of job security, job autonomy, perceived threat (dual-factor model), and empowerment. This latter factor shows that if clinicians perceive mHealth as a threat to their job autonomy or security or makes them feel that they are losing control of their work, they would resist its adoption [48,51,77,81].

We clustered the policy and regulations factors separately because they are external factors that go beyond the control of the organization. These are crucial in the health care setting, given their highly regulated nature but were hardly spelled out as specific constructs in most of the frameworks used, with the exception of CFIR. External policies (CFIR) and regulations were mentioned in eight studies and shed light on the importance of developing policies, clear guidelines, and legislative changes that would encourage adoption [15,35,39,53,62,76,78,85]. Reimbursement policies and funding were specifically mentioned in three studies [35,53,70], whereas two studies explained how the level of standardization in health care can facilitate or hinder adoption [58,78].

Patient-related factors were hardly defined as specific constructs in most of the frameworks used, despite the fact that various studies reported that they may impact the clinicians' adoption of mHealth. Factors such as patient condition and whether they are capable of using the tool [47,51,53], their engagement, or lack of it [41,47,67,78], whether the tool facilitates patient empowerment [47,53], and whether it impacts patient safety [48] may influence a clinician's decision on whether to adopt a specific mHealth tool. Furthermore, the tool's availability and accessibility to all patients equally was mentioned in seven studies as a factor that may impact clinicians' decision to adopt such a tool [35,40,47,52,55,70,82].

User engagement is another factor that was not defined as a distinct construct in the most frameworks used but specifically prevailed in five studies. The findings show that engaging the users in the design, development, and implementation of such tools can help encourage adoption [39,46,53,60,67].

Discussion

Framework-Based Versus Additional Factors

Expanded models allowed researchers in most of the included studies to examine potentially significant factors not specified in the original theories; nevertheless, some scholars have criticized such an approach of arbitrarily adding variables, as it leads to inconsistent use of theory [88].

Sometimes, the framework-based factors appear oversimplified or not precise enough. For example, if we say that usefulness is a key factor, we need to better understand the factors that influence the user's perception of usefulness; without such specificity, it is difficult to take specific actions. This is perhaps why many of the included studies operationalized such variables by breaking them down into more specific elements, for example, by asking whether the tool was useful to the job or increased job efficiency. Such examples can be seen in studies by Gagnon et al [83] and Orruno et al [73], both of which assumed that compatibility with existing work practice would impact the perception of tool usefulness. Moreover, Liu and Cheng [81] anticipated that a user's perception that a tool is a threat to job security would also impact their perception of usefulness. Similarly, Ray et al [45] proposed that there is, in fact, a whole cluster of contextual elements (organizational and individual factors) that influence perceived usefulness. The following sections discuss each of the frameworks used in more detail.

Technology Acceptance Model and Unified Theory of Acceptance and Use of Technology Expansions

The TAM was used in 34% of the studied sample; Davis developed it in the late 1980s [89] based on the principles from Fishbein and Ajzen's TRA [90]. The TAM assumes that perceived usefulness, and perceived ease of use are key predictors of the attitude toward using a new technology, which in turn determines the behavioral intention to use that can be translated into technology acceptance. Various scholars suggested extensions of the original TAM to overcome some of its limitations; notably, Holden and Karsh [91] suggested the addition of individual user factors, organizational readiness, or trust; Venkatesh and Davis [92] extended the model and referred to it as TAM2, which included factors that were considered to influence perceived usefulness, such as subjective norm, image, voluntariness of use, or job relevance. Venkatesh and Bala [93] extended the model further to TAM3 by adding computer anxiety or enjoyment.

There are many similarities between TAM and UTAUT, as the latter was based on the former. UTAUT was used in 12% of the sample; it was first published by Venkatesh et al [94] by analyzing and comparing TAM, TAM2, TRA, and DOI in an attempt to attain a unified TAM. The resulting model describes four key constructs:

1. Performance expectancy: the user's belief that the tool will be useful for their job, which matches perceived usefulness in TAM.

2. Effort expectancy: the belief that the tool is user friendly and easy to use, which matches perceived ease of use in TAM.
3. Social influence: the degree to which users believe that important others think they should use the tool.
4. Facilitating conditions: the degree to which users think that an organizational and technical infrastructure exists to support the tool's use.

Behavioral intention to use the technology is determined by performance expectancy, effort expectancy, and social influence. Actual usage is then determined by behavioral intention to use as well as the facilitating conditions. The model also adds moderating factors such as gender, age, experience, and voluntariness of use.

Most studies in our sample, where TAM or UTAUT were used, have added factors to extend the original frameworks to better adapt them to the context of health care. This demonstrates that despite their appeal, both frameworks frequently need some sort of extension to be applicable to complex health care settings. We briefly explain below the most significant expansion examples from our studied sample to emphasize the factors that researchers added to complement these 2 frameworks and their significance.

Orruno et al [73] and Gagnon et al [83] used very similar expansions of the TAM. They expanded the model with elements from the TIB and TRA to take into account the impact of the social environment and external variables as well as using Chau and Hu's model [95] to further subdivide factors into individual, technological, and organizational contexts. They also added compatibility and technological context along with habit and facilitators as proposed by TIB to account for automatized behavior and organizational infrastructure, respectively. They also included subjective norms from the TRA to assess whether users believe that people who are important to them will approve of their adoption decision. Both papers showed that facilitators were the most significant factor in the modified model, thus emphasizing the importance of taking the organizational context into account.

Saigi-Rubio [74,75] used similar expansions in their two papers, one expanded the TAM with elements from the TPB and TRA and the other added elements from the DOI and technology readiness. In their first paper [75], they added three key factors: optimism (the degree to which clinicians consider that technology will enable them to obtain benefits or reduce effort), the propensity to innovate (an individual's tendency to innovate in their daily work), and the level of information and communications technology (ICT) use (the degree to which an individual uses technology in their personal life). In their second paper [74], they also included personal technology use as a factor, calling it an *ICT user profile*, and added other factors such as security and confidentiality, improved quality, and reduced costs. They also considered the influence of patients, medical staff, and health care boards by assessing their impact on subjective norms. The papers showed that cost reduction, data security, and technology use in personal life were among the most important factors.

Asua [71] expanded the TAM with elements of the DOI and TIB; they justified this expansion with findings of some of their previous research, which showed that facilitating conditions were the most significant predictor of the clinicians' intention to use telemonitoring. Their expansion confirmed that the facilitators in the organizational context are the most essential elements to consider for boosting mHealth adoption by clinicians. They categorized their constructs based on Chau and Hu's model of telemedicine acceptance [95] that encompasses the individual, technological, and organizational contexts; whereas Ray et al [45] expanded the TAM using the organizational readiness for change model [96] and also shed some light on contextual factors. They added local elements such as meta-organizational, intraorganizational, and individual factors to their model. Their findings show that factors such as patient-specific education, clinical protocols for use, decreasing response times, and technology simplification may encourage adoption.

Liu and Cheng [81] expanded the TAM using the dual-factor model that considers both positive and negative factors by incorporating constructs that might hinder technological adoption, not only positive constructs such as TAM's usefulness and ease of use. They were guided by Walter and Lopez [97], who suggested that the perceived threat to professional autonomy can influence the clinicians' decision to adopt. They also added the factor of perceived mobility that was suggested by Huang et al [98], as they found that it can have a positive impact on perceived usefulness; this was also confirmed by Liu's findings.

Rho et al [70] expanded the TAM based on previous research by adding three key factors: accessibility of medical records and of patients as clinical factors, self-efficacy as an individual factor, and incentives as regulatory factors. Although Adenuga et al [66] expanded the UTAUT with the construct reinforcement factor to draw attention to the influence of direct incentives on adoption. In particular, financial incentives were reported as a significant factor in previous research, and their own research showed that reinforcement factors do indeed appear to have significant effects on the clinicians' decision to adopt.

Dünnebeil et al [58] used a combination of UTAUT and TAM and expanded the model with additional factors that may impact perceived usefulness and perceived ease of use of electronic health (eHealth). They based their additions on factors that were identified as significant in most of the published research on TAM. They added factors that can impact usefulness, such as the intensity of IT utilization, the importance of data security, and the importance of documentation. They also identified factors that could influence ease of use, such as eHealth knowledge and the importance of standardization and process orientation. Similarly, Sezgin et al [82] used a combination of UTAUT and TAM and expanded them with elements from TPB and DOI. They added factors such as compatibility and personal innovativeness in the domain of IT from the DOI and computer self-efficacy and anxiety from TAM3. They also redefined some of the constructs to make them more suitable for mobile apps.

Diffusion of Innovation Theory, Consolidated Framework for Implementation Research, and Sociotechnical Theory

In DOI theory, Rogers [99] goes beyond the technical aspects to argue that, ultimately, technologies that fit well into their context of use are more easily adopted than those that are not, it was the theoretical base for 16% of the studied sample and used together with TAM in another 6%. Rogers defines the five features of an innovation that will encourage its successful adoption; its relative advantage and the perception that it is better than the current processes, its compatibility with the adopters' needs and experiences, its complexity and difficulty of use, its trialability, and the extent to which it can be experimented with on a limited basis, and the ease of the observability of its results.

The fact that the DOI takes the context into account makes it an attractive expansion choice to the TAM in as much as it complements and compensates for the TAM's contextual gaps. In our sample, Putzer and Park [42,43] used elements from both frameworks in their two studies, guided by Kwon and Zmund, who suggested some modifications to the DOI with application research to make it more suitable for studying IT [100]. They added factors such as internal and external environment as well as moderating factors such as user demographics. They also removed trialability to reduce possible confusion with observability.

Other expansions of the DOI were also used. Han et al [76] expand the theory with Berwick's model [101] that assumes that there are three groups of factors that influence adoption; first, the perception of whether the innovation is beneficial; second, the *types* of adopters themselves such as innovators, early adopters, early majority, late majority, and laggards; and third, contextual factors such as the social setting. Abd Ghani and Jaber [62] expand the theory with the technology-organization-environment model [102], as it provides a good understanding of adoption while balancing the flexibility of identifying specific factors emerging from individual contexts. This particular expansion also includes the social exchange theory [103], as it takes into account elements such as power and trust, shedding light on the importance of factors such as top management support. In addition, taken into account are technological, organizational, environmental, and individual characteristics.

The CFIR [104,105], used by 6% of the studies, goes even further to include implementation factors. It is a comprehensive framework with a list of 39 constructs, clustered in five key areas: features of the tool or intervention, characteristics of the users, qualities of the organization in which the intervention is implemented, the wider community and the social setting within which organizations function, and the implementation process itself. Given the extensive nature of this framework, it was used without expansion. Possemato et al [41] chose this framework because of its multidimensional nature that considers organizational and patient factors, and Varsi et al [67] used it because of its comprehensiveness, which enables it to capture the complexity of technology implementation in a health care setting and its ability to help identify barriers and facilitators that influence such implementations. However, despite its comprehensibility, our findings show that there were still some shortcomings, such as the lack of very specific factors such as interoperability, reimbursement, and data-related issues in the initial list of constructs.

Sociotechnical theory [106,107] focuses on how the social and technical aspects of a workplace fit together, and it was used in 4% of the studies. Ahmad et al [55] and Casey et al [49] chose it to help them investigate the factors that come into play when technological tools are put into practice. This framework views organizations as systems that include interconnected social and technical subsystems that interact together, thus necessitating the integration and coordination of social and technical elements to achieve joint optimization.

Less Frequently Used Frameworks

There were also frameworks that only appeared in 1 publication. Nevertheless, many of these frameworks shed light on the importance of social and organizational aspects and employ theories that go beyond individual adoption decisions to also reflect the organizational context and the related implementation issues that may hinder adoption.

Various frameworks take into account the interactions that occur between the technical and social or organizational factors, highlighting the entanglement between those aspects. Bagot et al [15] employed the NPT [108,109], as it suggests a nonlinear understanding of adoption, taking into account the interdependent connections within an organizational context. The theory focuses on 4 key components: coherence (the process that people go through when attempting to understand a new set of practices), cognitive participation (the relational work that people do to build and sustain a new practice), collective action (the operational work that people do to enact a new practice), and reflexive monitoring (subsequent appraisal work undertaken to the effect of new practices). Their findings revealed that telemedicine necessitates both modifications in work practice and the development of new skills to enable successful implementation.

Similarly, Grünloh et al [77] wanted to not only understand clinicians' views on technology but also its impact on their work environment; therefore, they used the TF model [110] focusing on the *assumptions*, *expectations*, *and knowledge* people use to comprehend technologies in their organizational context. In this framework, Orlikowski and Gash [110] recognized three key frames that can help explain users' views: the nature of technology (users' understanding of the technology and its functionality), the technology strategy (the motivation behind technology adoption and its worth to the organization), and technology in use (how the technology is being used on a daily basis and the implication of its use). The model helped the researchers identify work-related aspects such as the importance of processes, workload, and control (concerns that mHealth may lead to patients monitoring and controlling the physicians) when studying clinicians' decision to adopt.

Some other frameworks had more emphasis on the role of the users and the inclusion of all relevant stakeholders. Bidmead and Marshall [47] used the Stakeholder Empowered Adoption

Model [111] to ensure the inclusion of all relevant stakeholder groups in their study. The model categorizes stakeholders into four groups: professional users, patient users, organizational management, and technology providers. The study captured the views of all stakeholder groups to better understand barriers to adoption and to help each group recognize the perspectives of the other relevant stakeholders. Their findings showed that privacy concerns, risk aversion, and data integration issues are the key barriers hindering adoption. Kuo et al [80] employed the TPB [112] using three key factors: attitude (users' feelings about using telemedicine), subjective norms (whether the user should or should not use telemedicine), and perceived behavioral control (the availability of the skills, opportunities, and resources required for using telemedicine). Their results show that the three factors have an impact on technological adoption.

Sharma et al [51] used Gidden's structuration theory [113-115] to help them tackle issues such as trust and ontological security, which are not really addressed in most other frameworks. They were guided in their thinking by Kouroubali's argument that a better understanding of such social aspects can help resolve potential conflict and contradiction, and enable successful implementation [116]. Their findings emphasized some key barriers such as insufficient training, support, and lack of information, highlighting the importance of gaining clinicians' trust and promoting a sense of security to encourage telehealth adoption.

Bramley et al [48] took a different approach by looking into the potential for change in the organizational culture that they were examining. They employed the theory of change [117] and investigated the factors that would help the key stakeholders be less resistant to change. They focused on three levels of stakeholders: the macrolevel, which focuses on how commissioning authorities work with technology providers and other organizations; the mesolevel, comprising providers working with clinicians and health care providers; and the microlevel, in which the cocreation and delivery of care packages occur. Their findings showed that elements such as staff buy-in, culture, privacy concerns, safety issues, leadership engagement, data analysis, and local setting are crucial for successful implementation. Similarly, Shaw et al [46] addressed the topic from an organizational change perspective employing the Weiner organizational theory of implementation effectiveness [96,118]. The model can help identify the factors impacting an organization's readiness for change, which refers to its members' collective agreement and the ability to implement a specific change. Such readiness depends on two key concepts: change valence, which refers to the organization's perception of factors such as the fit, advantage, and need for this change and informational assessment, which refers to elements such as information about the workload needs; and the availability of resources to implement the change. The study identified organizational barriers related to additional workload, lack of resources, and workflow integration.

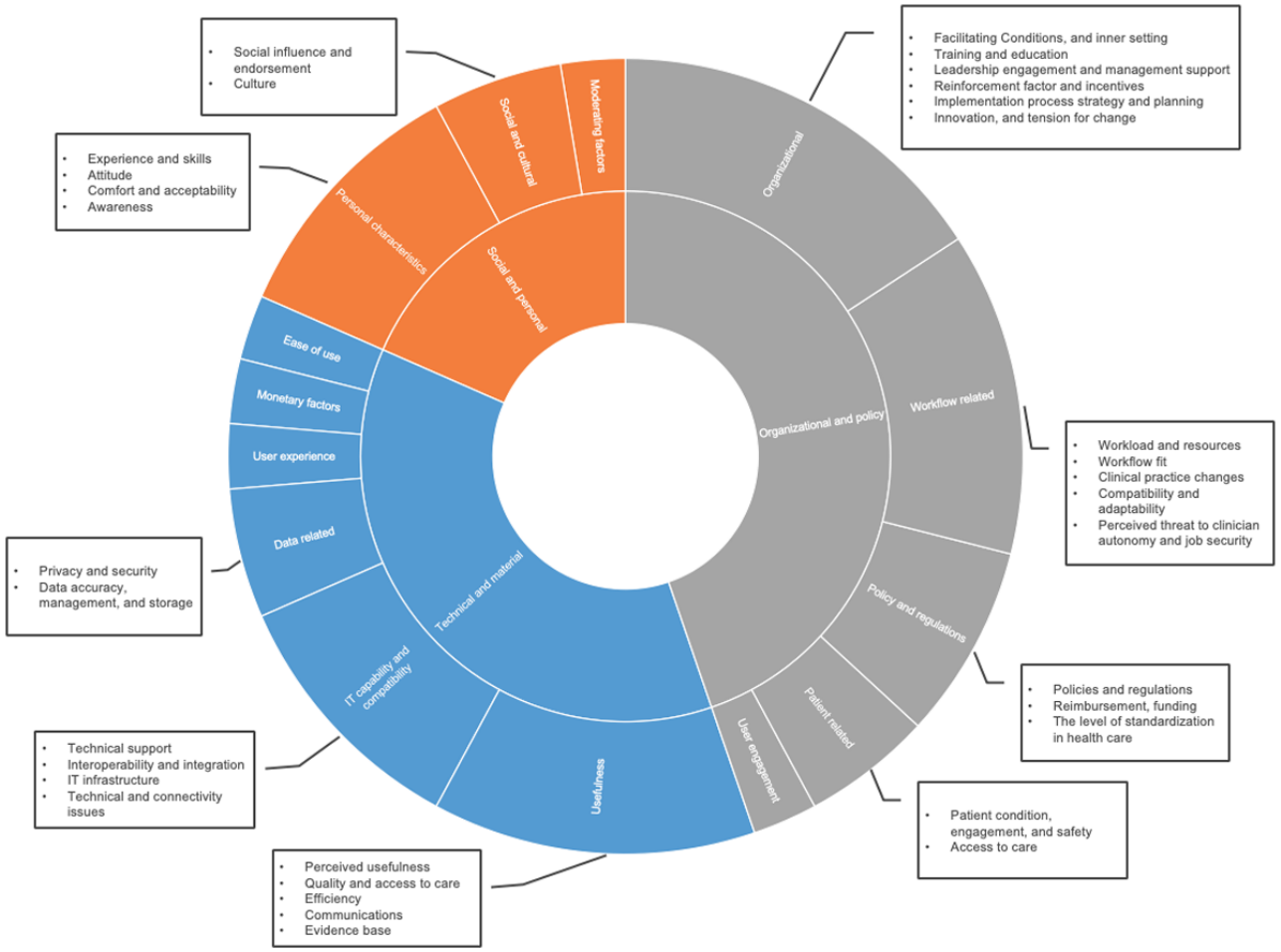
Furthermore, some theories had more focus on the design process itself. Egerton et al [52] used the APEASE framework [119], which is usually used to evaluate the design of new interventions and services. Their study identified some adoption barriers related to lack of information and support, disconnect with other programs, sense of loss of control, and a lack of familiarity with the new tool, resulting in resistance to change. Lapão [85], on the other hand, employed the DSRM [120,121] to help them study the link between research and practice by designing, implementing, and evaluating a tool that tackles a specific need. The methodology starts with problem identification and then moves through the definition of the objectives that result in the design and development of the tool, followed by testing and evaluation. Their findings show that the key barriers for implementation are the lack of time, skills, and clear role definition.

Quanbeck et al [44] focused on the implementation process itself and used the RE-AIM framework [122] to achieve that. They identified factors such as system integration, cost, and data management as key barriers for implementation. Furthermore, Okazaki et al [63] expanded the DeLone and McLean information system success model [123] to help them assess the success of a new technology in a specific organizational context. They used the original constructs of system quality, information quality, service quality, net benefits, and user satisfaction and added some additional factors such as privacy and security risks, ubiquitous control (time and place flexibility), and subjective norms from TPB. Their results showed that perceived value and net benefits are the most important factors impacting adoption.

Conclusions and Recommendations

Health care technologies are generally more complex than tools that address one specific user need, as they usually support patients with comorbidities that are typically treated by multidisciplinary teams, potentially working across more than one health care organization. The specific characteristics of how the health care sector operates (its highly regulated nature, ever-present budget deficits, and the interdependence between health care organizations) necessitate some crucial extensions to the existing generic frameworks for studying adoption. This paper sheds light on these specificities and makes recommendations regarding the important factors that should not be overlooked when working on mHealth adoption in health care. These include stringent data privacy, workload, workflow, communication, management support, policies, and complex external rules and regulations, as presented in our consolidated framework of emerging framework-based and additional factors in Figure 5.

Figure 5. Consolidated framework of the factors impacting clinicians' adaptation of mobile health. IT: information technology.



Most factors can be related to one framework or another, but no framework covers all factor categories without being extended. Although most models include relevant constructs, many of them do not break these down into their individual components to enable researchers to examine the specific reasons behind particular implementation or adoption issues and so tend not to help in the identification of specific solutions.

Some of the commonly used frameworks, such as TAM, present an oversimplified group of factors that need to be looked at in much more detail, and with greater specificity, if we are to develop and implement successful mHealth tools. Although it is sometimes claimed that it is simplicity that makes such frameworks useful [124], even in the health care context [91,125,126], our conclusions are aligned with the findings of researchers' critical frameworks such as TAM and UTAUT for overemphasizing individual user beliefs and perceptions [127]. Most of these frameworks were not developed within a health care setting and thus overlook its organizational and regulatory complexity [91,128,129]. Socio-organizational and cultural factors are not well covered by frameworks such as TAM and UTAUT [128,129]. We also note that many of the broadly used frameworks focus on tools that can be voluntarily used by individual adopters. In contrast, most health care settings involve an organizational-level decision to roll out technologies that are made available to all staff [128,130].

We propose a shift toward extended frameworks that take into account the complexity of health care organizations, their highly regulated nature, their interdependence, and the active role of the user in impacting how technology is being used in the work environment. Other researchers have also cited the technology-centered view of many of the broadly used models that mostly focus on the tool itself [128,129,131] and suggested a move to multidimensional theories that go beyond usability and consider the surrounding contexts and implementation issues [127,128,131-134]. Health care technology cannot be successfully adopted in isolation from the broader organizational context in which it is being used; therefore, we need to adopt theoretical frameworks that consider implementation challenges in light of the complexity of the sociotechnical structure and the interplay between the technical, social, and organizational aspects. This can be achieved by using our suggested consolidated framework that highlights the existing gaps that are not specifically covered in the most frequently used frameworks and focuses on complementing them using a sociotechnical approach that enables researchers to take all the contextual factors, and, crucially, the interplay between them, into account when studying adoption.

Limitations and Future Research

Although this study contributes to the understanding of the most used frameworks in understanding clinicians' adoption of mHealth, certain limitations must be acknowledged. This review may not have included related studies that were not indexed in the searched databases, written in a language other than English, and gray literature searches that could have also permitted the

identification of further relevant insights. However, this study focused on peer-reviewed scientific papers.

Moreover, this analysis only considered published studies, and no further contacts were made with the authors of the papers to obtain extra information or to validate our thematic analysis. Consequently, it is possible that other frameworks might have been missed. Future reviews could include studies in other languages and using other frameworks.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Phases of thematic analysis.
[PDF File (Adobe PDF File), 48 KB - mhealth_v8i7e18072_app1.pdf]

Multimedia Appendix 2
Summary of the included studies.
[XLSX File (Microsoft Excel File), 91 KB - mhealth_v8i7e18072_app2.xlsx]

Multimedia Appendix 3
Characteristics of the included studies.
[PNG File , 259 KB - mhealth_v8i7e18072_app3.png]

Multimedia Appendix 4
Overview of the most used frameworks.
[PNG File , 51 KB - mhealth_v8i7e18072_app4.png]

Multimedia Appendix 5
Definitions of the most prevalent framework-based themes.
[PDF File (Adobe PDF File), 103 KB - mhealth_v8i7e18072_app5.pdf]

References

1. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011.
2. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR Mhealth Uhealth 2019 May 3;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
3. Kotronoulas G, Papadopoulou C, MacNicol L, Simpson M, Maguire R. Feasibility and acceptability of the use of patient-reported outcome measures (PROMs) in the delivery of nurse-led supportive care to people with colorectal cancer. Eur J Oncol Nurs 2017 Aug;29:115-124. [doi: 10.1016/j.ejon.2017.06.002] [Medline: 28720258]
4. Kotronoulas G, Connaghan J, Grenfell J, Gupta G, Smith L, Simpson M, et al. Employing patient-reported outcome (PRO) measures to support newly diagnosed patients with melanoma: feasibility and acceptability of a holistic needs assessment intervention. Eur J Oncol Nurs 2017 Dec;31:59-68. [doi: 10.1016/j.ejon.2017.10.002] [Medline: 29173828]
5. Kotronoulas G, Papadopoulou C, Simpson MF, McPhelim J, Mack L, Maguire R. Using patient-reported outcome measures to deliver enhanced supportive care to people with lung cancer: feasibility and acceptability of a nurse-led consultation model. Support Care Cancer 2018 Nov;26(11):3729-3737. [doi: 10.1007/s00520-018-4234-x] [Medline: 29779057]
6. Schmalz O, Jacob C, Ammann J, Liss B, Iivanainen S, Kammermann M, et al. Impact of digital patient monitoring (DPM) on quality of clinical care of cancer immunotherapy (CIT)-treated patients (pts) with advanced/metastatic non-small cell lung cancer (a/mNSCLC). Ann Oncol 2019 Dec;30:xi30-xi31 [FREE Full text] [doi: 10.1093/annonc/mdz449.039]
7. Bostrom J, Sweeney G, Whiteson J, Dodson JA. Mobile health and cardiac rehabilitation in older adults. Clin Cardiol 2020 Feb;43(2):118-126 [FREE Full text] [doi: 10.1002/clc.23306] [Medline: 31825132]
8. Albrecht U, Framke T, von Jan U. Quality awareness and its influence on the evaluation of app meta-information by physicians: validation study. JMIR Mhealth Uhealth 2019 Nov 18;7(11):e16442 [FREE Full text] [doi: 10.2196/16442] [Medline: 31738179]
9. Wernhart A, Gahbauer S, Haluza D. eHealth and telemedicine: practices and beliefs among healthcare professionals and medical students at a medical university. PLoS One 2019;14(2):e0213067 [FREE Full text] [doi: 10.1371/journal.pone.0213067] [Medline: 30818348]
10. Strotbaum V, Piobruchin M, Schreiweis B, Wiesner M, Strahwald B. Your data is gold – data donation for better healthcare? IT Info Technol 2019;61(5-6):- [FREE Full text] [doi: 10.1515/itit-2019-0024]

11. Anderson K, Francis T, Ibanez-Carrasco F, Globerman J. Physician's perceptions of telemedicine in HIV care provision: a cross-sectional web-based survey. JMIR Public Health Surveill 2017 May 30;3(2):e31 [FREE Full text] [doi: 10.2196/publichealth.6896] [Medline: 28559226]
12. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to ehealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 5;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
13. Avey JP, Hobbs RL. Dial in: fostering the use of telebehavioral health services in frontier Alaska. Psychol Serv 2013 Aug;10(3):289-297. [doi: 10.1037/a0028231] [Medline: 22924798]
14. Bhatta R, Aryal K, Ellingsen G. Opportunities and challenges of a rural-telemedicine program in Nepal. J Nepal Health Res Counc 2015;13(30):149-153. [Medline: 26744201]
15. Bagot KL, Cadilhac DA, Bladin CF, Watkins CL, Vu M, Donnan GA, VSTASTUTE investigators. Integrating acute stroke telemedicine consultations into specialists' usual practice: a qualitative analysis comparing the experience of Australia and the United Kingdom. BMC Health Serv Res 2017 Nov 21;17(1):751 [FREE Full text] [doi: 10.1186/s12913-017-2694-1] [Medline: 29157233]
16. Brewster L, Mountain G, Wessels B, Kelly C, Hawley M. Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. J Adv Nurs 2014 Jan;70(1):21-33. [doi: 10.1111/jan.12196] [Medline: 23786584]
17. Kissi J, Dai B, Dogbe CS, Banahene J, Ernest O. Predictive factors of physicians' satisfaction with telemedicine services acceptance. Health Informatics J 2019 Dec 19:- epub ahead of print. [doi: 10.1177/1460458219892162] [Medline: 31854222]
18. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
19. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
20. Boonstra A, van Offenbeek M. Towards consistent modes of e-health implementation: structurational analysis of a telecare programme's limited success. Inf Syst J 2010;20(6):537-561. [doi: 10.1111/j.1365-2575.2010.00358.x]
21. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014 Jan 23;11(2):1279-1298 [FREE Full text] [doi: 10.3390/ijerph110201279] [Medline: 24464237]
22. Newbould J, Exley J, Ball S, Corbett J, Pitchforth E, Roland M. GPs' and practice staff's views of a telephone first approach to demand management: a qualitative study in primary care. Br J Gen Pract 2019 May;69(682):e321-e328 [FREE Full text] [doi: 10.3399/bjgp19X702401] [Medline: 31015225]
23. Doak S, Schwager A, Hensel JM. Exposure to and attitudes regarding electronic healthcare (e-health) among physician assistants in Canada: a national survey study. Health Soc Care Community 2020 May;28(3):1049-1057. [doi: 10.1111/hsc.12938] [Medline: 31899587]
24. Goffman E. The Presentation of Self in Everyday Life. Garden City, New York, USA: Doubleday Anchor; 1956.
25. Pinch T. The invisible technologies of Goffman's sociology from the merry-go-round to the internet. Technol Cult 2010;51(2):409-424. [doi: 10.1353/tech.0.0456]
26. Garfinkel H. Studies in Ethnomethodology. Englewoods Cliffs, NJ: Prentice Hall; 1967.
27. Leonardi PM. Methodological guidelines for the study of materiality and affordances. In: Raza M, Jain S, editors. The Routledge Companion to Qualitative Research in Organization Studies. New York, USA: Routledge; 2018:279-290.
28. Bélanger F, Watson-Manheim MB. Virtual teams and multiple media: structuring media use to attain strategic goals. Group Decis Negot 2006 Jul 11;15(4):299-321. [doi: 10.1007/s10726-006-9044-8]
29. Chudoba K, Wynn E, Lu M, Watson-Manheim M. How virtual are we? Measuring virtuality and understanding its impact in a global organization. Inf Syst J 2005;15(4):279-306. [doi: 10.1111/j.1365-2575.2005.00200.x]
30. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
31. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
32. Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, USA: Wiley; 2011.
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Merriam-Webster I. The Merriam-Webster Dictionary. Springfield, MA: Merriam-Webster Inc; 2008.
35. Jetty A, Moore MA, Coffman M, Petterson S, Bazemore A. Rural family physicians are twice as likely to use telehealth as urban family physicians. Telemed J E Health 2018 Apr;24(4):268-276. [doi: 10.1089/tmj.2017.0161] [Medline: 28805545]
36. Kowitlawakul Y. The technology acceptance model: predicting nurses' intention to use telemedicine technology (eICU). Comput Inform Nurs 2011 Jul;29(7):411-418. [doi: 10.1097/NCN.0b013e3181f9dd4a] [Medline: 20975536]
37. Smith PB, Buzi RS. Reproductive health professionals' adoption of emerging technologies for health promotion. Health Informatics J 2014 Dec;20(4):250-260. [doi: 10.1177/1460458213492525] [Medline: 25411221]

38. Williamson KM, Muckle J. Students' perception of technology use in nursing education. Comput Inform Nurs 2018 Feb;36(2):70-76. [doi: 10.1097/CIN.0000000000000396] [Medline: 29084028]
39. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a substance use recovery support mobile phone app in community settings: qualitative study of clinician and staff perspectives of facilitators and barriers. JMIR Ment Health 2016 Jun 28;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
40. Miller KE, Kuhn E, Owen JE, Taylor K, Yu JS, Weiss BJ, et al. Clinician perceptions related to the use of the CBT-I coach mobile app. Behav Sleep Med 2019;17(4):481-491. [doi: 10.1080/15402002.2017.1403326] [Medline: 29120247]
41. Possemato K, Kuhn E, Johnson EM, Hoffman JE, Brooks E. Development and refinement of a clinician intervention to facilitate primary care patient use of the PTSD Coach app. Transl Behav Med 2017 Mar;7(1):116-126 [FREE Full text] [doi: 10.1007/s13142-016-0393-9] [Medline: 27234150]
42. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the southeastern United States. Perspect Health Inf Manag 2012;9:1b [FREE Full text] [Medline: 22737094]
43. Putzer GJ, Park Y. The effects of innovation factors on smartphone adoption among nurses in community hospitals. Perspect Health Inf Manag 2010 Jan 1;7:1b [FREE Full text] [Medline: 20697467]
44. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
45. Ray KN, Felmet KA, Hamilton MF, Kuza CC, Saladino RA, Schultz BR, et al. Clinician attitudes toward adoption of pediatric emergency telemedicine in rural hospitals. Pediatr Emerg Care 2017 Apr;33(4):250-257. [doi: 10.1097/PEC.0000000000000583] [Medline: 26785087]
46. Shaw RJ, Kaufman MA, Bosworth HB, Weiner BJ, Zullig LL, Lee SD, et al. Organizational factors associated with readiness to implement and translate a primary care based telemedicine behavioral program to improve blood pressure control: the HTN-IMPROVE study. Implement Sci 2013 Sep 8;8:106 [FREE Full text] [doi: 10.1186/1748-5908-8-106] [Medline: 24010683]
47. Bidmead E, Marshall A. A case study of stakeholder perceptions of patient held records: the patients know best (PKB) solution. Digit Health 2016;2:2055207616668431 [FREE Full text] [doi: 10.1177/2055207616668431] [Medline: 29942567]
48. Bramley G, Mangan C, Conroy M. Using telemonitoring to support personal care planning for adults with learning disabilities. J Telemed Telecare 2019 Dec;25(10):602-610. [doi: 10.1177/1357633X18784419] [Medline: 30016895]
49. Casey M, Shaw S, Swinglehurst D. Experiences with online consultation systems in primary care: case study of one early adopter site. Br J Gen Pract 2017 Nov;67(664):e736-e743 [FREE Full text] [doi: 10.3399/bjgp17X693137] [Medline: 28993306]
50. L'Esperance ST, Perry DJ. Assessing advantages and barriers to telemedicine adoption in the practice setting: a MyCareTeam(TM) exemplar. J Am Assoc Nurse Pract 2016 Jun;28(6):311-319. [doi: 10.1002/2327-6924.12280] [Medline: 26178417]
51. Sharma U, Barnett J, Clarke M. Clinical users' perspective on telemonitoring of patients with long term conditions: understood through concepts of Giddens's structuration theory & consequence of modernity. Stud Health Technol Inform 2010;160(Pt 1):545-549. [Medline: 20841746]
52. Egerton T, Nelligan R, Setchell J, Atkins L, Bennell KL. General practitioners' perspectives on a proposed new model of service delivery for primary care management of knee osteoarthritis: a qualitative study. BMC Fam Pract 2017 Sep 7;18(1):85 [FREE Full text] [doi: 10.1186/s12875-017-0656-7] [Medline: 28882108]
53. Puszka S, Dingwall KM, Sweet M, Nagel T. e-Mental health innovations for Aboriginal and Torres strait islander Australians: a qualitative study of implementation needs in health services. JMIR Ment Health 2016 Sep 19;3(3):e43 [FREE Full text] [doi: 10.2196/mental.5837] [Medline: 27644259]
54. Hackl WO, Hoerbst A, Duftschmid G, Gall W, Janzek-Hawlat S, Jung M, et al. Crucial factors for the acceptance of a computerized national medication list: insights into findings from the evaluation of the Austrian e-Medikation pilot. Appl Clin Inform 2014;5(2):527-537 [FREE Full text] [doi: 10.4338/ACI-2014-04-RA-0032] [Medline: 25024766]
55. Ahmad F, Norman C, O'Campo P. What is needed to implement a computer-assisted health risk assessment tool? An exploratory concept mapping study. BMC Med Inform Decis Mak 2012 Dec 19;12:149 [FREE Full text] [doi: 10.1186/1472-6947-12-149] [Medline: 23253913]
56. Jarvis-Selinger S, Bates J, Araki Y, Lear SA. Internet-based support for cardiovascular disease management. Int J Telemed Appl 2011;2011:342582 [FREE Full text] [doi: 10.1155/2011/342582] [Medline: 21822430]
57. Zhang H, Cocosila M, Archer N. Factors of adoption of mobile information technology by homecare nurses: a technology acceptance model 2 approach. Comput Inform Nurs 2010;28(1):49-56. [doi: 10.1097/NCN.0b013e3181c0474a] [Medline: 19940621]
58. Dünnebeil S, Sunyaev A, Blohm I, Leimeister JM, Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012 Nov;81(11):746-760. [doi: 10.1016/j.ijmedinf.2012.02.002] [Medline: 22397989]

59. Sandholzer M, Deutsch T, Frese T, Winter A. Predictors of students' self-reported adoption of a smartphone application for medical education in general practice. BMC Med Educ 2015 May 21;15:91 [FREE Full text] [doi: 10.1186/s12909-015-0377-3] [Medline: 25994310]
60. Schmeer R, Behrends M, Kupka T, Meyenburg-Altwarg I, Marschollek M. Use and acceptance of mobile technology by hospital nurses in Germany. Stud Health Technol Inform 2016;225:944-945. [Medline: 27332422]
61. Tahamtan I, Pajouhanfar S, Sedghi S, Azad M, Roudbari M. Factors affecting smartphone adoption for accessing information in medical settings. Health Info Libr J 2017 Jun;34(2):134-145. [doi: 10.1111/hir.12174] [Medline: 28406547]
62. Ghani MK, Jaber MM. Willingness to adopt telemedicine in major Iraqi hospitals: a pilot study. Int J Telemed Appl 2015;2015:136591 [FREE Full text] [doi: 10.1155/2015/136591] [Medline: 26557848]
63. Okazaki S, Castañeda JA, Sanz S, Henseler J. Factors affecting mobile diabetes monitoring adoption among physicians: questionnaire study and path model. J Med Internet Res 2012 Dec 21;14(6):e183 [FREE Full text] [doi: 10.2196/jmir.2159] [Medline: 23257115]
64. van Houwelingen CT, Barakat A, Best R, Boot WR, Charness N, Kort HS. Dutch nurses' willingness to use home telehealth: implications for practice and education. J Gerontol Nurs 2015 Apr;41(4):47-56. [doi: 10.3928/00989134-20141203-01] [Medline: 25486115]
65. de Vries ST, Wong L, Sutcliffe A, Houÿez F, Ruiz CL, Mol PG, IMI Web-RADR Work Package 3b Consortium. Factors influencing the use of a mobile app for reporting adverse drug reactions and receiving safety information: a qualitative study. Drug Saf 2017 May;40(5):443-455 [FREE Full text] [doi: 10.1007/s40264-016-0494-x] [Medline: 28035492]
66. Adenuga KI, Iahad NA, Miskon S. Towards reinforcing telemedicine adoption amongst clinicians in Nigeria. Int J Med Inform 2017 Aug;104:84-96. [doi: 10.1016/j.ijmedinf.2017.05.008] [Medline: 28599820]
67. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
68. Wynn R, Bergvik S, Pettersen G, Fossum S. Clinicians' experiences with videoconferencing in psychiatry. Stud Health Technol Inform 2012;180:1218-1220. [Medline: 22874406]
69. Kim S, Lee K, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
70. Rho MJ, Choi IY, Lee J. Predictive factors of telemedicine service acceptance and behavioral intention of physicians. Int J Med Inform 2014 Aug;83(8):559-571. [doi: 10.1016/j.ijmedinf.2014.05.005] [Medline: 24961820]
71. Asua J, Orruño E, Reviriego E, Gagnon MP. Healthcare professional acceptance of telemonitoring for chronic care patients in primary care. BMC Med Inform Decis Mak 2012 Nov 30;12:139 [FREE Full text] [doi: 10.1186/1472-6947-12-139] [Medline: 23194420]
72. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
73. Orruño E, Gagnon MP, Asua J, Ben Abdeljelil A. Evaluation of teledermatology adoption by health-care professionals using a modified technology acceptance model. J Telemed Telecare 2011;17(6):303-307. [doi: 10.1258/jtt.2011.101101] [Medline: 21844171]
74. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jan;32(1-2):29-36. [doi: 10.1017/S0266462316000015] [Medline: 27472158]
75. Saigí-Rubió F, Torrent-Sellens J, Jiménez-Zarco A. Drivers of telemedicine use: comparative evidence from samples of Spanish, Colombian and Bolivian physicians. Implement Sci 2014 Oct 8;9:128 [FREE Full text] [doi: 10.1186/s13012-014-0128-6] [Medline: 25293651]
76. Han KJ, Subramanian R, Cameron GT. Listen before you leap: Sri Lankan health professionals' perspectives on m-health. Health Informatics J 2019 Sep;25(3):858-866. [doi: 10.1177/1460458217725903] [Medline: 28825329]
77. Grünloh C, Cajander A, Myreteg G. 'The record is our work tool!'-physicians' framing of a patient portal in Sweden. J Med Internet Res 2016 Jun 27;18(6):e167 [FREE Full text] [doi: 10.2196/jmir.5705] [Medline: 27349531]
78. Zhang Y, Koch S. Mobile health apps in Sweden: what do physicians recommend? Stud Health Technol Inform 2015;210:793-797. [Medline: 25991263]
79. Ehrler F, Ducloux P, Wu DT, Lovis C, Blondon K. Acceptance of a mobile application supporting nurses workflow at patient bedside: results from a pilot study. Stud Health Technol Inform 2018;247:506-510. [Medline: 29678012]
80. Kuo K, Talley PC, Lee C, Yen Y. The influence of telemedicine experience on physicians' perceptions regarding adoption. Telemed J E Health 2015 May;21(5):388-394. [doi: 10.1089/tmj.2014.0091] [Medline: 25764024]
81. Liu C, Cheng T. Exploring critical factors influencing physicians' acceptance of mobile electronic medical records based on the dual-factor model: a validation in Taiwan. BMC Med Inform Decis Mak 2015 Feb 7;15:4 [FREE Full text] [doi: 10.1186/s12911-014-0125-3] [Medline: 25889506]

82. Sezgin E, Özkan-Yildirim S, Yildirim S. Understanding the perception towards using mhealth applications in practice. Inf Dev 2016 Dec 19;34(2):182-200. [doi: 10.1177/0266666916684180]
83. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59 [FREE Full text] [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
84. Ehrler F, Wipfli R, Teodoro D, Sarrey E, Walesa M, Lovis C. Challenges in the implementation of a mobile application in clinical practice: case study in the context of an application that manages the daily interventions of nurses. JMIR Mhealth Uhealth 2013 Jun 12;1(1):e7 [FREE Full text] [doi: 10.2196/mhealth.2344] [Medline: 25100680]
85. Lapão LV, da Silva MM, Gregório J. Implementing an online pharmaceutical service using design science research. BMC Med Inform Decis Mak 2017 Mar 27;17(1):31 [FREE Full text] [doi: 10.1186/s12911-017-0428-2] [Medline: 28347304]
86. Kim J, Song Y, Park J, Lee J. Attitude of Korean primary care family physicians towards telehealth. Korean J Fam Med 2011 Sep;32(6):341-351 [FREE Full text] [doi: 10.4082/kjfm.2011.32.6.341] [Medline: 22745872]
87. Ferrari M, Ahmad F, Shakya Y, Ledwos C, McKenzie K. Computer-assisted client assessment survey for mental health: patient and health provider perspectives. BMC Health Serv Res 2016 Sep 23;16(1):516 [FREE Full text] [doi: 10.1186/s12913-016-1756-0] [Medline: 27663508]
88. Benbasat I, Barki H. Quo vadis TAM? J Assoc Inf Syst 2007 Apr;8(4):211-218. [doi: 10.17705/1jais.00126]
89. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
90. Ajzen I, Fishbein M. Attitude-behavior relations: a theoretical analysis and review of empirical research. Psychol Bull 1977;84(5):888-918. [doi: 10.1037/0033-2909.84.5.888]
91. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
92. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
93. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
94. Venkatesh V, Morris MG, Davis GD, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
95. Chau P, Hu P. Examining a model of information technology acceptance by individual professionals: An exploratory study. J Manag Inf Syst 2002 Mar;18(4):191-229. [doi: 10.1080/07421222.2002.11045699]
96. Weiner BJ. A theory of organizational readiness for change. Implement Sci 2009 Oct 19;4:67 [FREE Full text] [doi: 10.1186/1748-5908-4-67] [Medline: 19840381]
97. Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decis Support Syst 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
98. Huang J, Lin Y, Chuang S. Elucidating user behavior of mobile learning. Electron Libr 2007 Oct 9;25(5):585-598. [doi: 10.1108/02640470710829569]
99. Rogers EM. Diffusion of Innovations. Fifth Edition. New York, USA: Free Press; 2003.
100. Kwon T, Zmud R. Unifying the fragmented models of information systems implementation. Crit issues Inf Syst Res 1987;-(-):227-251. [doi: 10.5555/54905.54915]
101. Berwick DM. Disseminating innovations in health care. J Am Med Assoc 2003 Apr 16;289(15):1969-1975. [doi: 10.1001/jama.289.15.1969] [Medline: 12697800]
102. Mohanty R, Deshmukh S. Processes of Technological Innovation. Lexington, MA: Lexington Books; 1990.
103. Emerson RM. Social exchange theory. Annu Rev Sociol 1976 Aug;2(1):335-362. [doi: 10.1146/annurev.so.02.080176.002003]
104. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
105. Damschroder LJ, Hagedorn HJ. A guiding framework and approach for implementation research in substance use disorders treatment. Psychol Addict Behav 2011 Jun;25(2):194-205. [doi: 10.1037/a0022284] [Medline: 21443291]
106. Trist EL, Bamforth KW. Social exchange theorysome social and psychological consequences of the longwall method of coal-getting. Hum Relations 2016 Apr 22;4(1):3-38. [doi: 10.1177/001872675100400101]
107. Morgan G, Trist E, Murray H. The Social Engagement of Social Science, a Tavistock Anthology, Volume 3: The Socio-Ecological Perspective. Pennsylvania, USA: University of Pennsylvania Press; 1992.
108. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
109. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: normalization process theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
110. Orlikowski WJ, Gash DC. Technological frames. ACM Trans Inf Syst 1994 Apr;12(2):174-207. [doi: 10.1145/196734.196745]

111. Marshall A. Designing Telemedicine Apps That Health Commissioners Will Adopt. In: Proceedings of the 14th Conference of Open Innovation Association. 2013 Presented at: FRUCT'13; November 11-15, 2013; Espoo, Finland. [doi: 10.1109/fruct.2013.6737946]
112. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
113. Parsons HL, Giddens A. Central problems in social theory: action, structure and contradiction in social analysis. Philos Phenomenol Res 1980 Sep;41(1/2):246. [doi: 10.2307/2107414]
114. Giddens A. The Consequences of Modernity. Stanford, USA: Stanford University Press; 2012.
115. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Berkeley, CA: University of California Press; 1984.
116. Kouroubali A. Structuration Theory and Conception-reality Gaps: Addressing Cause and Effect of Implementation Outcomes in Health Care Information Systems. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. 2002 Presented at: HICSS'02; January 10, 2002; Big Island, HI, USA. [doi: 10.1109/hicss.2002.994120]
117. WK Kellogg Foundation. 2004. WK Kellogg Foundation Logic Model Development Guide URL: https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide [accessed 2020-05-14]
118. Weiner BJ, Lewis MA, Linnan LA. Using organization theory to understand the determinants of effective implementation of worksite health promotion programs. Health Educ Res 2009 Apr;24(2):292-305. [doi: 10.1093/her/cyn019] [Medline: 18469319]
119. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Surrey, UK: Silverback Publishing; 2014.
120. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst 2014 Dec 8;24(3):45-77. [doi: 10.2753/MIS0742-1222240302]
121. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q 2004;28(1):75. [doi: 10.2307/25148625]
122. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
123. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
124. Shachak A, Domb S, Borycki E, Fong N, Skyrme A, Kushniruk A, et al. A pilot study of computer-based simulation training for enhancing family medicine residents' competence in computerized settings. Stud Health Technol Inform 2015;216:506-510. [Medline: 26262102]
125. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
126. Garavand A, Mohseni M, Asadi H, Etemadi M, Moradi-Joo M, Moosavi A. Factors influencing the adoption of health information technologies: a systematic review. Electron Physician 2016 Aug;8(8):2713-2718 [FREE Full text] [doi: 10.19082/2713] [Medline: 27757179]
127. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315. [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
128. Ammenwerth E. Technology acceptance models in health informatics: TAM and UTAUT. Stud Health Technol Inform 2019 Jul 30;263:64-71. [doi: 10.3233/SHTI190111] [Medline: 31411153]
129. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
130. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
131. Ward R. The application of technology acceptance and diffusion of innovation models in healthcare informatics. Heal Policy Technol 2013 Dec;2(4):222-228. [doi: 10.1016/j.hlpt.2013.07.002]
132. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19(Suppl 3):i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
133. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
134. Karsh B. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]

Abbreviations

APEASE: affordability, practicability, effectiveness, acceptability, safety/side effects, and equity

CFIR: consolidated framework for implementation research
DOI: diffusion of innovation theory
DSRM: design science research methodology
eHealth: electronic health
IT: information technology
mHealth: mobile health
NPT: normalization process theory
STS: sociotechnical system
TAM: Technology Acceptance Model
TF: technological frames
TIB: theory of interpersonal behavior
TRA: theory of reasoned action
UTAUT: unified theory of acceptance and use of technology

Edited by A Tsanas; submitted 13.02.20; peer-reviewed by K Ray, T Muto, D Wu, K Blondon, T Vanbellingen; comments to author 27.03.20; revised version received 15.04.20; accepted 26.04.20; published 06.07.20.

Please cite as:
Jacob C, Sanchez-Vazquez A, Ivory C
Understanding Clinicians' Adoption of Mobile Health Tools: A Qualitative Review of the Most Used Frameworks
JMIR Mhealth Uhealth 2020;8(7):e18072
URL: https://mhealth.jmir.org/2020/7/e18072
doi:10.2196/18072
PMID:32442132

Review

Exclusively Digital Health Interventions Targeting Diet, Physical Activity, and Weight Gain in Pregnant Women: Systematic Review and Meta-Analysis

Alexandra Rhodes[1], MSc; Andrea D Smith[1], PhD; Paul Chadwick[1], PhD; Helen Croker[1*], PhD; Clare H Llewellyn[1*], PhD

University College London, London, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Alexandra Rhodes, MSc
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 207679 1720
Email: alexandra.rhodes.15@ucl.ac.uk

Abstract

Background: Interventions to promote a healthy diet, physical activity, and weight management during pregnancy are increasingly embracing digital technologies. Although some interventions have combined digital with interpersonal (face-to-face or telephone) delivery, others have relied exclusively on digital delivery. Exclusively digital interventions have the advantages of greater cost-effectiveness and broader reach and as such can be a valuable resource for health care providers.

Objective: This systematic review aims to focus on exclusively digital interventions to determine their effectiveness, identify behavior change techniques (BCTs), and investigate user engagement.

Methods: A total of 6 databases (Medical Literature Analysis and Retrieval System Online [MEDLINE], Excerpta Medica dataBASE [EMBASE], PsycINFO, Cumulated Index to Nursing and Allied Health Literature [CINAHL] Plus, Web of Science, and ProQuest) were searched for randomized controlled trials or pilot control trials of exclusively digital interventions to encourage healthy eating, physical activity, or appropriate weight gain during pregnancy. The outcome measures were gestational weight gain (GWG) and changes in physical activity and dietary behaviors. Study quality was assessed using the Cochrane Risk of Bias tool 2.0. Where possible, pooled effect sizes were calculated using a random effects meta-analysis.

Results: In total, 11 studies met the inclusion criteria. The risk of bias was mostly high (n=5) or moderate (n=3). Of the 11 studies, 6 reported on GWG as the primary outcome, 4 of which also measured changes in physical activity and dietary behaviors, and 5 studies focused either on dietary behaviors only (n=2) or physical activity only (n=3). The meta-analyses showed no significant benefit of interventions on total GWG for either intention-to-treat data (−0.28 kg; 95% CI −1.43 to 0.87) or per-protocol data (−0.65 kg; 95% CI −1.98 to 0.67). Substantial heterogeneity in outcome measures of change in dietary behaviors and physical activity precluded further meta-analyses. BCT coding identified 7 BCTs that were common to all effective interventions. Effective interventions averaged over twice as many BCTs from the *goals and planning*, and *feedback and monitoring* domains as ineffective interventions. Data from the 6 studies reporting on user engagement indicated a positive association between high engagement with key BCTs and greater intervention effectiveness. Interventions using proactive messaging and feedback appeared to have higher levels of engagement.

Conclusions: In contrast to interpersonal interventions, there is little evidence of the effectiveness of exclusively digital interventions to encourage a healthy diet, physical activity, or weight management during pregnancy. In this review, effective interventions used proactive messaging, such as reminders to engage in BCTs, feedback on progress, or tips, suggesting that interactivity may drive engagement and lead to greater effectiveness. Given the benefits of cost and reach of digital interventions, further research is needed to understand how to use advancing technologies to enhance user engagement and improve effectiveness.

(JMIR Mhealth Uhealth 2020;8(7):e18255) doi:10.2196/18255

KEYWORDS

gestational weight gain; digital interventions; behavior change techniques; user engagement; smartphone; mobile phone

Introduction

Background

Poor diet and lack of physical activity are 21st century problems contributing to the obesity crisis. In the United Kingdom, more than 50% of women of childbearing age are estimated to have overweight or obesity [1]. Pregnancy has been identified as a teachable moment [2], when women may be motivated to make lifestyle changes to improve their own health and the health of their unborn baby. Encouraging women to improve their diet and levels of physical activity is beneficial for not only supporting healthy gestational weight gain (GWG) and maternal health [3] but also developing behaviors that may potentially improve the health of the whole family.

In many countries, GWG is monitored, and women are given guidance for recommended levels of weight gain [4]. The most widely used GWG guidelines are the US Institute of Medicine (IOM) 2009 guidelines, where the recommended range of weight gain is based on a woman's pre-pregnancy BMI [5]. It is estimated that in many high-income countries, more than 50% of women gain excessive weight during pregnancy [6,7]. This is problematic because excessive GWG is associated with an increased risk of adverse health outcomes, such as gestational diabetes, large for gestational age babies, macrosomia, and cesarean section [6,8]. GWG is also associated with an increased risk of postpartum weight retention, which increases the likelihood of starting subsequent pregnancies with overweight or obesity [9]. The effects of excessive GWG are also believed to have an intergenerational impact, increasing the likelihood of overweight and obesity throughout the life of the baby [10,11]. In the United Kingdom, most women are not routinely weighed during pregnancy nor are they given specific advice on healthy levels of weight gain. However, they are provided with general advice to eat a healthy diet and participate in regular physical activity [12].

Interventions targeting diet, physical activity, or both to encourage healthier lifestyles during pregnancy and reduce rates of excessive GWG have been shown to be effective, with diet-only interventions leading to greater weight reductions than physical activity alone or combined interventions [13-15]. A recent review of systematic reviews and meta-analyses reported reductions in GWG between 0.7 and 1.8 kg, as well as positive effects on maternal and infant health outcomes [16]. The majority of these lifestyle interventions used interpersonal delivery, either in person or telephone. Several more recent interventions have embraced digital delivery methods, recognizing their advantages of significantly lower costs and broader reach [17]. A self-managed digital intervention that can be delivered anytime and anywhere and at a lower cost to the patient and provider can be a valuable resource for health care providers, provided it can affect positive behavior change.

In nonpregnant populations, digital interventions have been shown to be effective in changing nutritional behaviors [18], encouraging weight management [19], and improving levels of physical activity [20]. However, evidence of their effectiveness in improving lifestyle behaviors during pregnancy is mixed. In the past 3 years, 5 systematic reviews have explored various aspects of digital interventions to improve diet, increase physical activity, or manage weight during pregnancy [14,17,21-23]. Of the 4 including meta-analyses, 2 found no significant effect of the interventions [14,21], whereas one showed a significant result for limiting GWG, increasing physical activity, and reducing dietary energy intake in women with overweight or obesity [17] and another found a moderate effect on managing GWG among women of all BMIs [23]. However, except for Lau et al [17], who conducted a subgroup meta-analysis comprising 2 studies, these systematic reviews did not distinguish between interventions that combined digital with an interpersonal element of coaching or support (the majority of studies) and those that were exclusively digital. Moreover, although one review reported on usability [22], systematic reviews to date have not investigated user engagement, a vital component of self-managed digital interventions [24]. Finally, although previous systematic reviews have explored the behavior change techniques (BCTs) [25] used in this type of intervention [26], none have considered BCTs specifically within the context of digital interventions. As it cannot be assumed that BCTs have equal relevance to and effectiveness across different delivery methods, a review focused specifically on the role of BCTs in digitally delivered interventions for this population is a unique contribution to the literature.

Objectives

The aim of this systematic review was three-fold: (1) to determine the effectiveness of exclusively digital diet and physical activity interventions to improve lifestyle behaviors or avoid excessive weight gain during pregnancy; (2) to investigate user engagement with the interventions; and (3) to identify and assess the usage of BCTs within the interventions.

Methods

Review Protocol

This systematic review and meta-analysis follows the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [27] and is registered with the International Prospective Register of Systematic Reviews (CRD42019124838; see Multimedia Appendix 1 for the PRISMA checklist).

Search Strategy

A search of 6 databases (MEDLINE [Medical Literature Analysis and Retrieval System Online], PsycINFO, EMBASE, Cumulated Index to Nursing and Allied Health Literature [CINAHL] Plus, Web of Science, and ProQuest Dissertations and Theses) was conducted in February 2019 to identify relevant studies. Advanced searches of keywords and index terms covered 4 concept areas (pregnancy status, diet or physical activity intervention, digital technology, and study design) and were tailored according to each database (Multimedia Appendix 2). The Cochrane Library was also searched for related

systematic reviews. Their reference lists along with those of eligible studies were hand searched. Once duplicates had been removed, 2 authors (AR and PC) independently screened and assessed each study for eligibility based on title and abstract.

Inclusion and Exclusion Criteria

Studies were included in the study if they fulfilled the following population, interventions, comparators, and outcomes criteria.

Population

Pregnant women over the age of 18 years, of all BMIs, were included in this study. However, pregnant women with physical or mental health issues that would preclude them from participating in a diet- or physical activity–based intervention were excluded.

Interventions

Digital interventions targeting dietary behaviors or physical activity in pregnancy, with the aim of improving diet or physical activity during pregnancy or managing GWG, were included. Interventions aimed at increasing GWG were excluded. Interventions were exclusively digital and used text messages, apps or websites. Initial in-person or telephone study briefing sessions were deemed acceptable, as they seemed to reflect real-world situations in which health care professionals would introduce an intervention to pregnant women as a part of an antenatal care program. Interventions using interpersonal coaching or support beyond this were excluded, as were digital interventions delivered in a health care setting.

Comparators

Comparators were usual antenatal care, minimal interventions (ie, information only rather than active behavior change), or nondiet or physical activity interventions.

Outcomes

The primary outcomes were GWG (measured as total gain in kilos or compliance with IOM GWG guidelines [5]), changes in dietary behaviors, and changes in levels of physical activity. The secondary outcome was engagement, which was measured by intervention attrition rates and usage of the intervention features. BCTs were coded according to the BCT Taxonomy (version 1) [25].

Study Design

Only randomized controlled trials (RCTs) and randomized pilot studies were included in this review.

Data Extraction and Data Synthesis

Data extracted for the systematic review included author and date of publication, geographical region, study design, behaviors targeted and specific behavioral goals, sample size, participant information, the technology used, intervention features, the theory used, gestational week in which intervention started, length of intervention, nature of control, attrition rate, engagement levels, outcome measures, and outcomes. Data extraction was completed independently by 2 authors (AR and HC). In addition, 2 authors (AR and PC) independently coded the BCTs within each intervention according to the BCT Taxonomy (version 1) [25]. If available and required, study development papers and protocols were retrieved for this purpose. In most instances, the authors were contacted for additional information. BCTs were coded only if there was unequivocal evidence of their existence [25]. Disagreements were discussed to reach consensus.

Quality Assessment

Two authors (AR and AS) independently evaluated the risk of bias within studies using the Cochrane Collaboration Risk of Bias (RoB) 2.0 tool for assessing the risk of bias (the Cochrane Collaboration) [28]. The 5 domains evaluated were risk of bias arising from the randomization process, risk of bias because of deviations from the intended interventions, missing outcome data, risk of bias in the measurement of the outcome, and risk of bias in the selection of the reported result. Bias was classified as low risk, high risk, or some concerns according to predetermined criteria set by RoB 2.0. Rating discrepancies among the authors were resolved through discussion. The risk of bias across studies could not be evaluated because of the small number of studies included in the meta-analyses [29].

Data Analysis

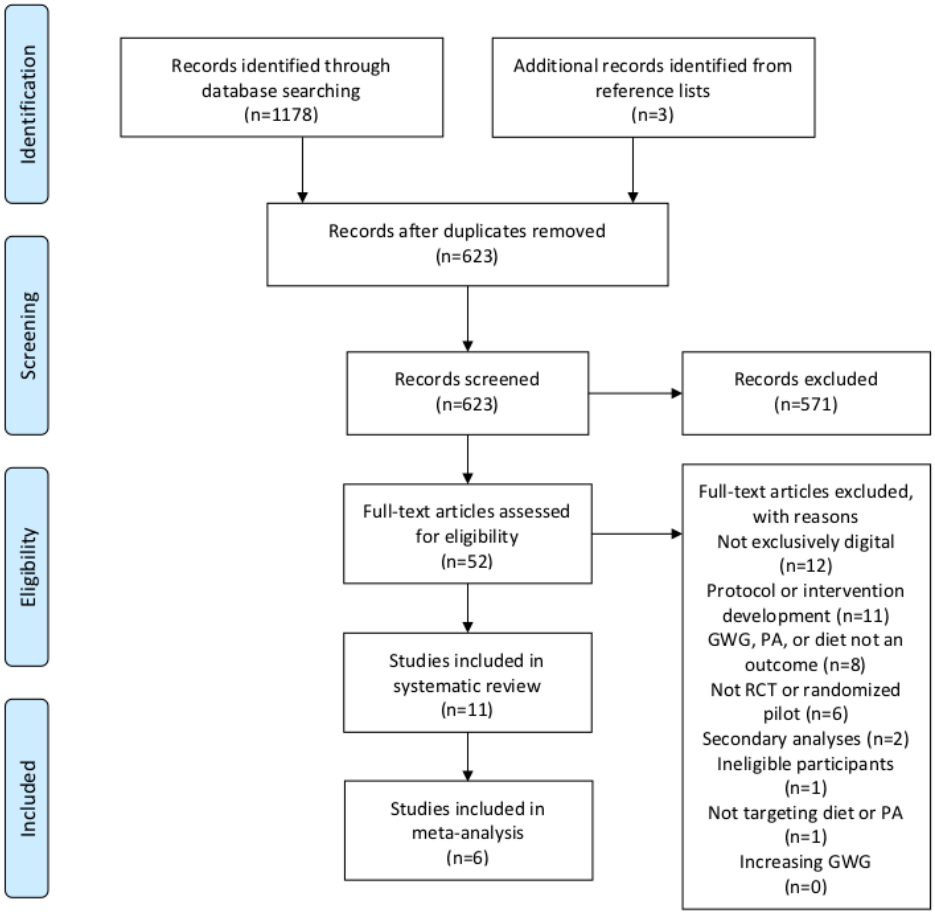
Given the substantial heterogeneity of reported outcome measures in the identified studies, data could only be quantitatively pooled for meta-analysis from studies measuring GWG. Only 4 studies used intention-to-treat (ITT) analysis; therefore, separate analyses were conducted for ITT data and per-protocol (PP) data. Meta-analysis was used to determine the differences in mean total GWG (in kg) from baseline to postintervention using the inverse variance method. The odds ratio (OR) was meta-analyzed for studies reporting GWG as a dichotomous outcome (proportion of women exceeding IOM guidelines) using the Mantel-Haenszel method. The test for the overall pooled effect estimate was assessed using Z - statistics at P=.04. Heterogeneity between studies was evaluated using the Cochran Q (chi-square test) and the I^2 statistics in the Review Manager 5.3 (the Cochrane Collaboration) [30]. Preplanned subgroup analyses were conducted comparing studies where BCTs could be identified in initial briefing sessions with those where none were apparent.

Results

Study Selection

The selection process is illustrated in Figure 1. A systematic search of 6 literature databases identified 623 nonduplicate study records. After the assessment of eligibility in accordance with the inclusion and exclusion criteria, 11 eligible studies were identified, of which 6 studies were included in subsequent meta-analyses.

Figure 1. Study selection process.



Study Characteristics

Table 1 summarizes the characteristics of the 11 studies included in this review. All studies were published between 2012 and 2019. Of the 11 studies, 7 were randomized pilot or feasibility studies [31-37] and 4 were RCTs [38-41]. Of the 4 RCTs, 2 reported being adequately powered [39,41], whereas the other 2 reported being underpowered as a result of a small starting sample [40] or low follow-up rate [38]. Overall, 9 studies took place in the United States [31-33,35,37-41] and 2 in Australia [34,36]. The sample sizes varied from 35 to 1689. In addition, 2 studies [31,38] targeted diet only, 4 studies [33,36,39,40] targeted physical activity only, and the remaining 5 studies [32,34,35,37,41] targeted both diet and physical activity. Seven studies reported on only one outcome: GWG [35,41], dietary [31,38], or physical activity [33,36,40] behaviors. The remaining 4 studies [32,34,37,39] reported on GWG, as well as changes in diet and physical activity. Three studies focused specifically on women with overweight or obesity [32,34,35], and 3 studies [33,39,40] focused on inactive or sedentary women.

The delivery method varied across studies with 4 using text messaging [31,32,38,40], 3 using an app [33,35,37], 3 using a website [36,39,41], and 1 combining text messaging with a website [34]. In total, 7 studies included an interpersonal briefing session at the start of the study [31-34,37,39,40]. In 4 studies [31,32,37,40], these sessions were for screening or study measures only, but 3 studies [33,34,39] included discussions with intervention participants about the intervention features. In one study [39], these discussions involved an in-person tutorial on how to use the website and its features and practice tracking physical activity. In another study [34], the researcher discussed individual GWG targets and weight monitoring and asked participants to set a physical activity or dietary goal. In a third study [33], a 30-min in-person session covered physical activity recommendations, goal setting, problem solving, social support, and planning for lapses.

Table 1. Study characteristics.

Authors, year	Country	Study design	Sample size, n	Technology used	Behaviors targeted	Outcomes measured	Participants
Evans et al, 2012 [31]	United States	Pilot	90	Text	Diet	Diet	Low-income, underserved pregnant women
Pollak et al, 2014 [32]	United States	Pilot	35	Text	Diet and physical activity	GWG[a], diet and physical activity	BMI 25-40 kg/m^2; gestation 12-21 weeks
Evans et al, 2015 [38]	United States	RCT[b]	996	Text	Diet	Diet	Military women; gestation <14 weeks
Smith et al, 2016 [39]	United States	RCT	51	Website	Physical activity	GWG, diet, and physical activity	Sedentary women; gestation 10-14 weeks
Choi et al, 2016 [33]	United States	Pilot	30	App	Physical activity	Physical activity	Physically inactive women; gestation 10-20 weeks
Willcox et al, 2017 [34]	Australia	Pilot	91	Text and website	Diet and physical activity	GWG, diet, and physical activity	BMI>25 kg/m^2; gestation 10-17.6 weeks
Redman et al, 2017 [35]	United States	Pilot	54	App	Diet and physical activity	GWG	BMI>25 kg/m^2; gestation 10.4-13.6 weeks
Hayman et al, 2017 [36]	Australia	Pilot	77	Website	Physical activity	Physical activity	Gestation 10-20 weeks
Huberty et al, 2017 [40]	United States	RCT	80	Text	Physical activity	Physical activity	Not meeting physical activity recommendations; gestation 8-16 weeks
Olson et al, 2018 [41]	United States	RCT	1689	Website	Diet and physical activity	GWG	BMI 18.5-35 kg/m^2; gestation <20 weeks
Dahl et al, 2018 [37]	United States	Pilot	87	App	Diet and physical activity	GWG	BMI≥18.5 kg/m^2; gestation <20 weeks

[a]GWG: gestational weight gain.

[b]RCT: randomized controlled trial.

Risk of Bias

Table 2 summarizes the study quality assessment. The overall study quality was variable. Five studies were deemed to have an overall *high risk* of bias, 3 had a *low risk* of bias, and 3 were classified as having *some concerns*.

Table 2. Risk of bias summary.

Study, year	Domain 1: risk of bias arising from the randomization process	Domain 2: risk of bias because of deviations from the intended interventions	Domain 3: missing outcome data	Domain 4: risk of bias in the measurement of the outcome	Domain 5: risk of bias in selection of the reported result	Overall risk of bias
Evans et al, 2012 [31]	High	High	High	Some concerns	Low	High
Pollak et al, 2014 [32]	High	Low	Low	Some concerns	Low	High
Evans et al, 2015 [38]	Low	Low	Low	Some concerns	Low	Some concerns
Smith et al, 2016 [39]	Low	High	High	High	Low	High
Choi et al, 2016 [33]	Low	Low	Low	Low	Low	Low
Willcox et al, 2017 [34]	Low	Some concerns	Low	Some concerns	Low	Some concerns
Redman et al, 2017 [35]	Low	Low	Low	Low	Low	Low
Hayman et al, 2017 [36]	Low	High	High	Low	Low	High
Huberty et al, 2017 [40]	Some concerns	Low	Low	Low	Low	Some concerns
Olson et al, 2018 [41]	Low	Low	Low	Low	Low	Low
Dahl et al, 2018 [37]	Some concerns	High	High	Some concerns	Low	High

Description of the Interventions

Multimedia Appendix 3 summarizes the intervention features, outcome measures, effectiveness, attrition, and engagement data. All interventions were theory-based, with social cognitive theory [42] being the most widely used (n=8). All trials started in the first or second trimester of pregnancy. The study duration varied considerably with one trial lasting 4 weeks [36], 2 trials lasting 12 weeks [33,37], one trial lasting 16 weeks [32], and the remaining trials lasting 20 weeks or more, completing at or close to term. Most studies compared interventions with usual care [31,32,34,35,38,39] or access to information-only aspects of the intervention [33,36,40,41]. In one study, the control was an equivalently structured intervention targeting stress reduction [37].

Effectiveness of Interventions

Of the 11 studies, 3 reported significant positive effects of their interventions on GWG [34,35] and physical activity [34,36] in comparison with control groups.

The 6 studies with GWG as the primary outcome varied in their measurement of total GWG. Three studies used the difference between last measured weight before delivery (34-37 weeks) and baseline weight (10-17 weeks) [34,35,41]. Two studies [37,39] used self-reported prepregnancy weight as the starting weight, and one study [32] used a model of estimated mean weights at 16 and 40 weeks. One study showed significantly lower total GWG among intervention participants [34], whereas another showed significantly fewer intervention participants exceeding the IOM guidelines [35]. The remaining studies [32,37,39,41] showed no significant difference between the intervention and control groups on any GWG measures. A meta-analysis of ITT data (n=3) showed a nonsignificant effect of the interventions, with a mean difference in total GWG of −0.28 kg (95% CI −1.43 to 0.87) using the inverse variance method and a fixed effects model (I^2=0%; P=.38; Figure 2). The mean difference in total GWG for PP data (n=4) was −0.65 kg (95% CI −1.98 to 0.67; I^2=53%; P=.10; Figure 3). The subgroup analyses revealed no significant change in this result (Multimedia Appendix 4). A meta-analysis of studies reporting PP percentages exceeding IOM guidelines showed no effect of interventions relative to comparators (OR 1.02, 95% CI 0.82-1.27; I^2=45%; P=.16; Figure 4).

Of the 7 studies reporting physical activity, 3 showed significant positive effects of the intervention on levels of physical activity [34,36,39]. Of these, one study relied on self-reported physical activity and showed significantly smaller reductions in total, light-intensity, and moderate-intensity physical activities in the intervention group compared with the control group [34]. Two other studies used smart technology to provide an objective measure of physical activity (Fitbit [36] and SenseWear Mini Arm Band [39]). One study reported a significant increase in moderate-to-vigorous physical activity in the intervention group compared with the control group, although over a 4-week period only [36]. Another study reported significantly greater levels of sustained physical activity for intervention participants compared with control participants in midpregnancy, but the

effect had disappeared by the end of the intervention [39]. Only one of the 6 studies reporting on dietary behaviors was effective in improving diet [37]. Using a self-report measure (the Rapid Eating and Activity Assessment for Participants Short Scale), intervention participants in this study scored significantly higher on healthy eating practices (measuring meal behaviors and serving frequencies) compared with the control participants.

Figure 2. Pooled analysis of digital interventions on total gestational weight gain (kg)—intention-to-treat studies.

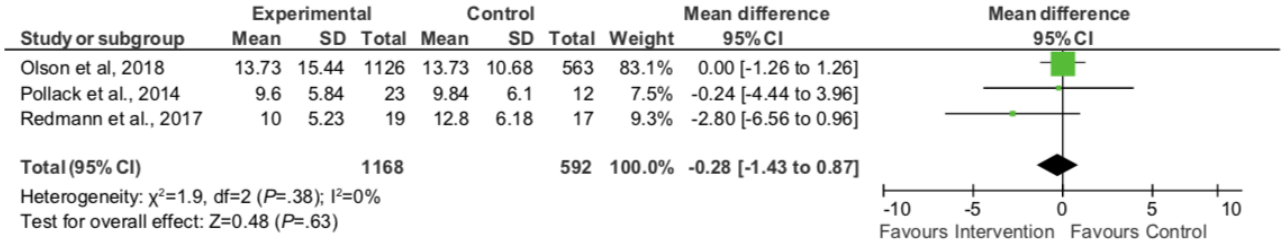


Figure 3. Pooled analysis of digital interventions on total gestational weight gain (kg)—per-protocol studies.

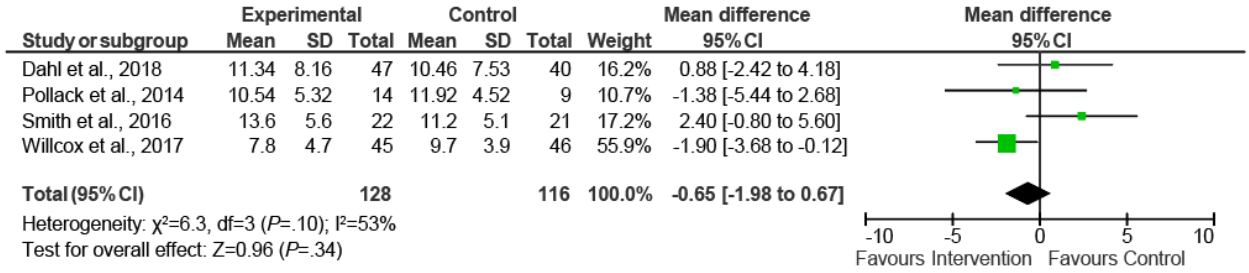
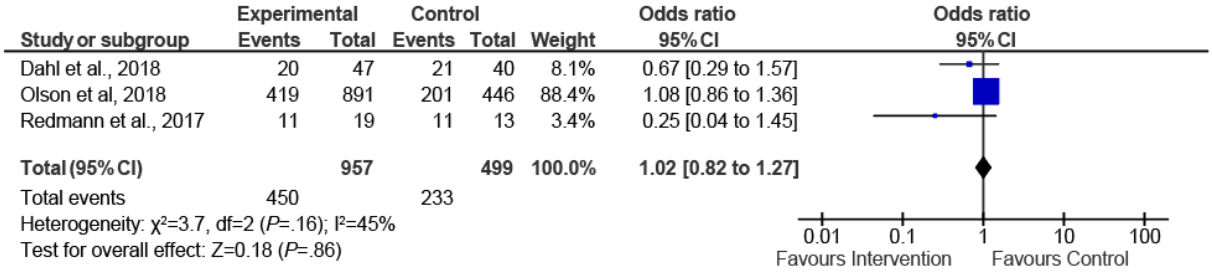


Figure 4. Pooled analysis of digital interventions on the percentage of women exceeding Institute of Medicine guidelines—per-protocol studies.



BCTs

Multimedia Appendix 5 summarizes the 23 different BCTs identified within the interventions. Only one study specified all the included BCTs [34], and this study reported the highest number of BCTs (n=17). In 2 interventions, only 1 BCT was evident [31,38]. In the remaining interventions, the number of BCTs ranged from 5 to 15. The 3 effective interventions used, on average, twice the number of BCTs compared with other interventions (mean 14, SD 2.9 vs mean 6.8, SD 4.1). Information about health consequences was the only BCT to be used in all interventions. Beyond this, goal setting (behavior) appeared in 8 interventions and problem solving and self-monitoring (behavior and outcome) in 7 interventions. Seven BCTs were common to the 3 interventions, showing a significant effect [34-36]. These were goal setting (behavior), problem solving, review of behavior goals, feedback on behavior, social support, information about health consequences, and information about emotional consequences. Review of the behavior goal was the only BCT used exclusively in the 3 effective interventions. The 3 information-only interventions that included no active or interactive BCTs, such as goal setting, self-monitoring, problem solving, or feedback, were ineffective [31,38,40].

Engagement With Interventions

Attrition rates were reported by all the studies, although 2 studies [31,40] only provided figures for all participants rather than separating intervention and control participants (Multimedia Appendix 3). Six studies reported intervention attrition rates of 10% or less [33-35,39-41]. These studies identified the reasons for dropping out of the study by distinguishing between medical and study-related reasons. Three of these studies [34,35,40] incentivized participants and 2 [33,39] introduced an element of self-selection by recruiting women who were motivated or willing to increase their physical activity. In the remaining 5 studies, intervention attrition was more than 30% and lost to follow-up reasons were not explained beyond being unable to recontact participants.

Six studies [32-36,41] reported intervention engagement levels using a variety of measures, including usage of self-monitoring,

goal setting, action planning, and social media features, response to texts, completion of tasks, and website logins (Multimedia Appendix 3). Four studies [32,34,36,41] evaluated participants' views of the intervention, although only one study [36] explored the user experience of the technology. Four studies [32,34-36] reported engagement levels over 70%, including the 3 interventions with significant effects [34-36]. A further study [33] started with a similarly high level of engagement, although it fell to below 50% over the course of the 12-week intervention. The final study [41] reported an engagement level of 46%. Five studies [32-36] integrated interactive elements to encourage engagement with the intervention. Two interventions [32,34] sent 4 or more text messages per week, encouraging self-monitoring and giving tailored feedback, and 3 studies [33,35,36] used in-app messaging to provide tailored feedback.

Discussion

Principal Findings

The aim of this systematic review was to determine the effectiveness of diet and physical activity interventions during pregnancy delivered using exclusively digital technology, and to provide insight into how BCTs and engagement with intervention features might be driving effectiveness. A total of 11 studies were identified, all of which were published from 2012 onward, with 6 studies published in 2017 and 2018. App and mobile-accessible website interventions appeared only in the last 2 years, reflecting the emergent nature of mobile health interventions to encourage healthy behaviors during pregnancy. Meta-analyses showed no significant benefit of exclusively digital interventions on total GWG. Substantial heterogeneity in outcome measures of change in dietary behaviors and physical activity precluded further meta-analyses. BCT coding identified 7 BCTs that were common to all effective interventions. Effective interventions averaged over twice as many BCTs from the *goals and planning* and *feedback and monitoring* domains as ineffective interventions. Six studies reported on user engagement, and their data indicated a positive association between high engagement with key BCTs and greater intervention effectiveness. Interventions using proactive messaging, such as reminding participants to engage in BCTs or providing feedback or tips, appeared to have higher levels of engagement.

Effectiveness

Meta-analyses of the digital interventions measuring GWG showed no effect on the total GWG or weight gain within the IOM guidelines. Although the majority of these studies were pilot RCTs and insufficiently powered to detect an effect, these findings indicate that exclusively digital interventions to manage GWG may be less effective than those using interpersonal delivery. Lack of consistency in outcome measures precluded meta-analyses of the effects of digital interventions on dietary behaviors and physical activity. Only 3 of the 7 studies measuring changes in physical activity reported significant effects of the intervention, suggesting that for physical activity interventions during pregnancy, digital delivery may similarly be less effective than interpersonal delivery [43].

The 11 interventions varied considerably in terms of not only the targeted behaviors but also the technologies, functionalities, and BCTs used. As such, it would be premature to conclude that exclusively digital delivery methods *per se* are less effective than interpersonal delivery methods for lifestyle interventions during pregnancy. Indeed, one of the included studies made a direct comparison of digital delivery and in-person delivery of the same intervention [35]. It found the intervention to be effective via both delivery methods, with digital delivery showing greater adherence and lower costs (for both participants and clinics) compared with in-person delivery.

BCTs

This systematic review aimed to identify the BCTs associated with effective interventions. The number of identifiable BCTs ranged from 1 to 17, with the 2 most effective interventions [34,35] using the highest number (n=17 and n=15). The average number of BCTs was 9 compared with approximately 5 reported in 2 earlier systematic reviews of lifestyle interventions targeting pregnant women [44,45]. It is unclear whether this increase reflects a trend toward greater intervention complexity, reflects the opportunity digital interventions afford to include more components, or is simply a matter of improved reporting of BCTs. Consistent with previous systematic reviews, this review found that effective interventions tended to report a greater number of BCTs [44,46]. A meta-analysis of 122 physical activity and healthy eating interventions (for all adults) showed effectiveness to be a function of not simply the number of BCTs but particular BCTs—*self-monitoring* and at least one other technique derived from control theory [47]. In this review, *self-monitoring* appeared in 7 interventions but was notably absent from one of the 3 interventions showing a significant effect [36]. The 3 effective interventions did however average over twice as many *goals and planning* and *feedback and monitoring* BCTs as ineffective interventions (mean 7.6, SD 2.1 and mean 3.4, SD 2.9, respectively).

There was considerable variation in the execution and delivery of BCTs. For example, in some studies, participants were invited to set a single goal, whereas in others, they were able to set multiple goals. In some instances, participants were encouraged to choose their own goal, whereas in others, the goal was prescribed. Similarly, some interventions required participants to submit self-monitoring data regularly, whereas others recommended and provided functionality for self-monitoring but did not make it obligatory. Four studies proactively messaged participants to remind them to self-monitor [32-34,37], whereas one messaged participants only if they failed to self-monitor [35]. In 3 of the studies that incorporated an initial in-person session for intervention participants [33,34,39], one or more BCTs were identifiable at this stage, raising the question as to whether the content of these sessions contained sufficient BCTs in their own right to bring about a change. The influence of these variations in the context, execution, and delivery of BCTs on intervention effectiveness could not be quantified by the methods used in this study. Given the interactive and dynamic nature of digital interventions, additional measures may be needed to capture the impact of features, such as the timing of delivery and degree of individual tailoring of BCTs.

Consistent with other systematic reviews of this type of intervention [44,48,49], *information about health consequences* was the most widely used BCT, which featured in all interventions. *Goal setting (behavior)* was the next most widely used BCT, appearing in all but 3 text message–only interventions [31,38,40]. *Problem solving*,*self-monitoring of behavior, self-monitoring of outcomes*, and *instructions on how to perform a behavior* all appeared in 7 interventions. *Feedback on behavior* was provided in 6 interventions, including the 3 [34-36] reporting significant effects of the interventions. The BCT *review behavior goal* was only present in the 3 effective interventions [34-36], suggesting that this may be a critical active ingredient in these digital interventions. It is possible that *review behavior goal* in combination with *self-monitoring of behavior* and *feedback on behavior* work together to support the self-regulation of energy balance behaviors during pregnancy.

Social support was present in 6 interventions [33-37,40], including the 3 effective interventions. Once again, the execution of *social support* varied, ranging from advice on how to seek support to online group forums for participants. There is no consensus on whether social support or interaction with other participants improves intervention effectiveness, and no clear pattern emerged from this review [14,23,50]. More research is needed to understand the type of social support that is most beneficial to digital interventions encouraging healthy behaviors during pregnancy.

Insufficient description of intervention components, coupled with a lack of systematic recording of BCTs, compromised the quality of the BCT analysis. Only 1 study provided details of all the BCTs [34] used in the intervention, whereas the presence of BCTs had to be inferred from descriptions of the interventions in all other studies. This raises the possibility that there may be additional but unreported BCTs in other studies. Previous studies that have coded BCTs used in gestational weight management trials have called for greater clarity and accuracy in the reporting of BCTs [26]. Without systematic reporting of active intervention ingredients, it is difficult to precisely determine which BCTs may be driving effectiveness.

Engagement

Six studies provided measures of user engagement [32-36,41]. These varied considerably, including the number of replies to texts, frequency of inputting weight monitoring data and logging onto and viewing web pages. Only one study [36] provided feedback on user experience. Given the importance of user engagement to the success or otherwise of self-managed digital interventions, more detailed and standardized measures could facilitate better evaluation and cross-study comparison [51]. Perski et al [52] proposed more comprehensive measures, including both the extent (ie, amount, frequency, duration, depth) of usage and the user experience. Reinforcing the need for a more holistic evaluation of engagement, Yardley et al [24] proposed identifying and reporting on *effective engagement* rather than simply higher levels of engagement. The combination of web analytics and survey feedback clearly offers the opportunity to develop specific and relevant indices of engagement [53].

The 3 effective interventions [34-36] all reported engagement levels of over 70% with key BCTs (*goal setting* [34,36], *self-monitoring* [34,35], and *action planning* [36]). Conversely, the study with the lowest engagement level, where only 46% of participants logged onto the website at least once every 45 days, and the use of *goal setting* and *self-monitoring* features was 35% and 23% of the participants respectively, reported no effect of the intervention [41]. These findings suggest that ineffectiveness may be partially a function of poor engagement with key BCTs rather than poorly designed interventions *per se*. Supporting this hypothesis, this study with low levels of engagement [41] conducted secondary analyses investigating whether usage patterns of the intervention features reduced the risk of excessive GWG and found frequent usage patterns were associated with lower total GWG [54,55]. In addition, the use of the dietary tool (*goal setting* and *self-monitoring*) was associated with improved GWG management for women with normal BMI, although not for those with high BMI.

One consistent feature of the interventions reporting the levels of engagement over 70% was regular in-app messaging or text messaging giving encouragement, reminders to self-monitor, or tailored feedback on progress. Prompts and reminders have been shown to promote engagement in digital interventions [56]. Similarly, tailoring messages to the characteristics and usage patterns of the individual has been shown to improve adherence [57]. Notably, the study [41] in which participants were sent a generic weekly email reported particularly low levels of engagement. The frequency and timing of these messages are also important [58]. One study [34] delivering 4 to 5 texts per week found that 79% of participants thought the frequency of messages was *about right*, although 21% thought it was *too high*. Another study [40] investigated the dose and timing of messages to promote physical activity by comparing 3 texts per week with daily texts and found daily texts to be less effective, indicating that too much messaging can be counterproductive. None of the studies referred to the use of gamification techniques to promote engagement, although elements of some of the interventions could potentially be classified as gamification, such as team challenges [37,59]. Incorporating gamification features, such as badges and challenges, has been shown to increase regular engagement and immersion in digital health interventions [60,61].

The final issue regarding engagement concerns who the interventions are reaching. Only one study reported (in follow-up analyses) on high versus low engagers, revealing significant differences according to ethnicity, income group, BMI, and partner status [54]. Often, it is those who would benefit most from behavior change who are least likely to engage in behavior change interventions [62]. Greater insight into who engages with the interventions could enhance learnings from these studies.

Strengths and Limitations

A strength of this systematic review is that it is the first to focus on exclusively digital interventions to promote healthy dietary behaviors, physical activity, or weight management during pregnancy. In addition to evaluating their effectiveness, this review conducted a rigorous assessment of BCTs and participant

engagement to provide detailed insight into what may be driving effectiveness—a crucial step if the cost and reach advantages of digital interventions are to be leveraged. However, there are several limitations to this systematic review. First, most of the studies included were pilot studies rather than RCTs and, as such, were not adequately powered to show effect sizes. Moreover, there was considerable heterogeneity of intervention features and outcome measures, and several studies reported the results from PP analyses rather than ITT. As such, the results of the meta-analyses should be interpreted with caution. Second, the risk of bias across the studies was moderate to high, with 5 studies scoring overall *high* and a further 3 scoring as *some concerns*, as assessed by RoB 2.0. Third, the timing of the interventions within pregnancy varied both in terms of the start point within the gestational window and duration. This, coupled with inconsistent measures of GWG and, in some cases, reliance on self-reported weight measures should be considered when appraising the findings. Finally, limited reporting of intervention features meant that not all BCTs were recorded. Providing more detailed descriptions of the interventions' design and content (in supplementary files) would augment shared learnings from these studies. Similarly, more detailed and consistent engagement measures would have enhanced the interpretation of user engagement data.

Conclusions

Meta-analyses show that the mean impact on GWG of exclusively digital interventions targeting dietary behaviors, physical activity, and weight management during pregnancy to be nonsignificant, meaning that the current exclusively digital interventions are less effective than interpersonal interventions in this field. There was considerable variation in intervention effectiveness across the 11 studies, with 3 studies from 2017 reporting significantly positive effects of their interventions. Limited data precluded confident identification of the ingredients of successful interventions, although this review suggests that variation in effectiveness could be partially explained by the BCTs used and levels of interactivity to encourage engagement with the intervention features. Effective interventions used more BCTs (particularly BCTs from *goals and planning* and *feedback and monitoring* domains) and reported higher levels of engagement with key BCTs. Effective interventions also used interactivity, in the form of messages of encouragement, personalized feedback, and prompts to remind participants to use key BCTs, such as *goal setting* and *self-monitoring*, to promote engagement.

There are several compelling reasons for considering using digital interventions to promote healthy energy balance behaviors during pregnancy: smartphone ownership is over 90% among women of childbearing age [63] and usage of pregnancy apps is pervasive [64]; digital interventions have broader reach and lower costs than interpersonal interventions [35,65]; and apps have been shown to be particularly successful in reaching those who may be less likely to engage with traditional antenatal health care [66]. Meanwhile, midwives frequently report that they have neither the time nor expertise to advise pregnant women on physical activity or healthy eating [67]. Future research needs to consider how to seize the opportunities presented by new technologies to enhance interactivity, improve user engagement, and bring greater effectiveness to these digital interventions.

Acknowledgments

This work was supported by the Economic and Social Research Fund (grant number ES/P000592/1) and UK charity, Best Beginnings. The authors would like to thank the authors of the included studies for the additional information they provided to help this study's analyses.

Conflicts of Interest

CL reports grants from the Economic and Social Research Fund, grants from Best Beginnings (UK charity), during the conduct of the study; personal fees from Yellow Kite; personal fees from the experiment; and personal fees from Diamond Inc, outside the submitted work. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
PRISMA Checklist.
[PDF File (Adobe PDF File), 52 KB - mhealth_v8i7e18255_app1.pdf]

Multimedia Appendix 2
Subject headings and key words for search.
[PDF File (Adobe PDF File), 59 KB - mhealth_v8i7e18255_app2.pdf]

Multimedia Appendix 3
Intervention features, engagement and effectiveness.
[PDF File (Adobe PDF File), 63 KB - mhealth_v8i7e18255_app3.pdf]

Multimedia Appendix 4
Subgroup analyses.

[DOCX File , 88 KB - mhealth_v8i7e18255_app4.docx]

Multimedia Appendix 5
Summary of BCTs.
[PDF File (Adobe PDF File), 57 KB - mhealth_v8i7e18255_app5.pdf]

References

1. Baker C. Obesity Statistics. House of Commons Library - UK Parliament. 2019. URL: https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN03336 [accessed 2019-07-12]
2. Phelan S. Pregnancy: a 'teachable moment' for weight control and obesity prevention. Am J Obstet Gynecol 2010 Feb;202(2):135.e1-135.e8 [FREE Full text] [doi: 10.1016/j.ajog.2009.06.008] [Medline: 19683692]
3. Raghavan R, Dreibelbis C, Kingshipp B, Wong Y, Abrams B, Gernand A, et al. Dietary patterns before and during pregnancy and maternal outcomes: a systematic review. Am J Clin Nutr 2019 Mar 1;109(Suppl_7):705S-728S. [doi: 10.1093/ajcn/nqy216] [Medline: 30982868]
4. Rasmussen K, Yaktine A. Weight Gain During Pregnancy: Reexamining the Guidelines. New York, USA: National Academies Press; 2009.
5. Scott C, Andersen CT, Valdez N, Mardones F, Nohr EA, Poston L, et al. No global consensus: a cross-sectional survey of maternal weight policies. BMC Pregnancy Childbirth 2014 May 15;14:167 [FREE Full text] [doi: 10.1186/1471-2393-14-167] [Medline: 24884985]
6. Goldstein R, Abell S, Ranasinha S, Misso M, Boyle J, Harrison C, et al. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med 2018 Aug 31;16(1):153 [FREE Full text] [doi: 10.1186/s12916-018-1128-1] [Medline: 30165842]
7. Crozier S, Inskip H, Godfrey K, Cooper C, Harvey N, Cole Z, Southampton Women's Survey Study Group. Weight gain in pregnancy and childhood body composition: findings from the Southampton women's survey. Am J Clin Nutr 2010 Jun;91(6):1745-1751 [FREE Full text] [doi: 10.3945/ajcn.2009.29128] [Medline: 20375187]
8. Hedderson MM, Gunderson EP, Ferrara A. Gestational weight gain and risk of gestational diabetes mellitus. Obstet Gynecol 2010 Mar;115(3):597-604 [FREE Full text] [doi: 10.1097/AOG.0b013e3181cfce4f] [Medline: 20177292]
9. Gilmore LA, Klempel-Donchenko M, Redman LM. Pregnancy as a window to future health: excessive gestational weight gain and obesity. Semin Perinatol 2015 Jun;39(4):296-303 [FREE Full text] [doi: 10.1053/j.semperi.2015.05.009] [Medline: 26096078]
10. Mamun AA, Mannan M, Doi SA. Gestational weight gain in relation to offspring obesity over the life course: a systematic review and bias-adjusted meta-analysis. Obes Rev 2014 Apr;15(4):338-347. [doi: 10.1111/obr.12132] [Medline: 24321007]
11. Josey MJ, McCullough LE, Hoyo C, Williams-DeVane C. Overall gestational weight gain mediates the relationship between maternal and child obesity. BMC Public Health 2019 Aug 7;19(1):1062 [FREE Full text] [doi: 10.1186/s12889-019-7349-1] [Medline: 31391077]
12. Recommendation 2 Pregnant Women. The National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ph27/chapter/1-Recommendations#recommendation-2-pregnant-women [accessed 2019-08-22]
13. Craemer K, Sampene E, Safdar N, Antony K, Wautlet C. Nutrition and exercise strategies to prevent excessive pregnancy weight gain: a meta-analysis. AJP Rep 2019 Jan;9(1):e92-120 [FREE Full text] [doi: 10.1055/s-0039-1683377] [Medline: 31041118]
14. Walker R, Bennett C, Blumfield M, Gwini S, Ma J, Wang F, et al. Attenuating pregnancy weight gain-what works and why: a systematic review and meta-analysis. Nutrients 2018 Jul 22;10(7):- [FREE Full text] [doi: 10.3390/nu10070944] [Medline: 30037126]
15. Shieh C, Cullen D, Pike C, Pressler S. Intervention strategies for preventing excessive gestational weight gain: systematic review and meta-analysis. Obes Rev 2018 Aug;19(8):1093-1109. [doi: 10.1111/obr.12691] [Medline: 29806187]
16. Farpour-Lambert N, Ells L, de Tejada BM, Scott C. Obesity and weight gain in pregnancy and postpartum: an evidence review of lifestyle interventions to inform maternal and child health policies. Front Endocrinol (Lausanne) 2018;9:546 [FREE Full text] [doi: 10.3389/fendo.2018.00546] [Medline: 30319539]
17. Lau Y, Klainin-Yobas P, Htun T, Wong S, Tan K, Ho-Lim T, et al. Electronic-based lifestyle interventions in overweight or obese perinatal women: a systematic review and meta-analysis. Obes Rev 2017 Sep;18(9):1071-1087. [doi: 10.1111/obr.12557] [Medline: 28544551]
18. Villinger K, Wahl D, Boeing H, Schupp H, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
19. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]

20. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
21. Sherifali D, Nerenberg K, Wilson S, Semeniuk K, Ali M, Redman L, et al. The effectiveness of ehealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
22. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
23. Chan KL, Chen M. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
24. 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]
25. 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]
26. Soltani H, Arden MA, Duxbury AM, Fair FJ. An analysis of behaviour change techniques used in a sample of gestational weight management trials. J Pregnancy 2016;2016:1085916 [FREE Full text] [doi: 10.1155/2016/1085916] [Medline: 27034836]
27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
28. RoB 2: A Revised Cochrane Risk-of-Bias Tool for Randomized Trials. Cochrane Methods. URL: https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials [accessed 2019-08-14]
29. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Training. 2019. URL: https://training.cochrane.org/handbook#how-to-cite [accessed 2020-01-10]
30. Cochrane RevMan. Cochrane Training. URL: https://community.cochrane.org/help/tools-and-software/revman-5 [accessed 2019-08-14]
31. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012 Nov 26;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
32. Pollak K, Alexander S, Bennett G, Lyna P, Coffman C, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. Patient Educ Couns 2014 Nov;97(2):256-260 [FREE Full text] [doi: 10.1016/j.pec.2014.07.030] [Medline: 25153313]
33. 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 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
34. Willcox J, Wilkinson S, Lappas M, Ball K, Crawford D, McCarthy E, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728 [FREE Full text] [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
35. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a novel ehealth intervention for management of gestational weight gain: randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Sep 13;5(9):e133 [FREE Full text] [doi: 10.2196/mhealth.8228] [Medline: 28903892]
36. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CM. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women - the Fit4Two randomised controlled trial. BMC Pregnancy Childbirth 2017 Mar 23;17(1):96 [FREE Full text] [doi: 10.1186/s12884-017-1277-9] [Medline: 28335767]
37. Dahl A. Healthy Motivations for Moms-To-Be (Healthy MoM2B) Study: A Mobile Health Intervention Targeting Gestational Weight Gain among US Women. Semantic Scholar. 2018. URL: https://www.semanticscholar.org/paper/Healthy-Motivations-for-Moms-To-Be-(Healthy-MoM2B)-Dahl/828d334dde24173f90099c03457c0674e0050740 [accessed 2020-06-16]
38. Evans W, Nielsen P, Szekely D, Bihm J, Murray E, Snider J, et al. Dose-response effects of the text4baby mobile health program: randomized controlled trial. JMIR Mhealth Uhealth 2015 Jan 28;3(1):e12 [FREE Full text] [doi: 10.2196/mhealth.3909] [Medline: 25630361]
39. Smith K, Lanningham-Foster L, Welch A, Campbell CK. Web-based behavioral intervention increases maternal exercise but does not prevent excessive gestational weight gain in previously sedentary women. J Phys Act Health 2016 Jun;13(6):587-593. [doi: 10.1123/jpah.2015-0219] [Medline: 26594820]

40. Huberty J, Buman M, Leiferman J, Bushar J, Hekler E, Adams MJ. Dose and timing of text messages for increasing physical activity among pregnant women: a randomized controlled trial. Transl Behav Med 2017 Jun;7(2):212-223 [FREE Full text] [doi: 10.1007/s13142-016-0445-1] [Medline: 27800565]
41. Olson C, Groth S, Graham M, Reschke J, Strawderman M, Fernandez I. The effectiveness of an online intervention in preventing excessive gestational weight gain: the e-moms ROC randomized controlled trial. BMC Pregnancy Childbirth 2018 May 9;18(1):148 [FREE Full text] [doi: 10.1186/s12884-018-1767-4] [Medline: 29743026]
42. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
43. Chan C, Yeung EA, Law B. Effectiveness of physical activity interventions on pregnancy-related outcomes among pregnant women: a systematic review. Int J Environ Res Public Health 2019 May 23;16(10):- [FREE Full text] [doi: 10.3390/ijerph16101840] [Medline: 31126153]
44. Hill B, Skouteris H, Fuller-Tyszkiewicz M. Interventions designed to limit gestational weight gain: a systematic review of theory and meta-analysis of intervention components. Obes Rev 2013 Jun;14(6):435-450. [doi: 10.1111/obr.12022] [Medline: 23534901]
45. Gardner B, Croker H, Barr S, Briley A, Poston L, Wardle J, UPBEAT Trial. Psychological predictors of dietary intentions in pregnancy. J Hum Nutr Diet 2012 Aug;25(4):345-353. [doi: 10.1111/j.1365-277X.2012.01239.x] [Medline: 22380723]
46. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
47. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
48. Katzmarzyk P, Barlow S, Bouchard P, Catalano D, Hsia DS, Inge TH, et al. An evolving scientific basis for the prevention and treatment of pediatric obesity. Int J Obes (Lond) 2014 Jul;38(7):887-905 [FREE Full text] [doi: 10.1038/ijo.2014.49] [Medline: 24662696]
49. Brown MJ, Sinclair M, Liddle D, Hill AJ, Madden E, Stockdale J. A systematic review investigating healthy lifestyle interventions incorporating goal setting strategies for preventing excess gestational weight gain. PLoS One 2012;7(7):e39503 [FREE Full text] [doi: 10.1371/journal.pone.0039503] [Medline: 22792178]
50. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digit Health 2015;1:2055207615595335 [FREE Full text] [doi: 10.1177/2055207615595335] [Medline: 29942544]
51. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the DBCI engagement scale. Transl Behav Med Internet 2019:30. [doi: 10.1093/tbm/ibz039]
52. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
53. 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]
54. Graham ML, Strawderman MS, Demment M, Olson CM. Does usage of an ehealth intervention reduce the risk of excessive gestational weight gain? Secondary analysis from a randomized controlled trial. J Med Internet Res 2017 Jan 9;19(1):e6 [FREE Full text] [doi: 10.2196/jmir.6644] [Medline: 28069560]
55. Olson CM, Strawderman MS, Graham ML. Use of an online diet goal-setting tool: relationships with gestational weight gain. J Nutr Educ Behav 2019 Apr;51(4):391-399. [doi: 10.1016/j.jneb.2019.01.024] [Medline: 30975376]
56. 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 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
57. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Sep;11:267-273 [FREE Full text] [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]
58. 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]
59. Hamari J, Koivisto J, Sarsa H. Does Gamification Work? - A Literature Review of Empirical Studies on Gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS'14; January 6-9, 2014; Waikoloa, HI, USA. [doi: 10.1109/hicss.2014.377]
60. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
61. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]

62. Dalton JA, Rodger D, Wilmore M, Humphreys S, Skuse A, Roberts CT, et al. The health-e babies app for antenatal education: feasibility for socially disadvantaged women. PLoS One 2018;13(5):e0194337 [FREE Full text] [doi: 10.1371/journal.pone.0194337] [Medline: 29768407]
63. Smartphone Ownership Penetration in the United Kingdom (UK) in 2012-2020, by Age. Statista. URL: https://www.statista.com/statistics/271851/smartphone-owners-in-the-united-kingdom-uk-by-age/ [accessed 2019-07-12]
64. Hughson JP, 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 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
65. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mhealth? A systematic review of economic evaluations of mhealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
66. Impact and Evaluation of Baby Buddy App. Best Beginnings. URL: https://www.bestbeginnings.org.uk/evidence-impact-and-evaluation [accessed 2019-07-12]
67. de Vivo M, Mills H. 'They turn to you first for everything': insights into midwives' perspectives of providing physical activity advice and guidance to pregnant women. BMC Pregnancy Childbirth 2019 Dec 3;19(1):462 [FREE Full text] [doi: 10.1186/s12884-019-2607-x] [Medline: 31795961]

Abbreviations

BCT: behavior change technique
GWG: gestational weight gain
IOM: Institute of Medicine
ITT: intention-to-treat
OR: odds ratio
PP: per-protocol
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 15.02.20; peer-reviewed by A Merkx, M Bestek, L König; comments to author 23.04.20; revised version received 30.04.20; accepted 14.05.20; published 10.07.20.

Please cite as:
Rhodes A, Smith AD, Chadwick P, Croker H, Llewellyn CH
Exclusively Digital Health Interventions Targeting Diet, Physical Activity, and Weight Gain in Pregnant Women: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2020;8(7):e18255
URL: https://mhealth.jmir.org/2020/7/e18255
doi:10.2196/18255
PMID:32673251

Original Paper

Use of Mobile Phone App Interventions to Promote Weight Loss: Meta-Analysis

Md Mohaimenul Islam[1,2,3], MSc; Tahmina Nasrin Poly[1,2,3], MSc; Bruno Andres Walther[4], PhD; Yu-Chuan (Jack) Li[1,2,3,5], PhD

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[2]International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei, Taiwan

[3]Research Center of Big Data and Meta-Analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[4]Department of Biological Sciences, National Sun Yat-Sen University, Kaohsiung, Taiwan

[5]Department of Dermatology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Yu-Chuan (Jack) Li, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
Taipei
Taiwan
Phone: 886 886 2 2736166
Email: jaak88@gmail.com

Abstract

Background: Obesity and lack of physical activity are major health risk factors for many life-threatening diseases, such as cardiovascular diseases, type 2 diabetes, and cancer. The use of mobile app interventions to promote weight loss and boost physical activity among children and adults is fascinating owing to the demand for cutting-edge and more efficient interventions. Previously published studies have examined different types of technology-based interventions and their impact on weight loss and increase in physical activity, but evidence regarding the impact of only a mobile phone app on weight loss and increase in physical activity is still lacking.

Objective: The main objective of this study was to assess the efficacy of a mobile phone app intervention for reducing body weight and increasing physical activity among children and adults.

Methods: PubMed, Google Scholar, Scopus, EMBASE, and the Web of Science electronic databases were searched for studies published between January 1, 2000, and April 30, 2019, without language restrictions. Two experts independently screened all the titles and abstracts to find the most appropriate studies. To be included, studies had to be either a randomized controlled trial or a case-control study that assessed a mobile phone app intervention with body weight loss and physical activity outcomes. The Cochrane Collaboration Risk of Bias tool was used to examine the risk of publication bias.

Results: A total of 12 studies involving a mobile phone app intervention were included in this meta-analysis. Compared with the control group, the use of a mobile phone app was associated with significant changes in body weight (−1.07 kg, 95% CI −1.92 to −0.21, *P*=.01) and body mass index (−0.45 kg/m2, 95% CI −0.78 to −0.12, *P*=.008). Moreover, a nonsignificant increase in physical activity was observed (0.17, 95% CI −2.21 to 2.55, *P*=.88).

Conclusions: The findings of this study demonstrate the promising and emerging efficacy of using mobile phone app interventions for weight loss. Future studies are needed to explore the long-term efficacy of mobile app interventions in larger samples.

(JMIR Mhealth Uhealth 2020;8(7):e17039) doi:10.2196/17039

KEYWORDS

mobile app; mHealth; obesity; physical activity; weight gain prevention

Introduction

Overweight (BMI ≥25 kg/m^2), obesity (BMI ≥30 kg/m^2), and physical inactivity are major preventable public health problems that are linked to increased risks of chronic diseases such as diabetes, high blood pressure, heart disease, and cancer [1]. In 2016, over 1.9 billion (39%) and over 650 million (13%) adults older than 18 years were overweight and obese, respectively, and the prevalence of overweight and obesity has nearly tripled since 1980 [2]. Reducing weight and improving physical activity are thus important priorities to minimize the burden associated with overweight and obesity-related comorbidities. A high number of studies have already demonstrated that a change in lifestyle can help reduce and maintain weight [3,4]. However, for many people, it is very difficult to change their lifestyle and maintain weight loss [5]. Therefore, an intervention that helps to motivate people to undertake a change in lifestyle, offers pragmatic goal settings, and offers feedback on activity rates can help maintain weight loss and greatly increase physical activity [6,7].

The mobile phone has become a very important medium of communication throughout the world [8], and approximately 75% of adults have used different kinds of mobile interventions [9,10]. Therefore, the use of a mobile phone intervention could be a promising approach for disease management and prevention that has a huge potential to reach out to the vast majority of the population [11]. Researchers have been using mobile phone interventions to support behavioral change by providing more interactive and timely access to relevant information and delivering context-specific prompt assistance [12]. Recently, the use of mobile apps has led to notable success in obesity control, weight reduction, physical activity increase, and quality of life improvement [13]. App-based interventions have already been shown to be cost-effective and to reduce the immense barriers associated with more traditional approaches. However, the effectiveness and success of a mobile app intervention also relies on how the intervention has been developed. A mobile app intervention delivery system has the capacity to reach each participant effectively and efficiently, and it is thus an innovative way to manage weight and increase physical activity.

The use of a mobile app is a rapidly expanding area of research for disease management and behavioral change. Therefore, the purpose of this systematic review and meta-analysis was to evaluate the current evidence for the feasibility of a mobile phone intervention. Our study updates and extends the scope of a prior systematic review and meta-analysis in three ways [14]. First, we included two more randomized controlled trials than the previous review. Second, we provided more subgroup analyses than the previous review. Third, we extended the sample size and characteristics, such as age, gender, geographic region, and features of the intervention, to evaluate mobile app efficacy for weight loss and increased physical activity.

Methods

Guidelines

This systematic review was conducted in accordance with the Meta-Analysis of Observational Studies in Epidemiology guidelines [15] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses standard (Multimedia Appendix 1) [16].

Literature Search

PubMed, Google Scholar, Scopus, EMBASE, and the Web of Science electronic databases were searched for studies published between January 1, 2000, and April 30, 2019, without language restrictions. The search was conducted by two experts (MMI and TNP) using combinations of relevant search terms and Boolean operators as follows: “mobile apps” AND “weight loss” OR “weight control” OR “obesity” OR “BMI” OR “body mass index” (Multimedia Appendix 2). There was no language or data restriction for the initial search. However, we did not consider any gray literature for unpublished studies (abstracts and conference proceedings) in the initial search. Such unpublished studies were not considered because they did not pass a proper peer-review process. We then used EndNote X7 (Thomson Reuters) to remove any duplicate publication.

Other Resources Search

We carefully checked all the retrieved systematic reviews and meta-analyses in order to find further relevant studies.

Eligibility Criteria

Two experts (MMI and TNP) independently screened all the titles and abstracts to find the most appropriate full-text studies for inclusion. They then further screened all full-text studies for quantitative synthesis of evidence and recorded the inclusion and exclusion criteria. Any disagreements over the inclusion and exclusion criteria that arose during this stage were subsequently resolved by the main investigator (YCL) of this study. We considered all studies if they met the following criteria: (1) published in English; (2) reported a mobile app intervention for change in body weight, BMI, or waist circumference; and (3) reported a mobile app for weight loss among children and adults compared with a control group.

Studies were excluded if they met the following criteria: (1) review or methodology study, short communication, or letter to the editor; (2) case report or case series; (3) no control group; (4) outcomes of interest were other diseases except for a diagnosis of obesity; and (5) other types of mobile phone interventions like text messaging.

Data Extraction

A predefined standard procedure was used to retrieve all the information from the included studies by the same two authors (MMI and TNP). They also used the Review Manager software (RevMan-5) to check the accuracy of the included studies. They collected the following information from all the included studies: (1) *methods*: number of studies, number of patients, age of participants, gender of participants or patients, study period, inclusion and exclusion criteria, and intervention and follow-up

duration; (2) *results*: study characteristics, target group outcome, intervention characteristics, type of intervention, change of participant's behavior, mean changes from baseline, variation that was reported as SD or 95% CI, and bias assessment; and (3) *discussion*: main findings, suggestions, intended recommendations, and limitations.

Study Quality Assessment

The methodological quality of the included observational design studies was evaluated in accordance with the Cochrane Handbook for Systematic Reviews of Intervention [17]. This guideline is used to evaluate how well-randomized controlled trials were conducted to avoid bias based on a total of seven criteria. The following elements are reported in these guidelines: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. Therefore, the total score was used to assess the quality of the included studies as low, uncertain, or high risk.

Statistical Analysis

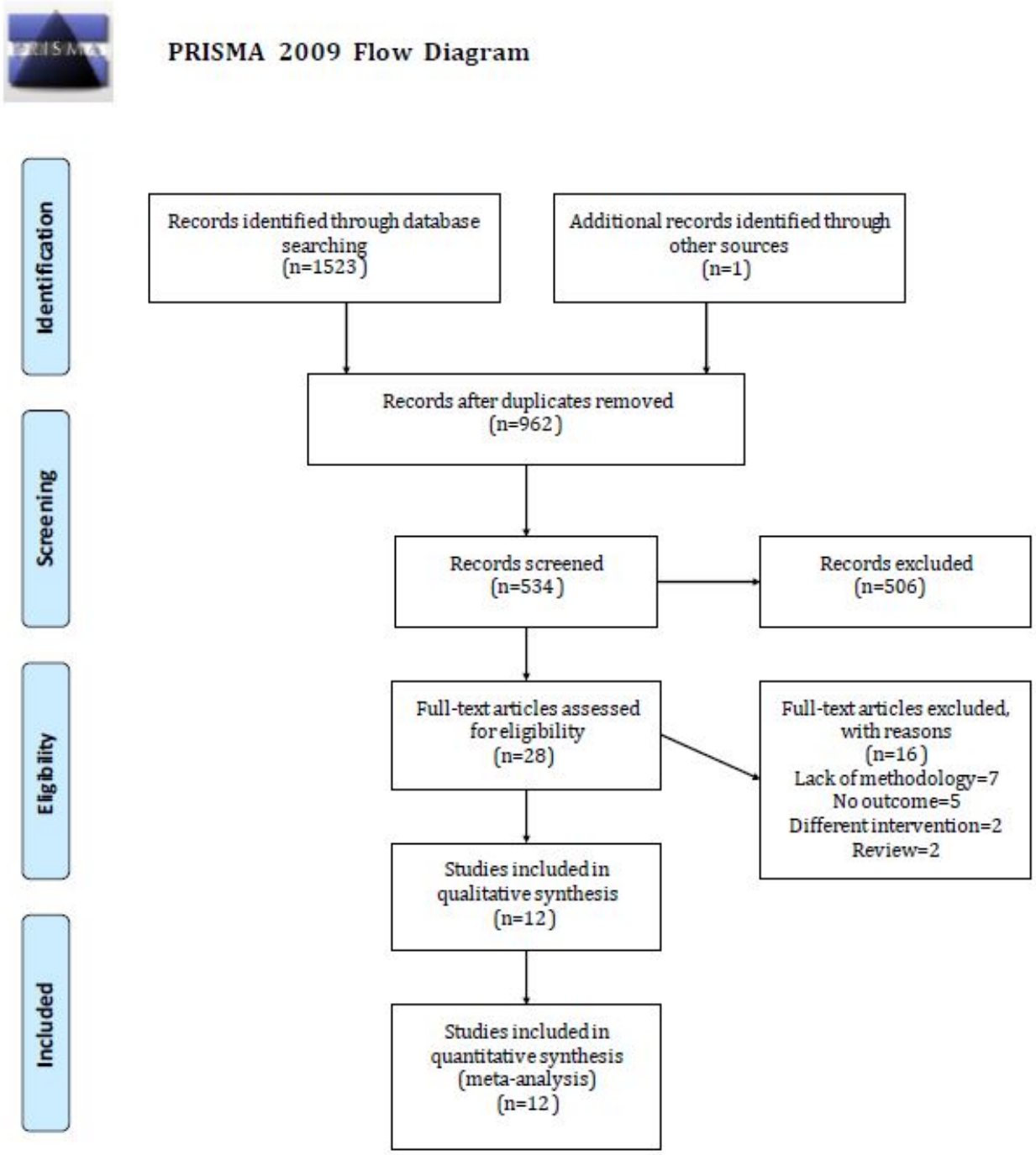
The meta-analysis was performed to evaluate mobile app interventions for weight loss. A random-effect model was used to obtain an overall effect size. We calculated the standardized mean difference between the experimental group (mobile app) and control group using the mean, standard deviation, and total number of individuals. Moreover, an effect size and standard error for outcome (weight loss, BMI, and physical activity) were also calculated. We calculated the effect size with a 95% CI, and statistical significance was considered at a *P* value <.05. The chi-square (Q) statistic and I^2 were also calculated to determine the sources of heterogeneity between the studies. If the *P* value of the chi-square test was >.05, the findings of the study were considered to be due to chance. Homogeneity between the studies was determined if the *P* value was >.05. The I^2 value quantitatively determines heterogeneity. The heterogeneity between studies was categorized as very low, low, medium, and high if the I^2 values were 0%-25%, 25%-50%, 50%-75%, and more than 75%, respectively [18]. A random-effect model was utilized because it is the appropriate model to calculate the effect size if the included studies are heterogeneous, and for homogenous studies, fixed-effect models are appropriate [19]. In addition, a funnel plot was generated to assess publication bias, and it was evaluated by the Egger method. All analyses were conducted using Comprehensive Meta-Analysis software (CMA) Version 2.

Results

Study Selection

The literature search of the electronic databases yielded 1523 studies. After our review of the titles and abstracts, a total of 1495 studies were excluded either because of duplication or because of the lack of adherence to our topic. Consequently, only 28 studies underwent full-text review. We also checked their reference lists for further relevant studies and retrieved one additional study. However, 16 more studies were then excluded because they did not meet the inclusion criteria mentioned previously. Finally, a total of 12 studies were included in our meta-analysis [13,20-30]. The flow chart of our systematic literature search is presented in Figure 1.

Figure 1. Flow chart of the study search and selection.



Study Characteristics

The studies included in our meta-analysis were 11 randomized controlled trials and 1 case-control study (Table 1). The publication year ranged from 2010 to 2019. The studies included child and adult participants, with a total of 792 experimental participants and 799 controls. Five studies were published in Australia, four in North America, two in Europe, and one in Asia. A total of 1714 participants were included, and the sample size ranged from 35 to 361. Most of the participants were female. The proportion of female participants ranged from 85% to 100%, and the mean age of the participants ranged from 12.7 to 44.9 years. The follow-up period of the included studies ranged from 6 weeks to 9 months. Moreover, age, gender, marital status, and education level were used as baseline variables, and mobile apps, email, etc were considered as study interventions (Table 2).

Table 1. Characteristics of the studies included in the meta-analysis.

First author (year)	Country	Study design	Study sample	Male, %	Age (years), mean	Study duration	Inclusion criteria	Exclusion criteria	Outcomes
Patel (2019)	USA	RCT[a]	100	16	42.7	3 months	Age 21-65 years with BMI 25-45 kg/m^2, and willingness to reduce weight through dietary change. Availability of an iPhone or Android smartphone and personal email address, and ability to read and write in English.	Enrollment in other weight loss programs, use of MyFitnessPal to track diet in the past 6 months, loss of ≥10 lb, or use of weight loss medication in the past 6 months. Moreover, pregnancy and disorders, such as cancer, eating disorders, uncontrolled hypertension, diabetes mellitus, cardiovascular events, and congestive heart failure.	Weight and BMI
Farinelli (2016)	Australia	RCT	258	40.7	28.1	9 months	Age 18-35 years, BMI 25.0-31.9 kg/m^2 or 23.0-24.9 kg/m^2 with reported weight gain greater than 2 kg over the previous 12 months. Fruit intake of less than two servings daily, vegetable intake of less than five servings daily, and SSB[b] intake of at least 1 L weekly. Energy-dense meals prepared away from home more than once per week. Owning a mobile phone capable of receiving text messages and having access to the internet at least once a week.	Pregnancy or plan for pregnancy within the next 9 months, enrollment in another mobile app–based weight loss program, weight reduction more than 10 kg voluntarily in the past 3 months, taking medications that cause more than 2 kg of weight gain, medical conditions that preclude following dietary or physical recommendations, history of disorders like eating disorders, and inability to read or write in English.	Weight, BMI, and PA[c]
Partridge (2015)	Australia	RCT	250	38	27.2	9 months	BMI 25.0-31.9 kg/m^2 or 23.0-24.9 kg/m^2 with reported weight gain greater than 2 kg over the previous 12 months, fruit intake of less than two servings daily, vegetable intake of less than five servings daily, SSB intake of at least 1 L weekly, energy-dense meals prepared away from home more than once per week, etc.	Pregnancy or plan for pregnancy within the study period, enrollment in other mobile app–based weight loss programs, reduction in weight greater than 10 kg in the past 3 months, use of medications that help to gain weight greater than 2 kg, other medical conditions that preclude following dietary or physical activity recommendations, and inability to speak English.	Weight, BMI, MPA[d], and VPA[e]
Laing (2014)	USA	RCT	212	27	43.3	6 months	Age ≥18 years, BMI ≥25 kg/m^2, and smartphone ownership.	Current, planned, or previous pregnancy within the last 6 months, hemodialysis, life expectancy less than 6 months, lack of interest in weight loss, or current use of other kinds of apps for weight loss.	Weight

First author (year)	Country	Study design	Study sample	Male, %	Age (years), mean	Study duration	Inclusion criteria	Exclusion criteria	Outcomes
Hebden (2014)	Australia	RCT	41	15	22.6	3 months	BMI 24.00-31.99 kg/m^2 with weight gain greater than 2 kg in the past 12 months, age 18-35 years, moderate intensity physical activity <60 min/day, SSB intake of at least 1 L weekly, fruit intake of less than two servings daily, vegetable intake of less than five servings daily, or at least two energy-dense takeaway meals weekly.	Inability to receive SMS messages or no regular internet access, a diet required for medical reasons, medical conditions that influence body weight or ability to comply with the intervention, intake of medications or herbal preparations that might influence body weight, enrollment in weight loss programs, pregnancy, or plan for pregnancy in the next 3 months.	Weight, BMI, and PA
Smith (2014)	Australia	RCT	361	100	12.7	7 months	Male students in their first year at the study schools completed a short screening questionnaire.	NR[f]	BMI and waist circumference
Glynn (2014)	Ireland	RCT	139	32	44	2 months	Age >16 years and active use of an Android smartphone.	No android smartphone, acute psychiatric illness, pregnancy, or inability to undertake moderate exercise.	Weight, BMI, and PA
Brindal (2013)	Australia	RCT	58	0	42	2 months	BMI >25 kg/m^2 and ability to measure weight at home.	Medical conditions that are likely to interfere with the ability to undertake the meal replacement program (eg, pregnancy, breastfeeding, active cancer, gastrointestinal disorders, and type 1 diabetes).	Weight
Carter (2013)	UK	RCT	128	23.3	41.2	6 months	BMI ≥27 kg/m^2, age 18-65 years, and willingness to commit the necessary time and effort to the study. Ability to read and write in English, ability to access the internet, and willingness to be randomized to one of three groups.	Pregnancy, breast feeding, plan for pregnancy, use of antiobesity medication or medication/insulin for diabetes, surgery for weight loss, and use of the antidepressant sertraline.	Weight and BMI
Allen (2013)	USA	RCT	35	22.1	44.9	6 months	Age 21-65 years, BMI 28-42 kg/m^2, ownership of an iPhone or Android phone, willingness to download the app to be used on their device.	History of myocardial infarction, angina, coronary artery bypass graft surgery, percutaneous transluminal coronary angioplasty, congestive heart failure, and diabetes. Current participation in other weight loss programs, pregnancy, plan for pregnancy in the next 6 months, use of weight loss medications, and history of psychiatric illness, alcohol, or substance abuse within the past 12 months.	Weight, BMI, and waist circumference

First author (year)	Country	Study design	Study sample	Male, %	Age (years), mean	Study duration	Inclusion criteria	Exclusion criteria	Outcomes
McGrievy (2011)	USA	RCT	96	24.7	44	6 months	Age 18-60 years and BMI 25-45 kg/m^2.	Smoking, unstable medical status, uncontrolled thyroid condition, inability to attend the three monitoring visits or improve the walking status, psychiatric illness, alcohol consumption, drug dependency, eating disorders, enrollment in another weight loss program, pregnancy, breast feeding, and plan for pregnancy within the next 6 months.	BMI and PA
Li (2010)	South Korea	CCS[g]	36	NR	28.5	6 weeks	Different ages and blood groups because of individual lifestyle and health effects according to blood group and the requirement of various amounts of calories based on gender.	NR	Weight and BMI

[a]RCT: randomized controlled trial.

[b]SSB: sugar-sweetened beverage.

[c]PA: physical activity.

[d]MPA: moderate physical activity.

[e]VPA: vigorous physical activity.

[f]NR: not reported.

[g]CCS: case-control study.

Table 2. Descriptions of baseline, interventions, apps, and findings of the included studies.

First author (year)	Baseline variables	Intervention type	App description	Control group treatment	Difference of the intervention group, mean (SD)	Difference of the control group, mean (SD)	Inference	Recommendation
Patel (2019)	Age, gender, marital status, race/ethnicity, education, employment status, annual household income, body mass index category, self-monitoring of diet frequency, and type of smartphone	App, email, MyFitnessPal, mobile, and internet	Weight loss goal, calorie goal, self-monitoring of body weight, dietary intake, rea-time feedback, skill training, and reminder of the goal	Self-regulation, email, and action plans via weekly email	−1.8 (1.53)	−2.55 (1.11)	The mobile app is an effective intervention for clinically meaningful weight loss.	Stand-alone digital health treatments may be a viable option for those looking for a lower intensity approach.
Farinelli (2016)	Age, gender, weight status, BMI, WHO-5 score, SES[a], ethnic background, education, fruit, vegetable, SSB[b], take-out meals, and physical activity	Mobile app, email, online weight tracker, physical activity planner, a blog facility for communication, and printable eating chart	Smart mobile apps for education and self-monitoring	Four text messages, one on each key behavior, and a two-page handout based on dietary guidelines.	−3.8 (4.9)	−0.80 (3.7)	The mHealth intervention has the potential to reduce weight and improve physical activity.	Replication of trials and widespread adoption of this model are needed.
Partridge (2015)	Age, gender, SES, ethnicity, education level, and weekly income	App, text messages, email, internet forum, a community blog, and usual care.	Educational program and self-monitoring	Mailed two-page handout, four text messages, and access to a website	−1.9 (2.84)	0.2 (2.99)	The app has huge potential for preventing weight gain with modest weight loss. It also helps to improve lifestyle behaviors.	Implementation of a large-scale study is needed.
Laing (2014)	Gender, self-reported race, education, annual income, and type of smartphone	App and usual care plan	MyFitnessPal app	Counseling and one-page educational handout for eating plan	−0.03 (4.64)	0.27 (4.64)	The app was an effective tool for reducing weight.	NR[c]
Hebden (2014)	Age, gender, SES, education, work history, lives with parents, and English proficiency	App, text messaging, email, internet forum, and usual care	Four types of behavior plans	10-page printed book	−1.6 (3)	−1.4 (3.18)	The app provided short-term positive changes in weight, nutrition, and physical activity.	More studies are needed to explore engagement and personalized support

First author (year)	Baseline variables	Intervention type	App description	Control group treatment	Difference of the intervention group, mean (SD)	Difference of the control group, mean (SD)	Inference	Recommendation
Smith (2014)	Age, English language, cultural background, socioeconomic position, weight, height, BMI, weight status, and waist circumference	App, parent newsletters, seminars, spot sessions, lunchtime physical activity monitoring, and teaching material	Fitness challenges, activity monitoring, and motivational messages	Traditional approaches	0.6 (1.21)	0.61 (1.07)	The app-based intervention helped to improve fitness, movement skill, and key weight-related behavior.	More studies require to capture objective data on app usage throughout the intervention period and find out the association. It is also important to add some features like gamification.
Glynn (2014)	Gender, age, systolic and diastolic blood pressure, weight, BMI, HADS[d], EQ-VAS[e], EQ-5D[f], and daily step count	App and usual care plan	Accupedo-Pro Pedometer app	Education program about the benefits of physical activity and exercise	−2.2 (3.4)	−1.5 (4.3)	The mobile app–based intervention had a positive impact on weight loss	NR
Brindal (2013)	Weight and dietary status	App and celebrity slim program	Support apps like my meals, my weight, and my task	Only celebrity slim program	−2.9 (6.4)	−2.1 (1)	The app intervention was useful for weight loss and psychological changes.	Integrating more dynamic stage-based tailoring, as behavioral changes of individuals may further enhance similar apps in the future.
Carter (2013)	Age, weight, BMI, body fat, gender, race, smoking status, occupation, and education	App	Self-monitoring	Food diary and a calorie-counting book	−4.6 (5.2)	−2.9 (5.85)	The mobile app was an acceptable and feasible weight loss intervention	More studies are needed to investigate the cost of implementing a smartphone app intervention compared with other types of interventions
Allen (2013)	Age, weight, BMI, waist circumference, education, and marital status	App and intensive counseling	Lose it!	Comprehensive counseling	−5.4 (4)	−2.5 (4.1)	The app intervention had a positive impact on weight loss and contributed to behavioral changes.	Need to conduct a large-scale population-based study.
McGrievy (2011)	NR	App + podcast + twitter	Diet plan and physical activity monitoring	Podcast only	−2.57 (2.6)	−2.45 (4.39)	NR	NR

First author (year)	Baseline variables	Intervention type	App description	Control group treatment	Difference of the intervention group, mean (SD)	Difference of the control group, mean (SD)	Inference	Recommendation
Li (2010)	Age, occupation, education, monthly income, smoking, drinking, and exercise history	Mobile app and usual care	Mobile apps that provided a personal diet profile based on gender and promoted knowledge about nutrition and physical activity	NR	−1.9 (2.3)	−0.9 (4.64)	Improved user satisfaction.	A more effective study to motivate participants and extend study duration is required.

[a]SES: socioeconomic status.

[b]SSB: sugar-sweetened beverage.

[c]NR: not reported.

[d]HADS: Hospital Anxiety and Depression Scale.

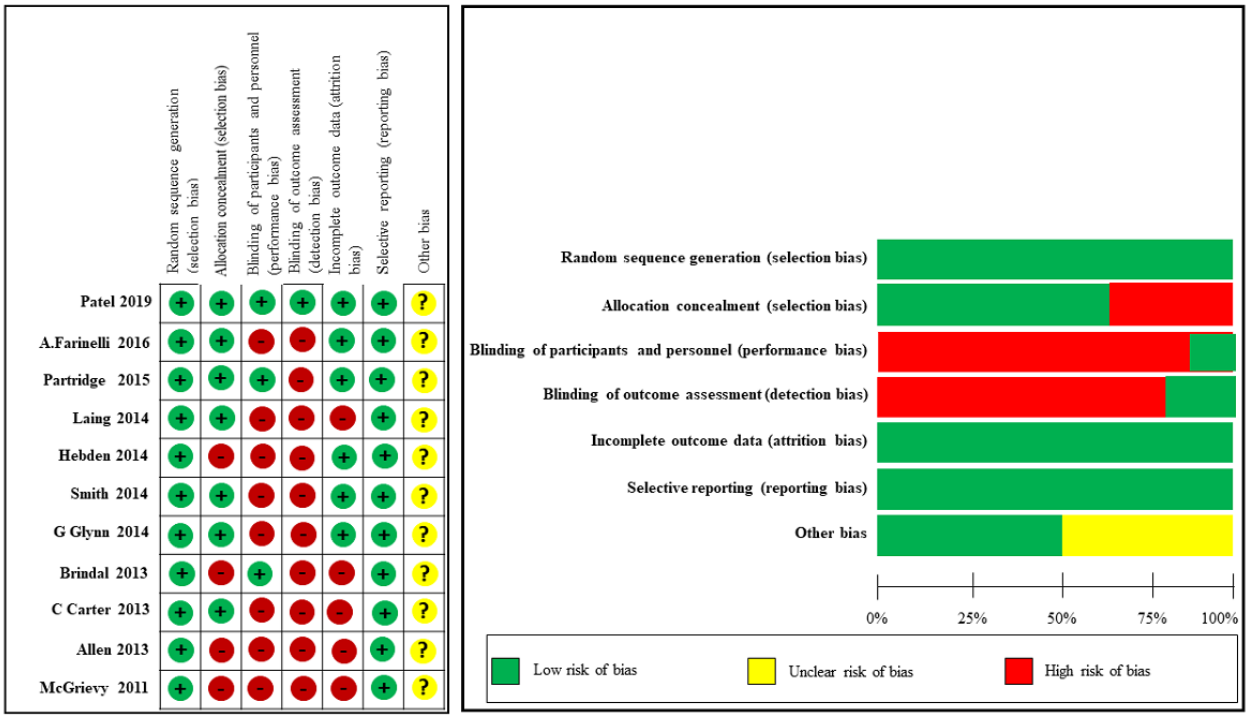
[e]EQ-VAS: EuroQol visual analogue scale.

[f]EQ-5D: EuroQol five-dimension scale.

Assessment of the Risk of Bias

Owing to the nature of mobile app interventions, participant blinding is not always feasible in trials. All the 12 studies reported random sequence generation that showed a low risk of bias. Overall, 10 out of the 12 studies were considered to constitute high-quality evidence. Only three studies had blinding of the outcome assessment, and two studies had blinding of the participants and personnel. A summary of the evaluation of the included studies is shown in Figure 2.

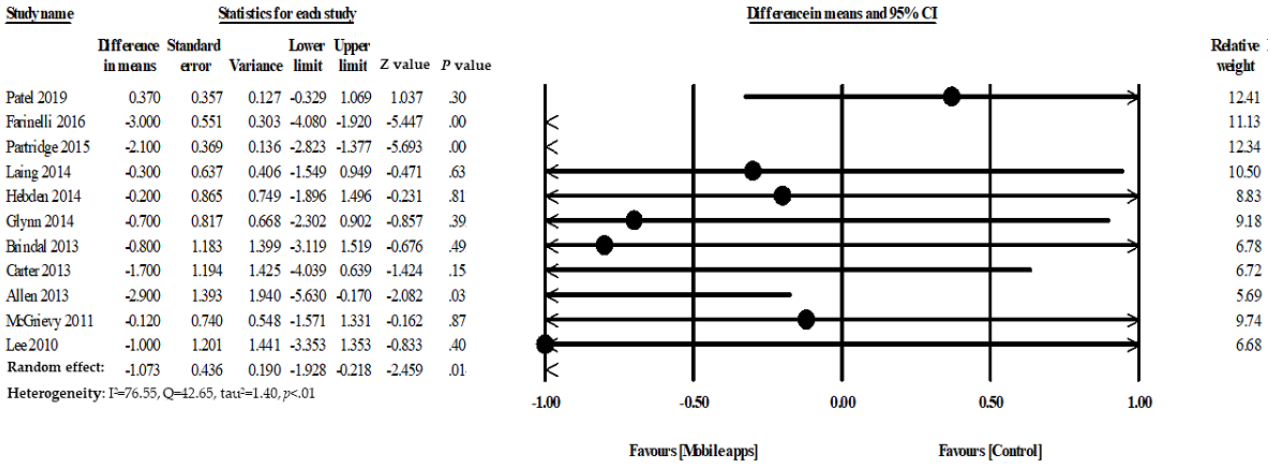
Figure 2. Risk bias assessment of the included studies.



Mobile App Intervention and Weight Loss

Eleven studies assessed the effectiveness of mobile phone app interventions for reducing body weight. Participants in the intervention group showed a decrease in their body weight (−1.07 kg, 95% CI −1.92 to −0.21) when compared with the control group (Figure 3). However, moderate heterogeneity was observed among the studies (heterogeneity I^2=71.55%, Q=42.65, P=.01, τ^2=1.40).

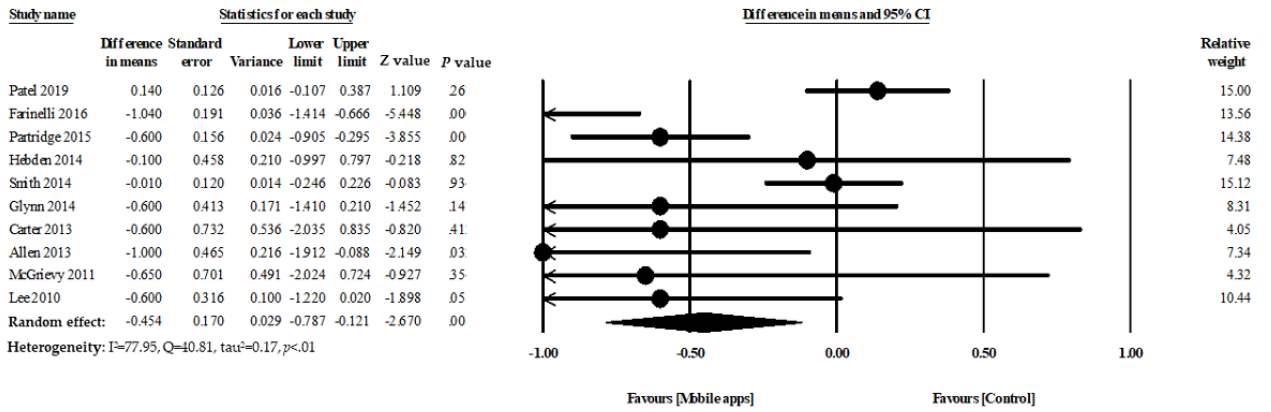
Figure 3. Forest plot of mobile phone app interventions and weight loss.



Mobile App Intervention and BMI

A total of 10 studies evaluated the efficacy of mobile apps for BMI reduction. The overall pooled findings showed a significant difference in BMI between participants in the mobile phone app intervention group and control group (−0.45 kg/m^2, 95% CI −0.78 to −0.12, *P*=.008, heterogeneity I^2=77.95%, Q=40.81, τ^2=0.17) (Figure 4).

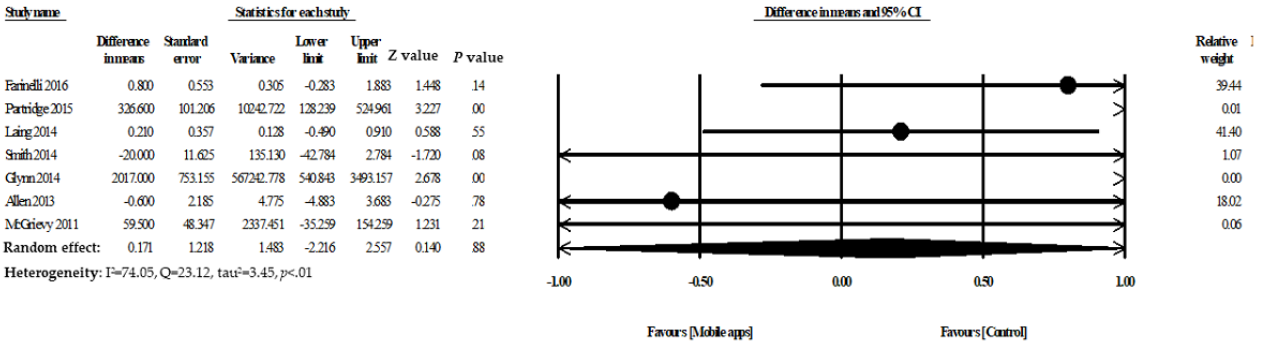
Figure 4. Forest plot of mobile phone app interventions and change in BMI.



Mobile App Intervention and Physical Activity

Seven studies assessed the effectiveness of mobile phone apps for increasing physical activity. The usability and effectiveness of mobile apps had promising but insignificant results (mean difference 0.17, 95% CI −2.21 to 2.55, *P*=.88). However, moderate heterogeneity was observed among the included studies (heterogeneity I^2=74.05%, Q=23.12, and τ^2=3.45) (Figure 5).

Figure 5. Forest plot of mobile phone apps for increased physical activity.



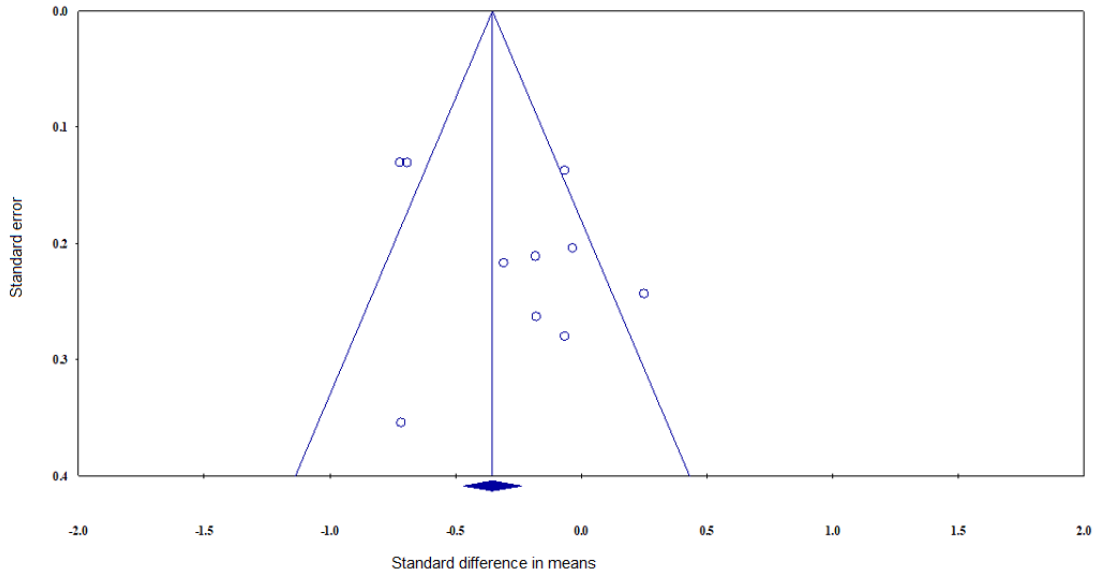
Sensitivity Analysis

Five studies evaluated the effectiveness of mobile apps used ≤3 months for weight loss. Pooled findings indicated no significant difference in weight between the mobile app intervention group and the control group (−0.004 kg, 95% CI −0.79 to 0.80, *P*=.99). Six studies evaluated the effectiveness of mobile apps used >3 months for weight loss. Pooled findings showed a significant difference in weight between the mobile app intervention group and the control group (−1.63 kg, 95% CI −2.64 to −0.61, *P*<.002). In addition, the pooled findings from only the randomized controlled trials showed a significant difference in weight between the mobile app intervention group and the control group (−1.03 kg, 95% CI −2.05 to −0.025, *P*=.04) (Multimedia Appendix 3, Multimedia Appendix 4, and Multimedia Appendix 5).

Publication Bias

The meta-analysis of the observational studies had some sort of publication bias. Egger regression test was used to calculate the publication bias, and a funnel plot was drawn to visualize it. The funnel plot in Figure 6 shows no relevant publication bias.

Figure 6. Funnel plot.



Discussion

Principal Findings

This meta-analysis evaluated the effectiveness of mobile app interventions for weight management. The meta-analysis showed a small but significant reduction in body weight (−1.07 kg) and a reduction in BMI (−0.45 kg/m^2). Our findings are more comprehensive across sensitivity analyses. Moreover, our findings showed that mobile app interventions promote additional positive health benefits through the maintenance of BMI and increased physical activity from baseline. The ubiquitous use of a mobile app intervention in any age group may therefore have great clinical value when compared with traditional interventions. A previous systematic review and meta-analysis reported pooled effects of app interventions compared with controls of −1.04 kg for weight loss and −0.45 kg/m^2 for BMI [14]. The magnitude of the mobile app intervention effect in our updated meta-analysis suggests that the use of mobile app interventions is effective for promoting body weight management following an initial weight loss when compared with other interventions.

Public Health Implications

Rising obesity and physical inactivity are associated with chronic diseases and increased health costs [31]. To reduce the health care burden, researchers have already pointed out the importance of effective health communication. Digital technology provides fast and interactive communication that is easy to use and cost-effective. Mobile phones have emerged as a potential medium for interventions to assist people in maintaining health, and they have shown promising results for weight loss and increased physical activity [32]. However, a mobile app can be tailored to the individual, and information could be delivered in a more effective way that may be more realistic and feasible than conventional ways to deliver information. Several studies have investigated the efficacy of weight management using text messaging interventions. A previous meta-analysis that included 14 studies with an intervention period ranging from 1 month to 2 years found that text messaging interventions can promote weight loss (−2.56 kg, 95% CI −3.46 to −1.65) [33].

Several studies have previously highlighted targeting behavioral change techniques that include dietary self-monitoring and reporting, behavior reinforcement through motivational messages, social support, setting and evaluating various goals, and setting reminders, and they are all key components to reduce and maintain weight successfully [34]. The use of mobile apps helps to improve self-awareness, provides valuable information, and can be an early indicator of health-related issues; this information and support can help to spur positive behavioral changes [35]. Results from this study suggest that even in the short term (<6 months), mobile app interventions provide a generally positive effect for reducing weight and maintaining BMI. In this meta-analysis, mobile apps were related to weight loss goals, calorie goals, self-monitoring of body weight, dietary intake, real-time feedback, educational content, behavioral change plans, nutrition, physical activity, several types of trainings, and reminders of goals. Moreover, the intervention types included apps, email, internet systems, online weight trackers, physical activity planners, blog facilities for communication, internet forums, parental newsletters, seminars, spot sessions, podcasts, twitter, and printable eating charts, which were all deemed to be effective for reducing weight and maintaining BMI.

Limitations

This meta-analysis shows that a mobile app intervention has a variety of uses in reducing weight and maintaining BMI with immense benefits. However, there are several limitations in this study that we need to address. First, the findings of this study should be interpreted with caution considering the sample sizes of the included studies and considering the studies that reported on short-term periods. To increase the generalizability of our findings, a larger sample size with a longer follow-up period (at least 1 year) in diverse racial or ethnic settings is needed. Second, although heterogeneity among the studies was not high, the results were nevertheless based on a relatively small number of studies. A small number of studies also prevented us from conducting a meta-regression analysis to evaluate other factors for reducing weight and maintaining BMI.

Conclusions

The findings of this study suggest that mobile app interventions appear to be feasible and acceptable for reducing weight, maintaining BMI, and increasing physical activity, although the overall effects might be relatively modest. However, the course averages hide some variation; therefore, such interventions are highly successful in some people and completely ineffective in some people. Public awareness of safety and the benefits of weight management and physical activity should be promoted, and more studies with a larger sample and longer follow-up are needed to evaluate the potential role of a mobile phone app intervention.

Acknowledgments

We would like to thank our colleague, who is a Native English speaker, for editing our manuscript. This research was funded in part by the Ministry of Education (MOE) under grant MOE 108-6604-001-400 and DP2-109-21121-01-A-01 and the Ministry of Science and Technology (MOST) under grant MOST 108-2823-8-038-002- and 109-2222-E-038-002-MY2.

Authors' Contributions

MMI contributed to the study design and acquisition and analysis of data, as well as led the writing of the manuscript. TNP was involved in the conceptualization of the study. MMI and TNP independently screened the literature, and YCL resolved discrepancies.

MMI, TNP, BAW, and YCL reviewed the manuscript and provided relevant feedback on the manuscript. All authors have read and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 17 KB - mhealth_v8i7e17039_app1.docx]

Multimedia Appendix 2
Search words.
[DOCX File , 153 KB - mhealth_v8i7e17039_app2.docx]

Multimedia Appendix 3
Intervention duration ≤3 months.
[DOCX File , 19 KB - mhealth_v8i7e17039_app3.docx]

Multimedia Appendix 4
Intervention duration >3 months.
[DOCX File , 81 KB - mhealth_v8i7e17039_app4.docx]

Multimedia Appendix 5
Only randomized controlled trials.
[DOCX File , 21 KB - mhealth_v8i7e17039_app5.docx]

References

1. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 2011 Aug 27;378(9793):804-814. [doi: 10.1016/S0140-6736(11)60813-1] [Medline: 21872749]
2. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism 2019 Mar;92:6-10. [doi: 10.1016/j.metabol.2018.09.005] [Medline: 30253139]
3. MacLean PS, Wing RR, Davidson T, Epstein L, Goodpaster B, Hall KD, et al. NIH working group report: Innovative research to improve maintenance of weight loss. Obesity (Silver Spring) 2015 Jan;23(1):7-15 [FREE Full text] [doi: 10.1002/oby.20967] [Medline: 25469998]
4. Franz MJ, Boucher JL, Rutten-Ramos S, VanWormer JJ. Lifestyle weight-loss intervention outcomes in overweight and obese adults with type 2 diabetes: a systematic review and meta-analysis of randomized clinical trials. J Acad Nutr Diet 2015 Sep;115(9):1447-1463. [doi: 10.1016/j.jand.2015.02.031] [Medline: 25935570]
5. Dombrowski SU, Knittle K, Avenell A, Araújo-Soares V, Sniehotta FF. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. BMJ 2014 May 14;348:g2646 [FREE Full text] [doi: 10.1136/bmj.g2646] [Medline: 25134100]
6. Ogilvie D, Foster CE, Rothnie H, Cavill N, Hamilton V, Fitzsimons CF, Scottish Physical Activity Research Collaboration. Interventions to promote walking: systematic review. BMJ 2007 Jun 09;334(7605):1204 [FREE Full text] [doi: 10.1136/bmj.39198.722720.BE] [Medline: 17540909]
7. Armit CM, Brown WJ, Marshall AL, Ritchie CB, Trost SG, Green A, et al. Randomized trial of three strategies to promote physical activity in general practice. Prev Med 2009 Feb;48(2):156-163. [doi: 10.1016/j.ypmed.2008.11.009] [Medline: 19100282]
8. Vodopivec-Jamsek V, de Jongh T, Gurol-Urganci I, Atun R, Car J. Mobile phone messaging for preventive health care. Cochrane Database Syst Rev 2012 Dec 12;12:CD007457 [FREE Full text] [doi: 10.1002/14651858.CD007457.pub2] [Medline: 23235643]
9. Sanou B. International Telecommunications Union Geneva. ICT Facts and Figures 2016 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [accessed 2017-01-01]
10. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text Messaging-Based Interventions for Smoking Cessation: A Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
11. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]

12. Ricci-Cabello I, Bobrow K, Islam SM, Chow CK, Maddison R, Whittaker R, et al. Examining Development Processes for Text Messaging Interventions to Prevent Cardiovascular Disease: Systematic Literature Review. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e12191 [FREE Full text] [doi: 10.2196/12191] [Medline: 30924790]
13. Turner-McGrievy G, Tate D. Tweets, Apps, and Pods: Results of the 6-month Mobile Pounds Off Digitally (Mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec 20;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]
14. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
15. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [doi: 10.1001/jama.283.15.2008] [Medline: 10789670]
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
17. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane. Cochrane Handbook for Systematic Reviews of Interventions Version 5 URL: https://training.cochrane.org/handbook [accessed 2015-03-25]
18. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
19. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. Int J Epidemiol 2008 Oct;37(5):1148-1157 [FREE Full text] [doi: 10.1093/ije/dyn065] [Medline: 18424475]
20. 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):e78 [FREE Full text] [doi: 10.2196/mhealth.5768] [Medline: 27335237]
21. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing Self-Monitoring Strategies for Weight Loss in a Smartphone App: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
22. Lee W, Chae YM, Kim S, Ho SH, Choi I. Evaluation of a mobile phone-based diet game for weight control. J Telemed Telecare 2010;16(5):270-275. [doi: 10.1258/jtt.2010.090913] [Medline: 20558620]
23. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
24. 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]
25. Brindal E, Hendrie G, Freyne J, Coombe M, Berkovsky S, Noakes M. Design and pilot results of a mobile phone weight-loss application for women starting a meal replacement programme. J Telemed Telecare 2013 Apr;19(3):166-174. [doi: 10.1177/1357633X13479702] [Medline: 23520213]
26. 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 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
27. 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]
28. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):e723-e731. [doi: 10.1542/peds.2014-1012] [Medline: 25157000]
29. 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 2014 Aug;27(4):322-332. [doi: 10.1111/jhn.12155] [Medline: 23992038]
30. Partridge SR, McGeechan K, Hebden L, Balestracci K, Wong AT, Denney-Wilson E, et al. Effectiveness of a mHealth Lifestyle Program With Telephone Support (TXT2BFiT) to Prevent Unhealthy Weight Gain in Young Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2015 Jun 15;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
31. Holmes WS, Moorhead SA, Coates VE, Bond RR, Zheng H. Impact of digital technologies for communicating messages on weight loss maintenance: a systematic literature review. Eur J Public Health 2019 Apr 01;29(2):320-328. [doi: 10.1093/eurpub/cky171] [Medline: 30239699]

32. Shaw R, Bosworth H. Short message service (SMS) text messaging as an intervention medium for weight loss: A literature review. Health Informatics J 2012 Dec;18(4):235-250 [FREE Full text] [doi: 10.1177/1460458212442422] [Medline: 23257055]
33. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
34. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
35. Liu F, Kong X, Cao J, Chen S, Li C, Huang J, et al. Mobile phone intervention and weight loss among overweight and obese adults: a meta-analysis of randomized controlled trials. Am J Epidemiol 2015 Mar 01;181(5):337-348 [FREE Full text] [doi: 10.1093/aje/kwu260] [Medline: 25673817]

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by E Bellei, LK Chen, A Griffin, C Carrion, K Tamura; comments to author 08.03.20; revised version received 16.03.20; accepted 07.04.20; published 22.07.20.

Please cite as:
Islam MM, Poly TN, Walther BA, (Jack) Li YC
Use of Mobile Phone App Interventions to Promote Weight Loss: Meta-Analysis
JMIR Mhealth Uhealth 2020;8(7):e17039
URL: https://mhealth.jmir.org/2020/7/e17039
doi:10.2196/17039
PMID:32706724

Review

mHealth and Engagement Concerning Persons With Chronic Somatic Health Conditions: Integrative Literature Review

Hanna Tuvesson[1], PhD, RN; Sara Eriksén[2], PhD; Cecilia Fagerström[3,4], PhD, RN

[1]Department of Health and Caring Sciences, Linnaeus University, Växjö, Sweden

[2]Blekinge Institute of Technology, Karlskrona, Sweden

[3]Department of Health and Caring Sciences, Linnaeus University, Kalmar, Sweden

[4]Blekinge Centre of Competence, Blekinge County Council, Karlskrona, Sweden

Corresponding Author:
Hanna Tuvesson, PhD, RN
Department of Health and Caring Sciences
Linnaeus University
Universitetsplatsen 1
Växjö
Sweden
Phone: 46 0480446915
Email: hanna.tuvesson@lnu.se

Abstract

Background: Chronic somatic health conditions are a global public health challenge. Being engaged in one's own health management for such conditions is important, and mobile health (mHealth) solutions are often suggested as key to promoting engagement.

Objective: The aim of this study was to review, critically appraise, and synthesize the available research regarding engagement through mHealth for persons with chronic somatic health conditions.

Methods: An integrative literature review was conducted. The PubMed, CINAHL, and Inspec databases were used for literature searches. Quality assessment was done with the guidance of Critical Appraisal Skills Programme (CASP) checklists. We used a self-designed study protocol comprising 4 engagement aspects—cognitive, behavioral and emotional, interactional, and the usage of mHealth—as part of the synthesis and analysis.

Results: A total of 44 articles met the inclusion criteria and were included in the analysis. mHealth usage was the most commonly occurring engagement aspect, behavioral and emotional aspects the second, cognitive aspects the third, and interactional aspects of engagement the least common aspect in the included articles. The results showed that there is a mix of enablers and barriers to engagement in relation to the 4 engagement aspects. The perceived meaningfulness and need for the solution and its content were important to create and maintain engagement. When perceived as meaningful, suitable, and usable, mHealth can support knowledge gain and learning, facilitate emotional and behavioral aspects such as a sense of confidence, and improve interactions and communications with health care professionals.

Conclusions: mHealth solutions have the potential to support health care engagement for persons with chronic somatic conditions. More research is needed to further understand how, by which means, when, and among whom mHealth could further improve engagement for this population.

(JMIR Mhealth Uhealth 2020;8(7):e14315) doi:10.2196/14315

KEYWORDS

engagement; eHealth; mHealth; somatic disease; integrative literature review; telehealth

Introduction

Background

The growing burden of chronic health conditions has been described as a challenge and threat to health worldwide [1]. Ischemic heart disease, cerebrovascular disease, chronic respiratory conditions, and diabetes are the leading causes of death [2] and are predicted to continue increasing through the year 2030, according to a projection made in 2006 [3]. To address these challenges, patient engagement in treatment and

care has been suggested as a key and powerful resource [4], with a shared understanding in the health care sector that engaged patients could improve quality of care and reduce unnecessary costs [5]. As part of ongoing developments in the field, health care organizations and national health care policies in many countries have recently begun focusing on engaging patients in the management of their health [6]. Various strategies and methods, such as electronic health (eHealth) and mobile health (mHealth), have been considered for facilitating and maintaining engagement with patients who have chronic somatic health conditions. Despite increased attention, however, there is still a lack of compiled information regarding engagement when using technological solutions as support in the care and management of health.

Although there seems to be a consensus concerning the importance of engagement in relation to living with and managing chronic health conditions, the concept of engagement in this context is underdefined, with several terms and definitions available to describe the phenomenon, and there is no shared understanding of how to conceptualize it [4,7]. Aside from *engagement*, commonly used terms are *commitment*, *activation*, *involvement*, and *adherence*. However, the concept of engagement conveys a somewhat divergent meaning from the other listed concepts, implying emotional commitment and involvement of the engaged individual. Moreover, depending on the discipline involved, the various preferences for which term to use and who should be engaged further complicate the understanding of the concept. Commonly used terms are *patient engagement* [4,8], *citizen engagement*, and *client engagement* [4]. *Consumer engagement* is another commonly used term [4,5], whereas in computer science, where the object and objective of engagement may often be more clearly delimited to a specific technological solution and its design process, *user engagement* is frequently used.

The American Health Information Management Association (AHIMA) [5] states that engagement comprises various activities such as interacting with health care professionals, seeking health information, maintaining a personal health record, and playing an active role in making decisions related to personal health care. Definitions of engagement in the literature have also been reviewed and categorized as *intraindividual factors* such as emotional, behavioral, and cognitive factors, and as a function of *interindividual factors* in terms of relationships and interactions. Many of the definitions have been suggested to be oversimplified and lacking in consideration of the progressive development of engagement over time as a process [4]. A recent review found that engagement in relation to digital behavior change interventions could be understood as 2 constructs: a subjective experience and behaviors [7]. Engagement in relation to digital and technological solutions has been described in terms of usage and patterns of usage of a product, such as the number of logs and time required to use the tool [9]. In this review engagement is understood as a multidimensional progressive process comprising cognitive, emotional, behavioral, relational, and interactional elements, as well as the usage of a product, service, or system.

Increasingly, various technological solutions are being suggested to support the health management of persons with chronic health conditions and to enhance their engagement. Mobile technology has, for example, been suggested to have an important role in facilitating patient engagement [10]. mHealth has been defined as the delivery of health care via the generation, aggregation, and dissemination of health information using mobile or wireless devices and the sharing of that information between patients and providers [5]. According to the World Health Organization [11], mHealth refers to medical and public health practices supported by mobile devices, such as mobile phones, patient monitoring devices, and personal digital assistants. In a systematic review [12], 3 distinct periods were identified with regard to the mobile devices used in mHealth research: before 2007, personal digital assistants dominated, while basic and feature phones took the lead during 2007-2012, after which smart devices came to dominate mHealth research. Over the past decade, there has been a considerable increase in the number and range of mHealth solutions that have been developed and implemented [11]. Because of their popularity, availability, portability, and technological capacity, smartphones and mHealth have enormous potential to impact chronic disease management around the globe [13]. However, so far, evaluations of mHealth solutions that focus on coverage, functionality, and impact on public health are few and far between [11]. In addition, in a systematic review of the impact of mHealth chronic disease management on treatment adherence and patient outcomes, Hamine et al. [13] argue that the impact of mHealth tools on adherence to treatment regimens may be overlooked because mHealth promoters are mainly focused on demonstrating their more direct effects on clinical outcomes (eg, morbidity, mortality, and biometric markers of clinical disease), while the long-term and more indirect effects of mHealth tools on adherence to treatment have not been in focus. Adherence to treatment, specifically adherence to treatment of chronic diseases, they argue, is critical to achieving improved health outcomes, quality of life, and cost-effective health care [13].

Various factors that are influential in the engagement process concerning mHealth solutions have been described; for example, in relation to the technological device/system (such as smartphone/computer access and technological shortcomings), the context (such as internet access and settings) and targeted behavior (such as expectations and meaningfulness) [5,7,10] are important factors. A known consumer/patient engagement challenge is that the patient may be uninterested in becoming engaged in the health care delivery process. Consumers are becoming more concerned with personal information security, including scenarios of hacking or identity theft, and in some cases, they have decided not to interact with online technology solutions. One of the major obstacles that providers face is that their patients' health information is not always accurately and easily exchanged from one provider to the next [5]. The ability for patients and caregivers to access their health information electronically has been described as facilitating engagement [6], and maintaining engagement could improve personal and public health, patient experiences, and cost-reduction efforts [5]. Therefore, overcoming barriers to engagement and finding ways for those with chronic somatic health conditions to become engaged in their personal health and maintain that engagement are crucial. At this point, it is essential to review the research

with regard to engagement through mHealth and to present the knowledge to identify the gaps and needs from the perspectives of those living with chronic somatic health conditions.

Objectives

The aim of this study was to review, critically appraise, and synthesize the available research regarding engagement through mHealth for persons with chronic somatic health conditions.

Methods

Literature Search and Selection

An integrative literature review was made following the description of Whittemore and Knafl [14]. Systematic literature searches were performed in 2017 using 3 electronic bibliographic databases: CINAHL, PubMed, and Inspec. Two blocks of keywords were used to build the search strings for the identification of studies regarding engagement and mHealth among those with chronic somatic health problems: (1) engagement and (2) mHealth. Different forms (ie, inflections of engagement) of the word *engagement* were taken into account and used as keywords for the first block of the search. For the second one, (1) a list of keywords that were synonymous with or related to mHealth and (2) relevant MeSH and CINAHL headings were used as keywords, in combination. The concept of mHealth is embedded in the MeSH and CINAHL headings telemedicine and telehealth, and these were therefore used as keywords. A search string was conducted by combining the engagement keywords and the mHealth keywords with the Boolean operator OR. These 2 blocks of keywords were then combined with AND to complete the search string. The keywords and MeSH and CINAHL headings are presented in Textbox 1.

A total of 1907 references were found in the 3 databases. After the removal of duplicates, studies not written in English, and studies that were obviously irrelevant due to content or article type, 501 studies remained for further detailed screening. These 501 studies were screened using the abstract to determine eligibility. Studies were included if they (1) included patients 18 years of age or older, (2) were written in English, and (3) reported empirical results concerning engagement with patients suffering from chronic somatic health problems using, or participating in the design or research of, mHealth solutions. In the screening process, studies describing solutions that could be accessed through mobile or wireless devices were considered as mHealth [5]. No restriction by publication year was made as we anticipated that articles focusing on mHealth would be published in recent years. Studies were excluded if they focused on infectious and psychiatric conditions or reported on various injuries. Studies with an entirely preventive focus were also excluded. Several articles were excluded as they did not report any empirical findings; for example, editorials, letters, and study protocols. We were also unable to retrieve 7 of the references in full-text versions.

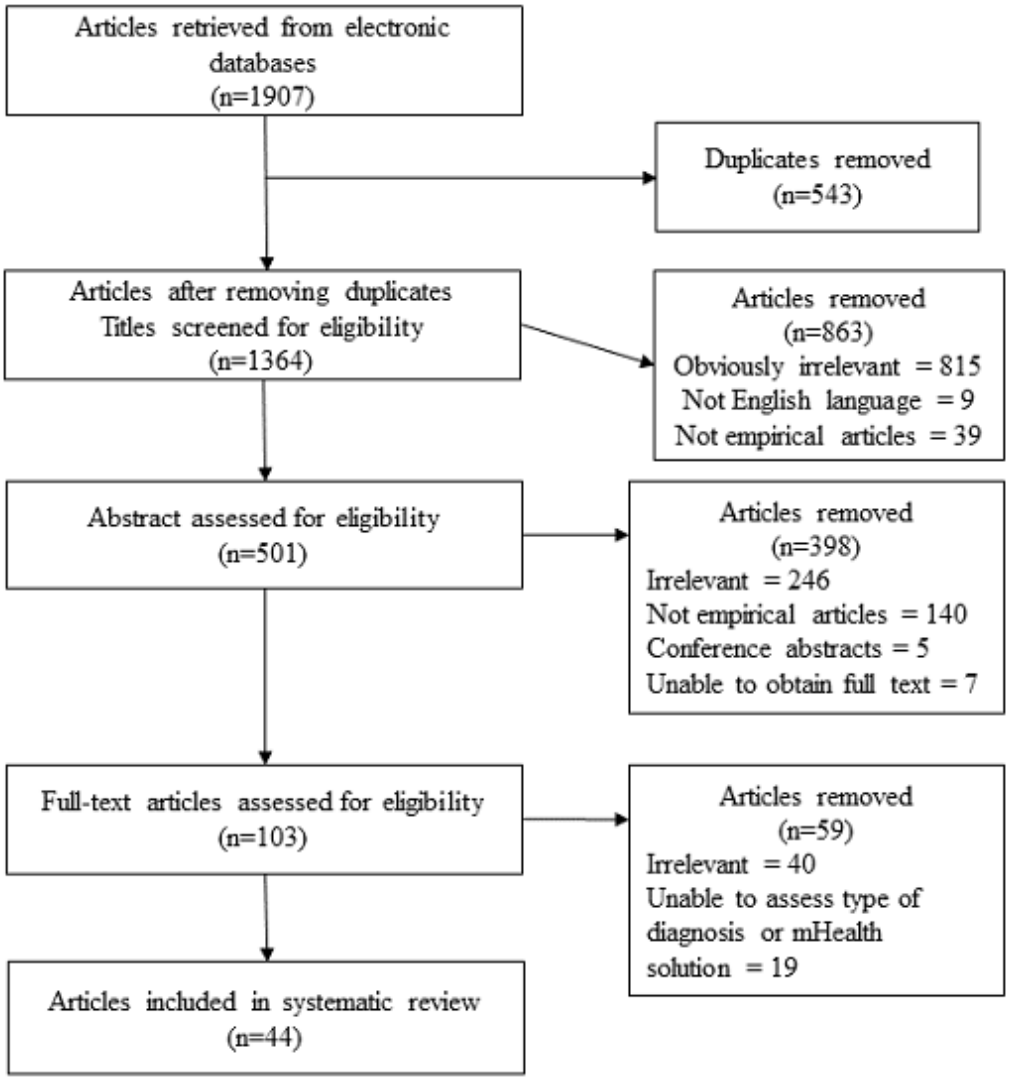
In the last step of the screening process, the full-text copies of 103 articles were reviewed by 2 authors (HT and CF), and the results from the reviews were compared and discussed until there was acceptable inter-reliability between the reviewers (κ=0.9). Studies with mixed diagnoses were excluded if the results could not be directly related to one specific chronic health condition or if there was difficulty determining what type of technological solution had been used. Studies were also excluded if the results did not report findings on engagement in relation to the mHealth solution. In the end, we identified 44 studies as suitable according to the inclusion criteria. A detailed description of the data screening process is shown in Figure 1.

Textbox 1. Search strategy and used search terms.

Search terms

- Engage OR Engaging OR Engaged OR Engagement AND Search terms (mHealth OR M health OR mobile health OR m-health OR mobile phone/-s OR mobile telephone/-s OR smart phone/-s OR cellular phone/-s OR cellular telephone/-s OR cell phone/-s OR cell telephone/-s OR telehealth OR Telemedicine)
- Engage OR Engaging OR Engaged OR Engagement AND MeSH and CINAHL headings (cell phones OR cellular phone OR smart phone OR smartphone OR telemedicine OR Telehealth)

Figure 1. Flowchart of the data selection process.



Quality Assessment

The 44 included studies were critically assessed for methodological quality using the CASP checklists for qualitative, randomized controlled trials, case–control, and cohort studies [15]. For the included mixed-methods studies, 2 checklists were used. The checklists, used worldwide, each comprise 10-12 questions with a numeric value representing each answer. All numeric values, based on subjective assessments, were summed to create a quality summary of each study. Because the quality assessment was part of the compilation, all studies screened for this integrative review were kept for further analysis regardless of the results of the quality assessment. A summary of the quality assessments for the 44 included studies is presented in Multimedia Appendix 1.

Data Analysis

Data were identified and extracted from the articles based on protocols designed for this review by 2 authors (HT and CF; Multimedia Appendix 1). The protocol comprised information about the studies in terms of country, year of publication, study methods, type of mHealth solution, type of chronic somatic health condition, number of study participants, mean age of participants, duration of project or intervention, and whether information about research ethics was provided. Research ethical aspects, such as informed consent, voluntariness, and if applicable, advice or approval from ethical committees, is important when researching health aspects in patients and we decided to add this aspect in the study protocol to provide an overview. The protocol also comprised 4 aspects of engagement to be identified in the studies. These aspects were chosen based on definitions and descriptions in the literature of engagement as a process [4,5,7,9]. They were (1) cognitive aspects (eg, learning, understanding, knowledge), (2) behavioral and emotional aspects (eg, action, motivation, confidence, self-management, involvement), (3) interactional aspects (eg, relationships, communication with and accessibility to health professionals), and (4) mHealth usage (eg, participation, responses, logs, duration). In case of overlapping meanings with regard to the engagement aspects in the same article, data were sometimes reused in more than one engagement aspect. The results of the data identification completed by 2 reviewing authors (HT and CF), independently and according to the protocols, were compared and showed high levels of agreement. The results of the identification are detailed in the study protocol shown in Multimedia Appendix 1.

In the next stage of the analysis, data related to the 4 engagement aspects were identified and retrieved from the articles and then assembled in a document by one author (HT). The retrieved data were then read and compared with the aim of this study by all 3 authors. Even at this stage, ambiguous data were discussed until consensus was reached. The data were compared, and patterns and contrasts noted, before the analysis was discussed and revised until consensus was reached among the authors. Verification was conducted by comparing the results of the data analysis with the original articles.

Results

Description of Studies

Of the 44 articles included in this review, half (n=22) were published in the years 2015 and 2016. The majority of the studies were conducted in the United States (n=26), and the designs were mainly mixed-method or randomized controlled trials. Several studies compared groups, such as in quasi-experimental studies with a study group and a control group. Some studies (n=5) were based only on an analysis of program or system data, whereas the rest were qualitative studies, surveys, or case studies. The included studies used a mix of different mHealth solutions, including apps, interactive voice response, short message service (SMS), telemonitoring, websites, and personal health records. The duration of the study period varied greatly, from a couple of hours in a laboratory experiment to 2 years of follow-up. No information about research ethical approval or ethical aspects of the research was given in 10 of the 44 included studies.

Study participants were between ages 28 and 88, with an overall mean age of 57. The majority of the studies included individuals with diabetes or cardiac diseases or both (n=36). Several studies focused on hypertension or pulmonary diseases. There were also articles involving patients with kidney disease, pain, spina bifida, osteoarthritis, and cystic fibrosis. A total of 20 studies reached high methodological quality according to CASP, whereas 21 reached medium quality, and 3 studies had low quality (Multimedia Appendix 1).

Engagement and mHealth

Cognitive Aspects of the Engagement Process

A total of 22 studies described cognitive aspects of the engagement process (Multimedia Appendix 1). Many of the studies reported that the mHealth solution supported engagement by providing persons with chronic health conditions with information about their health and their medical condition through personal health information, patient education in various forms, and a blend of feedback strategies [16-21]. This was, for example, described in ways such as seeking information on a website housed within a patient portal section of a public website [17] and gaining immediate and accurate information about one's health status and health management, for example, by the tracking of blood glucose level, weight, and nutritional information based on food intake [20,21]. In some studies, feedback from the mHealth product resulted in disengagement if it was perceived as confusing, inaccurate, or complicated [22,23].

Persons with chronic health conditions learned new skills and gained knowledge about their health and medical conditions as a form of engagement when using mHealth solutions [16,17,19,20,24-31]. The learning process as a form of engagement was expressed in various ways. For example, mHealth solutions with coaching and educational components, such as animated scenarios, combined audio and text instructions and quizzes, supported learning [16,28,29].

Cognitive aspects of the engagement process were also described as understanding and awareness. Receiving personal, immediate, and understandable feedback about their clinical results and health management, based on self-entered information or clinical responses or both after health care visits, increased understanding and awareness [18,22,26,27,32-35]. Feedback such as SMS text messages or illustrative graphs helped patients understand and be aware of the relationships between clinical values, trends, and early warning signs of potential problems [26,34,35]. For some persons, the understanding and reflection first began after receiving glycemic readings that were outside of the recommended range or were unexpected [35].

Behavioral and Emotional Aspects of the Engagement Process

A total of 26 studies in this review included behavioral and emotional aspects as a form of engagement, often in relation to self-management and self-care components. Several studies reported that the mHealth solution supported management or improved self-management (Multimedia Appendix 1). For example, in one study, 64% (18/28) of the participants with diabetes type 2 thought that an iPad (distributed to them with installed existing apps) had helped them manage their diabetes [20], 97.4% (793/814) of patients with hypertension felt more confident in taking their blood pressure after using a telehealth programme [32], and patients in cardiac rehabilitation perceived a smartphone app to have supported their adherence to cardiac rehabilitation activities [36].

Several studies reported on significant improvements in self-management and self-care. For example, in one telemedicine study, people with diabetes type 2 had significantly improved self-efficacy after 2 years [37]. Significant improvements in self-care have also been found when comparing a telehealth group and a control group [38], and in another study, participants with osteoarthritis who had 12 months of exposure to a freely available mHealth website with evidence-based information and self-management resources had significantly improved self-management compared with nonusers [39]. In one study where both the study group and the control group comprised patients with chronic heart failure, both groups showed significant improvements in the self-management of their disease [40].

In several studies, SMS text messages or interactive voice response calls were found to be helpful with behavioral aspects of engagement [22,33,41,42]. Some users perceived reminders and alarms as helpful for engaging in health-related activities [19,29], but they could also be seen as disengaging by being nagging [19]. Automatic feedback messages or system feedback has been used to support behavioral and emotional aspects of engagement. For many people, these messages were perceived

as beneficial in that system feedback reinforced positive changes in behavior and helped break negative cycles [43], helped patients establish and achieve goals [25,35], helped to establish routines [33], and reinforced compliance [36]. In at least one study, online stand-alone personal electronic health records were found to be not crucial for daily self-management [18].

Different mHealth solutions, such as telemonitoring and various apps, supported emotional components of the engagement process. For example, solutions were found to support or enhance motivation [19,23,29,31,43], provide a sense of control and mastery [20,23,27,29], boost confidence [20,21,29,35,40], and give users a sense of empowerment [20,21,26] through self-care and management of their illness.

Conversely, some mHealth solutions also created emotional hindrances to engagement. The tools could lead to less self-management after the novelty phase compared with that at the start of the program [20], and in some cases, users would experience worry and depressive feelings when readings were worse than normal, leading to a decline in self-management [27]. Still, other patients did not want to use and interpret readings, preferring to leave that to health professionals [27]. Further, patients who were already motivated before using an app or other tool did not necessarily become more motivated [31].

Interactional Aspects of the Engagement Process

The 18 studies describing interactional aspects of the engagement process were primarily focused on ways of communicating and sharing information, such as with regard to physical visits to health care professionals (Multimedia Appendix 1). The experiences of users included feelings of being continuously connected to health care professionals [26,27,33,36,44,45] and having increased access to care [26,46]. Some users thought that the mHealth solution had improved their visits to health care professionals [36], and a few studies found that the mHealth solution (including small-sized peripheral telehealth equipment with individual reading and teletherapy technology with tailored counseling) prevented admission to hospitals [26,47]. However, another study showed that hospital visits had increased in both the control group and the intervention group [48].

In one study, shared information was seen by some patients as beneficial for engagement, but most participants did not share their personal electronic health record with health care professionals [18]. Knowing that health care professionals were able to access their electronic health records and see recent test results or monitor readings made many users feel safe, peaceful, and watched over [26,27,49]. However, others felt uncomfortable or unsafe with their personal health information being shared in an mHealth system, feeling like they were under surveillance [29,49].

Many thought that communication through mHealth systems was insufficient in many ways, pointing to these tools' lack of face-to-face contact or a sense of *human touch* [21,26], while others were concerned about the quality of the communication due to, for example, the lack of access to body language or physical demonstrations with equipment [26,48], which were perceived as creating disengagement.

mHealth Usage

A majority of the included studies (37/44) described the engagement process in relation to usage of the mHealth solution through process measurements, for instance, where engagement was reduced due to the high number of activities involved, such as completion of messages, log-ins required, and system tracking of various sorts (Multimedia Appendix 1). High mHealth utilization rates were reported in a number of studies [21,22,35,36,38,44,45,47,48,50], but a few described limited utilization [17,24,41]. In several studies engagement was related to the time spent using the mHealth solution and if or when participants dropped out or declined to use the solution. The analyses show a great variation in engagement as described in these papers. In several studies, a decline in usage was reported, irrespective of whether the studies had shorter or longer durations [23,32,45,46,50,51]. There were some studies that found that the majority of participants stayed interested and continued to use the mHealth solution throughout the study period [36,44,47,48].

Some studies found that engagement in terms of usage varied depending on the user and his or her situation. Older patients with poor health, for example, were found to be less engaged in various mHealth solutions [42,52]. At the same time, having a need that could be satisfied with an mHealth solution, such as a need for self-management assistance, led some people to maintain engagement [23,53], whereas in another case, a ceiling effect was found when blood glucose levels were under control [46]. One study found that unmarried persons had lower engagement [52] rates, and in several studies, participants with informal caregivers had high engagement rates [42,50,52,53].

Several factors were described as influencing the engagement process. Lack of time and busy lifestyles, with constraints due to home obligations, childcare, work, and travel leading to difficulties in integrating the solution into daily life, often resulted in disengagement [18,23,28,34]. Technological problems such as glitches, errors, connectivity problems, and battery attrition were other factors described as influencing engagement [23,28,31,34]. Concerns about trust, privacy, and security were also reported to influence engagement and usage negatively [18,19,34,43]. Conversely, visually exciting and dynamic components in the mHealth solution were factors that could stimulate engagement [19].

Discussion

Principal Findings

This study aimed at reviewing, critically appraising, and then synthesizing the body of research regarding engagement with mHealth solutions by persons with chronic somatic health conditions. In total, we found 44 studies that met the inclusion criteria and described aspects of the engagement process in terms of cognitive, behavioral and emotional, or interactional aspects, along with those regarding the specific usage of an mHealth solution.

mHealth and the Engagement Process

Overall, the results of this review support the idea that engagement is a process and that cognitive, behavioral and emotional, interactional, and mHealth usage–related aspects of engagement go hand-in-hand and influence each other in many cases. The individual person and his or her situation, along with the type of health problem and the stage of some conditions, are important factors in mHealth use. The specific mHealth solution is also important in terms of content, usability, and visibility. When an mHealth solution is perceived as supportive and meaningful, the person's knowledge and understanding of his/her health and interaction with health care services can be facilitated. This, in turn, may result in more awareness of self-management of health as well as emotional empowerment in terms of confidence and control. Previous studies have found that when data transmissions and health information are being transferred to health care professionals and information about the data is asynchronously being sent back to the patient, the patient sometimes feels more involved and engaged in the interpretations of data [54,55]. Yet, this could also lead to passivity and disengagement as the person's personal interpretations of his or her health situation are being unattended in one-way communications [56]. mHealth solutions with self-care components convey an enhanced responsibility to someone with somatic health conditions, and therefore, these solutions need to be flexible when it comes to user preferences [54].

One interpretation of the results is that patient needs is one factor that could, alone, be a deal breaker for engagement. If the disease is under control, no new medicines have been added, or no new knowledge is needed, the engagement process may no longer be progressive in nature. The mHealth solution needs to serve an actual perceived need for the patient to maintain engagement with it. This was illustrated, for example, in the results with several studies that reported satisfying usage and engagement in the mHealth solution for the majority of the participants, but there were difficulties in maintaining engagement in the long term. This could possibly be related to ceiling effects and lack of perceived need to use the mHealth solution when knowledge gaps have been filled or when the disease is under control, and in this regard, a lack of engagement could be interpreted as something positive. These findings are in line with previous research where individuals ceased to use mHealth solutions when they had reached their goals, felt a lack of stimulation, or *outgrew* the solution [56], among other things. However, this insight could be used constructively when planning for mHealth interventions by targeting user groups in a more focused way, with the aim of reaching specifically those groups of people who are in a situation where the need for mHealth solutions is perceived as most pressing. Such *windows of opportunity* for achieving a high level of mHealth engagement could, for instance, be during the first 6 months after a person has been diagnosed with diabetes type 2, when the shock of the new situation, a lack of established habits and routines for dealing with it, and the need for information and deeper knowledge about the disease and how to manage it create a clearly perceived need for efficient mHealth support [57]. In addition, Zrebiec [58] found that people who have had diabetes for more than 2 years might have a need for information due to progression of the disease or the onset of diabetes-related complications or both. The literature suggests various options for maintaining engagement over time in mHealth solutions, such as with incentives. In a systematic review of incentive-driven mHealth solutions, the use of education, reminders, feedback, social aspects, alerts, gamification, and financial incentives were mechanisms found to engage and motivate users. The authors of that review discuss that although technological innovations change and develop, the incentives do not, and few new approaches have been presented [59]. The results of our review indicate that creative and intriguing mechanisms, such as artificial intelligence components (eg, customized SMS text message reminders, monitors, and avatars), could stimulate engagement. Furthermore, Liu and colleagues [60] found that loyalty rewards significantly influenced enrollment, but ongoing engagement was not influenced by the rewards. Stopping the usage of mHealth (nonusage attrition) or being lost to follow-up (dropout attrition) in mHealth research could be understood as the phenomenon of *the law of attrition* as formulated by Eysenbach [61]. According to Eysenbach, an innovation will be rejected if it is perceived as not providing any benefits or if there are usability problems. However, other factors are also important for understanding the engagement or disengagement process, such as demographics and study settings [62]. Furthermore, our review found that the context, characteristics of the person, and his or her life situation are crucial factors for engagement that need to be considered.

Of the 4 different aspects of engagement in mHealth among persons with chronic somatic health conditions as identified in this study, mHealth usage was the most commonly occurring aspect, identified in 37 of the 44 papers. The category of behavioral and emotional aspects was the second most common, identified in 26 of the 44. Cognitive aspects were identified in 22 articles, whereas interactional aspects of engagement were only identified in 18 articles included. A logical conclusion concerning this difference is that it indicates something about the history of research on engagement in mHealth; during the early years of mHealth studies, the focus would have primarily been on who was adopting the new technologies and what their associated characteristics (eg, demographic) were. Our hypothesis is thus that mHealth usage was the primary point of interest in the early years of mHealth studies and that studies of technology use have been historically heavily influenced by behaviorism and cognitive psychology. This could explain why behavioral and emotional aspects are the second most common aspects of mHealth engagement, as identified in the results of our study. Cognitive aspects come in third place, while interactional aspects come last of all, which is a bit surprising given that mobile technology has been developed primarily as a tool for communication and knowledge sharing. As such, it should be seen as supporting more efficient communication and interaction in health care. Why do we see relatively little focus on the interactional aspect of engagement in our study? Could this have historical roots in the health and medical sciences as well as in mHealth research, that is, could it be a reflection of a history heavily influenced by behaviorism and cognitive psychology and only more recently becoming influenced by social psychology, phenomenology, and social anthropology

as well as inspired by work practice theory, ethnographic and ethnomethodological studies, and participatory design? Such an interpretation resonates with recent and current discourse in human–computer interaction research on a perceived ongoing transition between second- and third-generation theory and technology [62,63], a transition characterized by a shift of focus from *human factors* to *human actors*, and a more human-centered and participatory approach to the design and use of technology [64].

It is also interesting in a study such as this to explore what might have been expected but appears to be missing in the results. Most of the articles included in this literature review related to diabetes and heart disease, concerned older patients, and were conducted in middle- and high-income countries. This might mean that the engagement process in youths or young adults, in patients with other chronic somatic health conditions, or in developing countries will involve different enablers and barriers that are not visible in this review. Thus, there appears to be a need for more research into mHealth tools as they relate to other chronic somatic health conditions, different populations, and across wider geographical areas.

Design and Quality of the Studies

The included studies varied in methodological design, which imposed a few challenges in terms of quality appraisal. For example, the methodological descriptions were sometimes limited due to mixed-method designs or for those involving larger ongoing projects where methodological information regarding the project had previously been published. Quality appraisal is a key component of a systematic review [65,66], but in mixed-method reviews, it can be difficult to find appropriate quality assessment guidelines that work for all included studies. Using mixed methods to understand complex phenomena and telehealth interventions is useful [67,68], yet threats to quality could be more difficult to identify compared with monomethod studies [68]. Because many of the studies in this review used a mix of methods in the same article, 2 CASP checklists were used for the same study in those cases.

No ethical aspects of research were mentioned in 10 of the 44 included studies. Arguably, research ethics must be considered the cornerstone of all research involving humans. In this review, we included research on persons with chronic somatic health conditions, a group that may be vulnerable or dependent on health care services. We realize that research ethics, laws, and standards vary among countries, but descriptions of particular aspects, such as protection from research-related harms, confidentiality, and informed consent, are important and applicable globally.

Strengths and Limitations

Strengths of this review include that each stage of the study, such as the identification of keywords, applicable articles, quality assessment, and analysis, was conducted by more than one author with regular discussions held between authors to ensure consistency. The authors of this review have a mix of methodological experiences, which was a strength when reviewing and synthesizing the data. At the start of this study, the initial idea for the literature search was to use synonymous words for the concept of engagement, for example, *commitment* and *involvement*. When using these synonyms in the first search strings, however, it became apparent that the majority of the studies did not involve what we were setting out to review in terms of engagement. Other related words such as *patient/consumer facing* and *motivation* were also excluded from use, owing to the same reasons. We therefore decided to limit the use of search terms to inflections of the word engagement and to create a protocol for engagement aspects that could guide the review and analysis process based on the description of the concept guiding this review. The concept and phenomenon of engagement are complex, and there may be a risk that relevant data were missed due to this approach, but we believe that it was primarily a strength in relation to the aim of this review.

In the literature search, we used a broad definition of mHealth and included studies that we understood were using mobile or wireless devices [5]. The use of MeSH and CINAHL headings, such as telemedicine and telehealth, resulted in a screening process in which mobile or wireless components needed to be identified in order to pass as mHealth. This process led to the exclusion of articles in case there was an uncertainty regarding the telehealth approach and may have led to potential missed data. The use of MeSH and CINAHL headings is, however, an advantage in literature reviews that we wanted to use.

The majority of engagement results related to aspects of usage of mHealth solutions. Several of these results were difficult to analyze and synthesize due to difficulties in conveying the actual meaning of the results, such as what a certain number of log-ins could imply in relation to engagement. Without knowing if this quantity was as expected or not, a reliable interpretation is difficult to draw without risk of oversimplifying or overestimating the results. Therefore, we decided to only synthesize results in studies where the meaning was clear. This may have influenced the results and led to the exclusion of some findings, but it was a necessary approach to create trustworthiness in the analysis process.

The included quantitative studies did not yield aggregated data for use in a meta-analysis and thus data were synthesized in a more narrative way together with those from other studies, which was considered a suitable alternative. We used 3 databases for the search of studies. These were chosen because they were deemed to be the largest and most relevant databases for this review, although we retrieved few appropriate studies with the use of the Inspec database, which was somewhat surprising. It is possible that the choice of these 3 databases may have limited the number of included studies.

Conclusion

Our integrative literature review revealed that engagement is a progressive process that comprises cognitive, behavioral and emotional, interactional, and mHealth usage–related aspects. To create and maintain sustainable engagement processes in mHealth research, enablers and barriers in all the engagement aspects might need to be acknowledged. mHealth solutions have the potential to support engagement in persons with chronic somatic health conditions when perceived as meaningful, suitable, and useful, but the research on engagement in relation

to the implementation and usage of such solutions is mixed. Further research is needed concerning other chronic somatic health problems besides diabetes and heart disease. In addition, research focusing on finding new and innovative incentive mechanisms to support engagement when using mHealth solutions for persons with chronic health conditions could improve mHealth solutions to better support engagement. Sometimes, disengagement in terms of nonusage attrition could be a positive aspect, and it might be important to consider when and where to stake efforts when striving for long-term engagement.

Acknowledgments

This literature review was initiated within the framework of a Swedish case study which was part of the Indo-Swedish research and development project entitled “Health in Hand –Transforming Health care Delivery” (VINNOVA Reg. No. 2013-046609) in collaboration between Blekinge Institute of Technology in Sweden and the Indian Institute of Technology Madras in India. The project, which started in January 2014 and ended in June 2017, was cofunded by the Swedish Governmental Agency for Innovation Systems (VINNOVA) and the Department of Biotechnology, Government of India, and had a multidisciplinary approach, with a focus on health and health promotion on one hand, and design and development of technology that facilitates the delivery of health promotion and health care on the other. We gratefully acknowledge the grant which made this study possible. The authors also thank the Department of Health and Caring Sciences at Linnaeus University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search protocol.
[DOCX File , 40 KB - mhealth_v8i7e14315_app1.docx]

References

1. Chan M. Ten Years in Public Health, 2007-2017. Geneva: World Health Organization; 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/255355/9789241512442-eng.pdf?sequence=1 [accessed 2020-07-01]
2. World Health Organization. Global Health Observatory (GHO) Data. Geneva: World Health Organization URL: https://www.who.int/gho/en/ [accessed 2020-07-01]
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
4. Barello S, Graffigna G. Engaging patients to recover life projectuality: an Italian cross-disease framework. Qual Life Res 2015 May;24(5):1087-1096. [doi: 10.1007/s11136-014-0846-x] [Medline: 25373927]
5. American Health Information Management Association (AHIMA). Consumer Engagement Toolkit. Chicago, IL: AHIMA; 2016.
6. Kadlec L, Kirby A, Shaffer H, Warren P. HIM Best Practices for Engaging Consumers in Their Overall Healthcare. J AHIMA 2015 Sep;86(9):52-56. [Medline: 26647597]
7. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
8. World Health Organization. Patient Engagement: WHO Technical Series on Safer Primary Care. URL: http://apps.who.int/iris/bitstream/handle/10665/252269/9789241511629-eng.pdf;jsessionid=9AA0FC083001B226174D9C2437D10FEB?sequence=12016 [accessed 2020-07-01]
9. 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]
10. Martin T. Assessing mHealth: opportunities and barriers to patient engagement. J Health Care Poor Underserved 2012 Aug;23(3):935-941. [doi: 10.1353/hpu.2012.0087] [Medline: 24212144]
11. World Health Organization. Global Diffusion of Ehealth - Making Universal Health Coverage Achievable: Report of the Third Global Survey On Ehealth (global Observatory For Ehealth Series). Geneva: World Health Organization; 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1 [accessed 2020-07-01]
12. Ali EE, Chew L, Yap KY. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan 05;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
13. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]

14. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
15. Critical Appraisal Skills Programme (CASP). CASP Checklist. 2018. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2020-07-01]
16. Athilingam P, Osorio RE, Kaplan H, Oliver D, O'neachtain T, Rogal PJ. Embedding Patient Education in Mobile Platform for Patients With Heart Failure: Theory-Based Development and Beta Testing. Comput Inform Nurs 2016 Feb;34(2):92-98. [doi: 10.1097/CIN.0000000000000216] [Medline: 26765655]
17. Brennan PF, Casper G, Kossman S, Burke L. HeartCareII: home care support for patients with chronic cardiac disease. Stud Health Technol Inform 2007;129(Pt 2):988-992. [Medline: 17911863]
18. Fuji KT, Abbott AA, Galt KA. A qualitative study of how patients with type 2 diabetes use an electronic stand-alone personal health record. Telemed J E Health 2015 Apr;21(4):296-300. [doi: 10.1089/tmj.2014.0084] [Medline: 25614996]
19. Hilliard ME, 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;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
20. Park S, Burford S, Nolan C, Hanlen L. The Role of Digital Engagement in the Self-Management of Type 2 Diabetes. Health Commun 2016 Dec;31(12):1557-1565. [doi: 10.1080/10410236.2015.1089468] [Medline: 27124817]
21. Whitten P, Bergman A, Meese MA, Bridwell K, Jule K. St. Vincent's Home telehealth for congestive heart failure patients. Telemed J E Health 2009 Mar;15(2):148-153. [doi: 10.1089/tmj.2008.0087] [Medline: 19292623]
22. Nelson LA, Mayberry LS, Wallston K, Kripalani S, Bergner EM, Osborn CY. Development and Usability of REACH: A Tailored Theory-Based Text Messaging Intervention for Disadvantaged Adults With Type 2 Diabetes. JMIR Hum Factors 2016 Sep 08;3(2):e23 [FREE Full text] [doi: 10.2196/humanfactors.6029] [Medline: 27609738]
23. Tatara N, Arsand E, Skrøvseth SO, Hartvigsen G. Long-term engagement with a mobile self-management system for people with type 2 diabetes. JMIR Mhealth Uhealth 2013;1(1):e1 [FREE Full text] [doi: 10.2196/mhealth.2432] [Medline: 25100649]
24. Capozza K, Woolsey S, Georgsson M, Black J, Bello N, Lence C, et al. Going mobile with diabetes support: a randomized study of a text message-based personalized behavioral intervention for type 2 diabetes self-care. Diabetes Spectr 2015 May;28(2):83-91. [doi: 10.2337/diaspect.28.2.83] [Medline: 25987806]
25. Fioravanti A, Fico G, Salvi D, García-Betances RI, Arredondo MT. Automatic messaging for improving patients engagement in diabetes management: an exploratory study. Med Biol Eng Comput 2015 Dec;53(12):1285-1294. [doi: 10.1007/s11517-014-1237-8] [Medline: 25564181]
26. Gorst SL, Coates E, Armitage CJ. "It's sort of a lifeline": Chronic obstructive pulmonary disease patients' experiences of home telehealth. Health Psychol 2016 Jan;35(1):60-68. [doi: 10.1037/hea0000246] [Medline: 26148189]
27. Huniche L, Dinesen B, Nielsen C, Grann O, Toft E. Patients' use of self-monitored readings for managing everyday life with COPD: a qualitative study. Telemed J E Health 2013 May;19(5):396-402. [doi: 10.1089/tmj.2012.0135] [Medline: 23531094]
28. Pekmezaris R, Schwartz RM, Taylor TN, DiMarzio P, Nouryan CN, Murray L, et al. A qualitative analysis to optimize a telemonitoring intervention for heart failure patients from disparity communities. BMC Med Inform Decis Mak 2016 Dec 24;16:75 [FREE Full text] [doi: 10.1186/s12911-016-0300-9] [Medline: 27343060]
29. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: A qualitative inquiry. J Telemed Telecare 2015 Jul 21:172-178. [doi: 10.1177/1357633X15595178] [Medline: 26199275]
30. Shane-McWhorter L, Lenert L, Petersen M, Woolsey S, McAdam-Marx C, Coursey JM, et al. The Utah Remote Monitoring Project: improving health care one patient at a time. Diabetes Technol Ther 2014 Oct;16(10):653-660 [FREE Full text] [doi: 10.1089/dia.2014.0045] [Medline: 24991923]
31. Tatara N, Årsand E, Bratteteig T, Hartvigsen G. Usage and perceptions of a mobile self-management application for people with type 2 diabetes: qualitative study of a five-month trial. Stud Health Technol Inform 2013;192:127-131. [Medline: 23920529]
32. Cottrell E, Cox T, O'Connell P, Chambers R. Implementation of simple telehealth to manage hypertension in general practice: a service evaluation. BMC Fam Pract 2015 Jul 17;16:83 [FREE Full text] [doi: 10.1186/s12875-015-0301-2] [Medline: 26183439]
33. Fischer HH, Moore SL, Ginosar D, Davidson AJ, Rice-Peterson CM, Durfee MJ, et al. Care by cell phone: text messaging for chronic disease management. Am J Manag Care 2012 Feb;18(2):e42-e47 [FREE Full text] [Medline: 22435883]
34. Koopman RJ, Wakefield BJ, Johanning JL, Keplinger LE, Kruse RL, Bomar M, et al. Implementing home blood glucose and blood pressure telemonitoring in primary care practices for patients with diabetes: lessons learned. Telemed J E Health 2014 Mar;20(3):253-260 [FREE Full text] [doi: 10.1089/tmj.2013.0188] [Medline: 24350806]
35. Mamykina L, Heitkemper EM, Smaldone AM, Kukafka R, Cole-Lewis H, Davidson PG, et al. Structured scaffolding for reflection and problem solving in diabetes self-management: qualitative study of mobile diabetes detective. J Am Med Inform Assoc 2016 Jan;23(1):129-136 [FREE Full text] [doi: 10.1093/jamia/ocv169] [Medline: 26769910]

36. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
37. Trief PM, Teresi JA, Eimicke JP, Shea S, Weinstock RS. Improvement in diabetes self-efficacy and glycaemic control using telemedicine in a sample of older, ethnically diverse individuals who have diabetes: the IDEATel project. Age Ageing 2009 Mar;38(2):219-225 [FREE Full text] [doi: 10.1093/ageing/afn299] [Medline: 19171951]
38. 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;15(12):e266 [FREE Full text] [doi: 10.2196/jmir.2699] [Medline: 24323283]
39. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G, et al. The Web-Based Osteoarthritis Management Resource My Joint Pain Improves Quality of Care: A Quasi-Experimental Study. J Med Internet Res 2015 Jul 07;17(7):e167 [FREE Full text] [doi: 10.2196/jmir.4376] [Medline: 26154022]
40. Evangelista LS, Lee J, Moore AA, Motie M, Ghasemzadeh H, Sarrafzadeh M, et al. Examining the effects of remote monitoring systems on activation, self-care, and quality of life in older patients with chronic heart failure. J Cardiovasc Nurs 2015;30(1):51-57 [FREE Full text] [doi: 10.1097/JCN.0000000000000110] [Medline: 24365871]
41. Nundy S, Dick JJ, Chou C, Nocon RS, Chin MH, Peek ME. Mobile phone diabetes project led to improved glycemic control and net savings for Chicago plan participants. Health Aff (Millwood) 2014 Feb;33(2):265-272 [FREE Full text] [doi: 10.1377/hlthaff.2013.0589] [Medline: 24493770]
42. Piette JD, Marinec N, Gallegos-Cabriales EC, Gutierrez-Valverde JM, Rodriguez-Saldaña J, Mendoz-Alevares M, et al. Spanish-speaking patients' engagement in interactive voice response (IVR) support calls for chronic disease self-management: data from three countries. J Telemed Telecare 2013 Feb;19(2):89-94 [FREE Full text] [doi: 10.1177/1357633X13476234] [Medline: 23532005]
43. Duggan GB, Keogh E, Mountain GA, McCullagh P, Leake J, Eccleston C. Qualitative evaluation of the SMART2 self-management system for people in chronic pain. Disabil Rehabil Assist Technol 2015 Jan;10(1):53-60. [doi: 10.3109/17483107.2013.845696] [Medline: 24112276]
44. Ong SW, Jassal SV, Miller JA, Porter EC, Cafazzo JA, Seto E, et al. Integrating a Smartphone-Based Self-Management System into Usual Care of Advanced CKD. Clin J Am Soc Nephrol 2016 Jun 06;11(6):1054-1062 [FREE Full text] [doi: 10.2215/CJN.10681015] [Medline: 27173169]
45. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015;3(2):e33 [FREE Full text] [doi: 10.2196/mhealth.3789] [Medline: 25842282]
46. Nijland N, van Gemert-Pijnen JEWC, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
47. Gellis ZD, Kenaley B, McGinty J, Bardelli E, Davitt J, Ten HT. Outcomes of a telehealth intervention for homebound older adults with heart or chronic respiratory failure: a randomized controlled trial. Gerontologist 2012 Aug;52(4):541-552. [doi: 10.1093/geront/gnr134] [Medline: 22241810]
48. Hanley J, Fairbrother P, Krishan A, McCloughan L, Padfield P, Paterson M, et al. Mixed methods feasibility study for a trial of blood pressure telemonitoring for people who have had stroke/transient ischaemic attack (TIA). Trials 2015 Mar 25;16:117 [FREE Full text] [doi: 10.1186/s13063-015-0628-y] [Medline: 25873155]
49. Pfaeffli L, Maddison R, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mHealth cardiac rehabilitation exercise intervention: findings from content development studies. BMC Cardiovasc Disord 2012;12:36 [FREE Full text] [doi: 10.1186/1471-2261-12-36] [Medline: 22646848]
50. Piette JD, Rosland A, Marinec NS, Striplin D, Bernstein SJ, Silveira MJ. Engagement with automated patient monitoring and self-management support calls: experience with a thousand chronically ill patients. Med Care 2013 Mar;51(3):216-223 [FREE Full text] [doi: 10.1097/MLR.0b013e318277ebf8] [Medline: 23222527]
51. Koufopoulos JT, Conner MT, Gardner PH, Kellar I. A web-based and mobile health social support intervention to promote adherence to inhaled asthma medications: randomized controlled trial. J Med Internet Res 2016 Dec 13;18(6):e122 [FREE Full text] [doi: 10.2196/jmir.4963] [Medline: 27298211]
52. Aikens JE, Zivin K, Trivedi R, Piette JD. Diabetes self-management support using mHealth and enhanced informal caregiving. J Diabetes Complications 2014;28(2):171-176 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.11.008] [Medline: 24374137]
53. Piette JD, Marinec N, Janda K, Morgan E, Schantz K, Yujra ACA, et al. Structured Caregiver Feedback Enhances Engagement and Impact of Mobile Health Support: A Randomized Trial in a Lower-Middle-Income Country. Telemed J E Health 2016 Apr;22(4):261-268 [FREE Full text] [doi: 10.1089/tmj.2015.0099] [Medline: 26352854]
54. 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]
55. Bardram J, Bossen C, Thomsen A. Designing for transformations in collaboration: a study of the deployment of homecare technology. New York, NY: ACM; 2005 Presented at: GROUP05: ACM 2005 International Conference on Supporting Group Work; November 6-9, 2005; Sanibel Island, FL p. 6-9.

56. Andersson L, Andreasen JO. Important considerations for designing and reporting epidemiologic and clinical studies in dental traumatology. Dent Traumatol 2011 Aug;27(4):269-274. [doi: 10.1111/j.1600-9657.2011.00992.x] [Medline: 21477103]
57. Hu Y, Eriksén S, Lundberg J, Tuvesson H, Nilsson L, Fagerström C, et al. Different ways of engaging the end-users in mHealth services. 2018 Presented at: eTELEMED The Tenth International Conference on eHealth, Telemedicine, and Social Medicine; March 25-29, 2018; Rome, Italy p. 102-101.
58. Zrebiec JF. Internet communities: do they improve coping with diabetes? Diabetes Educ 2005;31(6):825-830. [doi: 10.1177/0145721705282162] [Medline: 16288090]
59. de Ridder M, Kim J, Jing Y, Khadra M, Nanan R. A systematic review on incentive-driven mobile health technology: As used in diabetes management. J Telemed Telecare 2016 Feb 16. [doi: 10.1177/1357633X15625539] [Medline: 26888421]
60. Liu S, Hodgson C, Zbib AM, Payne AYM, Nolan RP. The effectiveness of loyalty rewards to promote the use of an Internet-based heart health program. J Med Internet Res 2014 Jul 02;16(7):e163 [FREE Full text] [doi: 10.2196/jmir.3458] [Medline: 24989982]
61. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
62. Bannon L. From human factors to human actors: the role of psychology and human-computer interaction studies in systems design. In: Greenbaum J, Kyng M, editors. Design at Work: Cooperative Design of Computer Systems. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991:25-44.
63. Bannon L. Reimagining HCI. Interactions 2011 Jul 01;18(4):50-57. [doi: 10.1145/1978822.1978833]
64. Malmborg L, Grönvall E, Messeter J, Raben T, Werner K. Mobilizing senior citizens in co-design of mobile technology. IJMHCI 2016;8(4):42-67.
65. Harris JD, Quatman CE, Manring MM, Siston RA, Flanigan DC. How to write a systematic review. Am J Sports Med 2014 Nov;42(11):2761-2768. [doi: 10.1177/0363546513497567] [Medline: 23925575]
66. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
67. Caffery LJ, Martin-Khan M, Wade V. Mixed methods for telehealth research. J Telemed Telecare 2017 Oct;23(9):764-769. [doi: 10.1177/1357633X16665684] [Medline: 27591744]
68. Fàbregues S, Paré M. Appraising the quality of mixed methods research in nursing: A qualitative case study of nurse researchers' views. Nurs Inq 2018 Oct;25(4):e12247. [doi: 10.1111/nin.12247] [Medline: 29927008]

Abbreviations

AHIMA: American Health Information Management Association
CASP: Critical Appraisal Skills Programme
SMS: short message service

Edited by G Eysenbach; submitted 09.04.19; peer-reviewed by A Fioravanti, K Fuji; comments to author 14.12.19; revised version received 30.04.20; accepted 14.05.20; published 24.07.20.

Please cite as:
Tuvesson H, Eriksén S, Fagerström C
mHealth and Engagement Concerning Persons With Chronic Somatic Health Conditions: Integrative Literature Review
JMIR Mhealth Uhealth 2020;8(7):e14315
URL: http://mhealth.jmir.org/2020/7/e14315/
doi:10.2196/14315
PMID:32706686

Original Paper

Using Internet of Things to Reduce Office Workers' Sedentary Behavior: Intervention Development Applying the Behavior Change Wheel and Human-Centered Design Approach

Yitong Huang[1,2], PhD; Steve Benford[3], PhD; Dominic Price[3], BSc; Roma Patel[3,4], PhD; Benqian Li[1,2], PhD; Alex Ivanov[1], PhD; Holly Blake[5,6], PhD

[1]School of Media and Communication, Shanghai Jiaotong University, Shanghai, China
[2]Institute of Intelligent Communication, Shanghai Jiaotong University, Shanghai, China
[3]School of Computer Science, University of Nottingham, Nottingham, United Kingdom
[4]Makers of Imaginary Worlds Ltd, Nottingham, United Kingdom
[5]School of Health Sciences, University of Nottingham, Nottingham, United Kingdom
[6]Nottingham Biomedical Research Centre, National Institute for Health Research, Nottingham, United Kingdom

Corresponding Author:
Yitong Huang, PhD
School of Media and Communication
Shanghai Jiaotong University
800 Dongchuan Road
Shanghai,
China
Phone: 86 57782665881
Email: huang.yitong@foxmail.com

Abstract

Background: Sedentary behavior (SB) is associated with various adverse health outcomes. The prevalence of prolonged sitting at work among office workers makes a case for SB interventions to target this setting and population. Everyday mundane objects with embedded microelectronics and ubiquitous computing represent a novel mode of delivering health behavior change interventions enabled by internet of things (IoTs). However, little is known about how to develop interventions involving IoT technologies.

Objective: This paper reports the design and development of an IoT-enabled SB intervention targeting office workers.

Methods: The process was guided by the behavior change wheel (BCW), a systematic framework for theory-informed and evidence-based development of behavior change interventions, complemented by the human-centered design (HCD) approach. Intervention design was shaped by findings from a diary-probed interview study (n=20), a stakeholder design workshop (n=8), and a series of theoretical mapping and collaborative technical design activities.

Results: The resulting intervention named *WorkMyWay* targets a reduction in office workers' prolonged stationary behaviors at work and an increase in regular breaks by modifying behavioral determinants in 11 theoretical domains with 17 behavior change techniques. The delivery technology consists of a wearable activity tracker, a light-emitting diode reminder device attached to a vessel (ie, water bottle or cup), and a companion Android app connected to both devices over Bluetooth. The delivery plan consists of a 2-week baseline assessment, a 30-min face-to-face action planning session, and 6-week self-directed use of the delivery technology.

Conclusions: This is the first study to demonstrate that it is possible to develop a complex IoT-enabled intervention by applying a combination of the BCW and HCD approaches. The next step is to assess the feasibility of *WorkMyWay* prior to testing intervention efficacy in a full-scale trial. The intervention mapping table that links individual intervention components with hypothesized mechanisms of action can serve as the basis for testing and clarifying theory-based mechanisms of action in future studies on *WorkMyWay*.

(JMIR Mhealth Uhealth 2020;8(7):e17914) doi:10.2196/17914

KEYWORDS

sedentary behavior; workplace; just-in-time adaptive intervention; internet of things

Introduction

Background

Sedentary behavior (SB) is "any waking behavior characterized by an energy expenditure of less than 1.5 metabolic equivalents (METs) while in a sitting, reclining, or lying posture" [1]. SB has negative and independent impacts on cardiometabolic health [2-6], and the health risks increase with prolonged bouts of sitting (ie, over 30 or even 60 min) [6,7]. Office workers often sit for long periods at work [8-11], which makes a strong case for SB interventions to target this setting and population [8].

The past two decades have seen an increasing number of digital technologies with various form factors (eg, personal computers, tablets, smartphones, wearables, service robots, and internet of things [IoT]) entering people's everyday lives; they are increasingly utilized to deliver digital behavior change interventions (DBCIs) for health [9,10]. However, DBCIs in the medical and health sciences literature mostly use screen-based multimedia. Attempts to explore emerging digital interfaces beyond screens for health have been sporadic yet encouraging. These include IoT-enabled persuasive designs that overlay or embed digital information on or in the physical environment to influence behaviors at both conscious [11-14] and unconscious [15-19] levels.

Our previous work [20] has systematically scoped DBCIs aimed at reducing office workers' SB and found that technological designs combining passive data collection with automated tailored feedback and scheduled prompts and involving several connected devices are promising for delivering just-in-time adaptive interventions. Positive user experiences are reported in studies where feedback and prompts are delivered with aesthetic and ambient media interfaces (eg, decorative objects and ambient light) that are seamlessly integrated with the physical environment and actuated over wireless connections based on real-time behavioral sensing [13,14]. However, very few of those novel modes of delivery emerging from the engineering and design fields have moved forward to the "evaluation phase" under the Medical Research Council's framework for developing and evaluating complex interventions [21].

A potential reason could be that the development of these IoT-based interventions rarely follows systematic theory-driven intervention design approaches, such as the intervention mapping approach [22] and the behavior change wheel (BCW) [23]. This makes it difficult to develop a theoretical understanding of the mechanisms of action underlying interventions, which, in turn, prevents novel interventions and emerging research from being utilized by the wider community of behavior scientists and health intervention researchers [21].

Another possible reason concerns the complexity of IoT systems and the additional challenges they pose for intervention designers compared with developing more traditional DBCIs that are web-based or app-based. For instance, in IoT development, requirements will need to be specified for not only software, but also hardware (ie, electronics) and industrial design (ie, the casing of electronics and integration with everyday objects). Moreover, as an IoT system usually involves multiple connected devices and interfaces that work in tandem, a key challenge lies in deciding what behavior change contents and functionalities should be delivered by each of the devices as part of an integrative intervention for optimal user experience, social validity, and behavior change outcomes in workplace settings. Design decisions as such are rarely documented but are highly valuable to inform and encourage future developments of high-quality interventions using similar technologies.

In view of the above, we develop and report an IoT-enabled SB intervention named *WorkMyWay*, which is systematically grounded in theories using the BCW and balanced with stakeholder requirements for acceptability in workplace contexts. In what follows, we first introduce the methodological frameworks that have guided or inspired our design process. We then detail a five-stage development process including the methods and outcomes of each stage that have fed into the final intervention design, followed by a description of the final intervention using the Template for Intervention Description and Replication (TIDieR). In the end, we reflect on both the intervention we developed and the development process to draw implications for future research.

Systematic Application of Theories to Intervention Development

More extensive use of theory in behavior change interventions is found to be associated with an increased effect size [9]. However, there are 83 theories of behavior change [24], many of which have overlapping constructs [25]. Therefore, the real challenge lies in selecting the most relevant theories in a systematic manner [26]. The BCW [23], developed by Michie and colleagues to support such processes, was therefore adopted to guide this research.

At the center of the BCW is the "COM-B" model of behavior, which breaks down behavioral determinants into three dimensions, with two subcomponents in each dimension, namely *capability* (psychological and physical), *opportunity* (physical and social), and *motivation* (automatic and reflective). It should be noted that the COM-B model is not a theory per se, but an abstraction of many theories about human cognition and behaviors, including the dual process model [27,28], modern habit theories [29], implementation intention [30], etc. The COM-B model can be expanded with the theoretical domain framework (TDF), which extracts 128 key theoretical constructs from 33 behavior change theories and organizes them into 14 theoretical domains [31]. The compatibility with TDF adds to the theoretical rigor of the BCW, as the TDF has been widely used and validated in studies identifying determinants of behavior change [32,33]. Using matrices from the BCW guide, intervention designers can translate the COM-B and TDF diagnosis into intervention options specified in terms of behavior

change techniques (BCTs), which are considered the irreducible "active ingredients" within any behavior change intervention [34].

Complementing Theories With a Human-Centered Design Approach

Intervention design is about not only theoretical soundness, but also appropriateness and relevance to the local context [23]. Hence, it is important to involve users (eg, office workers in the context of workplace health interventions) and other stakeholders (eg, managers of office workers) early in the design process and shape the design to their needs [35]. Moreover, the development, deployment, and upgrade of an IoT-based system are presumably more complex and therefore more costly than software systems, so it would be undoubtedly risky to exhaust a project's resources to implement an IoT delivery system without a thorough understanding of the context of use and stakeholders' preferences beforehand. Nonetheless, it is particularly hard for IoT novices to envision possible interactions with an imaginary IoT system beyond screen-based graphic user interfaces. This means conventional formative research approaches, such as interviewing and surveying stakeholders on "what features they want," are insufficient to elicit useful information for a potential design.

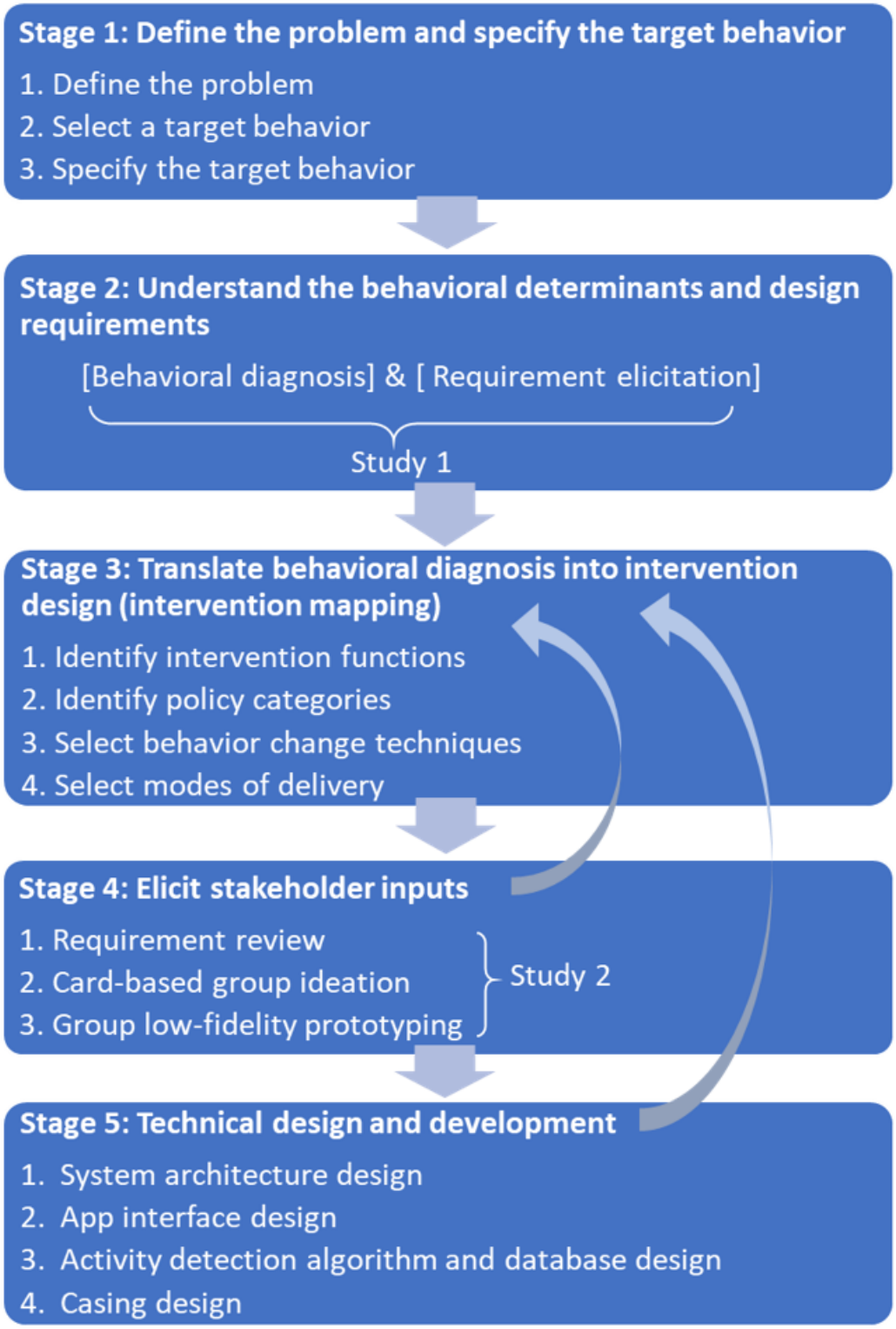
In view of this, we consulted the human centered design (HCD) methodology that originated from human-computer interaction, a field of research within the discipline of computer science concerned with designing and studying human interactions with computing systems by drawing on methods and theories from a range of other disciplines including psychology, sociology, and design. The HCD can be a valuable complement to the BCW-guided process, with its rich repertoire of "quick-and-dirty" design methods and tools for getting stakeholder inputs before developing a fully functional product [36].

Methods

Overview of the Design and Development Process

We report how we brought together these two approaches in *WorkMyWay* development, which included five stages (Figure 1) that were iterative, meaning outputs from each stage could be used to refine the outcome(s) of the previous stage(s) whenever needed. Both empirical studies (stage 2 and stage 4) were approved by the ethics committee at the School of Computer Science, University of Nottingham (January 14, 2016, and July 06, 2016).

Figure 1. WorkMyWay intervention and technology development process.



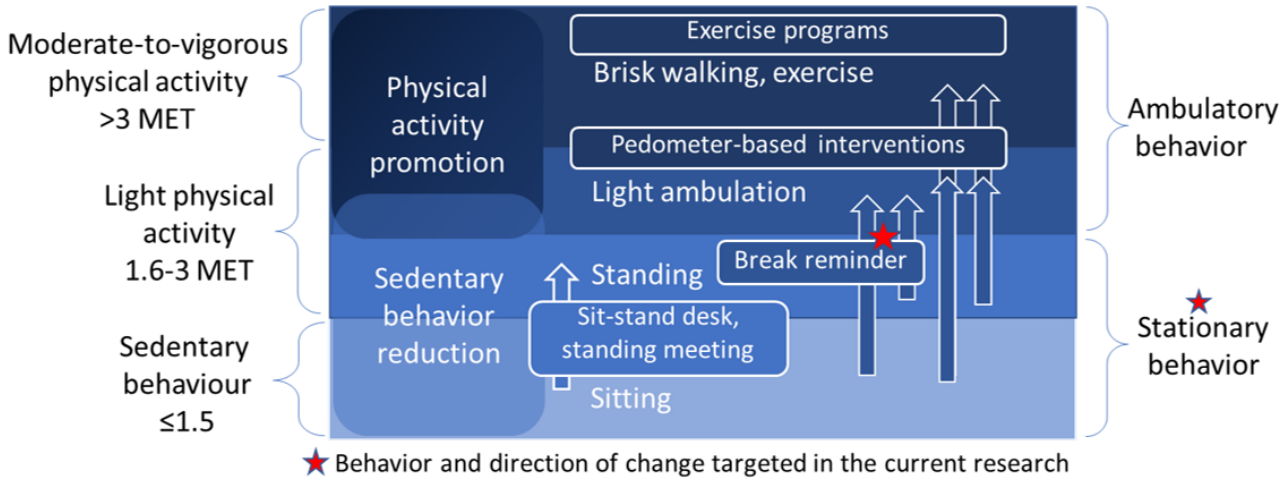
Stage 1: Define the Problem and Specify the Target Behavior

Stage 1 involved three BCW-guided analytic and decision steps. Step 1 entailed defining the problem in behavioral terms and listing all behaviors that might influence the behavior of our interest. Step 2 involved selecting a target behavior from the list of candidate behaviors based on the following four criteria: (1) the likely impact of the behavior change; (2) ease of change; (3) possible spillover effects on other related behaviors; and (4) ease of measurement. Step 3 required specification of the target behavior in terms of "what, who, where, when, how often, and with whom."

As shown in Figure 2, many published interventions have targeted physical activity (PA), although SB, on the lower end of the activity continuum, is increasingly targeted as a separate behavior in interventional studies. Targeting SB can be more effective in reducing sitting [37,38], but can have the undesirable

spillover effect of increasing prolonged standing and limited promise for a positive spillover effect on ambulatory time [39-41]. Therefore, a decision was made to target prolonged stationary behavior (a combination of sitting and standing [1]) by encouraging regular ambulatory breaks, a behavior with proven cardiometabolic benefits [42]. There is yet no consensus regarding the optimal interval and duration of breaks in sedentary work. Expert advice and interventions variably promoted targets that ranged from "a 2- to 3-min light physical activity (LPA) break every 30 min of sitting" [43] to "a 5-min LPA break every 60 min of sitting" [44]. Hence, we decided to support customizable target setting with our intervention.

Figure 2. The intervention continuum from sedentary behavior reduction to physical activity promotion.



Stage 2: Understand the Behavioral Determinants and User Requirements (Study 1)

As per the BCW, the next stage would entail a "behavioral diagnosis" using the COM-B and TDF to identify a full range of factors and processes requiring modification to achieve the desired behavior change. From an HCD perspective, we considered it equally important to understand the context of use and the needs of users prior to design, a step known as "requirement elicitation." Semistructured interviews and diary keeping are both common requirement elicitation methods, with the latter particularly useful for gaining a picture of how a future system can support the user in the context of everyday operations [36]. Hence, we integrated them with behavioral diagnostic interviews commonly used in intervention design.

Methods

We recruited 20 participants (12 female participants) via a mailing list and posters from the University of Nottingham and two local nongovernment organizations. The participants were self-identified sedentary office workers employed in diverse office roles including admin, communication, project management, filmmaking, research, and education.

We asked each participant to keep a diary for 2 workdays and interviewed them afterwards with diaries as probes. Interviews were audio recorded with consent. The diary template (Multimedia Appendix 1) allowed participants to record details of each break experience, including the trigger, timing, location, social context, activities, and experience of the break; working tasks prior to the break; and preferences on receiving a break reminder in that moment. Participants were told that a break was defined as "any interruption in sitting" in this study. To gather design inspirations from participants' perspectives for objects where electronics could be potentially embedded to make smart objects, one of the dairy items asked the participant to imagine potential interactions with an anthropomorphic everyday object by completing the sentence "*when… (context)…, I would like my …(object)….to say to me...(content of message)…*" Participants could also take and email photos to the researcher to illustrate the physical contexts of their breaks and the objects involved. Interview questions were constructed partly according to the COM-B and TDF to diagnose behavioral barriers and facilitators and partly around suggestions for a potential IoT delivery system (the full interview questioning route is presented in Multimedia Appendix 1).

All interviews were transcribed verbatim. Framework analysis [45] was applied to interview data for behavioral diagnosis under the COM-B and TDF, and thematic analysis [46] was applied for identifying design requirements. Content analysis was applied to diary data to identify common routines, physical environments, and objects that had triggered breaks and that could be potentially redesigned to prompt more breaks.

Results

Interviews ranged from 42 to 66 min. The behavioral diagnostic result from the study has been reported elsewhere [47] and hence will only be briefly summarized here. The following five COM-B components (with theoretical constructs specified in brackets) were identified as important determinants of office workers' SB: psychological capability (eg, knowledge, cognitive resources, and skills required for monitoring and regulating break patterns), automatic motivation (eg, prolonged sitting habit and effects or emotions associated with breaks), reflective motivation (eg, beliefs about the consequences of taking breaks, perceived behavioral control, the priority and accessibility of health-related goals at work, and the intention to break up sitting regularly), physical opportunity (eg, job demands, time pressure, and organizational climate), and social opportunity (eg, social norm of prolonged sitting versus regular breaks and direct social interactions that prompt or hinder breaks). This step laid the

foundation for identifying and clarifying the theoretical underpinnings for the resulting intervention.

As for the context of use, based on 291 break-related diary entries, the most common reasons that prompted participants to stand up were work-related (eg, walk between meetings and printing) (n=84), followed by the need to refill cups or water bottles (n=63), go to the toilet (n=53), do chores (eg, wash dishes after lunch and deliver envelopes) (n=48), and eat or snack (n=25). Accordingly, vessels like cups and water bottles appeared to be the objects most frequently seen in break activities.

Thematic analysis of interview data and object messages suggested by participants in diary entries further revealed user requirements for the proposed intervention delivery system, which are summarized in Textbox 1.

Textbox 1. User requirements elicited from study 1.

Reminders

- To be triggered when prolonged sitting is detected
- Timer should be automatically reset after I take a break
- Can be manually disabled in certain situations
- Should allow personalized settings
- Should not lock up the screen or enforce breaks when I want to work (ie, should let me retain control and autonomy)

Feedback

- Should provide visual feedback on my break pattern and support historical comparison and personalized goal setting
- Could provide social comparison with others (though the motivational value could vary across individuals and may encourage some workers to sit for even longer)

Manner of communication

- Perseverant yet flexible: the system should allow me to “snooze” it several times
- Factual and informational: the system should show me the sitting time
- Readily accessible but nonintrusive: feedback should always be there, but I can choose when to view it
- Credible and authoritative (so that it helps me justify my breaks)
- Difference between participants on preferred tone of voice: gentle and soft (eg, “maybe you wanna take a break”) versus forceful, telling off, and guilt-inducing (eg, “get up, lazy”)
- Difference between participants on preferred characteristics of the object: functional, utilitarian, and nongimmicky (eg, “just a beep or light would suffice”) versus caring, cute, or other anthropomorphic characters

Modalities of communication

- Tactile and visual prompts would be widely acceptable
- Audible prompts would be most noticeable but unacceptable in shared offices

Stage 3: Translate Research Insights Into Intervention Design

The next stage entailed translating research findings into intervention design by making four major decisions on the following: (1) the broad categories of means by which an intervention changed behaviors (ie, intervention functions), (2) the policy categories used by authorities to support and enact the interventions, (3) the BCTs that best served the interventions functions, and (4) the mode(s) of delivery for implementing the intervention contents.

The BCW [23] detailed links between the TDF domains and the potentially effective intervention functions, policy categories, and BCTs. However, as the previous step identified 11 out of 14 TDF domains as relevant to the office workers’ SB, we had a whole range of nine intervention functions as potentially effective options. We then assessed the appropriateness of each option by applying the APEASE (affordability, practicability, (cost-) effectiveness, acceptability, side-effects or safety, and equity) criteria [23]. From there, we excluded “coercion” (creating an expectation of punishment or cost), as it conflicted with the user requirement for agency, autonomy, and control over work break rhythms. We also excluded “restriction” (using rules to alter the opportunity to engage in the target or competing behavior), as it was impractical and unacceptable to restrict office workers’ access to seated workstations or prevent them from going to long-seated meetings.

We discussed the seven policy categories (ie, service provision, guidelines, fiscal measures, communication or marketing, environmental or social planning, regulation, and legislation) with human resource managers and staff wellbeing leads in several organizations. It was concluded that only service provision was relevant at the point of intervention design before firm evidence of effectiveness was established.

The process of selecting BCTs is detailed in Multimedia Appendix 2, with examples of mapping shown in Table 1. In a nutshell, we first adapted the matrix from the BCW guide [23] to map the relationships between intervention functions and the TDF-based behavioral diagnosis; we then identified potentially effective BCTs for each cell in the matrix based on the BCW guide [23]. This resulted in a range of BCTs selected for each cell with illustrative intervention contents. We then removed some BCTs (eg, restructuring the social or physical environment) that did not meet the APEASE criteria.

We set out to utilize novel modes of delivery, particularly IoT-enabled smart objects and wearables, as much as possible. However, considering the upfront development cost (the "affordability" criteria in APEASE) for digitizing the delivery of all BCTs and the high likelihood of making changes to the intervention design after the feasibility and piloting phase ("cost-effectiveness" criteria in APEASE), we decided to build a minimum viable IoT product with all the essential technological functions and complement it with face-to-face sessions and email communication for BCTs that required complex dialogue support and individualization (eg, goal setting and action planning). Examples of intervention mapping following the BCW are presented in Table 1 (see Multimedia Appendix 2 for full details).

Table 1. Examples of intervention mapping following the behavior change wheel.

Constructs/mechanisms of action targeted	Intervention function	BCTs[a]	Intervention components and *modes of delivery*
Memory, cognitive overload, and behavioral regulation	Enablement	Conserve mental resources, feedback on behaviors, and self-monitoring	Use *wearable* trackers to automatically monitor sitting time, and an *app* provides daily feedback to enable user to self-monitor day-to-day changes in break patterns.
Belief about capabilities	Education	Feedback on behaviors	An *app* presents daily summary of and feedback on the sit-break pattern.
Prospective memory, cognitive overload, and goal priming	Environmental restructuring	Conserve mental resources, prompts and cues, and add objects to the environment	Add or augment *objects* that facilitate the performance of breaks; use the *object* to cue breaks naturally associated with the object (eg, augment a cup to cue tea breaks).
Breaking habit, self-efficacy, and implementation intention (goal accessibility)	Enablement	Action planning	*Researcher* guides the person to set up plans to combat prolonged sitting by specifying the frequency and duration of breaks, including developing "if-then" rules that use an *IoT object* as the cue.
Habits and contingencies	Training	Habit formation	*Researcher* guides the person to develop automatic responses to the introduced *stimuli (the IoT object)* through repetition.

[a]BCT: behavior change technique.

Stage 4: Elicit Stakeholder Inputs (Study 2)

In this stage, we invited stakeholders (eg, managers and occupational health consultants working for organizations) to review both theory-based and user-centered design requirements generated from the previous stages in order to assess potential acceptability of the proposed design in various organizational contexts from managerial or experts' perspectives. We decided to conduct a design workshop, which is a common HCD method for bringing together a cross-section of stakeholders to not only identify issues that need to be addressed (similar to focus groups), but also produce design solutions collaboratively [36]. Card-based ideations and low-fidelity (lo-fi) prototyping activities are frequently employed in design workshops. The former usually relies on ideation decks created for specific design briefs with contents illustrating parameters directly relevant to the design problems of interest to prompt group creativity and discussions [48]. Lo-fi prototyping refers to creating noninteractive mock-ups of potential design solutions with little or no programming or engineering effort, which are useful to represent system requirements and to allow the design team and stakeholders to evaluate candidate solutions and identify problems with them before investments in technical development [36].

Methods

Our workshop was funded as an industry engagement activity by the Balance Network [49] and promoted by workplace health specialist networks and word of mouth. Many work health specialists expressed interest to learn about outcomes from the workshop, but only eight could make it on the workday when it was hosted. The eight delegates (two female and six male) represented large organizations, small-to-medium–size enterprises, and public and private sectors, and all had an interest in enhancing employee health and wellbeing.

Materials for the workshop are provided in Multimedia Appendix 3. The half-day workshop started with requirement

review, where each participant completed an individual worksheet that asked them to rate the COM-B and TDF behavioral determinants elicited from Study 1 in terms of *"to what extent does this statement reflects what you have observed in your workplace?"* and *"how important do you think this factor is in determining micro-break behaviors?"* We then presented the participants with a diagram illustrating the proposed intervention delivery system with the following three components: a wearable activity tracker, a smartphone app, and an IoT cup or water bottle (without delineating interactions or user interfaces in detail). We brought some commercially available IoT cups (eg, Cuptime, Moikit Ltd) and passed them around to give participants a more concrete idea of how embedded sensing, data processing, wireless connectivity, and different digital interfaces could fit together and what features such products were capable of delivering. Participants were encouraged to challenge the proposed system design and to raise potential deployment issues in workplace settings. We also encouraged participants to suggest alternative objects as the medium for delivery.

Thereafter, participants split into two groups to ideate system features by completing group worksheets. The process was supported by a deck of 25 persuasive IoT ideation cards that we specially designed for this project, which consisted of three categories of "opportunity" cards, namely physical, social, and sensing opportunities. The contents were created based on inspiration from previous IoT decks [50], persuasive design frameworks [51], and the BCT taxonomy [34].

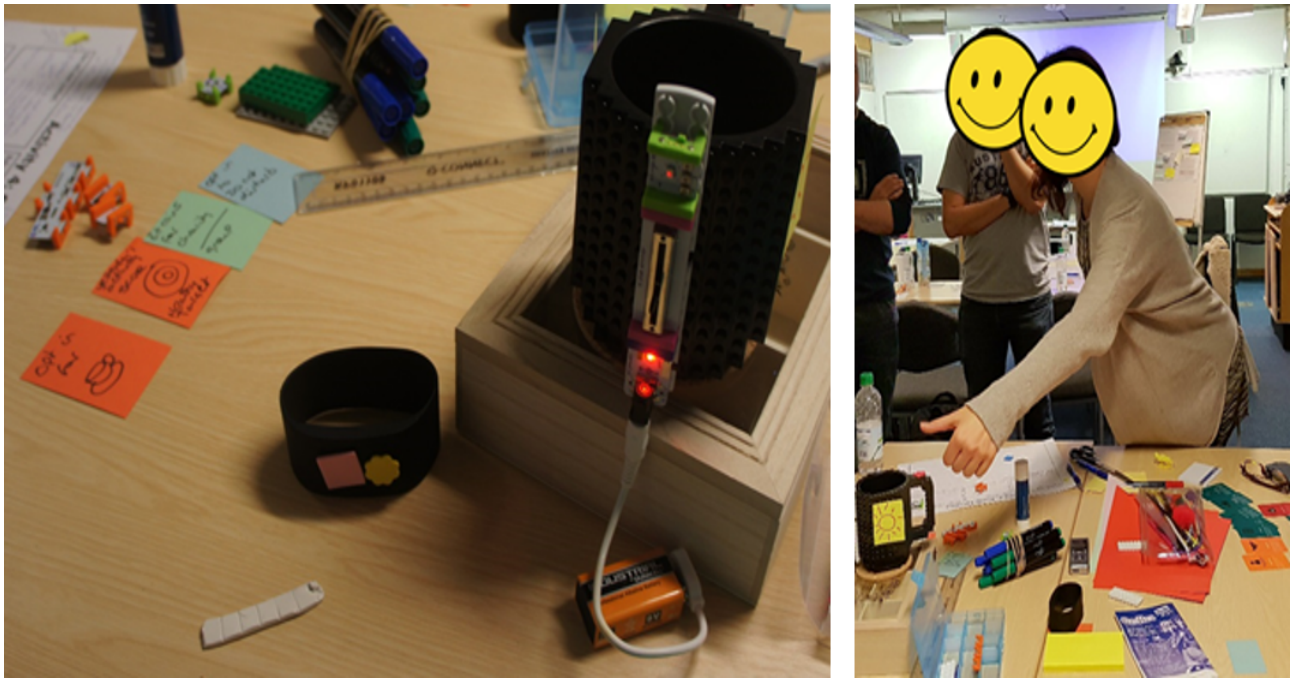
After the ideation, participants were introduced to the concept of lo-fi prototyping. In addition to common lo-fi prototyping materials like paper and Play-Doh, we supplied our participants with *LittleBits* (LittleBits Electronics Inc), an educational kit with modular electronics that could snap together with small magnets to make circuits attachable to Lego cups. This innovative combination of tools and materials was intended to help participants envision and evaluate various modalities (eg, light, vibration, and audio) through which they could be reminded by a potential smart cup in the workplace. The prototyping activity was concluded with a showcase session, where each group reported back on their design ideas and prototypes, with the other group asking questions and suggesting improvements to the design.

Results

The requirement review did not particularly challenge behavioral insights from Study 1; participants were receptive to the proposed intervention and technology, and raised questions mostly concerning technical feasibility (eg, *"can the system detect a break if the user doesn't not take the cup with him/her during the break?"*).

Group 1's design (Figure 3) validated the requirements for personalization and user autonomy elicited from Study 1 and revealed new requirements concerning specifics about the interaction flow and feedback interfaces. They prototyped a cup with an ambient display that could be personalized to individual users. They suggested the display, for instance, could gradually reveal a picture of the user's dog to suggest that the user was increasingly in need of a break (represented by the light-emitting diode [LED] dimmer switch in the prototype in Figure 3). The proposed design also included two buttons for the cup; one to allow the user to temporarily disable all reminders during meetings and the other to "snooze" reminders (the user could postpone a visual reminder for breaks for up to three times, after which a vibratory reminder would be triggered on one's wristband). Group 1 also created paper prototypes for a companion app, which illustrated features like automated tailored feedback and rewards for regular breaks. Finally, Group 1 preferred that all data be kept private online, although users could choose to share their app screens with friends and co-workers offline to foster competition. Group 2's design revealed more nuances in the balance between harnessing social influences and respecting individual privacy. The design was targeted at open plan offices with the culture of co-workers taking turns to bring back drinks from the kitchen for others so that most people could work for longer without interruptions. Group 2 decided to harness this existing culture by identifying and prompting the most sedentary person in the office to stand up for a break. In this way, each user would be motivated to break up sitting more proactively to prevent one's own cup from buzzing and embarrassing oneself in the office. In addition, Group 2 suggested fostering social cooperation and team competition by deploying, in the staff common room, a dashboard displaying each office's aggregate activity data. Nonetheless, Group 2 emphasized that the dashboard should not give away any behavioral data identifiable to individuals.

Figure 3. Design idea generated by Group 1.



Stage 5: Technical Design and Development

A range of tools, including diagrams, wireframes, mockups, and pseudocodes, were used to document the requirements throughout the previous stages. These were then compiled into a requirement specification document (Multimedia Appendix 4), which was frequently referred to and refined over the process of technical implementation.

It was decided at this stage that BCTs related to social influences (eg, social support, social comparison, and demonstration of the behavior by managers and workplace champions) would not be technically implemented for the following three reasons: (1) both Study 1 and Study 2 suggested that the technology would likely trigger conversations and competitions between office mates offline, even without explicit instructions for social comparisons in the app; (2) the integration of social functions would complicate the architecture design and increase development time; (3) both studies suggested potential ethical controversies associated with sharing data about an individual's break patterns in terms of surveillance on employees' work behavior and that it might impel some workers to sit for longer in an attempt to impress others.

Development Platform, System Architecture, and Application Programming Interfaces

We researched ways to digitally augment an everyday vessel and enable it to track the physical footprints of itself and its owner, and to deliver meaningful just-in-time adaptive interventions. After an audit and comparison of different IoT development platforms, we selected the *MetaWear* RG (MbientLab Inc) platform, which includes an accelerometer, a color LED, and a Bluetooth Low-Energy (BLE) module on-board, as well as an Android software development kit (SDK). The SDK was important as it allowed us to focus on software development (eg, streaming accelerometer data captured by the wearable device over a BLE connection in real-time) without much investment in hardware engineering. A diagram that illustrates the system architecture and application programming interfaces used is included in Multimedia Appendix 4.

Interface Design

Based on Android guides for user interface and navigation design [52], we created wireframes to illustrate the information architecture (eg, layout and navigation) and interaction flow for the app (Multimedia Appendix 4). Tabs were chosen for lateral navigation between three sibling sections, namely "track," "history," and "rewards," which were expected to be used most frequently. The infrequently used and discrete options ("about," "user setting," and "researcher setting") were accessible from a drop-down menu at the top right corner.

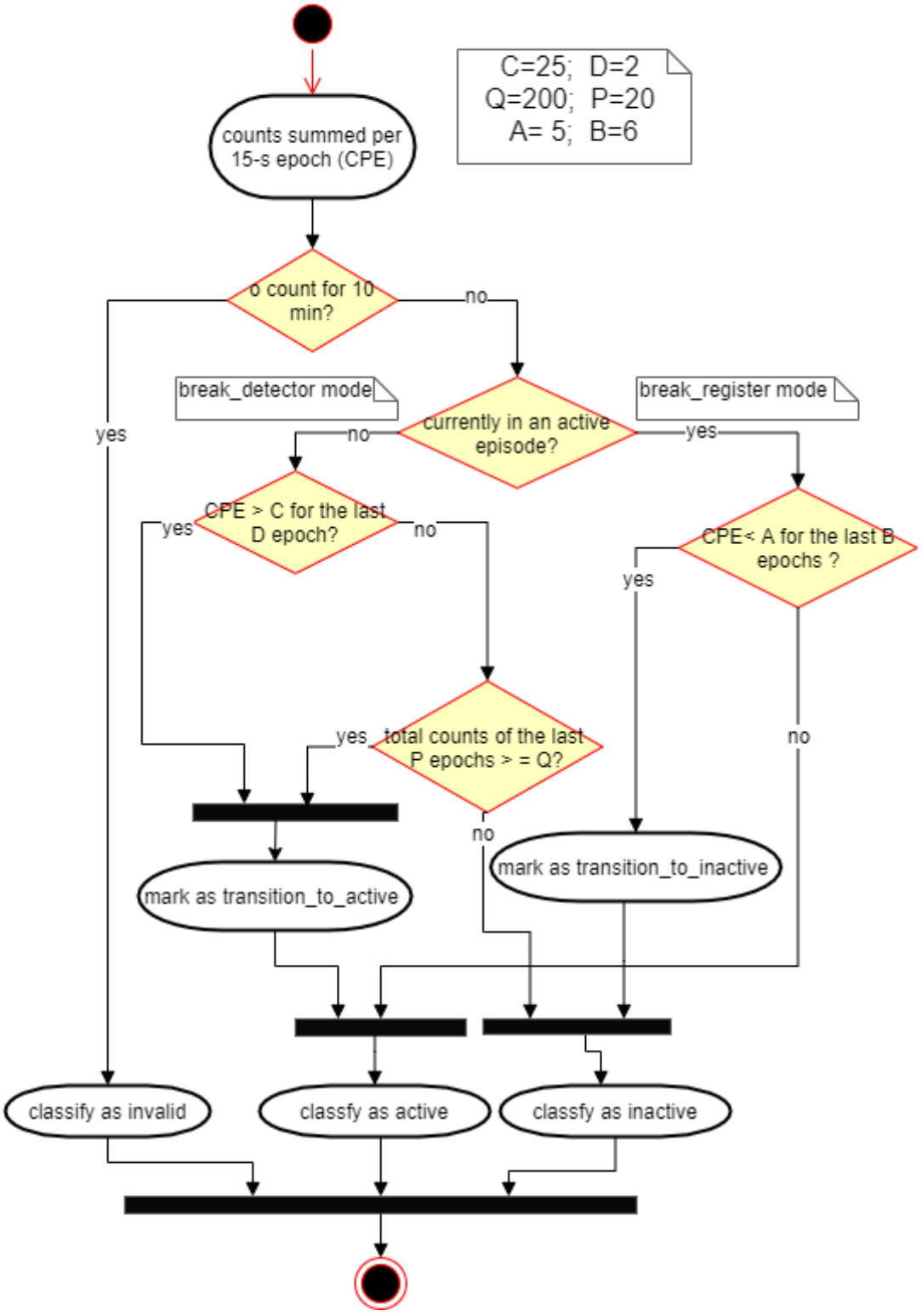
Algorithm and Database Design

Previous accelerometry-based activity classification algorithms predominantly applied thresholds to processed accelerometer output in the form of counts per epoch (CPE), which was indicative of activity intensity [53]. Owing to a lack of established cutoff points between SB and LPA for the MetaWear RG sensor, we conducted a series of structured data collection sessions, in which the wearer undertook various activities (eg, sitting while typing, sitting while writing, sitting while talking with hand gestures, sitting while doing torso twist, standing up, walking for five steps, and walking continuously) and identified the cutoff point between SB and LPA breaks based on visual inspection. The data revealed that sitting activities featured a CPE of 5 or less most of the time and 5-10 occasionally (eg, torso wrist), continuous walking featured a CPE of 23-30, and standing activities that involved mild ambulation (eg, standing up to open the window blinds and fetching a file in the same room) had an CPE of 10-25. We decided that the system should pick up two types of events that we wanted to consider and encourage as "breaks," including a burst of high-intensity movement (CPE >25) that signifies walking to a different room

and a continued period (>20 epochs or 5 min) of mild ambulation (CPE >10) that potentially involves doing chores in the room.

From there, we developed an algorithm that differentiated inactivity (ie, SB and standing still) and activity (ie, ambulatory breaks) based on a combination of activity intensity (ie, CPE cutoff points; parameters C, Q, and A in Figure 4) and temporality (ie, number of continuous epochs with CPE exceeding the cutoff points; parameters D, P, and B). Moreover, the algorithm featured a break detection mode and a break register mode to reduce frequent transitions between SB and breaks that were most likely caused by sporadic hand movements (eg, fidget and gesture).

Figure 4. Activity classification algorithm.



To support tailoring, we included a password-access menu in the app that allowed the researcher to change the default parameters to adjust the sensitivity of the break detector for individual participants if required. For instance, the value of C

could be raised by up to 5 CPE if the participant had particularly vigorous wrist movements in a sitting position that caused false positive break detections and the value of D could be lowered to 1 CPE if the break facility was close to the participant's office and if the participant preferred a quick trip to the break facility to be recognized as a break. The researcher could lower the value of P to 1 CPE, which would make the system not recognize mild movements as breaks at all.

Finally, while designing the algorithm, we considered the research needs for assessing fidelity and quantity of delivery for the *WorkMyWay* intervention. Hence, we designed the algorithm to detect an invalid tracking period caused by nonwear or technology failure based on the number of continuous epochs of zero counts. Moreover, we requested that the system log usage of the tracking and goal setting functions and record the timestamps of prompt delivery. In this way, fidelity could be operationalized as the percentage of tracking time that is valid and the percentage of prompts successfully delivered, and adherence to different functions could be operationalized as the number of days of use of each of the functions.

Casing Design

We discussed internally and consulted with product designers on several options to fix the electronics to the vessel, and we considered the ease of use for participants, as well as the ease and cost of production. We decided to three-dimensionally print cases for the *MetaWear* electronics using a template from the *MetaWear* manufacturer and make the printed case attachable to any vessel with a belt or Velcro tape. This would make the limited sets of *MetaWear* electronics reusable (ie, could be taken off and stuck to another vessel for the next participant without hygiene problems) and easy to handle for both researchers and users (ie, could be removed from the cup or bottle during meetings, for charging, and at the end of the study).

Results

Final WorkMyWay Intervention

We describe the resulting *WorkMyWay* intervention. To increase the quality of reporting, a TIDIeR checklist [54] is presented in Multimedia Appendix 5. Intervention materials are presented in Multimedia Appendix 6. The text below briefly describes the IoT-enabled delivery system and intervention protocol.

The Technological Delivery System

The resulting intervention delivery system consists of a PA monitor (called "wrist device") to be worn by the user, an LED reminder to be attachable to any cup or bottle (called "cup device") of the user's choice, and an Android app connected with both devices over Bluetooth (Figure 5). The wrist device automatically tracks movement and constantly syncs data with the app. The app differentiates activity and inactivity with the algorithm described above and actuates the LED according to the user's period of inactivity and prespecified rules (Figure 6A). At the end of each day, the app provides more detailed visual and numeric summaries of daily SB (Figure 6B) and rewards behavioral improvements and goal achievements with trophies and badges (Figure 6C). The "reward" section in the app allows the user to review previous rewards (Figure 6D) and adjust goals (Figure 6E). There is an "about" page with scientific facts about SB and information about the study, which is accessible from the drop-down menu. In addition, a *WorkMyWay* Lite app is made for baseline measurement, which works with the wrist device alone and merely provides "tracking" functionality (ie, no feedback).

Figure 5. The resulting intervention delivery system. LED: light-emitting diode.

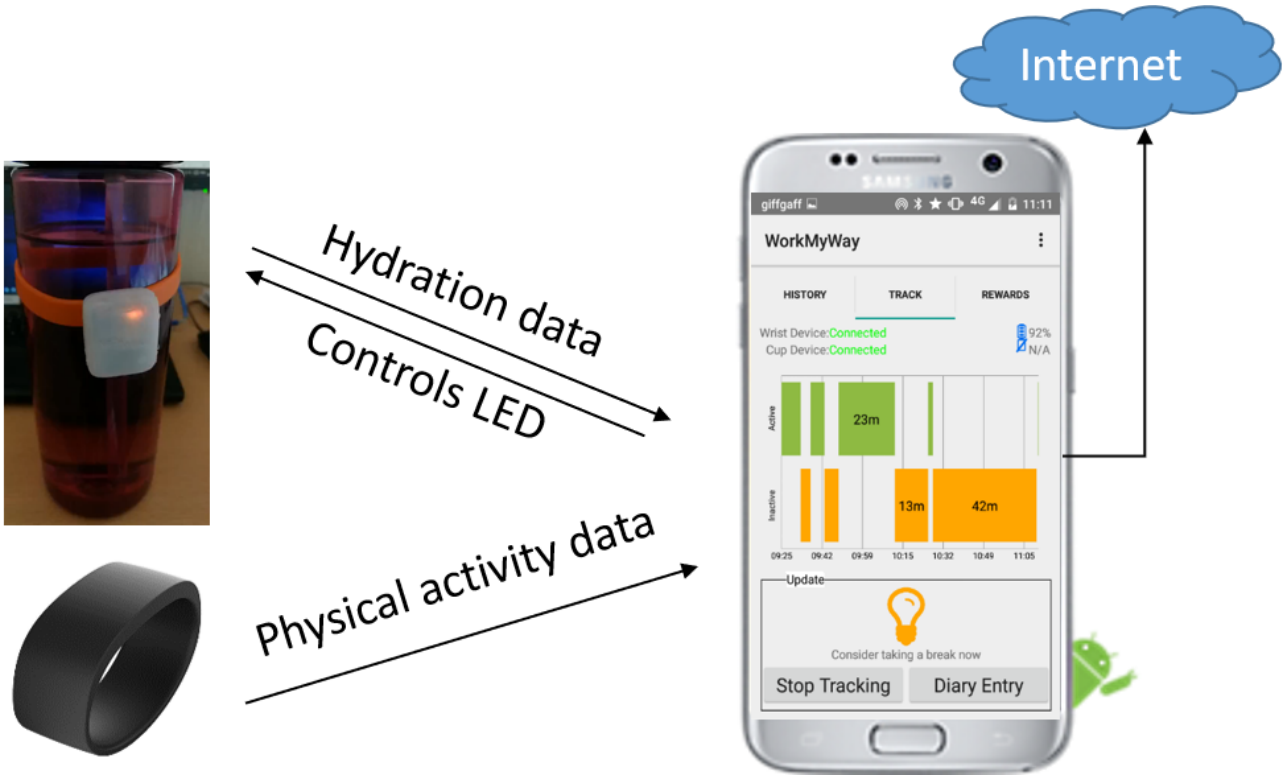
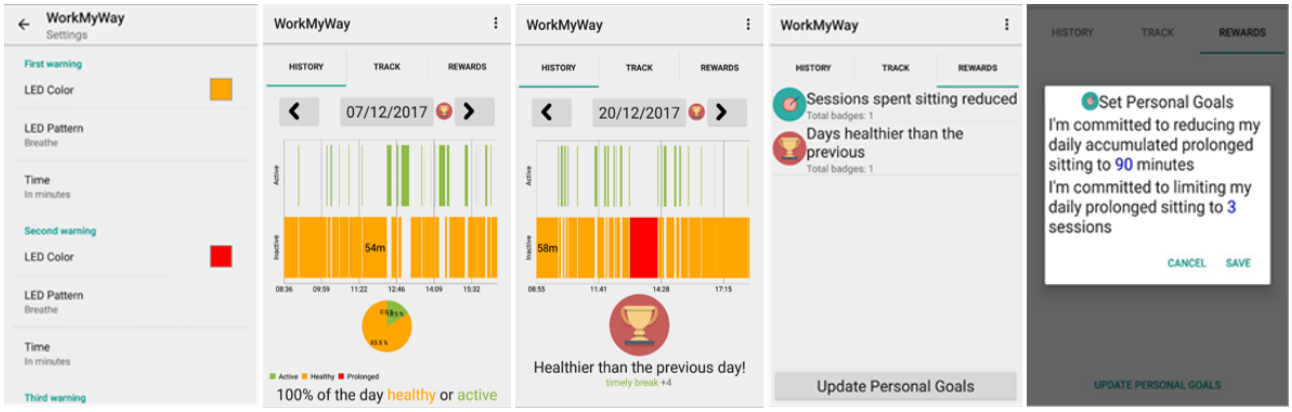


Figure 6. Screenshots of the WorkMyWay app.

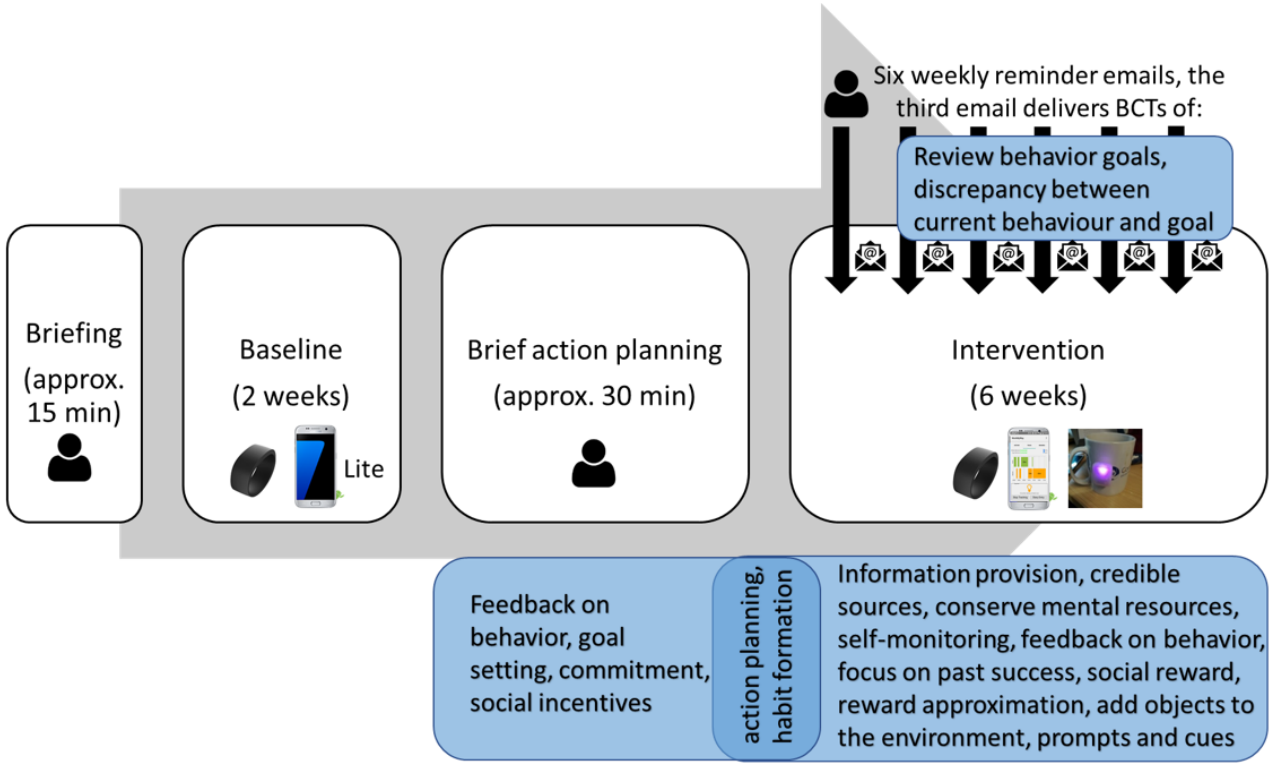


Intervention Delivery Protocol

The timing and dosage of delivery for different intervention components are summarized in Figure 7.

First, to set up personalized goals, the intervention requires information about individual participant's baseline SB, which needs to be collected with the *WorkMyWay* Lite system for 2 weeks. A trained researcher or occupational health consultant should help the participant set up the connection and give instructions on technology use at a face-to-face briefing session. A one-page two-sided "study cheat" sheet is also provided to each participant (Multimedia Appendix 6). After the 2-week baseline period, the researcher revisits the participant, replaces the Lite app with the full *WorkMyWay* app, fixes the "cup device" to a vessel of the participant's choice, and delivers a 30-min brief action planning (BAP) session to prepare the participant for the upcoming 6-week intervention period. BAP is a support technique that mirrors the spirits of motivational interview [55] and is aimed to facilitate participants in forming action plans that they feel confident to achieve. For instance, instead of prescribing a desired end state for the participant, the researcher can ask, "*now that we've looked at your baseline data and talked about prolonged sitting and health, is there anything you would like to do for your own health in the workplace in the next week or two?*" The participant will be guided to make specific action plans in terms of how often they would like to break up sitting in different contexts and to make use of the three configurable LED events to support execution of the action plans. The full BAP protocol is detailed in Multimedia Appendix 6.

Figure 7. The timing and dosage of delivery for each intervention component and the underlying behavior change techniques (BCTs). The underlying BCTs are specified in blue boxes, and the modes of delivery are indicated by icons (eg, wrist device, smartphone app, cup device, and researcher's email).



After the BAP session, the participants will use the full *WorkMyWay* app with the wrist and cup devices during their own office hours, with minimum intervention from the researcher for 6 weeks. To enhance adherence, a weekly reminder email is sent to the participants by the researcher on each Monday morning. The third weekly reminder email contains an extra message that completes the BAP protocol by prompting the participants to review their "history" in the app, compare performance against the goals initially set by them at the BAP session, and adjust the goals and reminder settings in the app if necessary. The email also asks the participants if they want to have the break detection sensitivity threshold adjusted to suit their individual work contexts and preferences.

After 6 weeks, the researcher contacts each participant to schedule a debriefing interview, where subjective experiences of the behavior change intervention and technology use can be discussed.

The researcher or practitioner needs to be trained in both troubleshooting the *WorkMyWay* system and applying the BAP techniques. All sessions need to take place at the participant's workplace.

Discussion

Principal Results

This paper describes the development process, as well as the final design of a novel SB intervention delivered with an IoT system. We have followed a systematic and comprehensive theory-driven development approach while drawing upon HCD methods to involve stakeholders in the design process. The resulting intervention draws upon a total of 17 BCTs to target behavioral determinants in 11 theoretical domains. We reflect on the novel design of our delivery system, as well as the development process to discuss lessons for future research.

Our *WorkMyWay* system is complex, combining the following three distinct types of devices that deliver different intervention components to target different behavioral determinants:

1. A screen-based component (the app) that focuses on delivering the intervention functions of education, persuasion, and incentivization. This approach has been extensively used and tested in existing DBCI research [56].
2. A wearable component that focuses on delivering the intervention function of enablement by automatically collecting and livestreaming PA data over a Bluetooth connection, which essentially offloads the human cognitive task of self-monitoring to the technology. This is also an established approach in both academic literature and commercial products [57].
3. An IoT component (the digitally augmented vessel) that is intended to deliver the intervention function of environmental restructuring and enablement, as it is seamlessly integrated into an office worker's working environment and daily routine to create awareness of sitting time with subtle prompts and cues for breaks. This is the novelty of our approach.

A key lesson from our technology design is therefore, first of all, to recognize the different modalities of communication afforded by different devices and carefully choose suitable media to deliver intervention contents with the appropriate level of obtrusiveness for different moments. In our case, the types of contents to be delivered range from a detailed timeline of activity episodes for end-of-day reflection, through "at-a-glance" environmental cues for in-situ awareness and actionable information, to unobtrusive and unnoticed data capture. Our various interfaces embody a range of interaction styles that operate in the foreground (ie, occurs in the center of attention) versus background (ie, attentional peripheral) [58], which has been debated substantially in human-computer interaction. The advocacy for designing background interactions dates back to Mark Weiser's vision [59] that technology will weave into everyday life and become unnoticed eventually, which was followed on by the call for calm computing [60]. Our design of the glanceable IoT component mirrors the ethos of calm computing by making it stay in the periphery of the user's attention most of the time without intruding on the user and only move to the center of attention when prolonged sitting is detected and just-in-time intervention is required. On the contrary, our app is designed to counter balance this embeddedness and calmness with the need to engage, stimulate, and provoke users to be reflective at the end of each workday [61]. This also highlights the importance of the intervention protocol that specifies when and how each technological component is to be used and for how long (ie, the required dosage of each component).

Furthermore, we argue that IoT-enabled objects should not be seen as a replacement for screen-based apps and wearables, but rather as an addition. Those smart objects are particularly suitable for delivering certain intervention functions from the BCW, such as enablement and environmental restructuring, whereas screen-based media are better reserved for the intervention functions of education and training. Hence, instead of designing a whole IoT intervention delivery system anew, designers might consider a more incremental approach, extending existing apps and wearables with complementary IoT objects that will eventually contribute to a more complete ecology of DBCI technologies.

Another aspect we want to reflect on is the design of data processing, which has implications for requirements like personalization, user autonomy, and privacy. Given sufficient training data, we recognize the potential to replace our "quick and dirty" algorithm with a more sophisticated machine learning algorithm that can learn about and adapt to individual patterns of behavior. However, it should be noted that our manually designed classification rule has the advantage of transparency, so that researchers and clinicians delivering the intervention can potentially adapt it for individuals by directly tweaking its various parameters, which is easier than adjusting a "black box" machine learning algorithm. We also noted the need for sharing data between the various technological components. This requires attention to reliable networking and a potentially encrypted data transmission protocol that is a hidden but often difficult aspect of technical development, with important implications for data security, privacy, and ownership. In the

context of PA tracking in the workplace, who has access to data is a nontrivial issue and needs to be handled with caution to minimize the chance of employer surveillance or peer pressure that can sabotage the health initiative.

Strengths

A major strength of this study was the application of the BCW and related frameworks (eg, TDF, the BCT taxonomy, and TIDIeR) to structure the development and description of the intervention, which helped clarify theoretical underpinnings, active ingredients, and the final design. There has been a call for more thorough reporting of intervention design and development processes [62], and papers documenting the design and development of DBCIs with those frameworks have emerged over the past few years [63-65]. However, these papers tended to report on the design and development of interventions delivered with less advanced technologies, such as web pages, smartphone apps, and SMS, rather than IoT technologies. To our knowledge, our study is the first to systematically apply all BCW steps to develop an IoT-enabled intervention and specifically reflect on this technological approach.

Another strength of this study was related to factoring in the need to assess fidelity and dosage of delivery at the time of designing the database. This meant the right type and structure of data were requested to be captured and stored to allow monitoring of user interactions with individual functions, as well as the whole intervention. The intervention mapping table (Multimedia Appendix 2) together with our monitoring data will be useful for separating and clarifying the effects of individual intervention components, which could contribute to the endeavor to establish links between BCTs and specific mechanisms of action in the field of DBCI research [66].

A third strength of the design process was the seamless integration of the bottom-up HCD approach into the top-down theory-driven intervention design process. Other studies drawing on similar approaches tended to implement BCW and HCD in distinct studies or phases [64,67]. Our approach was slightly different as we embedded HCD in BCW-guided studies. For instance, Study 1 served the dual purposes of behavioral diagnosis under the BCW framework and requirement elicitation under the HCD methodology. Study 2 could be seen as a public and patient involvement activity [68] in the context of health intervention design, though we moved beyond public and patient involvement to empower stakeholders with two innovative HCD methods. Ideation cards provide an accessible and "bite-sized" representation of design knowledge (including theory-informed BCTs and technological opportunities) for use by stakeholders from various backgrounds during collaborative design sessions. The cards essentially act as a bridging mechanism between theories and practical design solutions. Lo-fi prototyping, including the use of an IoT maker kit and loose materials, enables stakeholders to become more "hands on" in the design process and engage with emerging technologies without needing to acquire software or electronic engineering skills. We suggest that both methods can complement more traditional interviews and focus groups to elicit stakeholder requirements for novel futuristic modes of delivery (ie, smart objects) while grounding design solutions in theories.

Limitations and Future Work

First, our final design had a rather rough look, as we used the MetaWear hardware, SDK, and three-dimensional printing template to reduce production difficulty and cost. Developing a finely finished IoT product ideally requires a team of product designers and electronic engineers, as well as software and data engineers. With that said, IoT technologies evolve so rapidly that an IoT-based intervention likely needs to be improved over time after it is developed. Hence, we compromised by building a "minimum viable product" with all the essential components for a proof-of-concept study before investing heavily in developing a finely finished product. The Medical Research Council framework also suggests a phased approach to evaluating complex interventions, starting from feasibility studies targeted at each of the uncertainties in the design and moving on to a pilot and then a definitive trial [21]. Hence, we suggest deploying and evaluating the current version of the WorkMyWay intervention in real office-based workplaces on a small scale, with focus on assessing the acceptability and feasibility of the various components and the protocol, identifying barriers and facilitators to use, and clarifying the mechanisms of action prior to pilot and full-scale randomized controlled trials.

A second drawback of this research concerns the small and self-selected sample in both studies, which limits the generalizability of the findings. Self-selection filtered out those unconcerned about the issue of SB or lacking control of break behavior at work. The wider acceptability and effectiveness of WorkMyWay will need to be demonstrated by conducting evaluative studies with more representative samples in diverse office-based workplaces.

Conclusions

This paper documents the design and development of WorkMyWay, an IoT-enabled behavior change intervention to reduce workplace SB. The development process applied behavioral theories systematically while drawing on HCD methods. The resultant intervention, including its content, rationale, and delivery, is detailed to allow replication. Future studies are needed to evaluate the feasibility of the intervention in office-based workplaces and the efficacy of the intervention in improving office workers' behavioral and health outcomes.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (grant number EP/L015463/1), Horizon Centre for Doctoral Training at the University of Nottingham, and Unilever UK Ltd.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study 1 materials (stage 2).
[DOCX File , 135 KB - mhealth_v8i7e17914_app1.docx]

Multimedia Appendix 2
Mapping behavioral determinants to intervention components following the behavior change wheel (stage 3).
[DOCX File , 27 KB - mhealth_v8i7e17914_app2.docx]

Multimedia Appendix 3
Study 2 workshop materials (stage 4).
[DOCX File , 3137 KB - mhealth_v8i7e17914_app3.docx]

Multimedia Appendix 4
Requirement specification document (stage 5).
[DOCX File , 1684 KB - mhealth_v8i7e17914_app4.docx]

Multimedia Appendix 5
Template for Intervention Description and Replication (TIDieR) checklist.
[DOCX File , 31 KB - mhealth_v8i7e17914_app5.docx]

Multimedia Appendix 6
Intervention materials.
[DOCX File , 951 KB - mhealth_v8i7e17914_app6.docx]

References

1. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
2. Henson J, Yates T, Biddle SJ, Edwardson CL, Khunti K, Wilmot EG, et al. Associations of objectively measured sedentary behaviour and physical activity with markers of cardiometabolic health. Diabetologia 2013 May;56(5):1012-1020. [doi: 10.1007/s00125-013-2845-9] [Medline: 23456209]
3. Bankoski A, Harris TB, McClain JJ, Brychta RJ, Caserotti P, Chen KY, et al. Sedentary activity associated with metabolic syndrome independent of physical activity. Diabetes Care 2011 Feb;34(2):497-503 [FREE Full text] [doi: 10.2337/dc10-0987] [Medline: 21270206]
4. Lynch BM. Sedentary behavior and cancer: a systematic review of the literature and proposed biological mechanisms. Cancer Epidemiol Biomarkers Prev 2010 Nov;19(11):2691-2709 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-0815] [Medline: 20833969]
5. Dunstan DW, Howard B, Healy GN, Owen N. Too much sitting--a health hazard. Diabetes Res Clin Pract 2012 Sep;97(3):368-376. [doi: 10.1016/j.diabres.2012.05.020] [Medline: 22682948]
6. Brocklebank LA, Falconer CL, Page AS, Perry R, Cooper AR. Accelerometer-measured sedentary time and cardiometabolic biomarkers: A systematic review. Prev Med 2015 Jul;76:92-102 [FREE Full text] [doi: 10.1016/j.ypmed.2015.04.013] [Medline: 25913420]
7. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Breaks in sedentary time: beneficial associations with metabolic risk. Diabetes Care 2008 Apr;31(4):661-666. [doi: 10.2337/dc07-2046] [Medline: 18252901]
8. Buckley JP, Hedge A, Yates T, Copeland RJ, Loosemore M, Hamer M, et al. The sedentary office: an expert statement on the growing case for change towards better health and productivity. Br J Sports Med 2015 Nov;49(21):1357-1362. [doi: 10.1136/bjsports-2015-094618] [Medline: 26034192]
9. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
10. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]

11. Chi PY, Chen JH, Chu HH, Lo JL. Enabling calorie-aware cooking in a smart kitchen. In: Oinas-Kukkonen H, Hasle P, Harjumaa M, Segerståhl K, Øhrstrøm P, editors. Persuasive Technology. PERSUASIVE 2008. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2008:116-127.
12. Intille S. Designing a home of the future. IEEE Pervasive Comput 2002 Apr;1(2):76-82. [doi: 10.1109/mprv.2002.1012340]
13. Haller M, Richter C, Brandl P, Gross S, Schossleitner G, Schrempf A, et al. Finding the right way for interrupting people to posture guidance. In: Campos P, Graham N, Jorge J, Nunes N, Palanque P, Winckler M, editors. Human-Computer Interaction – INTERACT 2011. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2011.
14. Jafarinaimi N, Forlizzi J, Hurst A, Zimmerman J. Breakaway: an ambient display designed to change human behavior. In: CHI EA'05 Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2005 Presented at: International Conference on Human Factors in Computing Systems; April 2-7, 2005; Portland, Oregon, USA p. 1945-1948. [doi: 10.1145/1056808.1057063]
15. Ham J, Midden C. Ambient persuasive technology needs little cognitive effort: The differential effects of cognitive load on lighting feedback versus factual feedback. In: Ploug T, Hasle P, Oinas-Kukkonen H, editors. Persuasive Technology. PERSUASIVE 2010. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2010:132-142.
16. Lu S, Ham J, Midden C. Red radiators versus red tulips: the influence of context on the interpretation and effectiveness of color-based ambient persuasive technology. In: Meschtscherjakov A, De Ruyter B, Fuchsberger V, Murer M, Tscheligi M, editors. Persuasive Technology. PERSUASIVE 2016. Lecture Notes in Computer Science. Cham: Springer; 2016:303-314.
17. Van Ittersum K, Wansink B. Plate Size and Color Suggestibility: The Delboeuf Illusion's Bias on Serving and Eating Behavior. J Consum Res 2012 Aug 01;39(2):215-228. [doi: 10.1086/662615]
18. Pinder C. Nonconscious Behaviour Change Technology. In: CHI EA '17: Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017 Presented at: International Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, Colorado, USA p. 160-165. [doi: 10.1145/3027063.3048426]
19. Rogers Y, Hazlewood W, Marshall P, Dalton N, Hertrich S. Ambient Influence: Can Twinkly Lights Lure and Abstract Representations Trigger Behavioral Change. In: Proceedings of the 12th ACM Int Conf Ubiquitous Comput - Ubicomp'10. 2010 Presented at: The 12th ACM Conference on Ubiquitous Computing; September 26–29, 2010; Copenhagen, Denmark. [doi: 10.1145/1864349.1864372]
20. Huang Y, Benford S, Blake H. Digital Interventions to Reduce Sedentary Behaviors of Office Workers: Scoping Review. J Med Internet Res 2019 Feb 07;21(2):e11079 [FREE Full text] [doi: 10.2196/11079] [Medline: 30730294]
21. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions. Medical Research Council. 2019. URL: http://www.mrc.ac.uk/complexinterventionsguidance [accessed 2020-07-15]
22. Bartholomew LK, Parcel GS, Kok G, Gottlieb NH. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco, CA, USA: Jossey-Bass; 2006.
23. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback Publishing; 2014.
24. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
25. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
26. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
27. Deutsch R, Strack F. TARGET ARTICLE: Duality Models in Social Psychology: From Dual Processes to Interacting Systems. Psychological Inquiry 2006 Jul 08;17(3):166-172. [doi: 10.1207/s15327965pli1703_2]
28. Strack F, Deutsch R. Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 2004;8(3):220-247. [doi: 10.1207/s15327957pspr0803_1] [Medline: 15454347]
29. Wood W, Rünger D. Psychology of Habit. Annu Rev Psychol 2016;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]
30. Gollwitzer P. Implementation intentions: Strong effects of simple plans. American Psychologist 1999 Mar;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]
31. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
32. Francis JJ, Stockton C, Eccles MP, Johnston M, Cuthbertson BH, Grimshaw JM, et al. Evidence-based selection of theories for designing behaviour change interventions: using methods based on theoretical construct domains to understand clinicians' blood transfusion behaviour. Br J Health Psychol 2009 Nov;14(Pt 4):625-646. [doi: 10.1348/135910708X397025] [Medline: 19159506]

33. French S, Green S, O'Connor DA, McKenzie J, Francis J, Michie S, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. Implement Sci 2012 Apr 24;7:38 [FREE Full text] [doi: 10.1186/1748-5908-7-38] [Medline: 22531013]
34. 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]
35. Frauenberger C, Good J, Fitzpatrick G, Iversen OS. In pursuit of rigour and accountability in participatory design. Int J Hum Comput Stud 2015 Feb;74:93-106 [FREE Full text] [doi: 10.1016/j.ijhcs.2014.09.004] [Medline: 26109833]
36. Maguire M. Methods to support human-centred design. International Journal of Human-Computer Studies 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
37. Martin A, Fitzsimons C, Jepson R, Saunders DH, van der Ploeg HP, Teixeira PJ, EuroFIT consortium. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. Br J Sports Med 2015 Aug;49(16):1056-1063. [doi: 10.1136/bjsports-2014-094524] [Medline: 25907181]
38. Prince SA, Saunders TJ, Gresty K, Reid RD. A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. Obes Rev 2014 Nov;15(11):905-919 [FREE Full text] [doi: 10.1111/obr.12215] [Medline: 25112481]
39. Baker R, Coenen P, Howie E, Lee J, Williamson A, Straker L. A detailed description of the short-term musculoskeletal and cognitive effects of prolonged standing for office computer work. Ergonomics 2018 Jul;61(7):877-890. [doi: 10.1080/00140139.2017.1420825] [Medline: 29388500]
40. Waters TR, Dick RB. Evidence of health risks associated with prolonged standing at work and intervention effectiveness. Rehabil Nurs 2015;40(3):148-165 [FREE Full text] [doi: 10.1002/rnj.166] [Medline: 25041875]
41. Edwardson C, Yates T, Biddle S, Davies M, Dunstan D, Esliger D, et al. Effectiveness of the Stand More AT (SMArT) Work intervention: cluster randomised controlled trial. BMJ 2018 Oct 10;363:k3870 [FREE Full text] [doi: 10.1136/bmj.k3870] [Medline: 30305278]
42. Chastin SF, De Craemer M, De Cocker K, Powell L, Van Cauwenberg J, Dall P, et al. How does light-intensity physical activity associate with adult cardiometabolic health and mortality? Systematic review with meta-analysis of experimental and observational studies. Br J Sports Med 2019 Mar;53(6):370-376 [FREE Full text] [doi: 10.1136/bjsports-2017-097563] [Medline: 29695511]
43. Why we should sit less. NHS. 2016. URL: http://www.nhs.uk/Livewell/fitness/Pages/sitting-and-sedentary-behaviour-are-bad-for-your-health.aspx [accessed 2016-03-20]
44. Owen N, Bauman A, Brown W. Too much sitting: a novel and important predictor of chronic disease risk? Br J Sports Med 2009 Feb;43(2):81-83. [doi: 10.1136/bjsm.2008.055269] [Medline: 19050003]
45. Srivastava A, Thomson S. Framework analysis: A qualitative methodology for applied policy research. Journal of Administration & Governance 2009;4:72-79 [FREE Full text]
46. Braun A, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Huang Y, Benford S, Hendrickx H, Treloar R, Blake H. Office Workers's Perceived Barriers and Facilitators to Taking Regular Micro-Breaks at Work: a Diary-Probed Interview Study. In: de Vries P, Oinas-Kukkonen H, Siemons L, Beerlage-de Jong N, van Gemert-Pijnen L, editors. Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors. PERSUASIVE 2017. Lecture Notes in Computer Science. Cham: Springer; 2017:149-161 URL: https://doi.org/10.1007/978-3-319-55134-0_12 [doi: 10.1007/978-3-319-55134-0_12]
48. Golembewski M, Selby M. Ideation decks: a card-based design ideation tool. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems. 2010 Presented at: DIS' 10; August 16-20, 2010; Aarhus, Denmark p. 89-92. [doi: 10.1145/1858171.1858189]
49. Work-Life Balance in the Digital Age: Funding for networking events. Engineering and Physical Sciences Research Council. URL: https://epsrc.ukri.org/funding/calls/worklifebalanceindigitalage/ [accessed 2019-06-12]
50. Chen Y, Liang RH, Chiang CD. IoT deck: a digital card-based ideation game to inspire internet of things design. In: Proceedings of IASDR'2011. 2011 Presented at: The 4th World Conference on Design Research; October 31-November 4, 2011; Delft, The Netherlands.
51. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. CAIS 2009;24:485-500. [doi: 10.17705/1cais.02428] [Medline: 45267822]
52. Android Developer's Guide. Google. 2017. URL: https://developer.android.com/guide/navigation [accessed 2017-05-26]
53. Boerema S, Essink G, Tönis TM, van Velsen L, Hermens H. Sedentary Behaviour Profiling of Office Workers: A Sensitivity Analysis of Sedentary Cut-Points. Sensors (Basel) 2015 Dec 25;16(1) [FREE Full text] [doi: 10.3390/s16010022] [Medline: 26712758]
54. 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 Mar 07;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]

55. Gutnick D, Reims K, Davis C, Gainforth H. Brief action planning to facilitate behavior change and support patient self-management. Journal of Clinical Outcomes Management 2014;21(1):18-29.
56. Stephenson A, McDonough S, Murphy M, Nugent C, Mair J. 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 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
57. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for Self-Monitoring Sedentary Time or Physical Activity: A Scoping Review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
58. Buxton B. Integrating the Periphery and Context: A New Taxonomy of Telematics. In: Proceedings of Graph Interface'95. 1995 Presented at: Graphics Interface; May 17-19, 1995; Québec, Canada p. 239-246. [doi: 10.1145/1753326.1753499]
59. Weiser M. The Computer for the 21st Century. Sci Am 1991 Sep;265(3):94-104. [doi: 10.1038/scientificamerican0991-94]
60. Weiser M, Brown J. Designing Calm Technology. Calmtech. 1995. URL: https://people.csail.mit.edu/rudolph/Teaching/weiser.pdf [accessed 2020-07-15]
61. Rogers Y. Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences. In: Ubiquitous Computing. UbiComp 2006. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2006:404-421.
62. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. Implement Sci 2009 Jul 16;4:40 [FREE Full text] [doi: 10.1186/1748-5908-4-40] [Medline: 19607700]
63. Curtis KE, Lahiri S, Brown KE. Targeting Parents for Childhood Weight Management: Development of a Theory-Driven and User-Centered Healthy Eating App. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
64. Hendrie GA, James-Martin G, Williams G, Brindal E, Whyte B, Crook A. The Development of VegEze: Smartphone App to Increase Vegetable Consumption in Australian Adults. JMIR Form Res 2019 Mar 27;3(1):e10731 [FREE Full text] [doi: 10.2196/10731] [Medline: 30916653]
65. Neuhaus M, Healy GN, Fjeldsoe BS, Lawler S, Owen N, Dunstan DW, et al. Iterative development of Stand Up Australia: a multi-component intervention to reduce workplace sitting. Int J Behav Nutr Phys Act 2014 Feb 21;11:21 [FREE Full text] [doi: 10.1186/1479-5868-11-21] [Medline: 24559162]
66. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From Theory-Inspired to Theory-Based Interventions: A Protocol for Developing and Testing a Methodology for Linking Behaviour Change Techniques to Theoretical Mechanisms of Action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
67. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A Short Message Service Intervention to Support Adherence to Home-Based Strengthening Exercise for People With Knee Osteoarthritis: Intervention Design Applying the Behavior Change Wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
68. Barham L. Public and patient involvement at the UK National Institute for Health and Clinical Excellence. Patient 2011;4(1):1-10. [doi: 10.2165/11586090-000000000-00000] [Medline: 21766889]

Abbreviations

APEASE: affordability, practicability, (cost-) effectiveness, acceptability, side-effects or safety, and equity
BCT: behavior change technique
BCW: behavior change wheel
HCD: human-centered design
IoT: internet of things
LED: light-emitting diode
LPA: light physical activity
MET: metabolic equivalent
PA: physical activity
SB: sedentary behavior
TDF: theoretical domain framework

Edited by G Eysenbach; submitted 21.01.20; peer-reviewed by C Brakenridge, K De Cocker, T Muto, M Bestek; comments to author 11.02.20; revised version received 08.05.20; accepted 03.06.20; published 29.07.20.

Please cite as:
Huang Y, Benford S, Price D, Patel R, Li B, Ivanov A, Blake H
Using Internet of Things to Reduce Office Workers' Sedentary Behavior: Intervention Development Applying the Behavior Change Wheel and Human-Centered Design Approach
JMIR Mhealth Uhealth 2020;8(7):e17914
URL: http://mhealth.jmir.org/2020/7/e17914/
doi:10.2196/17914
PMID:32723716

Original Paper

Intra-Rater and Inter-Rater Reliability of Tongue Coating Diagnosis in Traditional Chinese Medicine Using Smartphones: Quasi-Delphi Study

Zhi Chun Wang[1], BSc; Shi Ping Zhang[1], PhD; Pong Chi Yuen[2], PhD; Kam Wa Chan[3*], PhD; Yi Yi Chan[1*], MPhil; Chun Hoi Cheung[1*], MPhil; Chi Ho Chow[1*], PhD; Ka Kit Chua[1*], PhD; Jun Hu[4], MD; Zhichao Hu[1*], BMed; Beini Lao[5*], MMed; Chun Chuen Leung[6*], MPhil; Hong Li[1*], MD; Linda Zhong[1*], PhD; Xusheng Liu[5*], MD; Yulong Liu[1*], MD; Zhenjie Liu[5*], MD; Xin Lun[1*], MPhil; Wei Mo[7*], MPhil; Sheung Yuen Siu[1*], BS, BCM; Zhoujian Xiong[8*], MD; Wing Fai Yeung[9*], PhD; Run Yun Zhang[10*], MPhil; Xuebin Zhang[1*], MD

[1]School of Chinese Medicine, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]School of Computing Science, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[3]Department of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)

[4]Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[5]Guangdong Provincial Hospital of Traditional Chinese Medicine, The Second Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine, Guangzhou, China

[6]Hong Zhi Tang Chinese Medicine Clinic, Hong Kong, China (Hong Kong)

[7]Guang Dong Second Traditional Chinese Medicine Hospital, Guangzhou, China

[8]Yongan Hospital, Foshan, China

[9]School of Nursing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[10]China Academy of Chinese Medical Sciences, Guang An Men Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Shi Ping Zhang, PhD
School of Chinese Medicine
Hong Kong Baptist University
Kowloon Tong, Kowloon
Hong Kong,
China (Hong Kong)
Phone: 852 3411 2466
Fax: 852 3411 2461
Email: spzhang@hkbu.edu.hk

Abstract

Background: There is a growing trend in the use of mobile health (mHealth) technologies in traditional Chinese medicine (TCM) and telemedicine, especially during the coronavirus disease (COVID-19) outbreak. Tongue diagnosis is an important component of TCM, but also plays a role in Western medicine, for example in dermatology. However, the procedure of obtaining tongue images has not been standardized and the reliability of tongue diagnosis by smartphone tongue images has yet to be evaluated.

Objective: The first objective of this study was to develop an operating classification scheme for tongue coating diagnosis. The second and main objective of this study was to determine the intra-rater and inter-rater reliability of tongue coating diagnosis using the operating classification scheme.

Methods: An operating classification scheme for tongue coating was developed using a stepwise approach and a quasi-Delphi method. First, tongue images (n=2023) were analyzed by 2 groups of assessors to develop the operating classification scheme for tongue coating diagnosis. Based on clinicians' (n=17) own interpretations as well as their use of the operating classification scheme, the results of tongue diagnosis on a representative tongue image set (n=24) were compared. After gathering consensus for the operating classification scheme, the clinicians were instructed to use the scheme to assess tongue features of their patients under direct visual inspection. At the same time, the clinicians took tongue images of the patients with smartphones and assessed

tongue features observed in the smartphone image using the same classification scheme. The intra-rater agreements of these two assessments were calculated to determine which features of tongue coating were better retained by the image. Using the finalized operating classification scheme, clinicians in the study group assessed representative tongue images (n=24) that they had taken, and the intra-rater and inter-rater reliability of their assessments was evaluated.

Results: Intra-rater agreement between direct subject inspection and tongue image inspection was good to very good (Cohen κ range 0.69-1.0). Additionally, when comparing the assessment of tongue images on different days, intra-rater reliability was good to very good (κ range 0.7-1.0), except for the color of the tongue body (κ=0.22) and slippery tongue fur (κ=0.1). Inter-rater reliability was moderate for tongue coating (Gwet AC2 range 0.49-0.55), and fair for color and other features of the tongue body (Gwet AC2=0.34).

Conclusions: Taken together, our study has shown that tongue images collected via smartphone contain some reliable features, including tongue coating, that can be used in mHealth analysis. Our findings thus support the use of smartphones in telemedicine for detecting changes in tongue coating.

(JMIR Mhealth Uhealth 2020;8(7):e16018) doi:10.2196/16018

KEYWORDS

mobile health; smartphone; traditional Chinese medicine; telemedicine; tongue image; machine learning; oral disease; Gwet AC2; COVID-19

Introduction

Traditional Chinese medicine (TCM) has been practiced for over 3000 years in China and it has recently embraced mobile health (mHealth) [1-3]. In TCM, the practitioners combine tongue diagnosis with other signs and symptoms collected through the 4 diagnostic methods to attain a holistic view of the patient's health status, so as to formulate strategies to adjust any imbalance in body functions [4]. Particularly, the color and texture of tongue coating or fur serves as a simple and effective means to aid syndrome differentiation in TCM [5]. Thus, tongue fur has been shown to be of significant importance in differentiation of TCM patterns associated with many diseases, such as asthma [6], gastric cancer [7], metabolic syndrome [8], hepatitis [9], gastritis [5], and, most recently, coronavirus disease (COVID-19) [10].

Anatomically, the dorsal surface of the tongue is divided by the sulcus terminalis into the anterior two-thirds (oral part), and the posterior one-third (pharyngeal part). TCM tongue diagnosis examines mainly the oral part, which is covered by a connective tissue core with overlying stratified squamous epithelium on the surface. The epithelium on the oral part of the tongue forms three types of papillae, and they have been named for their appearance: filiform, fungiform, and foliate papillae. In TCM, tongue fur refers to the keratinized tip of filiform papillae and dead epithelial cells, and its appearance is also affected by oral bacteria, blood metabolites, and salivary secretions from mucous and serous glands [11]. The fungiform papillae are highly vascularized and they provide the base color of the tongue body in TCM diagnosis. The foliate papillae are located at the edge of the tongue posteriorly and have an insignificant contribution to tongue coating.

In general, TCM diagnosis, including tongue diagnosis, has been found to be subjective with low inter-rater reliability [12-15]. Diagnosis made from tongue images collected by mobile devices may be more variable, due to variations in photographic techniques, camera settings, environmental lighting, light sources, display devices, etc [16]. These variations present barriers in the direct use of smartphone-collected tongue images in telemedicine and the application of machine learning for automatic tongue diagnosis in mHealth. Previously, we compared the results of tongue diagnosis from smartphone images and a commercially available tongue diagnostic device with those from direct inspection of human subjects. We found that the colors of tongue coating were the most consistently rated features between image inspection and direct subject inspection [17]. To further explore the characteristics of tongue diagnosis using smartphones, we set out to investigate the intra-rater and inter-rater reliability of tongue diagnosis using smartphone images. We first developed an operating classification scheme for tongue coating using a stepwise approach and a quasi-Delphi method. Subsequently, the study group used the same scheme to assess tongue features under direct subject inspection and with the tongue images of the same subjects. This was done to determine the reliability of tongue images in displaying features of the tongue. Finally, the intra-rater and inter-rater agreement of a representative tongue image set were evaluated to determine the reliability of tongue diagnosis using smartphone images. We hypothesized that certain tongue features would be recognized more consistently than others when examined through smartphone images.

Methods

Participants

First, 2 groups of TCM students (n=2 in each group) who had completed a TCM diagnostic course were employed to work independently on the initial classification of Image Set I in Experiment 1. To develop a working classification scheme, a total of 18 TCM clinicians, all of whom were qualified Chinese medicine practitioners, participated in a quasi-Delphi process, which involved Experiments 2 and 3 (Figure 1). The Delphi process group of clinicians were from several university clinics in Hong Kong and their respective teaching hospitals in mainland China, or were research collaborators associated with these universities, with a median of 14 years in clinical practice (95% CI 12-23). The same group of clinicians participated in the assessment of intra-rater and inter-rater reliability.

Figure 1. Flowchart of the study.

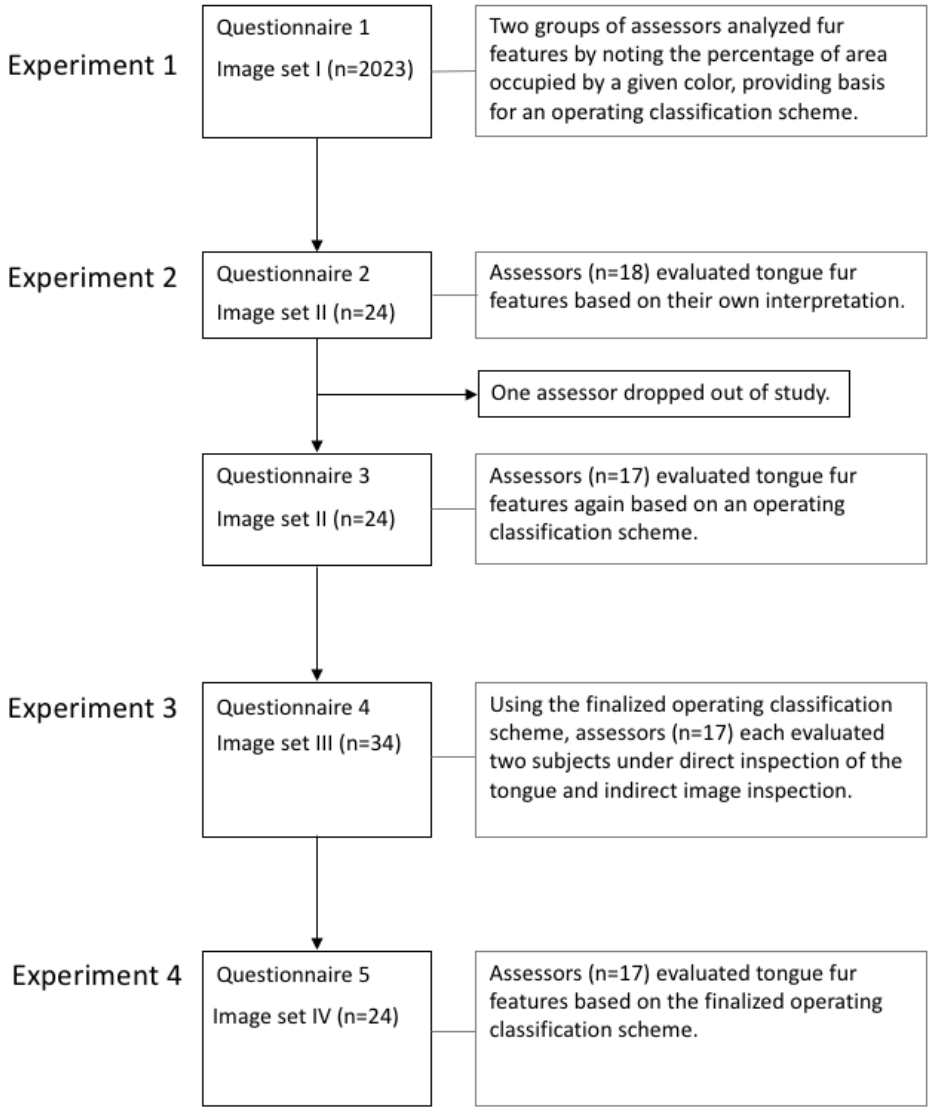


Image Collection

The study protocols for tongue image collection were designed in accordance with the Declaration of Helsinki and approved by the Committee on the Use of Human and Animal Subjects in Teaching and Research of Hong Kong Baptist University. Verbal informed consent was obtained from subjects after the nature of the study was explained. Tongue images were collected in various Chinese medicine clinics of Hong Kong Baptist University and its teaching hospitals in Guangzhou. Image set I consists of a total of 2023 tongue images taken with smartphones (iPhone [Apple Inc], HUAWEI [Huawei Technologies Co Ltd], Samsung [Samsung Group], and Xiaomi [Xiaomi Corp]; about 400 images per brand) and a computer-aided diagnostic system (DS01-A, Shanghai Daosheng Medical Technology Company; n=398) for previous studies [17]. Image set II consists of 24 representative images chosen from image set I. Image set III consists of 34 tongue pictures taken by 17 research participants using their own smartphones according to the standardized photographic protocol, and image set IV consists of 24 representative images chosen from image set III. The assessment of tongue image features was conducted online using the questionnaire builder Qualtrics to allow anonymous contributions by different participants from different locations (Figure 1).

Standardized Photographic Protocol

Images in image set III and IV were collected using a standardized photographic protocol, which consisted of the following instructions: (1) start the phone's camera function without using any filter or artificial intelligence (AI) function; (2), turn on the flash; (3) holding the phone directly in front of the subject, aim at their lips with a 45 degree angle with the phone 15-20 cm away; (4) tap on the screen to adjust focus when the subject protrudes the tongue and then take a picture; and (5) upload the full resolution image to the online questionnaire system. A 1-minute video demonstrating the photo-taking process was made available to all participants and written instructions were provided to indicate the resolution and orientation requirements for tongue images. The resultant images were checked for focus and orientation before being accepted into the study set. The images chosen for image set III were taken with iPhone (n=18), HUAWEI (n=11), and Xiaomi (n=5) smartphones.

Study Flow

To develop the operating classification scheme for tongue fur, 2 groups of assessors in Experiment 1 independently analyzed image set I for fur features by noting the percentage of area that was covered by one of the designated colors of fur (ie, white, yellow, gray, and black), without specifying a designated fur feature. They also analyzed tongue fur textures based on the standard description by the World Health Organization (WHO) [18]. Based on the results from Experiment 1, an operating classification scheme was formed for later experiments. In Experiment 2, 18 raters determined tongue fur features in image set II based on their own interpretation first, then again based on the operating classification scheme. The inter-rater agreements of the two assessments were compared to see if the results of the working classification scheme were comparable with conventional judgements. Additional feedback from participants was also taken into consideration for improving the scheme. For efficiency reasons, we only examined tongue fur in Experiment 2 because other tongue features, such as shape and color of the tongue body, were known to be highly variable when judged by images [17,19]. In Experiment 3, the operating classification scheme was modified based on feedback and this classification scheme was used to assess the intra-rater agreement between direct subject inspection and indirect image inspection using image set III. Participants were instructed to first observe the subject directly and then record their observations of tongue features on the online questionnaire. They then took a picture of the subject's tongue according to the standardized photographic protocol described above. Following that, they uploaded the image to the online questionnaire and answered the questions to describe the tongue features again based on their observation of the image. They were specifically instructed to not refer back to the rating based on direct inspection and to instead provide a new rating based on observation of the tongue image. In Experiment 4, which was conducted 24 hours after Experiment 3, the inter-rater agreement of image set IV was assessed with the operating classification scheme finalized by feedback from Experiment 3. The intra-rater agreement on tongue images between Experiments 3 and 4 was also assessed (Figure 1).

Statistical Analysis

The data were entered into SPSS Statistics Version 25.0 (IBM), where analysis was conducted. Results were presented as Cohen k, chi-square and Gwet AC2. Cohen k was computed for intra-rater agreements and for agreement between 2 raters. When comparing agreements among multiple raters, percentage was used and chi-square was computed for statistical comparisons between groups. Furthermore, the value of Gwet AC2 was computed to indicate the extent of agreement among multiple raters [19,20].

Results

Experiment 1

Figure 2 shows the results of analysis of 2023 images by 2 groups of independent observers in Experiment 1. The percentage of area that is covered by one of the designated fur colors is shown in Figure 2A-D. We then combined the percentage of area values into two categories: those between 1% and 50% and those over 50%. As expected, the κ values for the 2 observers increased after combining the percentages (Figure 2E-H).

Based on findings from Experiment 1, an operating classification scheme using a broad description for the percentage of coverage that determined the color of tongue fur was developed; Table 1 displays the finalized version of the scheme.

Figure 2. Results of tongue coating analysis on 2023 tongue images rated by 2 assessors. A-D) The original scoring results of white, yellow, gray, and black tongue fur, respectively. E-H) The scoring results after scores were sorted into two categories (1%-50% and 51%-100%). I) The results of tongue fur textures. The *P* value of κ is <.001.

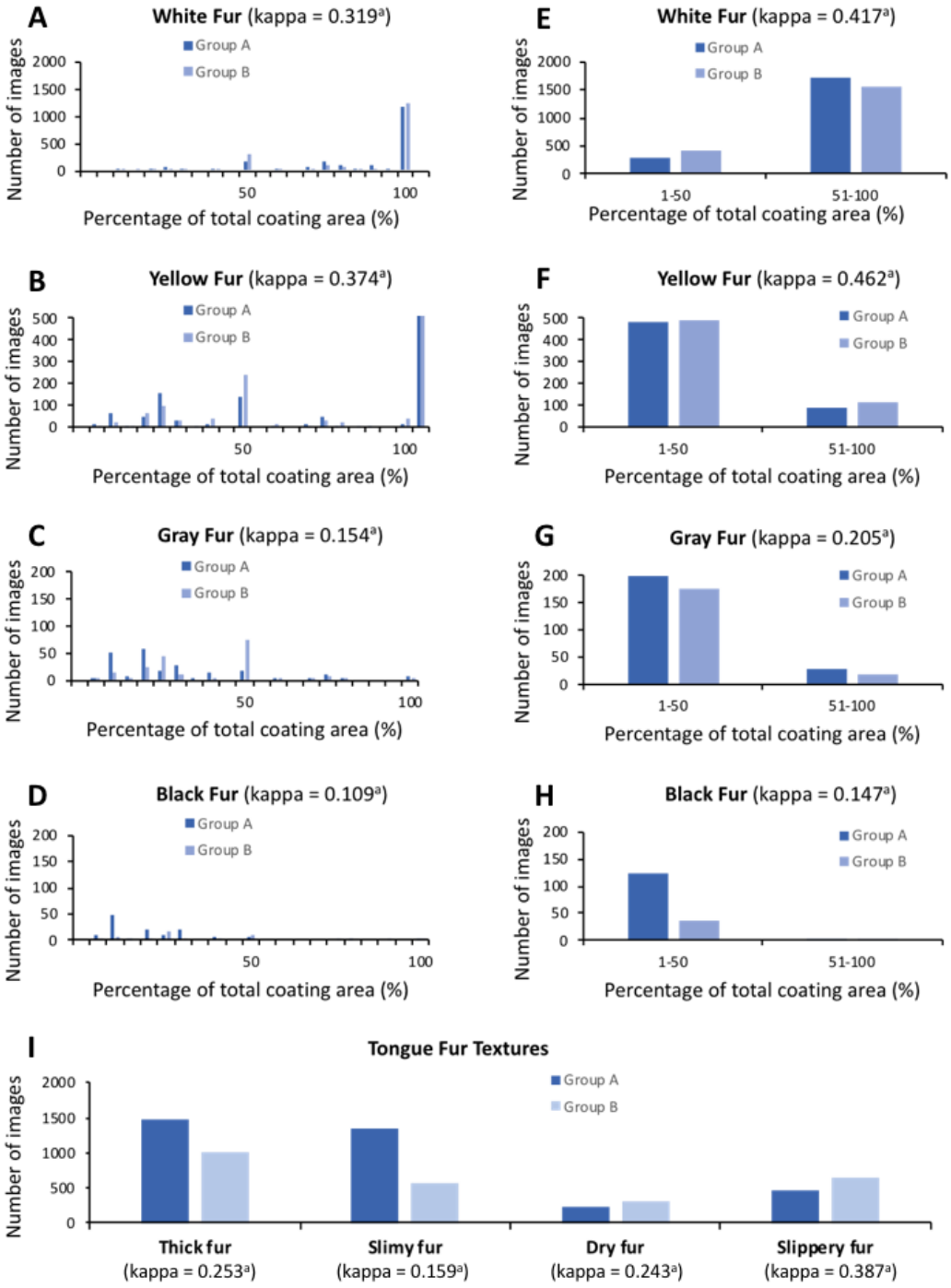


Table 1. Tongue fur feature definitions and finalized operating classification scheme.

Tongue fur features	World Health Organization (WHO) definition	Finalized operating classification scheme
White fur	Tongue coating white in color	Over 50% of the total area covered by coating is white, with the remaining coating showing no prominent coloring.
Yellow fur	Tongue coating yellow in color	Over 50% of the total area covered by coating is yellow, with the remaining coating showing no prominent coloring.
Gray fur	Tongue coating gray in color	Over 50% of the total area covered by coating is light dark, with the remaining coating showing no prominent coloring.
Black fur	Tongue coating black in color	Over 50% of the total area covered by coating is black, with the remaining coating showing no prominent coloring.
Yellow and white fur	N/A[a]	There is prominent yellow coating occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Gray and white fur	N/A	There is prominent light dark coating occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Black and white fur	N/A	There is prominent black coating occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Yellow and gray fur	N/A	There is yellow and light dark coating, with each color occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Yellow and black fur	N/A	There is yellow and black coating, with each color occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Gray and black fur	N/A	There is light dark and black coating, with each color occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Yellow, gray and black fur	N/A	There is yellow, light dark, and black coating, with each color occupying 50% or less of the total coating area, with the rest of the coating area being white or not obvious.
Mirror tongue	A completely smooth tongue free of coating, like a mirror	Absence of any coating.
Thick fur	A tongue coating through which the underlying tongue surface is not visible.	Apart from the pharyngeal part of the tongue, there is an area covered by coating through which the underlying color of tongue body is not visible.
Slippery fur	A moist tongue coating with excessive fluid, feels slippery.	The tongue surface is moist with a more than normal amount of fluid.
Dry fur	A tongue coating that looks dry and feels dry to the touch	Tongue coating is dry with a less than normal amount of fluid.
Slimy fur	A dense, turbid, slimy tongue coating, sticking on the tongue, hard to wipe off.	There are areas where the coating particles are fine, dense, evenly formed, and tightly attached to the tongue.
Thin fur	A tongue coating through which the underlying tongue surface is faintly visible	Apart from the pharyngeal part of the tongue, the tongue is covered by coating through which the underlying color of tongue body is visible.

[a]N/A: not applicable.

Experiment 2

Experiment 2 compared the results of tongue fur diagnosis based on the raters' own interpretation versus those determined using the operating classification scheme on image set II (n=24). It can be seen from Table 2 that, with the exception of black fur, thin fur, and dry fur, similar percentages of inter-rater agreement were found for both diagnostic processes (P>.05, chi-square test), which indicates that the operating classification scheme is comparable with conventional judgements.

Table 2. Inter-rater agreements according to own interpretation (Questionnaire 2) or the operating classification scheme (Questionnaire 3) on image set II (n=24).

Category	Questionnaire 2			Questionnaire 3		
Top 1 agreed features	Number of images	Average percentage of raters reaching agreement	Gwet AC2	Number of images	Average percentage of raters reaching agreement	Gwet AC2
Color of fur			**0.35**			**0.33**
White	3	56		4	60	
Yellow	4	57		9	61	
Black	2	61		1	88[a]	
Yellow and white	12	48		6	45	
Yellow and black	2	50		4	44	
Yellow and gray	1	44		1	30	
Mirror tongue	2	86		3	91	
Texture of fur			**0.39**			**0.42**
Thick	10	78		7	73	
Slippery	5	77		4	78	
Dry	4	68		N/A[b]	N/A	
Slimy	17	76		12	69	
Thin	4	82		7	66[c]	

[a]*P*=.04; chi-square test.

[b]N/A: not applicable.

[c]*P*=.01; chi-square test.

Experiment 3

We slightly modified the operating classification scheme according to comments from participants in Experiment 2, and then applied it to the assessment of direct inspection and indirect image inspection for the same subject. Several participants reported that more details of tongue features could be recognized from smartphone images than from direct inspection, which supported the use of the standardized photographic protocol. As seen in Table 3, there was very good agreement between direct inspection and image inspection for color of tongue fur, thick fur, dry fur, slimy fur, thin fur, and teeth-marked tongue (κ range 0.85-1.0), and good agreement on slippery fur and color of the tongue body (κ range 0.69-0.70). These results indicate that smartphone photos taken with the standardized photographic protocol retain many features of tongue fur.

Table 3. Intra-rater agreements between direct tongue inspection and image inspection, and between the first and second image inspection.

Tongue features	Direct and image inspection, n=34		First and second image inspection, n=24	
	K value	*P* value	K value	*P* value
Color of tongue fur	0.89	<.001	0.70	<.001
Thick fur	0.94	<.001	0.83	<.001
Slippery fur	0.70	<.001	0.10	.62
Dry fur	0.85	<.001	0.80	<.001
Slimy fur	0.88	<.001	0.75	<.001
Thin fur	0.93	<.001	1.00	<.001
Teeth-marked tongue	1.00	<.001	0.83	<.001
Color of tongue body	0.69	<.001	0.22	.06

Experiment 4

To allow the evaluation of inter-rater reliability using the finalized operating classification scheme, 24 representative images from image set III were selected to form image set IV (Table 4). The AC2 values for color of fur, texture of fur, color of tongue body, and tongue body feature were 0.49, 0.55, 0.34, 0.34, respectively. These findings indicate that there is a moderate level of inter-rater agreement on tongue fur features, but only a fair level of agreement on tongue body features.

Table 4. Inter-rater agreement based on the finalized operating classification scheme.

Category, Top 1 agreed features	Number of images	Average percentage of raters reaching agreement	Gwet AC2
Color of fur			**0.49**
White	7	65	
Yellow	6	80	
Yellow and white	4	50	
Yellow and black	3	51	
Yellow and gray	1	59	
Yellow and gray and black	1	47	
Mirror tongue	2	97	
Texture of fur			**0.55**
Thick	16	83	
Slippery	1	53	
Dry	1	82	
Slimy	16	86	
Thin	6	78	
Color of tongue body			**0.34**
Pale	4[a]	68	
Pale red	15[a]	64	
Red	7	52	
Features of tongue body			**0.34**
Teeth-marked	8	83	
Enlarged	7	71	

[a]In this instance, 2 images had the same number of raters choosing either Pale or Pale red as top 1 agreed feature in the category Color of tongue body, thus they were counted twice.

The intra-rater reliability on image set IV between the first and second assessments, separated by 24 hours or more, is shown in Table 3. It can be seen that the color of the tongue body and slippery fur showed no relationship between the assessments, whereas the other features show good to very good agreement.

Discussion

Principal Findings

To our knowledge, this is the first systematic study to evaluate intra- and inter-rater reliability of different tongue features using images from smartphones. Using a quasi-Delphi method, we developed an operating classification scheme to assess tongue fur features and then applied the scheme to direct tongue observation and indirect observation with tongue images. In addition, we have established a standardized photographic protocol for capturing tongue images with a smartphone. With the scheme and image collecting protocol in place, we found good to very good intra-rater agreement between the results of direct tongue inspection and those of smartphone images when rated immediately. This indicates that participants were generally satisfied with the quality of smartphone images in reflecting tongue features. However, when comparing the first and second assessments of the tongue images, the level of intra-subject agreement decreased, especially for the features of color of the tongue body and slippery fur. These results suggest that certain features are less reliable than others in tongue diagnosis using smartphone images. Finally, we found a moderate level of inter-rater agreement for tongue fur features and fair agreement for tongue body features.

Traditionally, tongue diagnosis is performed by direct visual inspection of the subject, but systematic evaluation of rater agreement by multiple practitioners on tongue features is difficult to achieve in clinical situations. It has been reported that agreement on overall TCM diagnoses is generally low; tongue diagnosis is one part of the criteria. Using tongue images, a number of studies examined different kinds of agreement in tongue diagnosis and the findings are quite variable, depending on the type of comparison [12,17,21]. Kim et al [21] found low levels of inter- and intra-rater practitioner agreements on tongue images and identified that inadequate operational definitions was the major contributing factor for the inconsistencies. To standardize operational definitions and procedures, in this study we developed an operating classification scheme and a standardized photographic protocol, which appear to have improved the level of inter-rater agreement, especially on tongue fur color. However, we found that the assessments of tongue body and slippery fur are unreliable, which is consistent with previous findings [15,17].

Previously, classification of tongue fur images has either been based on expert opinion or computer analysis, but no scheme has been accepted as the standard method. Zhen et al [19] recently studied 50 TCM experts and classified their diagnosis methods into 3 categories based on their internal consistency, external consistency, or both. This approach aimed to identify the experts who could provide the most consistent diagnosis. However, as the classification methods are not overtly stated, this approach remains a black-box approach. We are in the process of training AI for automatic tongue diagnosis, which requires labelling a large number of images for machine learning. For this purpose, it is not practical to find experts with consistent diagnosis skills to do the labelling task. Here we have employed a different approach, whereby we developed a simple classification scheme and evaluated the scheme as used by clinicians with different levels of experience. In so doing, we generated an operational framework and improved the definitions through feedback from participants. The advantage of our approach is that, once defined, the operating classification scheme can be used by any person with basic training in TCM tongue diagnosis to label tongue images, and the classification results are expected to produce similar inter-rater reliability.

Strengths and Limitations

There are a number of strengths as well as limitations in this study. The first strength is in its novelty with the use of smartphones without any color correction measures. It is generally believed that the use of smartphones instead of a purpose-built tongue imaging system would produce unreliable results, due to variations in factors such as camera settings, light source, and display settings [16]. However, we demonstrated, for the first time, that using our tongue photographic protocol, it is possible to obtain good quality tongue images with a good to very good level of intra-rater agreement against direct tongue inspection, and with a moderate level of inter-rater agreement on fur features. In this regard, it is interesting to note that the level of inter-rater agreement achieved using smartphone images in our study is comparable with those findings obtained through direct inspection of the tongue [15]. On the other hand, we found a reduced intra-rater reliability between the first and second assessments on tongue images, which suggests that human factors also play an important role in the variability of tongue diagnosis.

Another strength of this study is its practically oriented design. We tested the photographic protocol on a number of smartphones under uncontrolled situations, using practitioners with differing levels of clinical experience and tongue images of various fur colors. In addition, we kept both the classification scheme and the photographic protocol simple and practical. We believe that tongue images collected and evaluated in this way would have good external validity, suitable for inspection by TCM practitioners or development of a diagnostic app using AI. It is also interesting to note that a number of Delphi participants commented that, compared with routine clinical tongue inspection, more details of tongue features can be recognized within smartphone images. In fact, smartphone-captured tongue images have been used extensively in the recent TCM management of COVID-19, and it was believed that tongue coating diagnoses played an important role in the formulation of TCM treatment strategies [10]. Taken together, smartphone tongue images may be used as an initial screening in areas of telemedicine other than TCM, including common tongue conditions [22,23], diabetes [24], cancer [25], and COVID-19 [26].

One of the limitations of this study is that we were not able to assess the level of inter-rater agreement on direct tongue inspection. This would have been useful as a comparison to that of tongue image inspection. However, it was not feasible to have many assessors examine 24 subjects face-to-face repeatedly. Another limitation of this study is that the total number of tongue images in the final set (image set IV) for assessment of inter-rater agreement was relatively small. We restricted image set IV to 24 images because we received feedback from assessors that too many images could lead to mental fatigue and inaccurate labelling. For future studies, it would be advisable to separate the image set into smaller sets if a larger number of images were used.

Conclusions

By using a simple photographic protocol with smartphones, it is possible to obtain tongue images that contain many features of the tongue, especially those related to tongue coating. This study provides information on the inter-rater reliability of different tongue features, which can be used as future references for mHealth, including telemedicine by practitioners or automatic tongue diagnosis by AI. Taken together, our findings support the use of smartphone images for tongue coating analysis.

Acknowledgments

This study was supported by funds from a Hong Kong Baptist University Faculty Research Grant (FRG2/17-18/095) and a generous donation to Hong Kong Baptist University by Ms Flora Chan. The authors would like to thank Yang Xiaxin, Yeung Tsz Ching, Zhang Yumeng, Zhao Qiguang, and Anthony Yuen for their technical assistance. We would also like to thank Dr C Zaslawski from University of Technology Sydney for reviewing and editing the manuscript.

Conflicts of Interest

None declared.

References

1. Hao Y, Cheng F, Pham M, Rein H, Patel D, Fang Y, et al. A Noninvasive, Economical, and Instant-Result Method to Diagnose and Monitor Type 2 Diabetes Using Pulse Wave: Case-Control Study. JMIR mHealth uHealth 2019 Apr 23;7(4):e11959 [FREE Full text] [doi: 10.2196/11959] [Medline: 31012863]
2. Hu N, Lee C, Hou MC, Chen Y. A cloud system for mobile medical services of traditional Chinese medicine. J Med Syst 2013 Dec;37(6):9978. [doi: 10.1007/s10916-013-9978-8] [Medline: 24081559]
3. Kanawong R, Obafemi-Ajayi T, Liu D, Zhang M, Xu D, Duan Y. Tongue Image Analysis and Its Mobile App Development for Health Diagnosis. Adv Exp Med Biol 2017;1005:99-121. [doi: 10.1007/978-981-10-5717-5_5] [Medline: 28916930]
4. Jiang M, Lu C, Zhang C, Yang J, Tan Y, Lu A, et al. Syndrome differentiation in modern research of traditional Chinese medicine. J Ethnopharmacol 2012 Apr 10;140(3):634-642 [FREE Full text] [doi: 10.1016/j.jep.2012.01.033] [Medline: 22322251]
5. Jiang B, Liang X, Chen Y, Ma T, Liu L, Li J, et al. Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. Sci Rep 2012;2:936 [FREE Full text] [doi: 10.1038/srep00936] [Medline: 23226834]
6. Wang S, Xu Z, Wang Y. Tongue Characteristics of Different TCM Syndromes on Asthma. In: Proceedings of the 2016 International Conference on Biological Engineering and Pharmacy (BEP 2016).: Atlantis Press; 2016 Presented at: 2016 International Conference on Biological Engineering and Pharmacy (BEP 2016); December 9-11; Shanghai p. 202-205. [doi: 10.2991/bep-16.2017.42]
7. Hu J, Han S, Chen Y, Ji Z. Variations of Tongue Coating Microbiota in Patients with Gastric Cancer. Biomed Res Int 2015;2015:173729 [FREE Full text] [doi: 10.1155/2015/173729] [Medline: 26457297]
8. Lee T, Lo L, Wu F. Traditional Chinese Medicine for Metabolic Syndrome via TCM Pattern Differentiation: Tongue Diagnosis for Predictor. Evid Based Complement Alternat Med 2016;2016:1971295 [FREE Full text] [doi: 10.1155/2016/1971295] [Medline: 27313640]
9. Zhao Y, Gou X, Dai J, Peng J, Feng Q, Sun S, et al. Differences in metabolites of different tongue coatings in patients with chronic hepatitis B. Evid Based Complement Alternat Med 2013;2013:204908 [FREE Full text] [doi: 10.1155/2013/204908] [Medline: 23690837]
10. Yang HS, Li L, Gou CY, Zhang JY, Luo XL, Jin AH, et al. Bei Jing Di Qu Xin Xing Guan Zhuang Bing Du Fei Yan Zhong Yi Zheng Hou Ji Bing Ji Te Dian Chu Tan (A pilot exploration of TCM syndrome and pathogenesis of new coronavirus pneumonia in Beijing). Beijing Journal of Traditional Chinese Medicine in Chinese 2020 Feb 13 [FREE Full text]
11. Washio J, Sato T, Koseki T, Takahashi N. Hydrogen sulfide-producing bacteria in tongue biofilm and their relationship with oral malodour. J Med Microbiol 2005 Sep;54(Pt 9):889-895. [doi: 10.1099/jmm.0.46118-0] [Medline: 16091443]
12. Lo L, Chen Y, Chen W, Cheng T, Chiang JY. The Study on the Agreement between Automatic Tongue Diagnosis System and Traditional Chinese Medicine Practitioners. Evid Based Complement Alternat Med 2012;2012:505063 [FREE Full text] [doi: 10.1155/2012/505063] [Medline: 22924055]
13. Zhang GG, Lee W, Lao L, Bausell B, Berman B, Handwerger B. The variability of TCM pattern diagnosis and herbal prescription on rheumatoid arthritis patients. Altern Ther Health Med 2004;10(1):58-63. [Medline: 14727501]
14. Zhang GG, Lee W, Bausell B, Lao L, Handwerger B, Berman B. Variability in the traditional Chinese medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis. J Altern Complement Med 2005 Jun;11(3):415-421. [doi: 10.1089/acm.2005.11.415] [Medline: 15992224]
15. O'Brien KA, Abbas E, Zhang J, Guo Z, Luo R, Bensoussan A, et al. Understanding the reliability of diagnostic variables in a Chinese Medicine examination. J Altern Complement Med 2009 Jul;15(7):727-734. [doi: 10.1089/acm.2008.0554] [Medline: 19552599]
16. Hu M, Lan K, Fang W, Huang Y, Ho T, Lin C, et al. Automated tongue diagnosis on the smartphone and its applications. Comput Methods Programs Biomed 2019 Jun;174:51-64. [doi: 10.1016/j.cmpb.2017.12.029] [Medline: 29307471]
17. Parcus R, Chow K, Zhu H, Yu Q, Zhang S. Development of a mobile phone based tongue image acquisition system. 2018 Feb 27 Presented at: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); October 14-16; Shanghai p. 2017-2010. [doi: 10.1109/cisp-bmei.2017.8302299]
18. World Health Organization Regional Office for the Western Pacific. WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region. Manila: World Health Organization, Western Pacific Region; 2007.
19. Qi Z, Tu L, Luo Z, Hu X, Zeng L, Jiao W, et al. Tongue Image Database Construction Based on the Expert Opinions: Assessment for Individual Agreement and Methods for Expert Selection. Evid Based Complement Alternat Med 2018;2018:8491057 [FREE Full text] [doi: 10.1155/2018/8491057] [Medline: 30369958]
20. Gwet K. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters. A Handbook for Researchers, Practitioners, Teachers & Students. Gaithersburg: Advanced Analytics LLC; 2014.
21. Kim M, Cobbin D, Zaslawski C. Traditional Chinese medicine tongue inspection: an examination of the inter- and intrapractitioner reliability for specific tongue characteristics. J Altern Complement Med 2008 Jun;14(5):527-536. [doi: 10.1089/acm.2007.0079] [Medline: 18564955]
22. Reamy BV, Derby R, Bunt CW. Common tongue conditions in primary care. Am Fam Physician 2010 Mar 01;81(5):627-634 [FREE Full text] [Medline: 20187599]

23. Tiffany B, Blasi P, Catz SL, McClure JB. Mobile Apps for Oral Health Promotion: Content Review and Heuristic Usability Analysis. JMIR mHealth uHealth 2018 Sep 04;6(9):e11432 [FREE Full text] [doi: 10.2196/11432] [Medline: 30181114]
24. Tomooka K, Saito I, Furukawa S, Maruyama K, Eguchi E, Iso H, et al. Yellow Tongue Coating is Associated With Diabetes Mellitus Among Japanese Non-smoking Men and Women: The Toon Health Study. J Epidemiol 2018 Jun 05;28(6):287-291 [FREE Full text] [doi: 10.2188/jea.JE20160169] [Medline: 29311441]
25. Han S, Yang X, Qi Q, Pan Y, Chen Y, Shen J, et al. Potential screening and early diagnosis method for cancer: Tongue diagnosis. Int J Oncol 2016 Jun;48(6):2257-2264 [FREE Full text] [doi: 10.3892/ijo.2016.3466] [Medline: 27035407]
26. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int J Oral Sci 2020 Feb 24;12(1):8 [FREE Full text] [doi: 10.1038/s41368-020-0074-x] [Medline: 32094336]

Abbreviations

AI: artificial intelligence
COVID-19: coronavirus disease
mHealth: mobile health
TCM: traditional Chinese medicine
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.08.19; peer-reviewed by Y Wang, E van der Velde; comments to author 26.09.19; revised version received 14.10.19; accepted 23.03.20; published 09.07.20.

Please cite as:
Wang ZC, Zhang SP, Yuen PC, Chan KW, Chan YY, Cheung CH, Chow CH, Chua KK, Hu J, Hu Z, Lao B, Leung CC, Li H, Zhong L, Liu X, Liu Y, Liu Z, Lun X, Mo W, Siu SY, Xiong Z, Yeung WF, Zhang RY, Zhang X
Intra-Rater and Inter-Rater Reliability of Tongue Coating Diagnosis in Traditional Chinese Medicine Using Smartphones: Quasi-Delphi Study
JMIR Mhealth Uhealth 2020;8(7):e16018
URL: https://mhealth.jmir.org/2020/7/e16018
doi:10.2196/16018
PMID:32459647

Original Paper

Comparison of Geographic Information System and Subjective Assessments of Momentary Food Environments as Predictors of Food Intake: An Ecological Momentary Assessment Study

Katherine G Elliston[1], BBSc; Benjamin Schüz[2,3], PhD; Tim Albion[4], BE; Stuart G Ferguson[1], PhD

[1]College of Health and Medicine, University of Tasmania, Hobart, Australia

[2]Institute of Public Health and Nursing Science, University of Bremen, Bremen, Germany

[3]Leibniz Science Campus, Digital Public Health, Bremen, Germany

[4]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

Corresponding Author:
Katherine G Elliston, BBSc
College of Health and Medicine
University of Tasmania
17 Liverpool Street
Hobart, 7000
Australia
Phone: 61 362264259
Email: katherine.elliston@utas.edu.au

Abstract

Background: It has been observed that eating is influenced by the presence and availability of food. Being aware of the presence of food in the environment may enable mobile health (mHealth) apps to use geofencing techniques to determine the most appropriate time to proactively deliver interventions. To date, however, studies on eating typically rely on self-reports of environmental contexts, which may not be accurate or feasible for issuing mHealth interventions.

Objective: This study aimed to compare the subjective and geographic information system (GIS) assessments of the momentary food environment to explore the feasibility of using GIS data to predict eating behavior and inform geofenced interventions.

Methods: In total, 72 participants recorded their food intake in real-time for 14 days using an ecological momentary assessment approach. Participants logged their food intake and responded to approximately 5 randomly timed assessments each day. During each assessment, the participants reported the number and type of food outlets nearby. Their electronic diaries simultaneously recorded their GPS coordinates. The GPS data were later overlaid with a GIS map of food outlets to produce an objective count of the number of food outlets within 50 m of the participant.

Results: Correlations between self-reported and GIS counts of food outlets within 50 m were only of a small size (r=0.17; P<.001). Logistic regression analyses revealed that the GIS count significantly predicted eating similar to the self-reported counts (area under the curve for the receiver operating characteristic curve [AUC-ROC] self-report=0.53, SE 0.00 versus AUC-ROC 50 m GIS=0.53, SE 0.00; P=.41). However, there was a significant difference between the GIS-derived and self-reported counts of food outlets and the self-reported type of food outlets (AUC-ROC self-reported outlet type=0.56, SE 0.01; P<.001).

Conclusions: The subjective food environment appears to predict eating better than objectively measured food environments via GIS. mHealth apps may need to consider the type of food outlets rather than the raw number of outlets in an individual's environment.

(JMIR Mhealth Uhealth 2020;8(7):e15948) doi:10.2196/15948

KEYWORDS

ecological momentary assessment; mHealth; geographic information systems; food intake; mobile phone

Introduction

Background

Consistent with the notion of stimulus control, momentary environments are key correlates of a range of health-risk behaviors. For example, studies have shown associations between exposure to smoking-friendly environments and smoking [1], being in an abandoned space and illicit drug use [2] and being closer to fast-food outlets and an increase in discretionary food intake [3]. As such, being aware of what is in an individual's momentary environment could provide a means for issuing just-in-time adaptive interventions [4]. For example, when entering environments known to trigger health risk behaviors, mobile health (mHealth) technology could generate interventions and support to individuals [5], thereby minimizing the risk of engaging in health-damaging behaviors [6].

However, for effective, just-in-time, and geofenced intervention designs, it is crucial to know the components of momentary environments that are most reliably related to risk behaviors. In particular, it is an open question whether the subjective perceptions (eg, the number of food outlets an individual perceives as close by) or the objective indicators of food environments (eg, a geographic information system [GIS]–based count of the number of food outlets in a given radius around an individual) are more reliably associated with health risk behaviors, such as high-calorie snacking.

Previous real-time studies have typically favored self-reported measures, requiring a user to manually input details surrounding their affect, activities, and environment [1,3,7]. For example, many studies ask participants to indicate their current environment from several prespecified locations (eg, work, home, restaurant or bar). Intensive self-report is desirable in the context of research studies, but such monitoring is burdensome and, hence, likely to be unfeasible for the long-term usage that is necessary to achieve a lasting behavioral change. Although self-reported data might generate richer data sets, for example, by allowing researchers to gather data on unobservable psychological processes and motivations, this needs to be balanced against the possibility of missing data through noncompliance with monitoring protocols. Another option is to passively monitor an individual's environment using location, movement, or biometric sensors. In the case of location, for example, this could be achieved by combining GPS data from individual devices with GIS data, which could then be used to create targeted geofence-based mHealth interventions. Being passive, such monitoring is likely better suited for long-term monitoring than relying on self-reported information.

Passive monitoring, however, is not without its potential drawbacks. Of particular concern is that passively collected GPS data and self-reported data may capture differential aspects of the environment that might be relevant for behavior change. For example, although passive monitoring may be objectively accurate, individuals may not always be aware of—or influenced by—cues in their surrounding environment. It is possible that being actively aware of environmental cues is crucial to the initiation of health risk behaviors; therefore, passively monitoring locations may not be an appropriate way to target context-sensitive interventions. Indeed, some studies explicitly ask individuals to report on their behavioral triggers using a *cues to action* scale [8,9], thereby implying that the individuals are aware of the environmental cues that trigger their behavior. Previous environmental interventions have been shown to improve health behavior, such as food safety behaviors [10], suggesting that consciously perceived cues can trigger behaviors. However, other behaviors, such as eating, maybe prompted by the automatic processing of environmental cues, such as advertisements and brand logos [11,12]. This is consistent with stimulus control theory as it does not specifically require conscious awareness of cues. Therefore, in this study, we obtained both passive and active measures of the environment and compared the associations of both with food choices, a behavior shown to be influenced by environmental cues [3,13-15]. Comparing potentially different effects of passive and actively collected location information will allow us to examine the automatic and deliberate processing of cues that may prompt individuals to eat.

Although the role of environmental determinants on eating behavior has been previously examined [16-18], these studies typically conceptualize a static notion of the environment by relying on postcode information to calculate estimates of food outlets—which can be viewed as a proxy measure for food availability—in the neighborhood food environment corresponding to the residential address of a particular person. However, each day, people move between different neighborhoods and do not always shop in their residential areas [19,20]. Therefore, studies need to consider environmental food exposures using individuals' daily travel patterns (their *activity space* [21]). On the other hand, studies that examine fluctuating environmental exposures have captured the food environment using self-reported measures [3,21]. However, with developments in technology, it is increasingly possible to use GIS data to provide an objective measure of the environments to which individuals are exposed to throughout the day.

Objectives

As ecological momentary assessment (EMA) [22] allows for real-time assessment of an individual's environment, it might be a particularly useful technique for examining environmental exposures to food intake. This study, therefore, used EMA to examine the GPS coordinates of individuals as they go about their daily lives. As previous studies have supported the role of environmental cues prompting eating, this study aimed to extend this work by investigating whether objectively collected information on momentary environmental exposures (through automatic GPS reports) predict food intake as effectively as subjectively reported environmental cues.

Methods

Overview

This study was a part of a larger project designed to examine the relationship among attentional bias, stimulus control, and obesity and to explore BMI-related differences among individuals' eating behaviors [23]. It used EMA methods to explore the feasibility of using GIS data to predict snacking.

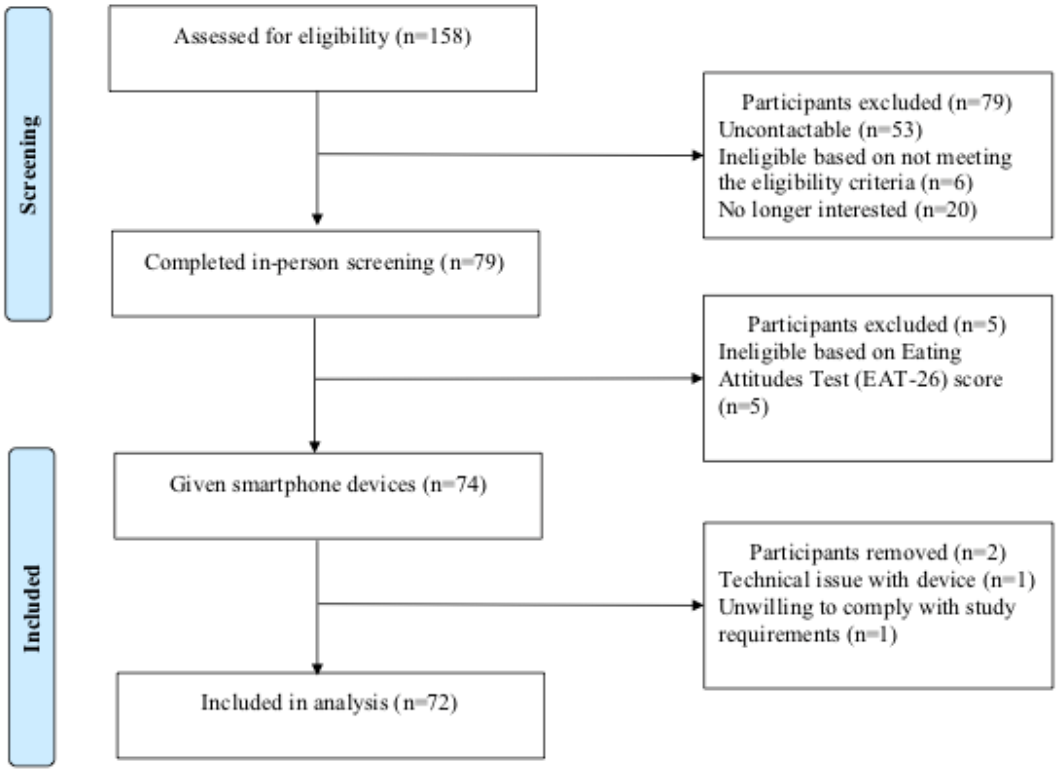
The participants carried a study-issued smartphone for 2 weeks to self-report their food and drink intake in real-time and respond to randomly timed prompts throughout the day (see Measurement Instruments below, for assessment details). During assessments, participants self-reported on environmental exposures, including describing the number and type of food retail outlets nearby. In addition to these self-reported responses, the smartphone logged the participants' GPS locations. The participants' GPS locations were then overlaid on a GIS map of known food outlets. Thus, the study obtained both objective (GIS) and self-reported information about the participants' environment at each time point. A comparison of the environments logged in the food reports with random prompts allowed for the examination of environmental cues to eating.

Participants

Seventy-nine participants were recruited for this study by looking at everyday food choices through social media advertising and a university staff newsletter in Tasmania. The eligibility criteria included being above 18 years of age, not currently dieting, and having no history of an eating disorder. BMI was stratified to obtain equal groups of participants in the healthy weight range (BMI ≥18.5-24.9) and the overweight and obese (BMI ≥25) range. Upon the completion of the study, the participants received an AUD $60 (US $39.3) shopping voucher as reimbursement for their time. Ethics approval was obtained from the Tasmanian Social Science Human Research Ethics Committee (reference number H0017015).

Five participants were excluded from the study because of screening scores exceeding 20 on the Eating Attitudes Test [24], indicating concerns regarding body weight, shape, and eating (Figure 1). In addition, 2 more participants were removed (1 participant was removed because of technical issues with his or her electronic device resulting in missing GPS stamps and 1 participant withdrew from the study). This left a total of 72 eligible participants, 71% were females (51/72; mean age 33.72 years, SD 12.08). BMI ranged from 18.59 to 40.22 (mean 26.67, SD 5.62). Most participants (62/72, 86%) were white. Over half, (43/72, 60%) of the participants had graduated from university, and 28% (20/72) participants had completed at least some university or were currently studying at a university. All participants lived in areas classified as urban [25].

Figure 1. Participant flow diagram.



Procedure

Participants attended 3 study visits during the 14-day monitoring period. During the first visit, participants provided informed consent, were weighed, and their height was measured by study staff to calculate their BMI (kg/m^2). Participants also completed a baseline survey assessing demographic information and their general dietary intake and received training on how to use the electronic diaries. Participants began recording their food intake; situational cues, such as their environment; and their affect levels immediately after this visit. During the participants' second visit (around day 2-3 of monitoring), participants' EMA data were uploaded and retraining was provided as necessary. During the third visit, after 14 days of monitoring, participants returned their study devices, were debriefed, and received reimbursement for their participation.

For the duration of the 2-week monitoring period, participants logged their food and drink intake and responded to the randomly timed prompts using a specially programmed smartphone. To reduce the participants' burden, a random

subsample (approximately 60%) of the food reports were followed by a set of questions assessing perceptions of the local food environment and contextual cues, such as the participants' affect level and food cravings. In addition to the food reports, participants were issued a series of randomly timed prompts, occurring approximately 4 to 5 times per day. During the randomly timed prompts, participants received the same assessment questions as the food reports. The randomly timed prompts served as a comparison of situational and contextual details regarding eating versus noneating times. All the participants' reports were time, date, and geographically stamped using a combination of GPS and mobile phone transmitter triangulation. Participants received an AUD $60 (US $39.3) shopping voucher upon completion of the study and the return of their EMA device, but they were not given additional payment for completing the randomly timed prompts.

Measurement Instruments

Food intake was measured via participants' self-reports. Participants reported—by tapping a button on the Android device—whether their food intake was a main meal or a snack.

Current environmental exposures were assessed via both subjective (self-reports) and objective reports (GPS stamps with subsequent GIS integration)—collected during participants' randomly timed prompts and food reports. For the *self-reported food outlets*, participants were asked to report the number of food outlets nearby. Participants were presented with the question, "From where you are now, how many food outlets can you see?" Then, they were given a list of 6 types of food outlets: (1) fast-food and takeaway stores, (2) restaurants and cafes, (3) supermarket and corner store, (4) specialty food stores, (5) discount stores, and (6) other. Participants entered a number ranging from 0 to 5+ corresponding to each type of food outlet nearby (total possible range 0-30+). For model 2 in the analysis, the total number of self-reported food outlets within sight were summarized. For model 3 in the analysis, each self-reported outlet type was dichotomized (0=absent and 1=present), and all outlet types were simultaneously entered into the model.

For the *objective measure of food outlets*, the participants' electronic devices automatically recorded their GPS coordinates every time they completed a report. The GPS location for each outlet and the participants' locations were first split into latitude and longitude coordinates. The distance between the participants and the food outlets was calculated by overlaying their GPS coordinates with a combination of 3 local city council food outlet maps using Environmental Systems Research Institute's ArcMap [26]. Local council food outlet maps were obtained, with each council providing the outlets' names, addresses, and type of each food outlet. The councils classified food outlets as being a bakery, butcher shop, café, canteen, caterer, delicatessen, eatery, fish shop, food van, hotel, meat premises, restaurant, sports club, supermarket, takeaway, vessels selling food, or other. However, the classification of food outlets was not consistent across councils, which meant that the study was unable to separate food outlets into venue types. As a result, this study used an indicator of any food outlet near participants for analyses. Using council-reported latitude and longitude coordinates of local food businesses, food outlets within a 50-m radius of a participant's GPS location were identified using the Buffer tool from the Analysis Tools Proximity toolbox [26]. The number of food outlets near a participant at the time of each report was then summarized and used in the analyses using the GIS measures.

Analytical Procedure

To examine whether passively collected GPS reports correspond to the self-reported food environment measures, a repeated measures correlation between the GIS-derived counts and self-reported counts of nearby food outlets was calculated using the R package rmcorr [27]. Next, both GPS-derived food outlet and self-reported food outlet measures were used in participant-level logistic regression models to determine if the number of food outlets within the immediate environment discriminated between eating and noneating reports. Consistent with previous studies [28], the days on which random prompt compliance was below 50% were excluded from the analysis (total 145 days). Poor random prompt compliance may indicate instances of disengagement from the study protocol or systematic biases within the data and are, therefore, removed from further analysis.

While accounting for individual differences in eating, logistic analyses were conducted on each participant's data to gauge the effect of the local food environment on food intake. First, a series of within-subject univariate logistic regression models using the area under the curve for the receiver operating characteristic curve (AUC-ROC) analyses were run. During the randomly timed assessments, each model examined if the odds of eating were higher when the density of food outlets was higher. For each model, food intake (yes or no) was the outcome variable and environmental measures (1) 50-m GIS food outlet count, (2) self-reported food outlet count, and (3) self-reported food outlet type were predictors. The study chose 50-m as it was a rough approximation of the line of sight typical for urban settings; thus, this radius was deemed as a reasonable approximation of the self-reported measure. AUC-ROC values can range from 0.5 (random guessing; no prediction) to 1.0 (perfect prediction), indicating the probability of identifying an eating event (versus a randomly timed prompt).

After generating an AUC-ROC for each participant for models 1 to 3 of the environment, the mean for each model was compared with 0.5 (ie, no predictability, at $P<.05$ threshold) using weighted *t* tests. This was used to determine the environmental measures that could accurately differentiate between eating and noneating (ie, randomly timed) assessments. Observations were weighted by the inverse of the SE of the AUC-ROC scores to allow more precise estimates to receive greater weight [29,30]. If the AUC-ROC score was significantly different from 0.5 at the $P<.05$ threshold, the model was able to accurately differentiate between eating instances and randomly timed prompts.

Next, 3 *t* tests were run to compare the food count models with each other and compare each count model with the self-reported outlet type. The *t* tests used each participant's AUC-ROC score for the comparisons. Bonferroni adjustments were applied (at $P=.02$ level) to account for the inflation of type 1 errors with multiple comparisons. Finally, the correlations between

GIS-derived measures and the self-reported measures were analyzed. This enabled the determination of the passively collected (ie, GIS-derived) environmental information was comparable with the environmental exposure information generated through self-reports. In addition, the counts between GIS-derived assessments of the food environment for both 50-m and 100-m surrounding an individual were compared, and the same basic outcomes were found. The results of the 50-m GIS count of only food outlets are presented below. All analyses were conducted in R version 3.3.1.

Results

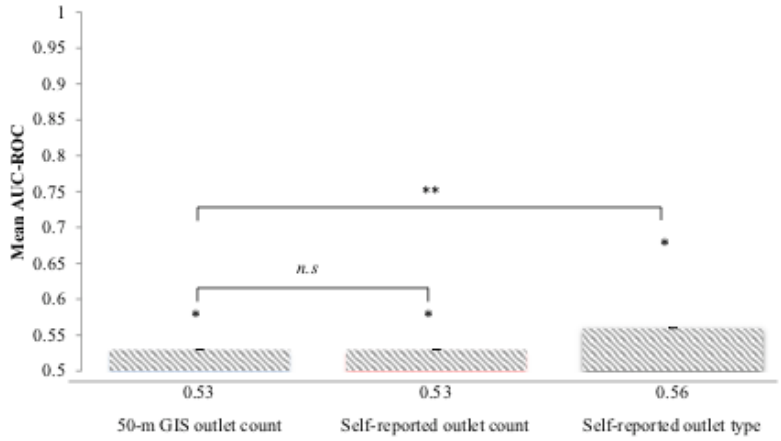
Overview

Seventy-two participants completed between 3 and 21 days of EMA monitoring and were retained in the analysis: mean 14.74 (SD 2.58) monitoring days per person. In total, 1061 days of food intake and the immediate food environment were recorded. GIS measures recorded 2097 food outlets within a 50-m radius of the participants, and the participants self-reported a total of 1756 food outlets. Over the monitoring period, participants completed 3302 food reports, and 36.86% (1217/3302) of those were snacks. Participants reported between 2 and 10 food intakes (meals and snacks) per day (mean 4.42, SD 1.47). The snack intake ranged from 1 to 8 (mean 2.02, SD 1.24) per person per day. Participants received between 0 and 11 randomly timed assessments each day (mean 3.28, SD 1.73), and the compliance with the randomly timed assessments ranged from 35% to 100%. Overall, the compliance with the randomly timed assessments was excellent [31] (mean 78.75%, SD 14.75).

Geographic Information System–Derived Measures of Food Outlets

The GIS-derived AUC-ROC values ranged from 0.50 to 0.87 and yielded similar AUC-ROC values for the self-reported food outlet count (AUC-ROC for 50-m GIS food outlet count=0.53, SE 0.00; AUC-ROC for the self-reported food outlet count=0.53, SE 0.00; Figure 2). Weighted *t* tests showed that the GIS-derived model had AUC-ROC values significantly higher than 0.50 (the null value; *P*<.001), indicating that the presence of food outlets within a 50-m radius of an individual is significantly better than chance at discriminating between eating and noneating instances.

Figure 2. Mean area under the curve for the receiver operating characteristic curve (AUC-ROC) for each measure of the local food environment. The AUC-ROC value represents the probability of accurately differentiating between eating and noneating instances. The * symbol denotes t tests, where the mean AUC-ROC was significantly different from 0.50. The ** symbol denotes significant differences in AUC-ROC values (based on t-tests with alpha set to .02). n.s denotes models where the AUC-ROC values are not significantly different. Error bars indicate the SE for each model.



Self-Reported Measures of Food Outlets

The AUC-ROC values for the self-reported count of the number of food outlets within sight ranged from 0.50 to 0.62 and had a similar AUC-ROC value for the 50-m GIS count (AUC-ROC for the self-report food outlet count=0.53, SE 0.00; AUC-ROC for the 50-m GIS food outlet count=0.53, SE 0.00; Figure 2), indicating that both measures of food outlets in the environment are significant predictors of food intake. Results from a paired sample *t* test showed no significant difference between the 50-m GIS count and the self-reported food outlet count on the participants' AUC-ROC scores: t_{71}=0.82, *P*=.41, and *d*=0.00.

The AUC-ROC for the self-reported type of food outlets in the environment ranged from 0.50 to 0.75. Model 3 showed that the self-reported type of food outlet was also a significant predictor of eating (AUC-ROC=0.56; Figure 2). A paired sample *t* test showed that there was a significant difference between the 50-m GIS model and the type of food outlets on the participants' AUC-ROC scores: t_{71}=−2.71, *P*<.001, and *d*=0.40. Similarly, there was a significant difference between the self-reported food outlet count and the type of food outlets on the participants' AUC-ROC scores: t_{71}=−5.16, *P*<.001, and *d*=0.48.

Correlations Between Environmental Measures

A repeated measures correlation between the 50-m GIS food outlet count and the self-reported food outlet count was significant but weak (*r*=0.17; *P*<.001), indicating that assessments of the food environment derived through GIS are similar to the self-reported environmental measure. As the local councils had slightly different classifications of food outlets, the study was unable to compare the self-reported types of food outlets with a GIS-derived assessment.

Discussion

Principal Findings

This study used EMA methods to compare the assessments of momentary food environment using subjective and location stamp (GPS and GIS) data. The study found that the GIS-derived counts and self-reported counts of food outlets performed worse than the self-reported type of food outlets at predicting eating. These results suggest that subjective assessments of food outlet type are better predictors of momentary food intake and that the objective and subjective counts of food outlets may capture conceptually different aspects of the food environment compared with the subjective outlet type.

The finding that the type of food outlet nearby influences eating is consistent with findings from previous literature [32,33] and is likely to be evidence that food outlet density is a proxy measure for the availability of food. For example, living within one mile of a grocery store has been associated with increased fruit and vegetable intake [18] and having numerous supermarkets in one's neighborhood is associated with lower BMI [34]. However, other types of food outlets are associated with increased unhealthy eating. For example, greater access to fast-food restaurants has been associated with a higher likelihood of fast-food purchasing [35] and a higher risk of overweight and obesity [16,34,36]. Overall, this suggests that the type of food outlet in the environment influences individuals' diet and weight. Importantly, however, much of this previous research has relied on static assessments of individuals' environments, that is, their residential addresses. This research, however, examined momentary environments, thereby accounting for the fluctuations in the environments to which individuals are exposed to throughout the day, each day.

The finding that subjectively reported food outlet counts and objectively reported food outlet counts are equally predictive of behavioral indicators (ie, eating) is novel. In the domain of physical activity, research has examined static environments and found inconsistencies between the availability and accessibility of parks to an individual and engagement in park-based physical activity [37,38]. Assessing park proximity and acceptability (ie, transport to parks, park paths or trails, and park cleanliness) differ based on whether the assessments are subjectively or objectively reported [38]. Given that the assessments of the environment are differentially associated with park-based physical activity based on the measurement type, it is possible that objectively and subjectively reported information may be tapping into conceptually different exposures. In terms of the momentary food environment as examined in this research, triggers to the craving and subsequent food intake may depend on the type of food outlet in an individual's immediate environment; such information is not captured through counts of nearby food outlets. Certain food outlets (eg, fast-food restaurants) may be more likely to trigger cravings than other food outlets (eg, supermarkets) as the sights and smells from these outlets are associated with highly palatable food [39]. Therefore, subjectively reported food outlets—specifically, the type of food outlets nearby—may be better predictors of eating than count-based assessments of the food environment.

Despite finding a small correlation between the self-reported food outlet count and the 50-m GIS count, there is minimal difference between subjective and objective measures of the number of food outlets within the environment. Overall, the results of this study suggest that the type of food outlet nearby is a better predictor of eating (versus noneating) than the density or number of food outlets. The difficulty with this is that there is a lack of standardization with the classification of food outlets. For example, an outlet can be classified as a butcher shop in one council and as a meat premise in another. For this reason, the study was unable to calculate GIS-based assessments of food outlet types. Therefore, passively collected data with subjective assessments of various food outlet types on eating could not be compared. Importantly, neither the objective nor the subject measure can be considered a truly *accurate* measure of food outlet density; both measures involve a degree of measurement error. As such, although it can be concluded that the 2 measures are aligned, the differences between the measures as evidence for over- or under-reporting of the subjective values cannot be used.

If the information on the food outlet types were measured consistently across councils, mHealth apps may be able to passively monitor an individual's location and proactively issue interventions before dietary lapses occur. This could be useful given this study's finding that subjectively reported food outlet type is a better predictor of momentary food intake than either of the count-based measures. Alternatively, mHealth apps may be able to create personalized GIS maps of environmental triggers to eating by relying, at least initially, on subjective user input. Users could report their eating locations, and the corresponding GPS reports could be used to determine the locations where the users are most likely to consume unhealthy foods. When locations are repeatedly associated with unhealthy food intake, mHealth apps could then deliver just-in-time adaptive interventions to users.

The presence of restaurants, in particular, maybe a target for mHealth dietary apps using geofencing techniques. The energy content from meals consumed at restaurants has been found to contribute to most daily energy requirements [40]; thus, the presence of restaurants may be an appropriate target to reduce daily energy intake. Furthermore, some individuals may be particularly susceptible to eating unhealthy foods only when out [41]. Although this study did not examine the within-person differences in the healthiness or energy intake derived from food intake when out, it was able to examine how eating can be prompted by cues in the immediate environment. Future studies should examine person-specific traits that increase vulnerability to unhealthy eating when out.

Overall, the findings of this study suggest that an individual's eating can be predicted based on his or her momentary environment. Although the self-reported type of food outlet nearby was the superior model in predicting eating, it only differentiated instances of eating versus noneating 56% of the time. It is possible that geofencing-based information may not be the best way to predict eating. However, research has

demonstrated a relationship between the immediate food environment and individuals' food intake; therefore, the examination of whether subjectively reported environmental information is comparable with GIS-derived data provides a starting point toward creating simple user-friendly mHealth dietary interventions.

Although using GIS data for mHealth dietary interventions passively collects data and is, therefore, less burdensome for users, it is particularly time-consuming to code, placing the burden instead on the app developers. However, once GIS data have been coded, the process of data collection becomes automated, whereas subjectively reported information will continue to require manual intervention from the user. In addition, GIS maps can be calculated once and rolled out across multiple studies and numerous sites. Such wide-scale use of location information is easier with automated GIS data than subjectively reported data. Nevertheless, the costs and benefits of each method must be balanced between users and app developers.

The finding that the overall predictive ability of the presence of food outlets on predicting eating was modest is consistent with the idiosyncratic nature of how cues come to be associated with behaviors. For example, eating could be highly related to a particular cue for one person, but different cues will be important for other people. On the basis of these findings, for relevant individuals, it may be beneficial to issue personalized dietary interventions when they enter environments where they are most at risk of overeating or unplanned eating. Indeed, similar geofenced interventions have been successfully trialed in the literature for smoking (eg, the Q-Sense app [42]). Q-Sense delivered support to users based on a 100-m geofence from a location where the user reported smoking on at least 4 occasions [42]. It appears that mHealth apps may need to rely (at least initially) on user input to create relevant geofenced risk areas and, subsequently, generate place-based interventions. Importantly, research to date demonstrates that environmental interventions are feasible, and users report no privacy concerns with location-based data monitoring [42].

Strengths, Limitations, and Future Research

This study has several strengths. To the researchers' knowledge, this study is the first to integrate 2 ways of assessing the effect of an individual's immediate food environment on his or her food choices. By using both objective measurements of the environment and subjective reports, we were able to compare how momentary environmental exposures influence real-time eating decisions. Such information provides a greater understanding of how individuals' dietary choices may be influenced by momentary environmental cues.

The use of EMA to assess eating and food environment enabled the examination of real-time environmental exposures and how they influence eating decisions. Previous studies [20,43] have highlighted the need to use spatial data to examine environmental exposures and develop precise estimates of where individuals travel and purchase foods. The use of GIS data in this study allowed for a better understanding of how fluctuations in the momentary food environment shape an individual's food choices. Furthermore, repeatedly assessing an individual's environmental exposures allows for in-depth information on the environmental antecedents and consequences of overeating and dietary lapses.

The use of real-time reporting of food intake means that the participants in this study reported their current situation, activities, and environmental exposures and were, therefore, less prone to biases associated with recall [44]. Once behaviors are examined in real-time, an effective way of managing health-risk behaviors may be through issuing just-in-time adaptive interventions [6]. Just-in-time adaptive inventions may be able to utilize real-time cues, such as GPS-based information to identify individuals entering high-risk situations that require intervention and behavioral support (eg, the A-CHESS app [45]). The real-time aspect of this study is, therefore, the first step in identifying ways to conceptualize the environment to inform just-in-time adaptive interventions and mHealth apps.

Despite these strengths, there are some methodological limitations to this study. First, calculating GIS counts of food outlets from local council areas is difficult, and the GIS data are not sufficiently detailed to illustrate what types of food outlets exist. Furthermore, the local councils included in this study had different classification systems for recording food outlets, which meant that comparison among various council districts was feasible only by looking at the summary rather than the type of food outlets. Ideally, the best way to geocode food outlets would be to use a combination of council data, Google maps data, and by visiting neighborhoods of interest to identify the type of food outlets present. Despite this being the ideal way to assess food outlets within the local environment, it would be extremely time-consuming and perhaps impractical in large cities with numerous food outlets. Future studies should explore different ways to classify the food environment so that the best and simplest measures are identified.

Second, by relying on food outlet counts (either GIS-derived or self-reported), the understanding regarding exactly what aspects of food outlets influence food choice was limited. Furthermore, this study did not separately examine the effect of each type of food outlet. Food choice is likely to be shaped by factors that are independent of food outlets, such as individual taste preferences and social norms [46], as well as the availability and affordability of foods [47], none of which are captured by assessing the counts of nearby food outlets. Further investigation into the availability and other choice determinants associated with food selection are warranted to investigate the aspects of food outlets that influence food choice.

Third, by focusing on GIS counts of food outlets, this study was unable to determine the food outlets and food-related cues that the individuals could see. There may have been times when the food outlets were in close proximity to the participants but were hidden from view. For example, there may be food outlets between buildings or hidden within lanes or buildings. As individuals' decisions relating to food choice are thought to be shaped by momentary exposures to food cues [48], in situations where individuals cannot see nearby food outlets, they are unlikely to be influenced by their presence. Future research should consider other environmental exposures, such as

advertising and food smells, in addition to the presence of food outlets in prompting individuals' food choices.

Fourth, as noted earlier, this study chose 50 m as the radius as it was a rough approximation of the line of sight typical for urban settings. What someone can see from their current position will vary from place to place; this would have introduced error into this measure. Further work is required to determine the *optimal* unit of measurement; furthermore, it may prove fruitful to vary this measure from location to location based on the characteristics of the site.

Finally, the food outlet data from the local councils may not have been up-to-date. It is possible that there may have been a discrepancy between the GIS-derived food outlet count and the food outlets that were around and open during the time the study was conducted. Encapsulating the most recent and accurate information on the presence of food outlets is necessary to examine the association between the presence of food outlets and eating. Furthermore, this study did not consider the availability of food within each outlet. Factors like product availability and opening hours are likely to influence individuals' food options and, subsequently, their eating decisions. mHealth apps that require user input on environmental eating triggers will likely circumvent this issue. At present, mHealth interventions are unable to achieve targeted place–based information with passively collected data.

Conclusions

Examining the food outlets within one's environment is an important step in understanding how the built environment influences eating. This study found that although passively knowing an individual's environment can predict eating, knowing what type of food outlets are nearby is the best way for mHealth apps to create geofenced dietary interventions. Future advances in technology may enable passive calculation of the type of food outlets within a given geographical region. Such information would be integral to the success of geofenced interventions in mHealth dietary apps. In the meantime, mHealth apps will likely need to continue relying on users' self-reported information about their food environment to generate tailored geofenced dietary interventions.

Authors' Contributions

KE contributed to idea formulation, data analysis, and manuscript development. BS contributed to idea formulation and manuscript refinement. TA worked on GIS development and analysis of food outlets and contributed to manuscript refinement. SF contributed to idea formulation, data analysis, and manuscript refinement. All authors have reviewed and approved the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Jahnel T, Ferguson SG, Shiffman S, Thrul J, Schüz B. Momentary smoking context as a mediator of the relationship between SES and smoking. Addict Behav 2018 Aug;83:136-141 [FREE Full text] [doi: 10.1016/j.addbeh.2017.12.014] [Medline: 29273313]
2. Linas BS, Latkin C, Westergaard RP, Chang LW, Bollinger RC, Genz A, et al. Capturing illicit drug use where and when it happens: an ecological momentary assessment of the social, physical and activity environment of using versus craving illicit drugs. Addiction 2015 Feb;110(2):315-325 [FREE Full text] [doi: 10.1111/add.12768] [Medline: 25311241]
3. 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 Apr;36(4):337-345. [doi: 10.1037/hea0000439] [Medline: 27669177]
4. Naughton F. Delivering 'just-in-time' smoking cessation support via mobile phones: current knowledge and future directions. Nicotine Tob Res 2017 Mar;19(3):379-383. [doi: 10.1093/ntr/ntw143] [Medline: 27235703]
5. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
6. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
7. Duresso SW, Bruno R, Matthews AJ, Ferguson SG. Khat withdrawal symptoms among chronic khat users following a quit attempt: an ecological momentary assessment study. Psychol Addict Behav 2018 May;32(3):320-326. [doi: 10.1037/adb0000368] [Medline: 29771560]
8. Booker L, Mullan B. Using the temporal self-regulation theory to examine the influence of environmental cues on maintaining a healthy lifestyle. Br J Health Psychol 2013 Nov;18(4):745-762. [doi: 10.1111/bjhp.12015] [Medline: 23279265]
9. Todd J, Mullan B. The role of self-monitoring and response inhibition in improving sleep behaviours. Int J Behav Med 2014 Jun;21(3):470-477. [doi: 10.1007/s12529-013-9328-8] [Medline: 23813124]

10. Mullan B, Allom V, Fayn K, Johnston I. Building habit strength: a pilot intervention designed to improve food-safety behavior. Food Res Int 2014 Dec;66:274-278. [doi: 10.1016/j.foodres.2014.09.027]
11. Cohen DA, Babey SH. Contextual influences on eating behaviours: heuristic processing and dietary choices. Obes Rev 2012 Sep;13(9):766-779 [FREE Full text] [doi: 10.1111/j.1467-789X.2012.01001.x] [Medline: 22551473]
12. Kremers SP, de Bruijn GJ, Visscher TL, van Mechelen W, de Vries NK, Brug J. Environmental influences on energy balance-related behaviors: a dual-process view. Int J Behav Nutr Phys Act 2006 May 15;3:9 [FREE Full text] [doi: 10.1186/1479-5868-3-9] [Medline: 16700907]
13. Pitt E, Gallegos D, Comans T, Cameron C, Thornton L. Exploring the influence of local food environments on food behaviours: a systematic review of qualitative literature. Public Health Nutr 2017 Sep;20(13):2393-2405. [doi: 10.1017/S1368980017001069] [Medline: 28587691]
14. Cannuscio CC, Tappe K, Hillier A, Buttenheim A, Karpyn A, Glanz K. Urban food environments and residents' shopping behaviors. Am J Prev Med 2013 Nov;45(5):606-614. [doi: 10.1016/j.amepre.2013.06.021] [Medline: 24139774]
15. Trapp GS, Hickling S, Christian HE, Bull F, Timperio AF, Boruff B, et al. Individual, social, and environmental correlates of healthy and unhealthy eating. Health Educ Behav 2015 Dec;42(6):759-768. [doi: 10.1177/1090198115578750] [Medline: 25842383]
16. Li F, Harmer P, Cardinal BJ, Bosworth M, Johnson-Shelton D. Obesity and the built environment: does the density of neighborhood fast-food outlets matter? Am J Health Promot 2009;23(3):203-209 [FREE Full text] [doi: 10.4278/ajhp.071214133] [Medline: 19149426]
17. Thornton LE, Kavanagh AM. Association between fast food purchasing and the local food environment. Nutr Diabetes 2012 Dec 3;2:e53 [FREE Full text] [doi: 10.1038/nutd.2012.27] [Medline: 23208414]
18. Zenk SN, Lachance LL, Schulz AJ, Mentz G, Kannan S, Ridella W. Neighborhood retail food environment and fruit and vegetable intake in a multiethnic urban population. Am J Health Promot 2009;23(4):255-264 [FREE Full text] [doi: 10.4278/ajhp.071204127] [Medline: 19288847]
19. Chaix B, Kestens Y, Perchoux C, Karusisi N, Merlo J, Labadi K. An interactive mapping tool to assess individual mobility patterns in neighborhood studies. Am J Prev Med 2012 Oct;43(4):440-450. [doi: 10.1016/j.amepre.2012.06.026] [Medline: 22992364]
20. Thornton LE, Crawford DA, Lamb KE, Ball K. Where do people purchase food? A novel approach to investigating food purchasing locations. Int J Health Geogr 2017 Mar 7;16(1):9 [FREE Full text] [doi: 10.1186/s12942-017-0082-z] [Medline: 28270150]
21. Zenk SN, Schulz AJ, Matthews SA, Odoms-Young A, Wilbur J, Wegrzyn L, et al. Activity space environment and dietary and physical activity behaviors: a pilot study. Health Place 2011 Sep;17(5):1150-1161 [FREE Full text] [doi: 10.1016/j.healthplace.2011.05.001] [Medline: 21696995]
22. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
23. Franja S, Ferguson S, Matthews A, Elliston K. Center for Open Science. FoodChoices 3: Exploring the relationship between attentional bias, stimulus control and obesity URL: https://osf.io/pmxbj/ [accessed 2020-05-21]
24. Garner DM, Garfinkel PE. The eating attitudes test: an index of the symptoms of anorexia nervosa. Psychol Med 1979 May;9(2):273-279. [doi: 10.1017/s0033291700030762] [Medline: 472072]
25. Australian Bureau of Statistics. 2011. 1270.0.55.004 - Australian Statistical Geography Standard (ASGS): Volume 4 - Significant Urban Areas, Urban Centres and Localities, Section of State, July 2016 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.004 [accessed 2016-12-15]
26. ArcGIS Desktop: Documentation. 2019. ArcMap URL: http://desktop.arcgis.com/en/arcmap/ [accessed 2019-07-23]
27. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol 2017;8:456 [FREE Full text] [doi: 10.3389/fpsyg.2017.00456] [Medline: 28439244]
28. 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]
29. Shiffman S, Dunbar MS, Ferguson SG. Stimulus control in intermittent and daily smokers. Psychol Addict Behav 2015 Dec;29(4):847-855 [FREE Full text] [doi: 10.1037/adb0000052] [Medline: 25706335]
30. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997 Jul;30(7):1145-1159. [doi: 10.1016/S0031-3203(96)00142-2]
31. Schüz N, Walters JA, Frandsen M, Bower J, Ferguson SG. Compliance with an EMA monitoring protocol and its relationship with participant and smoking characteristics. Nicotine Tob Res 2014 May;16(Suppl 2):S88-S92. [doi: 10.1093/ntr/ntt142] [Medline: 24052500]
32. Lucan SC, Mitra N. Perceptions of the food environment are associated with fast-food (not fruit-and-vegetable) consumption: findings from multi-level models. Int J Public Health 2012 Jun;57(3):599-608. [doi: 10.1007/s00038-011-0276-2] [Medline: 21773839]
33. Sanchez-Flack JC, Anderson CA, Arredondo EM, Belch G, Martinez ME, Ayala GX. Fruit and vegetable intake of us Hispanics by food store type: findings from NHANES. J Racial Ethn Health Disparities 2019 Feb;6(1):220-229 [FREE Full text] [doi: 10.1007/s40615-018-0517-7] [Medline: 30027480]

34. Morland K, Roux AV, Wing S. Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. Am J Prev Med 2006 Apr;30(4):333-339. [doi: 10.1016/j.amepre.2005.11.003] [Medline: 16530621]
35. He M, Tucker P, Gilliland J, Irwin JD, Larsen K, Hess P. The influence of local food environments on adolescents' food purchasing behaviors. Int J Environ Res Public Health 2012 Apr;9(4):1458-1471 [FREE Full text] [doi: 10.3390/ijerph9041458] [Medline: 22690205]
36. Kestens Y, Lebel A, Chaix B, Clary C, Daniel M, Pampalon R, et al. Association between activity space exposure to food establishments and individual risk of overweight. PLoS One 2012;7(8):e41418 [FREE Full text] [doi: 10.1371/journal.pone.0041418] [Medline: 22936974]
37. Cerin E, Conway TL, Adams MA, Barnett A, Cain KL, Owen N, et al. Objectively-assessed neighbourhood destination accessibility and physical activity in adults from 10 countries: an analysis of moderators and perceptions as mediators. Soc Sci Med 2018 Aug;211:282-293 [FREE Full text] [doi: 10.1016/j.socscimed.2018.06.034] [Medline: 29966823]
38. Zhang R, Wulff H, Duan Y, Wagner P. Associations between the physical environment and park-based physical activity: a systematic review. J Sport Health Sci 2019 Sep;8(5):412-421 [FREE Full text] [doi: 10.1016/j.jshs.2018.11.002] [Medline: 31534816]
39. Ferriday D, Brunstrom JM. 'I just can't help myself': effects of food-cue exposure in overweight and lean individuals. Int J Obes (Lond) 2011 Jan;35(1):142-149. [doi: 10.1038/ijo.2010.117] [Medline: 20548302]
40. Roberts SB, Das SK, Suen VM, Pihlajamäki J, Kuriyan R, Steiner-Asiedu M, et al. Measured energy content of frequently purchased restaurant meals: multi-country cross sectional study. Br Med J 2018 Dec 12;363:k4864 [FREE Full text] [doi: 10.1136/bmj.k4864] [Medline: 30541752]
41. Spanakis G, Weiss G, Boh B, Lemmens L, Roefs A. Machine learning techniques in eating behavior e-coaching. Pers Ubiquit Comput 2017 Jun 8;21(4):645-659. [doi: 10.1007/s00779-017-1022-4]
42. Naughton F, Hopewell S, Lathia N, Schalbroeck R, Brown C, Mascolo C, et al. A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR Mhealth Uhealth 2016 Sep 16;4(3):e106 [FREE Full text] [doi: 10.2196/mhealth.5787] [Medline: 27637405]
43. Jiao J, Moudon AV, Kim SY, Hurvitz PM, Drewnowski A. Health implications of adults' eating at and living near fast food or quick service restaurants. Nutr Diabetes 2015 Jul 20;5:e171 [FREE Full text] [doi: 10.1038/nutd.2015.18] [Medline: 26192449]
44. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
45. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
46. Lytle LA, Sokol RL. Measures of the food environment: a systematic review of the field, 2007-2015. Health Place 2017 Mar;44:18-34. [doi: 10.1016/j.healthplace.2016.12.007] [Medline: 28135633]
47. Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? Int J Behav Nutr Phys Act 2006 Jan 25;3:2 [FREE Full text] [doi: 10.1186/1479-5868-3-2] [Medline: 16436207]
48. Bailey RL. Influencing eating choices: biological food cues in advertising and packaging alter trajectories of decision making and behavior. Health Commun 2017 Oct;32(10):1183-1191. [doi: 10.1080/10410236.2016.1214222] [Medline: 27589082]

Abbreviations

AUC-ROC: area under the curve for the receiver operating characteristic curve
EMA: ecological momentary assessment
GIS: geographic information system
mHealth: mobile health

Edited by G Eysenbach; submitted 21.08.19; peer-reviewed by S Smith , DR Wahl; comments to author 23.10.19; revised version received 18.01.20; accepted 26.04.20; published 22.07.20.

Please cite as:
Elliston KG, Schüz B, Albion T, Ferguson SG
Comparison of Geographic Information System and Subjective Assessments of Momentary Food Environments as Predictors of Food Intake: An Ecological Momentary Assessment Study
JMIR Mhealth Uhealth 2020;8(7):e15948
URL: https://mhealth.jmir.org/2020/7/e15948
doi:10.2196/15948
PMID:32706728

Original Paper

Wearable Technology to Quantify the Nutritional Intake of Adults: Validation Study

Sarah M Dimitratos[1], BSc, RD; J Bruce German[1], DPhil; Sara E Schaefer[1], DPhil

Foods for Health Institute, University of California, Davis, CA, United States

Corresponding Author:
Sara E Schaefer, DPhil
Foods for Health Institute
University of California
2141 Robert Mondavi Institute, North Building
1 Shields Ave
Davis, CA, 95616
United States
Phone: 1 530 574 0797
Email: seschaefer@ucdavis.edu

Abstract

Background: Wearable and mobile sensor technologies can be useful tools in precision nutrition research and practice, but few are reliable for obtaining accurate and precise measurements of diet and nutrition.

Objective: This study aimed to assess the ability of wearable technology to monitor the nutritional intake of adult participants. This paper describes the development of a reference method to validate the wristband's estimation of daily nutritional intake of 25 free-living study participants and to evaluate the accuracy (kcal/day) and practical utility of the technology.

Methods: Participants were asked to use a nutrition tracking wristband and an accompanying mobile app consistently for two 14-day test periods. A reference method was developed to validate the estimation of daily nutritional intake of participants by the wristband. The research team collaborated with a university dining facility to prepare and serve calibrated study meals and record the energy and macronutrient intake of each participant. A continuous glucose monitoring system was used to measure adherence with dietary reporting protocols, but these findings are not reported. Bland-Altman tests were used to compare the reference and test method outputs (kcal/day).

Results: A total of 304 input cases were collected of daily dietary intake of participants (kcal/day) measured by both reference and test methods. The Bland-Altman analysis had a mean bias of −105 kcal/day (SD 660), with 95% limits of agreement between −1400 and 1189. The regression equation of the plot was $Y=-0.3401X+1963$, which was significant ($P<.001$), indicating a tendency for the wristband to overestimate for lower calorie intake and underestimate for higher intake. Researchers observed transient signal loss from the sensor technology of the wristband to be a major source of error in computing dietary intake among participants.

Conclusions: This study documents high variability in the accuracy and utility of a wristband sensor to track nutritional intake, highlighting the need for reliable, effective measurement tools to facilitate accurate, precision-based technologies for personal dietary guidance and intervention.

(JMIR Mhealth Uhealth 2020;8(7):e16405) doi:10.2196/16405

KEYWORDS

wearable technology; mobile health; mobile phone; food intake; validation study

Introduction

Diet and health guidelines are based on preventing or treating illnesses in the general population. Technological advances and enhanced understanding of systems biology are guiding scientists to pursue personalized interventions for disease prevention and treatment. As scientists are quantifying the elasticity of human health and its diversity, opportunities to intervene in human health are broadening to include precision control of phenotypic performance. Precision or personalized health is the approach of using quantified information on individual characteristics to develop tailored products and services aimed at guiding the underlying processes of health

[1-5]. The breadth of precision interventions includes the measurement of individuals' characteristics, genetics, immunity, metabolism, physiology, medical history, and more [6-8]. Personalizing the content and delivery of approaches also require alignment with individuals' behaviors, preferences, goals, and barriers to modification as an integral aspect of achieving lasting behavior change [2,9,10].

Precision health is made possible by modern tools, technologies, and platforms that provide increasingly diverse, mechanistic, and accurate assessments of the human body [11,12]. Health measurement research encompasses the breadth of phenotypic differences between individuals that contribute to health status. Advancements in the -omics sciences highlight how many factors individually and interactively affect health, including genetics, lifestyle, life stage, diet, and microbial diversity. Many health metrics are assessed statically, but others must be captured dynamically using specific challenges, such as with insulin sensitivity and acquired immunity. These scientific breakthroughs are guiding the development of measurement technologies that interrogate individuals beyond disease diagnostics, including mobile and wearable body sensors that enable more spatially and temporally specific measures of a broader range of phenotypic factors [4,13-15].

The most important change in the science of diet and health is as much philosophical as mechanistic. The focus of nutrition research is shifting from the study of individual foods and ingredients and their effects on entire populations to the study of individual humans in response to entire diets.

Precision Nutrition: Challenges and Breakthroughs

Bringing accurate health monitoring technologies to the market provides a public service that reduces people's uncertainty about how day-to-day choices affect their individual health [15,16]. More precise and predictive dietary guidance follows the understanding in nutritional sciences that identifying the single *best diet* for human health is no longer scientifically defensible. It is now understood that different people respond differently to foods and nutrients, warranting personalized approaches to nutrition interventions and services [3,17-21]. National dietary guidelines are intended to prevent deficiency and maintain health for the majority of the population. Using evidence-based science to create diets for individuals requires an understanding of what humans share with regard to dietary needs as well as how, when, and why needs differ. On a fundamental level, all people require a diet sufficient in calories to support normal body weight and all essential nutrients to support life. However, nutrient requirements to prevent deficiency and sustain life are just the first step in understanding the role of diet in human health [22].

A fundamental challenge in nutrition research is the accurate quantification of food intake and its interpretation as precise diet quality. Currently, the gold standard of dietary assessment is the 24-hour in-patient study, yet major limitations include cost, reduced physical activity, boredom, depression, and weight loss because of reduced dietary freedom and food options deviating from one's personal routine. In epidemiologic and clinical nutrition, dietary assessment typically relies on researcher-facilitated or autonomous participant recall using methods such as 24-hour recall, food frequency questionnaires, and food diary inventories. These memory-based assessment methods have demonstrated poor validity because of human under- or overestimation of intake and intentional or unintentional alteration of intake patterns [23]. Each traditional assessment method is a reflection of the individual's perceived intake rather than an accurate measure of true intake. Furthermore, such assessment methods are nonfalsifiable, as what the participant reports must be accepted as truth, despite knowledge of likely incongruence. Moreover, other assessment methods rely on photograph analysis of foods consumed, conducted either by trained personnel or software analysis [24]. Although more closely reflective of nutrient intake in a free-living situation, the remote food photography method is still limited by the inability to record in *true* real time, difficulty in estimating portion sizes, the necessity for simultaneous use of a backup analysis method, difficulty analyzing culturally unique foods, and analyzing mixed dishes via photographs alone. The United States Department of Agriculture (USDA) Food Composition Database is the gold standard for nutrient analysis; albeit comprehensive in scope, this tool cannot possibly account for inherent variability in climate, soil quality, geographic location, item ripeness, and cooking method, all of which may significantly alter the nutrient composition of food. Even nutrient and energy quantification by way of nutrition facts label analysis is error prone, as the Food and Drug Administration allows for certain margins of error in nutrient reporting on packaged food labels. Nutrition fact labels, therefore, provide an educated estimation of packaged food nutrient content. Consumer-focused dietary tracking methods use databases that are often crowdsourced. These errors, compounded with the aforementioned human misreporting of dietary intake, demonstrate that more precise methods of dietary assessment and analysis are needed.

Standard approaches for recording dietary intake do not account for inherent nutrient losses in absorption and metabolism, the transformative processes by which food becomes usable energy for the body. Realistic and precise quantitative assessment remains challenging because of energy losses involved at every step of transforming a food matrix into bioavailable energy: absorption, distribution, metabolism, and excretion. The rate of breakdown and net usable energy vary depending on macronutrient composition (ie, a mixed meal high in fiber, protein, and fat will digest much more slowly than a meal high in simple carbohydrates) [25]. Furthermore, interindividual differences in metabolic rate, gastrointestinal health, and previous meals consumed all contribute to discrepancies between measured intake and bioavailable energy.

Emerging commercial and medical technologies designed to detect a person's physiological fluctuations claim to capture more dynamic aspects of cardiometabolic health [14]. For example, continuous glucose monitors are designed to provide more precise tracking of glucose levels for diabetic patients compared with standard blood sampling methods, the goal being to more precisely guide disease management [26,27]. No technologies are available that can effectively assess dietary intake directly, although some methods are claiming the ability to estimate dietary intake by assessing the physiological response of the body to food intake and bioavailable energy. In

all cases, rigorous testing is necessary to determine the accuracy, precision, utility, and validity of candidate devices. We sought to answer the question, "can wearable technologies measure aspects of metabolic performance and cardiometabolic health of a normal range of adult human phenotypes?" The objectives of this paper were to describe (1) the development and implementation of a reference method to estimate the nutritional intake of free-living study participants and (2) the accuracy and utility of a wristband technology for tracking nutritional intake (kcal/day).

Methods

Overview

A study was designed to assess the ability of wearable technology to estimate the nutritional intake of individuals. The wristband (GoBe2; Healbe Corp) intends to provide users with automatic tracking of daily energy intake (calories) and macronutrient intake (grams of protein, fat, and carbohydrates). The technology uses computational algorithms to convert bioimpedance signals into measured patterns of extracellular and intracellular fluids associated with the influx of glucose and essential nutrients into the body. From changes in fluid concentration, the technology estimates calories congruent with glucose absorption into the bloodstream. Time series data such as these, which capture postprandial processes, have the potential to inform phenotypic discernment of digestion, absorption, metabolism of foods, and their influence on health.

A sample of free-living adult participants (N=25) was sought to validate the technology over 2 data collection periods of 14 days each (28 days total). A reference method was designed to measure dietary intake; all meals were prepared, calibrated, and served at a campus dining facility and consumed under the direct observation of a trained research team. Approval for the research study and protocol was obtained from the University of California, Davis (UC Davis), institutional review board.

Participants

Participants aged 18 to 50 years were recruited from the UC Davis campus using emails and flyers. Those interested were screened by phone for inclusionary and exclusionary criteria. The exclusion criteria included historical or current diagnosis of chronic disease (including diabetes or prediabetes, cancer, asthma, hypertension, cardiovascular disease, stroke, kidney, thyroid, or autoimmune disease), known food allergies, current dieting or restricted dietary habits (ie, vegetarian, ketogenic, reduced calorie), pregnancy or lactation, smoking, drug or alcohol addiction, excessive exercise or athletic training, and taking medications impacting digestion or metabolism. In-person screenings were conducted at the Ragle Human Nutrition Center on the UC Davis campus. Participants who qualified after the phone screening were invited for in-person screening to complete a fasting blood draw, blood pressure, and anthropometric measurements. Copies of approved, signed consent forms were obtained from all participants at screening. All female participants completed urine pregnancy tests. Blood pressure measurements were obtained using a Nellcor pulse oximeter with OxiMax technology from Welch Allyn. For anthropometry, a digital scale by Scale-Tronix was used to weigh participants to the nearest 0.1 kg, and a wall stadiometer was used to measure height to the nearest 0.1 cm. Anthropometric measurements were used to calculate baseline BMI (weight [kg]/[height (m)]^2). As the wristband was intended to measure nutrient intake in a weight-stable population over the study duration, individuals with fluctuating weight (>5 lbs over the previous month) were excluded. All anthropometric measurements were conducted by the principal investigator (PI) using methods defined in the anthropometric standardization reference manual [28]. Participants were assigned a study ID on enrollment, and all data collected were maintained private and deidentified. Monetary compensation was offered to each participant who completed the screening (US $10), phase 1 (US $125), and phase 2 (US $150).

For metabolic screening, blood was drawn into ethylenediaminetetraacetic acid and plasma separation lithium heparin blood collection tubes and immediately placed on ice. Within 2 hours of collection, blood samples were centrifuged at 1800×g for 15 min at 4°C to separate blood from plasma and frozen at −20°C until laboratory analyses were performed in batches. Blood samples were analyzed, and individuals who tested abnormally for metabolic health indicators including complete blood count, fasting blood glucose, hemoglobin A_{1c}, erythrocyte sedimentation rate, serum protein, creatine, alkaline phosphatase, potassium, and carbon dioxide were excluded. Tests were performed according to the manufacturer's instructions and quality controls by UC Davis Health System Medical Diagnostics.

Between August 2018 and September 2018, 76 adults were screened, and 35 met the inclusion criteria for enrollment in phase 1 of the study that would take place from September 25 to October 9, 2018. The initial sample included 20 women and 15 men, with an ethnic distribution of 38% white, 41% Asian, and 21% Hispanic, an average age of 25.3 (SD 6.4) years, and a mean BMI of 24.2 (SD 5.1) kg/m^2. Three participants dropped out during the first week of phase 1 because of time constraints that prohibited multiple visits to the campus dining facility each day. Phase 1 was completed by 32 participants, of which 24 enrolled in phase 2 (October 30 to November 13, 2018). During phase 2, 2 participants completed 10 of 14 days because of scheduling conflicts and were included in the analyses.

Data Collection

Participants were assigned a GoBe2 (Figure 1) and instructed to use the latest version of the accompanying app synchronized to the wrist unit. The technology translates sensor signals into energy intake and expenditure outputs over a 24-hour period, in accordance with the rate of nutrient absorption. Participants received an explanation on how the wristband estimates personal calorie intake and expenditure throughout the day and over the week as well as its other functions, including heart rate, sleep, hydration, and stress measurement. Participants were instructed to synchronize the wrist unit with the app twice daily, in the morning and at night, and to collect screenshots from within the app that captured the previous day's final energy (kcal) estimations. The screenshots were collected by research staff as records of daily caloric outputs, including daily intake, expenditure, and total balance.

Figure 1. Healbe GoBe2 smartband.

Quantification of Dietary Intake

A reference method was developed to quantify the daily food, calorie, and macronutrient intake of participants during the 2 study periods. The project team collaborated with UC Davis Dining Commons (DC), a series of dining facilities where campus residents primarily eat but are also open to the campus community and public. A strategy was developed to carry out the nutrition study within the university dining facility. In this approach, a specific *project menu* was created in coordination with the facility's existing cycle menu serving all dining patrons. In this way, the dining facility's normal operations were minimally perturbed, and the study team used the facility's existing food prepared in accordance with standardized recipes from which nutritional information was readily derived. Meal cards were purchased for study participants and swiped on their arrival at each meal to deduct the meal price from the card. Student research assistants were trained to carry out food measurement at each meal, nutrient analysis, and data entry.

Menu Planning

A registered dietitian (RD) on the research team collaborated with the dining facility's primary chef to design the project menu. Menu items were selected to serve to study participants at breakfast, lunch, and dinner, using the following criteria: balanced macronutrients at each meal per USDA MyPlate guidelines and minimal multi-ingredient *mixed* dishes (ie, no casseroles, lasagna, pizza, etc). Mixed dishes were avoided to reduce error in calculating calories and macronutrients that were served at each meal. When necessary, menu modifications were requested to fit the study menu criteria (ie, sauces served on the side and sandwich ingredients served separately). Separating ingredients allowed the staff to weigh foods more precisely and calculate energy and macronutrient profiles accordingly.

Energy and Macronutrient Analysis for Onsite and Offsite Food Consumption

Overall, 2 research staff were trained and designated to analyze each project menu item for energy (kcal) and macronutrient content. Items were analyzed using a combination of the USDA Food Composition Database and the dining facility's nutritional database. In the latter, menu items were previously analyzed and recorded by the DC's RD using either product nutritional labels (when available) or the USDA Food Composition Database. Each menu item was analyzed for serving size, calories, grams of protein, fat, and carbohydrate content per serving and scaled to 100 g.

The RD determined a *standard* serving size for each menu item (eg, 1 cup cooked oats, 1 cup vegetables, half cup beans, 4 oz lean protein, or three-fourth cup grain). Participants were not restricted to the standard serving sizes and were free to request more or less food portions to meet their individual dietary needs. All deviations from standard portions were recorded by the research staff for each participant.

The primary chef coordinated study meals according to the study menu preference. Each meal was prepared in a commercial kitchen on the UC Davis campus by trained food service personnel following a stringent hazard analysis critical control points (HACCP) protocol. All food was delivered to the designated research study area of the facility and received by a team of research staff for onsite portioning and serving to study participants. The study leads inspected each delivered menu item for accuracy, noting any deviations as needed.

Study participants arrived at the dining facility during scheduled breakfast, lunch, and dinner mealtimes. On arrival, they were greeted by research staff, and the meal was paid for at the door using preloaded meal cards. Each morning at breakfast, research staff collected daily information from participants, including paper records of offsite foods consumed in the previous 24 hours and details of wristband use (charging, removals, and reported problems). A brief daily in-person interview was conducted each morning to collect details on exercise, any perceived stress, water intake, defecation, and continuous glucose monitoring (CGM) skin contact in the previous 24 hours. At each meal, participants could request either the standard meal offering or certain menu items in more or fewer portions according to

preference. The participants' meal choices were recorded on paper meal slips that were delivered to research staff responsible for food portioning, plating, and weighing.

All project staff were trained by the RD in appropriate food handling and safety, food weighing, and meal recording duties. Before each meal, a team of research staff was briefed on how to portion and serve each menu item. Individual menu items were weighed and recorded (0.0 g) using calibrated food scales, portioned using standardized tools, and served at each onsite meal. Each dish with multiple food components was deconstructed into individual items and was weighed and recorded individually (ie, burgers were deconstructed to individually weigh patty, bun, cheese, ketchup, mustard, and tomato). Staff assumed various roles to ensure optimal meal-time efficiency (ie, *menu collector*, *food weigher*, and *data recorder*). After recording the weight of each food item and time of meal (00:00), the plate was served to the appropriate participant. Participants were encouraged to consume all food served at each meal, but this was not mandatory. The plate waste from each participant was deconstructed by ingredient and individually weighed at the end of each meal period.

After each participant finished eating, the research staff weighed and recorded each individual item left on the plate. The gram weight of each food item consumed was quantified and entered into an electronic database. Energy and macronutrient profiles of each menu item were obtained from the dining facility's recipe, the food label, or the USDA Food Composition Database and calculated according to the gram weight consumed.

Consuming foods outside of the study facility was discouraged but not prohibited to minimize the changes made to the participants' usual habits and metabolism. If food was consumed outside of the dining facility, participants were instructed to follow a specified procedure of self-reporting, including only consuming packaged foods, weighing and recording the weight of each individual food item, and providing the food label from the package. To minimize the miscalculation of nutrient intake of offsite foods, participants were provided with various packaged foods of known nutritional content (protein bars, jerky sticks, ramen noodles, fruit leather, and chocolate bars). They were asked to consume these foods; if this was not possible, they were required to photograph the food and record the food item, brand, time of consumption, and food weight (g) using a calibrated food scale and recording in a food diary. Offsite food diaries were collected daily from participants.

Participants unable to report to the dining facility for a scheduled meal time received alternative options and selected a prepackaged meal from a convenience market managed by the dining facility. Nutritional information from the item was extracted from the nutrition label and recorded. A team of staff recorded and analyzed the nutritional values of all offsite foods consumed by the participants. Information from the food intake data of 1 participant, as measured by the study reference method, is presented in Table 1.

Table 1. Daily food intake record of 1 study participant.

Time of meal (00:00)	Menu item	Amount consumed (g)	Energy intake (kcal)	Source of nutrition information
9:34 AM	Scrambled eggs	53	69	Product label[a]
9:34 AM	Cooked oatmeal	189	105	Product label[a]
9:34 AM	Blueberries	69	39	USDA[b] Food Composition Database
9:34 AM	Bacon	15	56	USDA Food Composition Database
9:34 AM	Milk 1%	0	0	Product label
9:34 AM	Coffee, fresh brewed	246	0	N/A[c]
9:34 AM	Granulated sugar	10	23	USDA Food Composition Database
2:11 PM	Bun	90	218	Product label[a]
2:11 PM	Beef, ground, cooked	125	156	Product label[a]
2:11 PM	Sauce	40	53	Product label[a]
2:11PM	Mixed greens	57	16	USDA Food Composition Database
2:11 PM	Artichoke hearts, canned	21	6	USDA Food Composition Database
2:11 PM	Cherry tomatoes	44	12	USDA Food Composition Database
2:11 PM	Cucumbers, sliced	51	5	USDA Food Composition Database
2:11 PM	Carrots, shredded	25	10	Product label
2:11 PM	Olive oil	12	96	Product label
2:11 PM	Balsamic vinegar	19	17	Product label
6:46 PM	Chicken tamales	304	669	Product label[a]
6:46 PM	Vegetables, roasted	250	143	Product label[a]
6:46 PM	Rice, cooked	61	105	Product label[a]
6:46 PM	Milk 1%	0	0	Product label
11 AM	Energy bar	52	210	Product label
1 PM	Energy bar	48	190	Product label
8 PM	Dehydrated soup	64	290	Product label
10 PM	Energy bar	68	290	Product label
N/A	N/A	1913[d]	2753[d]	N/A

[a]Nutritional information of food items prepared by the University of California, Davis, Dining Commons.

[b]USDA: United States Department of Agriculture.

[c]N/A: not applicable.

[d]Final row contains column totals where applicable.

Quality Assurance

Before this study, the PI conducted a series of small pilot trials over 1 year to inform this study design and data collection procedures using the wristband technology. During these pilot trials, it was observed that the form factor of the technology was the main barrier to collecting consistent, uninterrupted data during the postprandial digestion period that lasts several hours beyond each meal. Practically, any signal interruption during the meal or in the hours following it would result in loss of data and underestimation of calorie intake by the technology. Unfortunately, signal interruption occurred often and for a variety of reasons in this study; for example, periodic loss of contact with the skin was likely depending on the user's wrist size and shape. In addition, the wristband required an hour each day to obtain a full charge; any loss of charge would disable

data collection accordingly. Several strategies were used to mitigate these challenges with the form factor. Participants were instructed to charge the wristband fully before any meal to avoid missing food intake and its subsequent digestion (ie, charging band in the morning before consuming food for the day). It was acceptable to charge the wristband at any point during the day as long as no food had been consumed for 3 hours prior. On arriving at the first meal of the day, the research staff visually confirmed that the wristband was positioned on each participant, such that the sensor was in complete and constant contact. Research staff used a third-party site (Dietitian's Cabinet) to access participants' deidentified data up to the minute from which the frequency of contact interruptions could be assessed. Those who had significant interruptions were targeted for individual solutions to improve sensor contact with the wrist, for example, tightening the wristband to achieve optimal sensor positioning.

Continuous blood glucose was monitored as a strategy to measure and account for nonadherence to the study's dietary intake reporting protocols. The FreeStyle Libre (FSL) Pro System (Abbott Diabetes Care Inc) CGM system includes a unit with a water-resistant sensor that attaches to the back of the user's upper arm. Within the unit is an Enlite sensor that consists of a wire containing glucose oxidase at the tip that is inserted subcutaneously with a dedicated inserting device. Glucose oxidase catalyzes a biochemical reaction in the presence of glucose and oxygen, which transfers electrons to a receiving molecule and creates a current that can be measured and converted into a glucose concentration [27]. The FSL Pro System collects up to 14 days of glucose readings, with recordings every 15 min. A single reader can be used to activate glucose data recording and download reports from multiple devices simultaneously. One study showed that the FSL's mean absolute relative difference compared with measured capillary blood glucose levels was 13.2% (95% CI 12.0% to 14.4%) [29].

CGM sensors were secured to the tricep or rotator cuff region of participants' arms on day 1 of the study, in the morning before consuming food or beverages. During the 14-day test period, units would occasionally become detached. The research staff downloaded data files from the participants' sensors every 2 days to minimize any data loss. Text file reports were exported through the LibreView software program (Abbott Diabetes Care Inc, 2018) and a secure cloud-based system. CGM data were analyzed to assess the adherence of individual participants to reference dietary intake reporting protocols. Significant glucose increases (>20 mg/dL per 30 min) occurring outside of specified study mealtimes or not reported in food intake diaries were flagged for further examination.

Statistical Analysis

The Bland-Altman analysis was conducted to compare daily energy intake (kcal/day) estimated by both the reference method and the wristband technology. Regression analyses were used to examine trends in the data and sample characteristics. Statistics were conducted in Microsoft 2008 (version 12.3.1) and Prism 8 2019 (version 8.3.1).

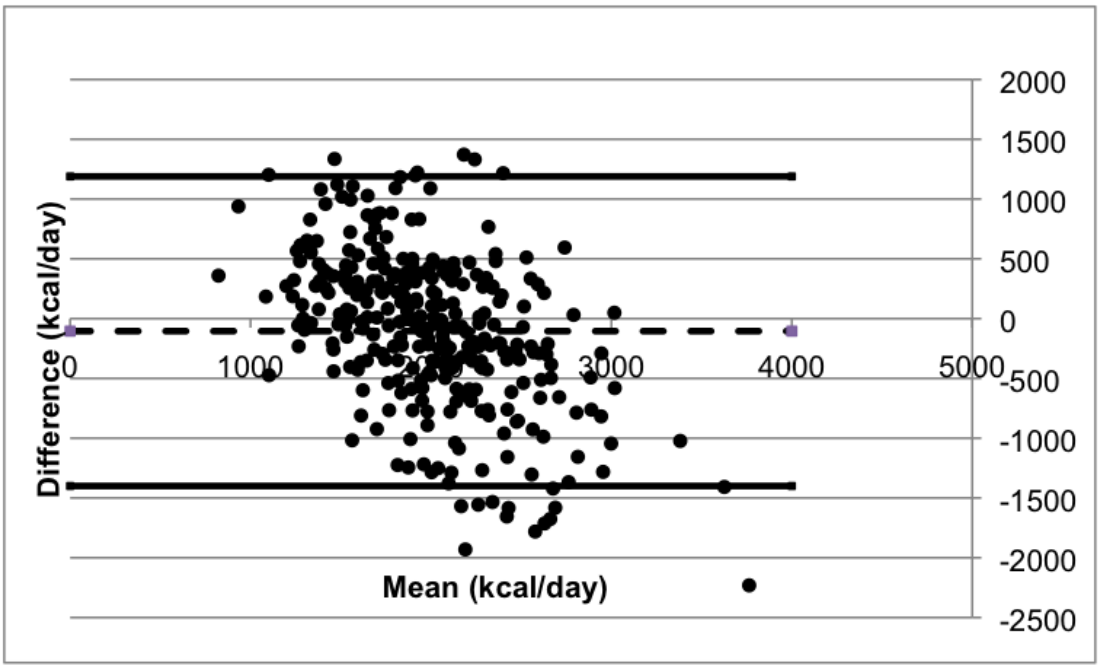
Results

This study developed a dietary intake reference method to evaluate a wearable sensor with the potential to generate objective and precise data on the dietary intake of adult individuals. The data accuracy and practical application of the current GoBe2 model was interrogated over two 14-day test periods in an intended sample of 25 participants. Of the 35 participants who were originally enrolled in phase 1 of the study, 304 measurements (kcal/day) collected from 24 participants were retained from phase 2 after data cleaning to remove missing or aberrant values.

Of the total cases, 10.9% (33/304) were excluded because they lacked an accompanying set of complete CGM data for the 24-hour period. Of the remaining cases, 22.1% (60/271) had at least one event per day of rapid blood glucose increase that was inconsistent with the recorded meal time. Of those, 68.3% (41/60) were attributed to reported bouts of exercise or other physical activity. Although CGM was used to measure nonadherence to dietary intake reporting protocols, these data were not incorporated in the present dataset.

As depicted in Figure 2, a Bland-Altman analysis showed a mean bias of −105 (SD 660) kcal/day, with 95% limits of agreement (LoA) between −1400 and 1189. Pearson correlation coefficient between the 2 methods was r=−0.496 (95% CI −0.576 to −0.406; P<.001). Linear regression analysis on the Bland-Altman plot revealed a regression equation of Y=−0.3401X+1963 that was significant (P<.001). A multiple regression analysis was conducted with the participants' age, sex, and BMI classification as independent confounding variables, but no significant effects were seen on the bias. Analysis of variance tests were conducted to assess the effects of the participants' age, sex, and BMI classification on bias, and the effects were not significant (P=.15, .18, and .12, respectively).

Figure 2. Bland-Altman (mean difference) plot of estimated nutrient intakes (kilo/day) by the test and reference method (N=304). Solid lines represent upper-lower limits of agreement, and the dashed line represents bias.



Discussion

Negative bias in the Bland-Altman analysis indicated a general underestimation of daily calorie intake by the wristband compared with the reference method. Despite a relatively small bias, the LoAs were wide, making the results of the comparison ambiguous. Regression analyses indicated a tendency for the wristband to systematically overestimate for lower calorie intake and underestimate for higher intake.

Our preliminary validation results indicate that although the ability of GoBe2 to make phenotypic discernments responsive to diet by noninvasive means has wide-reaching utility in research and practice, notable feasibility challenges were observed for free-living study participants to reliably use the technology to achieve accurate and precise measurements. These challenges were largely attributed to limitations in the technology's form factor. In observation, when positioned correctly on the arm and fully charged, the wristband's calorie intake estimates generally appeared accurate and provided interesting visuals pertaining to the body's rate of nutrient absorption. However, to achieve precise detection and accurate estimation of dietary intake, the unit's sensor required adequate skin contact be maintained at all times. Achieving this proved to be a considerable challenge for several reasons, including (1) battery life, as the unit required an hour of charging each day, which required removal of the device, preventing the detection of calories ingested within several hours before removal; (2) the wristband's bulky size, dimensions, and/or appearance were challenges for some users to maintain comfort and position on the arm; and (3) the user's own wrist size and shape; for example, small or tapered wrists were likely to result in inconsistent sensor contact. As described previously, several strategies were included in the study design to prevent data loss, such as targeting problematic cases early, checking in with participants, and monitoring sensor position daily. However, data loss from poor sensor contact was a significant barrier to the technology's ability to reliably detect calorie intake. Separate analyses, not reported in this study, further examine the technology's efficacy using data collected only during periods of protocol adherence concerning food reporting and technology use.

Establishing reliable adherence or compliance protocols is a widespread goal in measuring the dietary intake of human subjects [23]. Continuous glucose monitors were used to measure the participants' adherence to food intake recording protocols. Although CGM data do not provide a direct measure of dietary intake, its measurement of the body's relative physiological response to food intake can serve as a proxy to identify inconsistencies in reported intake data and blood glucose activity. Examination of CGM data confirmed that although a few participants (n=2) were likely nonadherent with dietary intake reporting protocols, aberrant increases in blood glucose levels could be attributed to multiple factors including exercise or other bouts of physical activity. The authors concluded that complex outcomes on CGM measurement and the participants' adherence would be appropriately detailed in the context of measuring or impacting compliance in nutrition research. Some challenges to using the CGM devices to collect data over continuous 14 days were also related to form factor limitations. The sensor included an adhesive material attached to the skin, but some devices became dislodged during the 14-day study period (13/72, 18% CGMs attached), causing complete or partial data loss. Of the 24 participants, 2 (8%) had repeated CGM sensor displacement, which was more likely to occur during physical activity (biking, gym workout, and weight lifting) and/or excessive sweating. In these cases, a skin adhesive (Skin-Tac) was useful in reinforcing the CGM attachment. As the FSL Pro System did not include individual readers with each unit, participants were blinded to their personal glucose data. At the end of the study period, data reports summarizing

glucose patterns were generated and distributed to participants. Readouts included daily blood glucose averages (g/ dL) across each 24-hour period, average glucose trend lines across each 24-hour period, and likelihoods of hypoglycemia or hyperglycemia during specified windows. A total of 7 days' worth of daily blood glucose trend lines were color coded and superimposed onto summary graphs. Participants were provided with general guidance from the RD to interpret numerical data into a relevant and actionable context for health and diet.

By collaborating with the university facility, this study used existing food production operations, resources, and personnel to carry out an extensive dietary observation study. Despite numerous strengths in the study design and utilization of a novel research environment, limitations were revealed during project implementation. For example, in the food facility where dishes were prepared for high throughput mass consumption, the exact quantity of nutrients in each portion could not be consistently and routinely ensured using these methods alone. In addition, considering that the project targeted students on a university campus, protocol adherence was less than anticipated, particularly with regard to meal attendance. Of the 42 total study meals offered to each of the 24 participants during the second 14-day testing period (1008 total meals), 56% of the scheduled meals were attended (565 meals). To improve adherence in the future, stricter enforcement of meal attendance is recommended. Studies excluding offsite food consumption may help improve the accuracy of nutrient intake reporting, with strategies in place to account for protocol adherence. Given that numerous factors were involved with intermittent data loss from the wristband technology, 2 weeks was defined as the minimum period required to gather continuous data from 25 free-living participants for validation purposes. Longer study periods could affect adherence issues without stricter guidelines around participant meal attendance.

This study validated participants' calorie intake as recorded by the wearable device, in comparison with a reference diet. The deviations in and between methods could be explained by any combination of the following factors: form factor limitations (skin contact/battery); the participants' nonadherence to dietary protocol (ie, consuming and failing to report ingested food or drink); interindividual differences in measured intake versus actual nutrient absorption and metabolism; human error in calculating food intake using the USDA Database; potential deviations from the standardized recipe during the meal preparation process; inability of the USDA Database account for nutrient variation depending on food ripeness, geographic location, and soil quality; inherent data loss because of required 1-hour daily device charging periods; and inaccuracies pertaining to technology algorithm development. Future studies should incorporate these suggestions for improvement to further interrogate the potential of wearable devices to accurately capture caloric and macronutrient intake. Ongoing engineering adjustments are recommended to accurately estimate the energy and nutrient intakes of individuals consuming various diets.

Tools are urgently needed to obtain accurate and precise measurements of diet and nutrition. Enhancing knowledge about individual phenotypes allows for more precise and predictive dietary guidance and intervention, and this has the potential to transform how people make informed diet and lifestyle choices. In today's personalized marketplace, we routinely use sophisticated technology to acquire personalized step-by-step guidance that assures arrival at nearly any physical destination (eg, satellite navigation). In accordance with the natural diversity of humans as unique phenotypes, this concept could also be applicable to the realm of food and diet. In other words, there is a need for sensitive and specific devices to deliver step-by-step directions to any desired *health destination*. This requires the tools able to quantify health status and progress over important time scales and adjust trajectories according to biofeedback. Smartphones are the cornerstone of the customization and precision of modern life, incorporating precise personal information with global databases accessible through cloud storage and applying straightforward computational algorithms to guide decisions. This basic principle and its applications offer a sophisticated and diverse range of possibilities for enhancing our individual experience, whether through personalized navigation, physical activity tracking, tailoring fitness routines, and identifying a song or even a face. However, to date, the app market does not offer reliable solutions for automating the quantification of dietary intake that would significantly impact individualized quality of life decisions. Measurement and tracking devices provide practical utility for discerning phenotypic traits and defining progressive *roadmaps to personalized health destinations*. Automated nutrient tracking devices could precisely inform diet and lifestyle choices appropriate to health status and guide individuals toward desired goals, including everything from diet planning to cardiometabolic performance. Validation and effectiveness testing of candidate devices are essential steps to be taken for the use of precision technologies to inform personalized diet and lifestyle guidance.

Conclusions

This study documented high variability in both the utility and accuracy of a wristband sensor to track nutritional intake (kcal/day). The researchers acknowledge that because dietary intake measurement of individuals has inherent challenges related to accuracy and variability, achieving precision of reference methods is a notable challenge. This study highlights the need for innovative measurement tools that are precise, reliable, and efficacious to facilitate accurate personalized dietary measurement.

Acknowledgments

The authors would like to thank UC Davis DC for their collaboration, including Felipe Becerra, Marci Ofina, Leah Beck, RD, head chefs Roger Thompson, Cesar Cienfuegos, and staff at Cuarto and Segundo DC. The authors would also like to thank undergraduate research interns, including Melissa Vilas, Mengyang Lu, Xianyu Zhu, Melanie Hercules, Adaeze Ezeagwula,

Jewel Esparza, Anna Liu, Fariba Osidary, Manvir Dhindsa, Alison Peng, Eileen Dihardja, Haley Adel, Brianna Bado, Macenzie Nielson, Taylor Janoe, Elena Chai, and Nealah Lee for their commitment to implementing research protocols, including dietary intake assessment and data entry. This research and publication were supported by unrestricted gifts to the Foods for Health Institute from Healbe, LLC.

Conflicts of Interest

This research and publication were supported by unrestricted gifts to the Foods for Health Institute from Healbe, LLC.

References

1. Betts JA, Gonzalez JT. Personalised nutrition: what makes you so special? Nutr Bull 2016 Nov 15;41(4):353-359 [FREE Full text] [doi: 10.1111/nbu.12238]
2. Ordovas J, Ferguson L, Tai E, Mathers JC. Personalised nutrition and health. Br Med J 2018 Jun 13;361:k2173 [FREE Full text] [doi: 10.1136/bmj.k2173] [Medline: 29898881]
3. Celis-Morales C, Lara J, Mathers JC. Personalising nutritional guidance for more effective behaviour change. Proc Nutr Soc 2015 May;74(2):130-138. [doi: 10.1017/S0029665114001633] [Medline: 25497396]
4. O'Sullivan A, Henrick B, Dixon B, Barile D, Zivkovic A, Smilowitz J, et al. 21st century toolkit for optimizing population health through precision nutrition. Crit Rev Food Sci Nutr 2018;58(17):3004-3015 [FREE Full text] [doi: 10.1080/10408398.2017.1348335] [Medline: 28678528]
5. German JB, Zivkovic AM, Dallas DC, Smilowitz JT. Nutrigenomics and personalized diets: what will they mean for food? Annu Rev Food Sci Technol 2011;2:97-123 [FREE Full text] [doi: 10.1146/annurev.food.102308.124147] [Medline: 22129377]
6. Parr EB, Camera DM, Burke LM, Phillips SM, Coffey VG, Hawley JA. Circulating microRNA responses between 'high' and 'low' responders to a 16-wk diet and exercise weight loss intervention. PLoS One 2016;11(4):e0152545 [FREE Full text] [doi: 10.1371/journal.pone.0152545] [Medline: 27101373]
7. Piccolo B, Keim N, Fiehn O, Adams SH, van Loan MD, Newman JW. Habitual physical activity and plasma metabolomic patterns distinguish individuals with low vs high weight loss during controlled energy restriction. J Nutr 2015 Apr;145(4):681-690 [FREE Full text] [doi: 10.3945/jn.114.201574] [Medline: 25833772]
8. Schork NJ. Personalized medicine: time for one-person trials. Nature 2015 Apr 30;520(7549):609-611. [doi: 10.1038/520609a] [Medline: 25925459]
9. Woolf SH, Purnell JQ. The good life: working together to promote opportunity and improve population health and well-being. J Am Med Assoc 2016 Apr 26;315(16):1706-1708. [doi: 10.1001/jama.2016.4263] [Medline: 27063639]
10. Bonn SE, Löf M, Östenson CG, Lagerros YT. App-technology to improve lifestyle behaviors among working adults-the health integrator study, a randomized controlled trial. BMC Public Health 2019 Mar 7;19(1):273 [FREE Full text] [doi: 10.1186/s12889-019-6595-6] [Medline: 30845949]
11. Glasgow RE, Kwan BM, Matlock DD. Realizing the full potential of precision health: the need to include patient-reported health behavior, mental health, social determinants, and patient preferences data. J Clin Transl Sci 2018 Jun;2(3):183-185 [FREE Full text] [doi: 10.1017/cts.2018.31] [Medline: 30370072]
12. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
13. Schaefer S, Ching C, Breen H, German J. Wearing, thinking, and moving: testing the feasibility of fitness tracking with urban youth. Amer J Health Educ 2016 Jan 8;47(1):8-16 [FREE Full text] [doi: 10.1080/19325037.2015.1111174]
14. Vigneshvar S, Sudhakumari CC, Senthilkumaran B, Prakash H. Recent advances in biosensor technology for potential applications-an overview. Front Bioeng Biotechnol 2016;4:11 [FREE Full text] [doi: 10.3389/fbioe.2016.00011] [Medline: 26909346]
15. 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]
16. Dorn D, Gorzelitz J, Gangnon R, Bell D, Koltyn K, Cadmus-Bertram L. Automatic identification of physical activity type and duration by wearable activity trackers: a validation study. JMIR Mhealth Uhealth 2019 May 23;7(5):e13547 [FREE Full text] [doi: 10.2196/13547] [Medline: 31124470]
17. Nording ML, Yang J, Georgi K, Karbowski CH, German JB, Weiss RH, et al. Individual variation in lipidomic profiles of healthy subjects in response to omega-3 fatty acids. PLoS One 2013;8(10):e76575 [FREE Full text] [doi: 10.1371/journal.pone.0076575] [Medline: 24204640]
18. Zivkovic A, Wiest M, Nguyen U, Nording ML, Watkins SM, German JB. Assessing individual metabolic responsiveness to a lipid challenge using a targeted metabolomic approach. Metabolomics 2009;5(2):209-218 [FREE Full text] [doi: 10.1007/s11306-008-0136-0] [Medline: 20442800]

19. Morris C, O'Grada C, Ryan M, Roche HM, Gibney MJ, Gibney ER, et al. Identification of differential responses to an oral glucose tolerance test in healthy adults. PLoS One 2013;8(8):e72890 [FREE Full text] [doi: 10.1371/journal.pone.0072890] [Medline: 23991163]
20. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. Cell 2015 Nov 19;163(5):1079-1094 [FREE Full text] [doi: 10.1016/j.cell.2015.11.001] [Medline: 26590418]
21. Bakker G, van Erk MJ, Pellis L, Wopereis S, Rubingh C, Cnubben N, et al. An antiinflammatory dietary mix modulates inflammation and oxidative and metabolic stress in overweight men: a nutrigenomics approach. Am J Clin Nutr 2010 Apr;91(4):1044-1059. [doi: 10.3945/ajcn.2009.28822] [Medline: 20181810]
22. Mozaffarian D, Ludwig DS. Dietary guidelines in the 21st century-a time for food. J Am Med Assoc 2010 Aug 11;304(6):681-682. [doi: 10.1001/jama.2010.1116] [Medline: 20699461]
23. Archer E, Marlow M, Lavie C. Controversy and debate: memory-based methods paper 1: the fatal flaws of food frequency questionnaires and other memory-based dietary assessment methods. J Clin Epidemiol 2018 Dec;104:113-124. [doi: 10.1016/j.jclinepi.2018.08.003] [Medline: 30121379]
24. 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]
25. Wrangham R. The curiously long absence of cooking in evolutionary thought. Learn Behav 2016 Jun;44(2):116-117. [doi: 10.3758/s13420-016-0223-4] [Medline: 27059233]
26. Mehrotra P. Biosensors and their applications-a review. J Oral Biol Craniofac Res 2016;6(2):153-159 [FREE Full text] [doi: 10.1016/j.jobcr.2015.12.002] [Medline: 27195214]
27. Dungan K, Verm N. Monitoring Technologies – Continuous Glucose Monitoring, Mobile Technology, Biomarkers of Glycemic Control. South Dartmouth, MA: MDText; 2000.
28. Lohan T, Roche AF, Martorell R. Anthropometric Standardization Reference Manual. Champaign, Illinois: Human Kinetics; 1988.
29. Ólafsdóttir AF, Attvall S, Sandgren U, Dahlqvist S, Pivodic A, Skrtic S, et al. A clinical trial of the accuracy and treatment experience of the flash glucose monitor freestyle libre in adults with type 1 diabetes. Diabetes Technol Ther 2017 Mar;19(3):164-172 [FREE Full text] [doi: 10.1089/dia.2016.0392] [Medline: 28263665]

Abbreviations

CGM: continuous glucose monitoring
DC: dining commons
FSL: FreeStyle Libre
LoA: limits of agreement
PI: principal investigator
RD: registered dietician
UC Davis: University of California, Davis
USDA: United States Department of Agriculture

Edited by G Eysenbach; submitted 26.09.19; peer-reviewed by S Bonn, S Haghayegh; comments to author 18.11.19; revised version received 13.01.20; accepted 10.04.20; published 22.07.20.

Please cite as:
Dimitratos SM, German JB, Schaefer SE
Wearable Technology to Quantify the Nutritional Intake of Adults: Validation Study
JMIR Mhealth Uhealth 2020;8(7):e16405
URL: https://mhealth.jmir.org/2020/7/e16405
doi:10.2196/16405
PMID:32706729

Original Paper

Derivation of Breathing Metrics From a Photoplethysmogram at Rest: Machine Learning Methodology

Joseph Prinable[1], BEng; Peter Jones[1], PhD; David Boland[1], PhD; Cindy Thamrin[2], PhD; Alistair McEwan[1], PhD

[1]School of Electrical and Information Engineering, The University of Sydney, Darlington, Australia

[2]The Woolcock Institute of Medical Research, The University of Sydney, Glebe, Australia

Corresponding Author:
Joseph Prinable, BEng
School of Electrical and Information Engineering
The University of Sydney
Room 402, Building J03
Maze Crescent
Darlington, 2006
Australia
Phone: 61 404035701
Email: joseph.prinable@sydney.edu.au

Abstract

Background: There has been a recent increased interest in monitoring health using wearable sensor technologies; however, few have focused on breathing. The ability to monitor breathing metrics may have indications both for general health as well as respiratory conditions such as asthma, where long-term monitoring of lung function has shown promising utility.

Objective: In this paper, we explore a long short-term memory (LSTM) architecture and predict measures of interbreath intervals, respiratory rate, and the inspiration-expiration ratio from a photoplethysmogram signal. This serves as a proof-of-concept study of the applicability of a machine learning architecture to the derivation of respiratory metrics.

Methods: A pulse oximeter was mounted to the left index finger of 9 healthy subjects who breathed at controlled respiratory rates. A respiratory band was used to collect a reference signal as a comparison.

Results: Over a 40-second window, the LSTM model predicted a respiratory waveform through which breathing metrics could be derived with a bias value and 95% CI. Metrics included inspiration time (–0.16 seconds, –1.64 to 1.31 seconds), expiration time (0.09 seconds, –1.35 to 1.53 seconds), respiratory rate (0.12 breaths per minute, –2.13 to 2.37 breaths per minute), interbreath intervals (–0.07 seconds, –1.75 to 1.61 seconds), and the inspiration-expiration ratio (0.09, –0.66 to 0.84).

Conclusions: A trained LSTM model shows acceptable accuracy for deriving breathing metrics and could be useful for long-term breathing monitoring in health. Its utility in respiratory disease (eg, asthma) warrants further investigation.

(JMIR Mhealth Uhealth 2020;8(7):e13737) doi:10.2196/13737

KEYWORDS

photoplethysmogram; respiration; asthma monitoring; LSTM

Introduction

There has been increasing interest in monitoring health using wearable sensors. However, very few technologies have focused on the breathing signal. The ability to monitor breathing may be beneficial for general health and particularly for asthma, which is a health condition that affects over 300 million people globally [1]. Monitoring of lung function using specialized metrics such as peak expiratory flow has been shown to be useful for predicting risk of an asthma episode [2]; however, this can be difficult to perform for patients as it involves forced maneuvers. It remains to be seen whether continuous monitoring of simple breathing metrics such as the interbreath interval (IBI) and the inspiration-expiration (I:E) ratio could provide further information on asthma control [3] and disease status [4].

The availability of a noninvasive sensor that measures breathing continuously and in an ambulatory manner would facilitate studies to establish clinical utility. One sensor of interest is the pulse oximeter that is commonly used in a clinical setting to measure both arterial blood oxygen saturation (SPO_2) and heart rate. A tidal breathing method exists that also shows promise for clinical prediction [5]; however, these methods are unsuitable

for continuous monitoring (eg, during walking or exercise). It was recently shown that a pulse oximeter can also be used to continuously monitor respiratory rate in a clinical setting [6]. This is possible because breathing periodicity [6,7] and effort [8] modulate photoplethysmogram (PPG) amplitude, frequency, and baseline wander [9,10]. Filtering and feature-based signal processing approaches can be applied to the PPG signal to extract a surrogate respiratory signal. This in turn can be processed to derive breathing rate (BR) with varying degrees of accuracy [7].

Unfortunately, there is poor amplitude correlation between the surrogate respiratory waveform and a gold standard respiratory trace. This poor correlation may make the I:E ratio difficult or impossible to derive using existing methods. In this work, we sought to address this using machine learning. In a previous pilot study [11], we demonstrated how a long short-term memory (LSTM) approach could predict a respiratory waveform from which BR could be derived. LSTM is a type of a recurrent neural network that can capture long-term, time-based dependencies in data [12]. Through the LSTM, we showed that the Pearson correlation coefficient between the derived respiratory waveform and a pneumotachograph trace had similarly high r values (r>0.8) to existing methods. In this paper, we built on this study by investigating the accuracy to which IBI, I:E ratio, and BR respiratory metrics can be attained from a PPG-derived surrogate respiratory waveform using an LSTM. We show that, in comparison to existing approaches, we can derive breathing metrics to a higher degree of accuracy from a pulse oximeter.

Methods

Datasets

Data Collection

Measurements were recorded from a group of 10 healthy participants who provided informed consent. The protocol for this study was approved by Northern Sydney Local Health District Human Research Ethics Committee (LNR/16/HAWKE/99 ethics approval). Participants conducted 5 randomized breathing serials at a rate of 6, 8, 10, 12, or 14 breaths per minute (BPM). Each serial was conducted for 5 minutes. Each participant was coached to breath one full inhalation and exhalation in time with a visual prompt.

An Alice PDx (Philips Respironics, Murrysville, PA) portable sleep diagnostic system was used to measure physiological signals during this study. The supplied pulse oximeter was attached to the index finger of the nonmaster hand, allowing the capture of a raw PPG trace, SPO_2, and pulse rate data. The Alice PDx reported calculated values for SPO_2 and pulse rate 3 times per second. PPG signals were sampled at 75 Hz.

Respiratory inductance plethysmography is a method to measure relative tidal volume (RTV) as a function of the chest and abdominal wall movement [13]. In this study, inductance bands were placed around the abdomen and ribcage according to the manufacturer's guidelines, allowing RTV to be estimated as the weighted sum of the chest and abdominal wall inductance signals. The Alice PDx system reported an RTV signal based on the contribution of both respiratory bands and was captured at 100 Hz.

Description of Available Features

The Alice PDx system outputs three independent time series: PPG, SPO_2, and pulse rate. The SPO_2, processed PPG, and pulse rate signals were up-sampled to 25 Hz while the RTV was down-sampled to 25 Hz before normalizing between ±1. The sampling rate of 25 Hz was selected to ensure respiratory rate accuracy [7,14] and so that all time series data had the same time scale.

In addition to the three time series given by the Alice PDx system, a bandpassed PPG time series was generated by passing the original PPG signal through a sixth order Butterworth bandpass filter with a center frequency corresponding to the respiratory rate of the signal with a bandwidth of 0.002 Hz. This additional time series was included because our previous findings suggested that this feature could improve model prediction [11].

Altogether, the available features used within our model are as follows:

- Feature 1: PPG
- Feature 2: bandpassed PPG
- Feature 3: SPO_2
- Feature 4: pulse rate

We previously determined experimentally that the inclusion of SPO_2 and pulse rate values helped inform the network when decoupling between the pulse signal and respiratory signal occurs [11]. The exact underlying physiological mechanisms are unclear.

Derivation of a Respiratory Waveform Time Series

For comparison purposes, RRest toolbox [15] was used to extract respiratory waveforms from a PPG using 10 feature-based and filter-based algorithms as shown in Figure 1. The resulting respiratory waveforms were temporally aligned to correspond with the reference respiratory waveform in the test set for comparison purposes. The techniques used to derive the respiratory waveforms, as well as our LSTM method, are described in Table 1.

Figure 1. Using existing filter-based and feature-based methods, 10 relative respiratory waveforms were derived from a photoplethysmogram (PPG) signal, and another relative respiratory waveform was derived using a long short-term memory (LSTM) that accepts PPG, arterial blood oxygen saturation (SPO2), band-passed (BP) PPG, and pulse rate inputs. BR: breathing rate; I:E: inspiration-expiration ratio; IBI: interbreath interval.

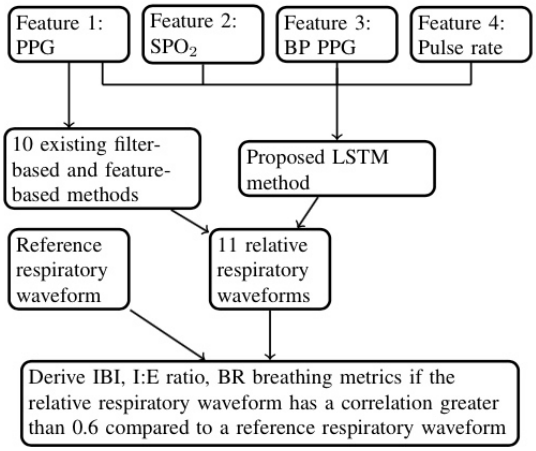


Table 1. Techniques for the extraction of respiratory signals from a photoplethysmogram (adapted from Charlton et al [15]).

Respiratory signal	Description
Filter-based	
X_{A1}	Bandpass filter between plausible respiratory frequencies
X_{A2}	Maximum amplitude of the CWT[a] within plausible cardiac frequencies (30-220 beats per minute) [16]
X_{A3}	The frequency corresponding to the maximum amplitude of the CWT within plausible cardiac frequencies [16]
Feature-based	
X_{B1}	Mean amplitude of troughs and proceeding peaks [7]
X_{B2}	Difference between the amplitudes of troughs and proceeding peaks [17]
X_{B3}	Time interval between consecutive troughs [17]
X_{B4}	Mean signal value between consecutive troughs [18]
X_{B5}	Peak amplitude [17]
X_{B6}	Trough amplitude [18]
X_{B10}	PPG[b] pulse width estimation using a wave boundary detection algorithm [19]
Machine learning–based	
X_{LSTM}[c]	Proposed LSTM method

[a]CWT: continuous wavelet transform.

[b]PPG: photoplethysmogram.

[c]LSTM: long short-term memory.

LSTM Architecture and Parameters

We propose the use of an LSTM model as an alternative to the signal processing methods described in Table 1. In this section, we discuss our training and validation procedures to determine an appropriate LSTM architecture to predict a respiratory waveform.

The core component of an LSTM architecture is a memory cell whose characteristics allow long-term data dependencies to be captured. A single LSTM cell uses gate mechanisms to forget irrelevant parts of a previous state, selectively update the current cell state, and to output the cell state [12]. Each cell contains a number of hidden units that define the dimensionality of both the current and output states. Increasing hidden units within a model may lead to overfitting. Conversely, reducing hidden units below a certain threshold will not allow a model to be trained.

Hyperparameter Search

We first conducted a structured, though nonexhaustive, hyperparameter search to determine suitable values for our final

LSTM architecture. We then performed more extensive training to maximize the performance of our final architecture.

Hyperparameter Exploration

An open-source Python 3.5 library called TensorFlow r1.3 was used to train the LSTM model on a Dell Optiplex D810 (i7, 32 GB RAM; Dell Inc, Round Rock, TX) and two Titan Xp (Nvidia Corp, Santa Clara, CA) graphics processing units (GPUs).

The AdamOptimizer class of Tensorflow was used to train the LSTM using a learning rate of 0.0005 for 100 epochs with a batch size of 128. We explored the effect of changing the amount of cells (100, 300, 500), hidden units within a cell (500, 1500, 2500), and layers (1, 2, 3) and compared the results against a default model containing 100 cells, 500 hidden units, and a single layer. For this study, cells were layered sequentially two times to improve model accuracy and robustness [20]. The dropout layer was placed between each layer with a dropout rate of 0.5 to reduce overfitting [21]. There was a single dense, fully connected layer at the end.

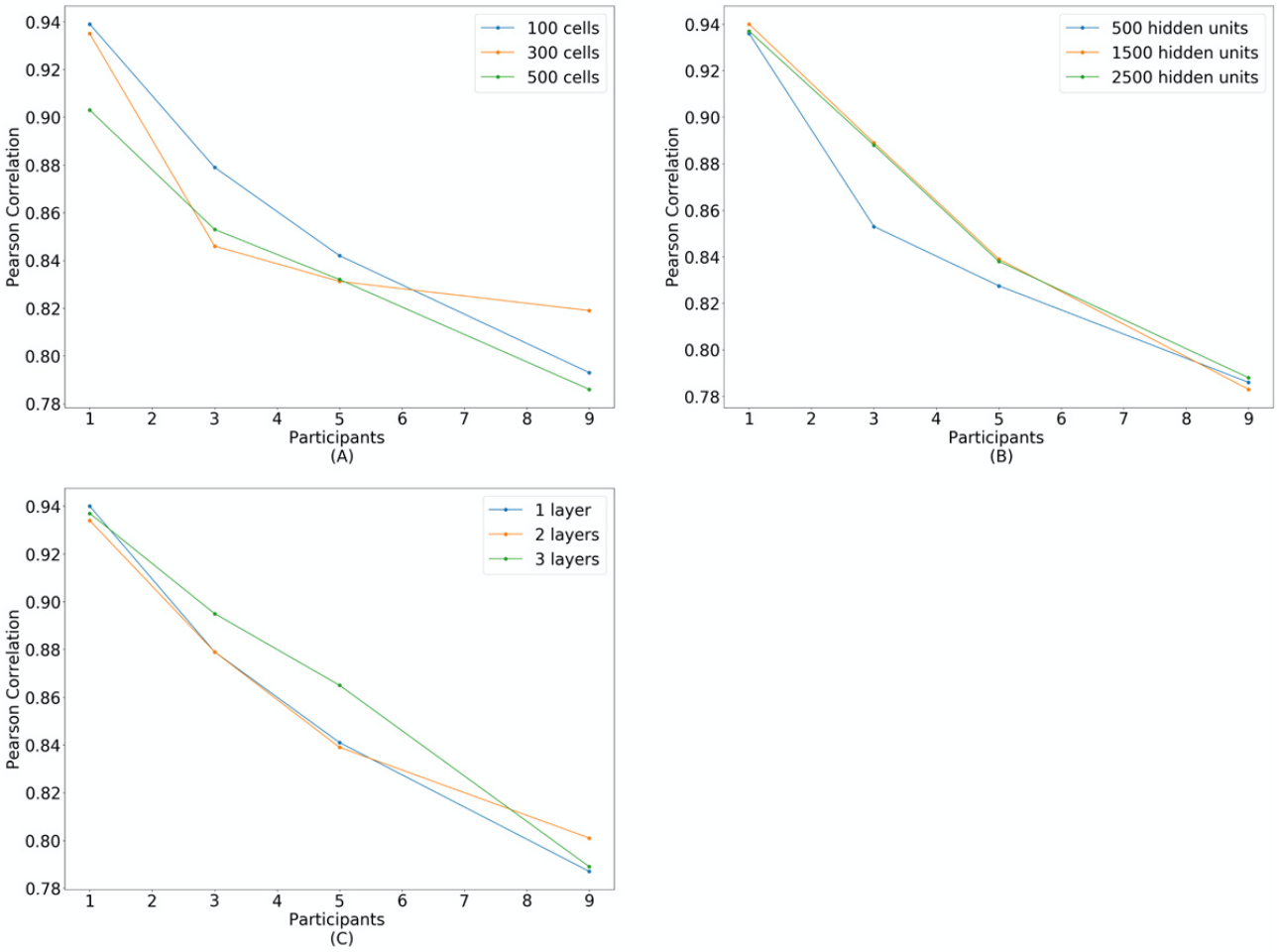
To minimize training time for hyperparameter exploration, 4 smaller training datasets were created from the original 45 unique datasets (9 participants, each with 5 breathing serials). These datasets contained data from 1 participant (7), 3 participants (3, 5, 7), 5 participants (1, 3, 5, 7, 9), or 9 (1, 2, 3, 4, 5, 6, 7, 8, 9) participants. This allowed us to compare model performance as the number of participants increased for the various configurations. To further reduce training time, each dataset was reduced to 1 minute of data, splitting 70%, 15%, and 15% into training, validation, and test sets, respectively. To assess the performance of the model, we conducted 5-fold cross validation. To reduce computational time that typically results in higher error bias but lower variability, we chose 5 folds over 10 folds [22]. We investigated permutations of the available features and found that accuracy increased with the number of features with a minimal cost in terms of execution time.

Table 2 shows the training time in minutes as a function of participants and the hyperparameter. The Pearson correlation coefficients between the derived and reference respiratory waveforms are plotted as a function of increasing number of participants for the chosen cell values (Table 2) in Figure 2A, hidden unit values in Figure 2B, and layer values in Figure 2C. The highest correlation was achieved with 300 cells and 2 layers for 9 participants. For hidden units, the correlation was similar between the quantities, with 2500 hidden units only slightly better than 500 (0.786 vs 0.788). Due to the minimal difference, the latter was selected as it required significantly less training time (211 minutes vs 1213 minutes) for comparable performance.

Table 2. Training time (minutes) for the hyperparameter search.

Hyperparameters	1 participant	3 participants	5 participants	9 participants
Cells				
100	24	75	110	208
300	54	200	272	505
500	84	313	542	932
Hidden units				
500	24	69	150	211
1500	52	161	264	482
2500	131	402	665	1213
Layers				
1	24	65	116	220
2	35	125	190	366
3	48	158	271	470

Figure 2. Pearson correlation values between derived and reference respiratory waveforms, given a dataset containing n participants, for a long short-term memory (LSTM) with (A) cells of size 100, 300, and 500; (B) hidden units of size 500, 1500, and 2500; (C) layers of size 1, 2, and 3.



Final Model Training

The final model had around 8257 trainable parameters consisting of 300 cells, 2 layers, and 16 hidden layers. To train our final model, we used AdamOptimizer with an initial learning rate of 0.02 and batch size of 256. We conducted 5-fold cross validation with approximately 223,786 training examples per fold with early stopping.

Extraction of Breathing Metrics

We defined a valid window when the Pearson correlation coefficient was >0.6 between the gold standard respiratory waveform and derived tidal volume waveform (TVW) in the window. For test sets that contained valid windows, we extracted peaks and troughs in MATLAB R2016b (MathWorks Inc, Natick, MA). To find the maximum points, 'findpeaks' was used, and we used a linear search algorithm to find the global minimum between 2 consecutive peaks. Using the peak and trough data, we extracted the following: IBI (the period in seconds between 2 consecutive peaks within the TVW signal), inspiration time (period in seconds between a trough and peak within the TVW signal), expiration period (period in seconds between a peak and trough within the TVW signal), and I:E (ratio between consecutive inspiration time and expiration period).

We then evaluated the Bland-Altman agreement [13] between the derived respiratory metrics to reference metrics.

Additionally, the root mean square error between hypothesized RTV signal y(t) and the true RTV Y(t) was calculated for each person and respiratory rate and subsequently averaged across the 5 folds.

Results

Data Collection

Data were acquired from 10 healthy subjects. One subject was excluded because of incomplete recordings due to an SD card save error on the Alice PDx. Therefore, data for 9 subjects were analyzed. The median (lower, upper quartiles) age of the analyzed subjects was 28 years (24.5 to 33.0 years). Median BMI was 23.59 kg/m^2 (21.28 to 30.04 kg/m^2), and 3 subjects (3/9, 33%) were female. In total, we recorded 3.75 hours of data, consisting of 5 minutes * 5 breathing rates * 9 participants.

Model Validation

The weights and biases were saved for each epoch during training. Training was stopped when the validation error diverged to avoid overfitting. Early stopping occurred when the validation cost did not improve for 5 epochs.

Derivation of Breathing Metrics

In total, 225 unique test sets were created from 9 participants, at 5 respiratory rates, over 5 folds. Each test set was a window of 1000 samples (40 seconds) in length. We plotted the number of valid windows as a function of increasing Pearson correlation coefficients between derived and reference respiratory waveforms in 0.2 increments in Figure 3. For a Pearson correlation coefficient ≥0.6, our approach, X_{LSTM}, was valid for 191/225 (85%) windows, while the next highest performing algorithm, X_{A1}, was valid for 128/225 (57%) windows, followed by X_{A2}, which was valid for 119/225 (53%) windows. Other algorithms were excluded from further analysis due to a small percentage of valid windows: 21/225 (9%) for X_{A3}, 56/225 (25%) for X_{B1}, 38/225 (17%) for X_{B2}, 36/225 (16%) for X_{B3}, 23/225 (10%) for X_{B4}, 65/225 (29%) for X_{B5}, 52/225 (23%) for X_{B6}, and 11/225 (5%) for X_{B10}.

Breathing metrics were averaged over each 40-second test set. The mean (SD) between derived and gold standard metrics and their associated *t* test results are shown in Table 3. The Bland-Altman agreement between derived and gold standard metrics for all subjects and respiratory rates are reported in Table 4. In the case of X_{LSTM}, a Savitzky-Golay filter was used to smooth the derived respiratory waveform prior to extracting the breathing metrics.

The Bland-Altman plot for the derived breathing metrics of inspiration time, expiration period, IBI, BR, and I:E across all participants (1-9) and all respiratory rates (6, 8, 10, 12, 14) using X_{LSTM} is shown in Figure 4. For comparison purposes, we report the Bland-Altman plot for derived respiratory rate across all participants and all respiratory rates using the highest performing algorithm found by Charlton et al [7] in Figure 5.

Figure 3. Number of valid windows as a function of increasing Pearson correlation coefficients between derived and reference respiratory waveforms in 0.2 increments. For an explanation of the variables please refer to Table 1.

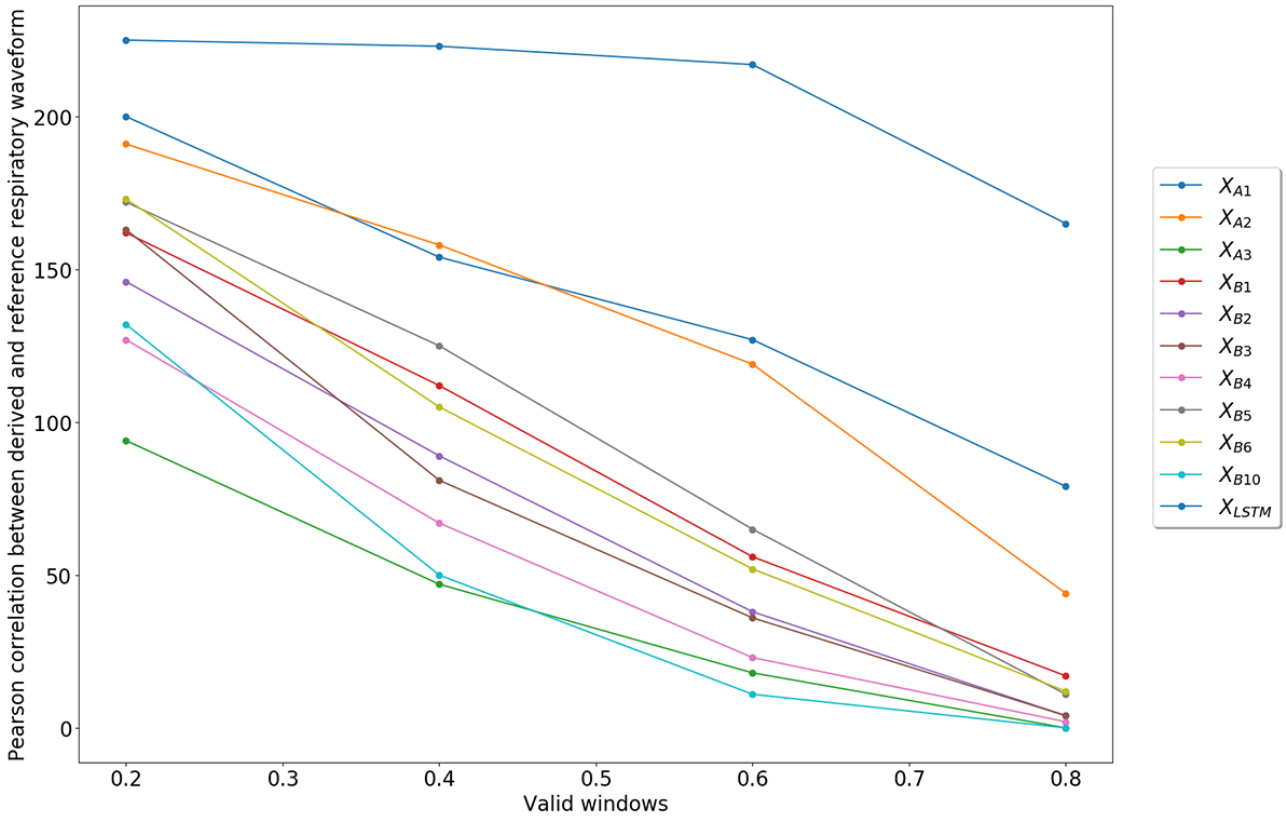


Table 3. Breathing metrics for the reference respiratory band, X_{LSTM}, X_{A1}, and X_{A2} methods, with their associated paired t test results.

Breathing metrics	Respiratory band, mean (SD)	X_{LSTM}[a]			X_{A1}			X_{A2}		
		Mean (SD)	t_{180} test	*P*	Mean (SD)	t_{126} test	*P*	Mean (SD)	t_{118} test	*P*
Tinsp[b] (seconds)	3.28 (1.29)	3.14 (1.15)	2.92	0.004	3.46 (1.31)	1.65	0.103	3.40 (1.42)	0.02	0.99
Texp[c] (seconds)	3.13 (1.01)	3.19 (1.05)	–1.68	0.095	3.38 (1.09)	–3.24	0.002	3.10 (0.95)	–1.44	0.152
BR[d] (BPM[e])	10.28 (2.72)	10.41 (2.74)	–1.39	0.167	9.69 (2.73)	1.93	0.056	10.35 (2.95)	–0.46	0.649
IBI[f] (seconds)	6.40 (1.98)	6.33 (1.96)	1.12	0.262	6.84 (2.09)	–2.18	0.031	6.50 (2.09)	–1.73	0.086
I:E[g]	1.01 (0.36)	1.09 (0.43)	–3.09	0.002	1.03 (0.40)	–2.68	0.008	1.00 (0.29)	–1.50	0.135

[a]LSTM: long short-term memory.

[b]Tinsp: inspiration time.

[c]Texp: expiration period.

[d]BR: breathing rate.

[e]BPM: breaths per minute.

[f]IBI: interbreath interval.

[g]I:E: inspiration:expiration ratio.

Table 4. Derived breathing metrics using the X_{LSTM}, X_{A1}, and X_{A2} methods and associated statistical analyses.

Method	Bland-Altman r^2	*P*	Absolute Bias	95% LoA[a]	Relative Bias (%)	95% LoA
Tinsp (seconds) [b]						
X_{LSTM}[c]	0.70	<.001	–0.16	–1.64 to 1.31	–3.70	–38.44 to 31.05
X_{A1}	0.74	<.001	–0.11	–1.51 to 1.30	–2.35	–35.65 to 30.95
X_{A2}	0.74	<.001	-0.01	–1.46 to 1.46	–0.22	–33.34 to 32.90
Texp (seconds)[d]						
X_{LSTM}	0.54	<.001	0.09	–1.35 to 1.53	2.35	–31.84 to 36.55
X_{A1}	0.41	<.001	0.25	–1.45 to 1.95	6.41	–32.82 to 45.63
X_{A2}	0.43	<.001	0.10	–1.39 to 1.59	2.70	–36.34 to 41.73
BR[e] **(BPM**[f]**)**						
X_{LSTM}	0.83	<.001	0.12	–2.13 to 2.37	1.22	–23.63 to 26.07
X_{A1}	0.92	<.001	–0.13	–1.68 to 1.41	–1.38	–18.42 to 15.65
X_{A2}	0.88	<.001	0.04	–1.94 to 2.02	0.14	–19.35 to 19.62
IBI[g] **(seconds)**						
X_{LSTM}	0.82	<.001	–0.07	–1.75 to 1.61	–0.98	–22.62 to 20.66
X_{A1}	0.88	<.001	0.14	–1.31 to 1.60	2.08	–16.55 to 20.70
X_{A2}	0.91	<.001	0.10	–1.13 to 1.33	1.37	–16.20 to 18.94
I:E[h]						
X_{LSTM}	0.30	<.001	0.09	–0.66 to 0.84	9.91	–63.89 to 83.70
X_{A1}	0.11	<.001	0.09	–0.68 to 0.87	6.65	–61.43 to 74.73
X_{A2}	0.04	<.001	0.05	–0.62 to 0.71	3.41	63.89 to 70.72

[a]LoA: limits of agreement.
[b]Tinsp: inspiration time.
[c]LSTM: long short-term memory.
[d]Texp: expiration period.
[e]BR: breathing rate.
[f]BPM: breaths per minute.
[g]IBI: interbreath interval.
[h]I:E: inspiration:expiration ratio.

Figure 4. Bland-Altman plots for (A) inspiration time (seconds), (B) expiration time (seconds), (C) interbreath interval (seconds), (D) breathing rate (breaths per minute), and (E) inspiration:expiration ratio using the LSTM method.

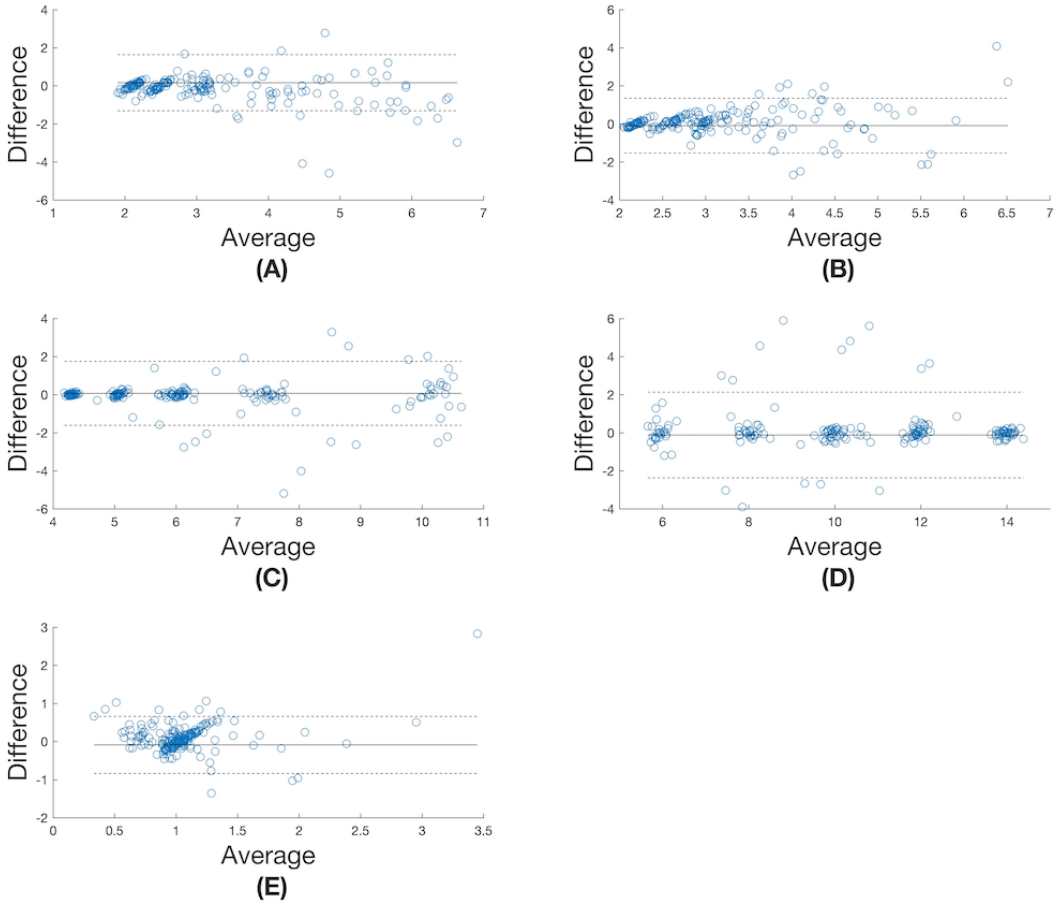
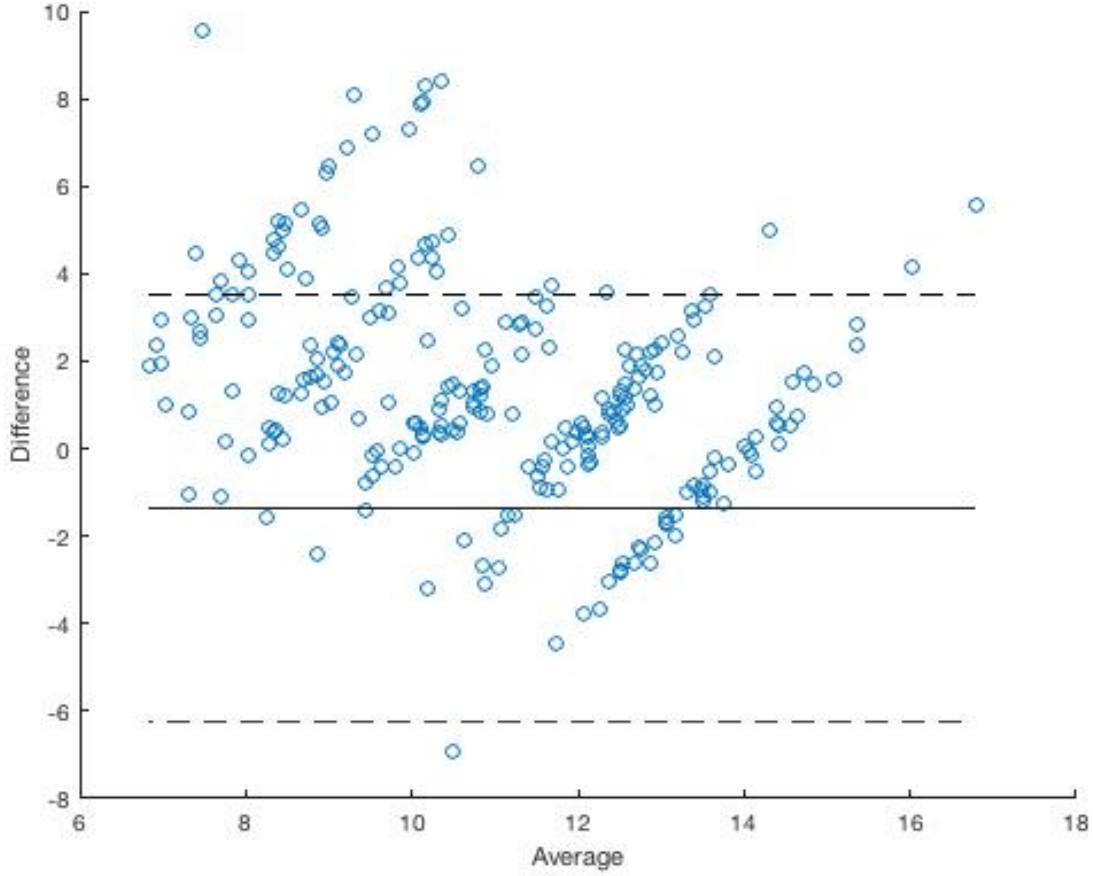


Figure 5. Bland-Altman plot for the highest performing algorithm ($X_{B1,2,3}E_{T4}F_{M1}$) found by Charlton et al [7].



Our model consistently performed comparably to the other methods, showing similar agreement (lower bias) and variability (narrower limits of agreement). The relative bias for our model was <4% for all breathing metrics examined except for I:E ratio (at 9.9%), which is within the limits of accuracy of existing standards on the estimation of breathing metrics using conventional methods [5], although the limits of variability are wide.

The differences for inspiration time are bound within the 95% CIs for average inspiration periods <4 seconds. Distinct clustering can be seen around an inspiration period of 2 seconds (Figure 4A). The differences for expiration period are bound within the 95% CIs for average expiration periods of 2-3 seconds (Figure 4B). For IBI, 4 distinct clusters occur corresponding to intervals of 4, 5, 6, and 7 seconds; however, the clustering weakens above 9 seconds (Figure 4C). For BR, 5 distinct clusters are formed corresponding to expected BRs of 6, 8, 10, 12, and 14 BPM (Figure 4D). There is noticeable clustering for I:Es of 0.8-1 (Figure 4E).

To quantify the accuracy of our model and provide a metric for future comparisons, we report the root mean square error over all participants and respiratory rates for X_{LSTM} for inspiration time (0.77 seconds), expiration period (0.74 seconds), IBI (0.8377 seconds), BR (0.86 BPM), and I:E (1.15).

Discussion

Principal Findings

In this work, we were interested in determining the feasibility of finding continuous measures of inspiration time, expiration period, IBI, BR, and I:E metrics from a PPG. We showed how an LSTM architecture could be used to predict these metrics for 191/225 (85%) test sets comprised of 9 participants at a respiratory rate of 6, 8, 10, 12, or 14 BPM. We conducted Bland-Altman analyses and found the LSTM was able to predict the average inspiration time of –0.16 seconds (–1.64 to 1.31 seconds) and expiration period of 0.09 seconds (–1.35 to 1.53 seconds) over a 40-second window. The LSTM was able to predict an I:E ratio of 0.09 (–0.66 to 0.84), although this was poorly correlated with reference values. However, this is the first time this metric is being reported in the literature as measured from a pulse signal.

The LSTM model was trained to minimize the error between derived and reference respiratory waveforms and was then able to generalize the breathing characteristics of 9 subjects and predict future respiratory waveforms based on PPG data. The ability to “see and learn” a reference signal presents a distinct advantage over existing methods. Through this approach, it was possible to determine the continuous average breathing metrics of inspiration time, expiration period, IBI, and BR for the majority of time (85%), exceeding a Pearson correlation threshold of 0.6. In contrast, these breathing metrics could only be derived, at best, around half the time (56% in the case of

X_{A1}) using existing feature-based and filter-based algorithms that did not rely on any previous reference data. While we directly compared the performance of X_{A1} and X_{A2} to the LSTM method, other methods were excluded from this analysis due to the fact that the correlation between the derived respiratory waveform and the gold standard was <0.6 more than 80% of the time. Feature-based techniques (X_{B1}-X_{B6}, X_{B10}) have performed well in previous respiratory rate algorithm assessments by Charlton et al [7] and would likely have similar performance on this dataset. In the cases where breathing metrics could be extracted for $X_{A1}$and X_{A2}, we found that the metrics of inspiration time, expiration period, and I:E were poorly correlated with the reference metrics, as shown in Table 4.

We conducted Bland-Altman analysis on the highest performing algorithm $X_{B1,2,3}E_{T4}F_{M1}$ found by Charlton et al [7] in his comparison of classical signal processing algorithms for PPG. The bias in our dataset compared to those in the dataset used by Charlton et al [7] was higher (–1.12 vs 1). However, the 95% limits of agreement (BPM) was lower (–2.4 to 2.1 vs –5.1 to 7.2). X_{LSTM} compares favorably to $X_{B1,2,3}E_{T4}F_{M1}$ with similar bias (0.12 vs –1.10) and a smaller 95% limits of agreement (BPM; –2.13 to 2.37 vs –2.63 to 2.44). The bias in our model compares well against existing standards on breathing metric estimation using conventional methods, which stipulate an accuracy of at least 2% for respiratory rate. It is worth noting that the standards are formulated for infant populations who breath faster. The wide variability seen in our model could be improved, although it is lower than that obtained from other methods examined. The high degree of variability could arise from differences in accuracy with different respiratory rates. While there is insufficient data from this study to ascertain this, it justifies use of longer-term data collection for further investigation.

The hyperparameters for the LSTM model were chosen in a structured, although non-exhaustive, manner by comparing a change in the number of cells, hidden units, or layers to a fixed model. Figures 2-4 show a decreasing trend in the correlation between the LSTM-derived respiratory waveform and the reference waveform as the number of participants increased. This trend occurred irrespective of the number of cells, hidden units, or layers. This may be accounted for, in part, by the complexity for which the LSTM model must account as the participant population increases. In the specific case of 300 cells, the correlation curve decreased quasi-exponentially. However, in the case of hidden units and layers, the correlation curve decreased quasi-linearly. It remains to be seen if the minimum correlation is bound between derived and reference respiratory waveforms for a given population. The findings of this paper show that our previous network parameter was much larger than required [11].

In this work, we used the following 4 features: PPG, filtered PPG, SPO_2, and pulse rate. We did not conduct feature selection, which may have helped to improve the overall model performance. It would be useful to see the effect of removing the filtered PPG signal feature to reduce additional preprocessing time and computational power.

Due to a limited participant population, we did not conduct leave-one-out participant cross validation. The shape of each respiratory waveform varied from person to person, and it is unlikely that the LSTM model derived in this work would be able to predict respiratory metrics from an unseen participant. However, with a larger training population, the LSTM model may be exposed to enough data to enable the accurate prediction of respiratory metrics in an unseen participant. Previously, we found that participants would prefer that a wearable sensor device have a watch form factor [23]. In this paper, we did not look at the feasibility of implementing an LSTM in this type of form factor. Currently, LSTM training requires GPU-grade computational power. With current low-power Bluetooth low energy devices [11,24,25], it may be possible to acquire PPG data and stream real-time data to a cloud-based GPU server to run online training. Once the weights and biases of the LSTM architecture are found, it may also be possible for an embedded platform to perform the required processing to obtain real-time breathing metric predictions. At present, field-programmable gate arrays can be used for real-time predictions and benefit from low latency and low power consumption [26]. Additionally, the field-programmable gate array architecture is reconfigurable. This would allow any potential device to be individually tailored to a specific model.

Conclusion

This paper presents the feasibility of monitoring simple breathing metrics such as the IBI, BR, inspiration time, expiration period, and I:E for a person at rest. We hope this proof-of-concept paper will inspire future research to collect further data and develop more powerful machine learning algorithms. In the future, it may also be possible to derive these metrics from a wristworn device that contains a pulse oximeter and accelerometer for a person at rest and support potential longitudinal studies to determine if these metrics can provide further information on asthma type [3] and provide any clinical utility [4].

Conflicts of Interest

None declared.

References

1. World Health Organization. World health statistics 2010. URL: https://www.who.int/whosis/whostat/2010/en/ [accessed 2020-05-15]

2. Frey U, Brodbeck T, Majumdar A, Taylor DR, Town GI, Silverman M, et al. Risk of severe asthma episodes predicted from fluctuation analysis of airway function. Nature 2005 Dec 01;438(7068):667-670. [doi: 10.1038/nature04176] [Medline: 16319891]
3. Raoufy MR, Ghafari T, Darooei R, Nazari M, Mahdaviani SA, Eslaminejad AR, et al. Classification of Asthma Based on Nonlinear Analysis of Breathing Pattern. PLoS ONE 2016 Jan 29;11(1):e0147976. [doi: 10.1371/journal.pone.0147976]
4. Kesten S, Maleki-Yazdi R, Sanders BR, Wells JA, McKillop SL, Chapman KR, et al. Respiratory rate during acute asthma. Chest 1990 Jan;97(1):58-62. [doi: 10.1378/chest.97.1.58] [Medline: 2403901]
5. Gobbi A, Dellacá RL, King G, Thamrin C. Toward Predicting Individual Risk in Asthma Using Daily Home Monitoring of Resistance. Am J Respir Crit Care Med 2017 Jan 15;195(2):265-267. [doi: 10.1164/rccm.201603-0537LE] [Medline: 28084825]
6. Bergese SD, Mestek ML, Kelley SD, McIntyre R, Uribe AA, Sethi R, et al. Multicenter Study Validating Accuracy of a Continuous Respiratory Rate Measurement Derived From Pulse Oximetry. Anesthesia & Analgesia 2017;124(4):1153-1159. [doi: 10.1213/ane.0000000000001852]
7. Charlton PH, Bonnici T, Tarassenko L, Clifton DA, Beale R, Watkinson PJ. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. Physiol Meas 2016 Apr 30;37(4):610-626 [FREE Full text] [doi: 10.1088/0967-3334/37/4/610] [Medline: 27027672]
8. Addison PS. Respiratory effort from the photoplethysmogram. Med Eng Phys 2017 Mar;41:9-18 [FREE Full text] [doi: 10.1016/j.medengphy.2016.12.010] [Medline: 28126420]
9. Bailon R, Sornmo L, LagunaCG-derived RFE. Advanced methods and tools for ECG data analysis. the University of Michigan: Artech House; 2006:E.
10. Meredith DJ, Clifton D, Charlton P, Brooks J, Pugh CW, Tarassenko L. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. J Med Eng Technol 2012 Jan 20;36(1):1-7. [doi: 10.3109/03091902.2011.638965] [Medline: 22185462]
11. Prinable J, Jones P, Thamrin C, McEwan A. Using a recurrent neural network to derive tidal volume from a photoplethsmograph. 2007 Dec 13 Presented at: IEEE Life Sciences Conference; 2017; Sydney, Australia p. 218-221. [doi: 10.1109/lsc.2017.8268182]
12. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997 Nov;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735]
13. Myles P, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. Br J Anaesth 2007 Sep;99(3):309-311 [FREE Full text] [doi: 10.1093/bja/aem214] [Medline: 17702826]
14. Choi A, Shin H. Photoplethysmography sampling frequency: pilot assessment of how low can we go to analyze pulse rate variability with reliability? Physiol Meas 2017 Mar 23;38(3):586-600. [doi: 10.1088/1361-6579/aa5efa] [Medline: 28169836]
15. Charlton PH, Bonnici T, Tarassenko L, Alastruey J, Clifton DA, Beale R, et al. Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants. Physiol Meas 2017 May 03;38(5):669-690. [doi: 10.1088/1361-6579/aa670e] [Medline: 28296645]
16. Addison P, Watson J. Secondary wavelet feature decoupling (SWFD) and its use in detecting patient respiration from the photo- plethysmogram. 2003 Sep 17 Presented at: Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. Vol. 3. IEEE, 2003; 2003; Cancun, Mexico. [doi: 10.1109/iembs.2003.1280448]
17. Karlen W, Raman S, Ansermino JM, Dumont GA. Multiparameter Respiratory Rate Estimation From the Photoplethysmogram. IEEE Trans. Biomed. Eng 2013 Jul;60(7):1946-1953. [doi: 10.1109/tbme.2013.2246160]
18. Ruangsuwana R, Velikic G, Bocko M. Methods to extract respiration information from ECG signals. 2010 Jun 28 Presented at: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing; March 14-19, 2010; Dallas, TX. [doi: 10.1109/ICASSP.2010.5495584]
19. Lázaro J, Gil E, Bailón R, Mincholé A, Laguna P. Deriving respiration from photoplethysmographic pulse width. Med Biol Eng Comput 2013 Feb;51(1-2):233-242. [doi: 10.1007/s11517-012-0954-0] [Medline: 22996834]
20. Cui Z, Ke R, Pu Z, Wang Y. Deep Stacked Bidirectional and Uni-directional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. 2016 Presented at: In 6th International Workshop on Urban Computing (UrbComp); 2017; Chicago, USA.
21. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 2014;15(1):1929-1958.
22. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist. Surv 2010;4:40-79. [doi: 10.1214/09-ss054]
23. Prinable JB, Foster JM, McEwan AL, Young PM, Tovey E, Thamrin C. Motivations and Key Features for a Wearable Device for Continuous Monitoring of Breathing: A Web-Based Survey. JMIR Biomed Eng 2017 Jun 26;2(1):e1. [doi: 10.2196/biomedeng.7143]
24. Choi C, Ko B, Lee J, Yoon S, Kwon U, Kim S, et al. PPG pulse direction determination algorithm for PPG waveform inversion by wrist rotation. 2017 Sep 14 Presented at: In Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE; 2017; Seogwipo, South Korea p. 4090-4093. [doi: 10.1109/embc.2017.8037755]

25. Wolling F, Van Laerhoven K. Fewer Samples for a Longer Life Span: Towards Long-Term Wearable PPG Analysis. 2018 Sep 20 Presented at: In Proceedings of the 5th international Workshop on Sensor-based Activity RecognitionInteraction (p. 5). ACM; 2018; Chicago, USA. [doi: 10.1145/3266157.3266209]
26. Chang A, Culurciello E. Hardware accelerators for recurrent neural networks on FPGA. 2017 Presented at: IEEE International symposium on circuits and systems (ISCAS); 2017; Baltimore, USA p. 1-4. [doi: 10.1109/iscas.2017.8050816]

Abbreviations

BPM: breaths per minute.
BR: breathing rate.
I:E: inspiration:expiration ratio.
IBI: interbreath interval.
GPU: graphics processing unit.
LoA: limits of agreement
LSTM: long short-term memory.
PPG: photoplethysmogram.
RTV: relative tidal volume.
Texp: expiration period.
Tinsp: inspiration time.
TVW: tidal volume waveform.

Edited by G Eysenbach; submitted 18.02.19; peer-reviewed by P Charlton, B Askarian, F Tabei; comments to author 18.05.19; revised version received 26.12.19; accepted 22.02.20; published 31.07.20.

Please cite as:
Prinable J, Jones P, Boland D, Thamrin C, McEwan A
Derivation of Breathing Metrics From a Photoplethysmogram at Rest: Machine Learning Methodology
JMIR Mhealth Uhealth 2020;8(7):e13737
URL: http://mhealth.jmir.org/2020/7/e13737/
doi:10.2196/13737
PMID:32735229

Original Paper

A Novel Smartphone App for the Measurement of Ultra–Short-Term and Short-Term Heart Rate Variability: Validity and Reliability Study

Yung-Sheng Chen[1], PhD; Wan-An Lu[2], MD, PhD; Jeffrey C Pagaduan[3], MSc; Cheng-Deng Kuo[4,5], MD, PhD

[1]Department of Exercise and Health Sciences, University of Taipei, Taipei, Taiwan

[2]Institute of Cultural Asset and Reinvention, Fo-Guang University, Yilan, Taiwan

[3]College of Health and Medicine, School of Health Sciences, University of Tasmania, Launceston, Australia

[4]Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan

[5]Tanyu Research Laboratory, Taipei, Taiwan

Corresponding Author:
Cheng-Deng Kuo, MD, PhD
Department of Medical Research
Taipei Veterans General Hospital
Number 201, Section 2, Shipai Rd
Beitou District
Taipei, 112
Taiwan
Phone: 886 932981776
Email: cdkuo23@gmail.com

Abstract

Background: Smartphone apps for heart rate variability (HRV) measurement have been extensively developed in the last decade. However, ultra–short-term HRV recordings taken by wearable devices have not been examined.

Objective: The aims of this study were the following: (1) to compare the validity and reliability of ultra–short-term and short-term HRV time-domain and frequency-domain variables in a novel smartphone app, Pulse Express Pro (PEP), and (2) to determine the agreement of HRV assessments between an electrocardiogram (ECG) and PEP.

Methods: In total, 60 healthy adults were recruited to participate in this study (mean age 22.3 years [SD 3.0 years], mean height 168.4 cm [SD 8.0 cm], mean body weight 64.2 kg [SD 11.5 kg]). A 5-minute resting HRV measurement was recorded via ECG and PEP in a sitting position. Standard deviation of normal R-R interval (SDNN), root mean square of successive R-R interval (RMSSD), proportion of NN50 divided by the total number of RR intervals (pNN50), normalized very-low–frequency power (nVLF), normalized low-frequency power (nLF), and normalized high-frequency power (nHF) were analyzed within 9 time segments of HRV recordings: 0-1 minute, 1-2 minutes, 2-3 minutes, 3-4 minutes, 4-5 minutes, 0-2 minutes, 0-3 minutes, 0-4 minutes, and 0-5 minutes (standard). Standardized differences (ES), intraclass correlation coefficients (ICC), and the Spearman product-moment correlation were used to compare the validity and reliability of each time segment to the standard measurement (0-5 minutes). Limits of agreement were assessed by using Bland-Altman plot analysis.

Results: Compared to standard measures in both ECG and PEP, pNN50, SDNN, and RMSSD variables showed trivial ES (<0.2) and very large to nearly perfect ICC and Spearman correlation coefficient values in all time segments (>0.8). The nVLF, nLF, and nHF demonstrated a variation of ES (from trivial to small effects, 0.01-0.40), ICC (from moderate to nearly perfect, 0.39-0.96), and Spearman correlation coefficient values (from moderate to nearly perfect, 0.40-0.96). Furthermore, the Bland-Altman plots showed relatively narrow values of mean difference between the ECG and PEP after consecutive 1-minute recordings for SDNN, RMSSD, and pNN50. Acceptable limits of agreement were found after consecutive 3-minute recordings for nLF and nHF.

Conclusions: Using the PEP app to facilitate a 1-minute ultra–short-term recording is suggested for time-domain HRV indices (SDNN, RMSSD, and pNN50) to interpret autonomic functions during stabilization. When using frequency-domain HRV indices (nLF and nHF) via the PEP app, a recording of at least 3 minutes is needed for accurate measurement.

(JMIR Mhealth Uhealth 2020;8(7):e18761) doi:10.2196/18761

KEYWORDS

heart rate variability; smartphone; reproducibility; limits of agreement; autonomic nervous function

Introduction

Background

Smartphone apps are recognized as convenient tools for our daily life activities in modern society. For health and fitness issues, there is an increasing number of smartphone users that utilize multiple free mobile phone apps to assess biosignals [1,2], psychological functions [3,4], and social behaviors in daily routines. Specific to healthy lifestyle promotion for cardiovascular functions, using a smartphone or smartwatch to monitor autonomic nervous system activities through heart rate (HR) and HR variability (HRV) is accessible and economical [5,6].

HRV is a physiological marker of cardiac autonomic responses that can be detected by recording heartbeat intervals over time. Assessment of daily HRV can provide useful information for understanding cardiac health with regards to labor force workload [7], mental conditions [8,9], and fitness status [10,11]. In general, the conventional methodology involves recording a 5-minute short-term HRV measurement, followed by a 5-minute stabilization [12].

Ultra–Short-Term HRV Studies

Recently, ultra–short-term recordings for HRV assessment have received notable attention in cardiovascular medicine [13-15], metabolic disease [16], cognitive function [8,9], exercise testing [17-19], and sports training [11,20] studies due to the time efficiency it offers to both patients and practitioners. Ultra–short-term recording only requires R-R intervals of less than 60 seconds. Excellent limits of agreement and reproducibility of 1-minute ultra-short recordings of root mean square of successive R-R intervals (RMSSD) measurements were observed during a 5-minute stabilization period in an athletic population [11,21]. However, the methodological considerations of ultra–short-term HRV assessment have not been extensively explored in the literature. For example, a shorter time segment of less than 1 minute tended to increase measurement errors when RMSSD was log-transformed (lnRMSSD) [18].

Study Objectives

Today, several HRV smartphone apps have been developed to evaluate autonomic health by using photoplethysmography [19,22,23]. However, the compatibility of photoplethysmographic detection is limited by physical contacts between recording locations and mobile sensors. Thus, our research group recently developed a free mobile app, Pulse Express Pro (PEP), which is compatible with wearable HR sensors and has Bluetooth functionality. The wireless app might provide an option to clients and practitioners using mobile phone–based HRV assessment. Therefore, the first aim of this study was to compare the degree of validity and reliability of ultra–short-term and short-term HRV recordings of the time-domain (standard deviation of normal R-R intervals [SDNN], RMSSD, and the proportion of NN50 divided by the total number of RR intervals [pNN50]) and frequency-domain (normalized very-low–frequency power [nVLF], normalized low-frequency power [nLF], and normalized high-frequency power [nHF]) variables with standard 5-minute assessment using a novel smartphone app, PEP. The second aim of this study was to determine the agreement of ultra–short-term and short-term HRV assessments by electrocardiogram (ECG) and PEP. We hypothesized that ultra–short-term HRV indices would show less valid and reliable measurements than that of short-term HRV indices for frequency-domain variables but not for time-domain variables.

Methods

Recruitment

In total, 60 healthy adults were recruited for this study (aged 22.3 [3.0] years; 168.4 [SD 8.0] cm tall; body weight: 64.2 [11.5] kg). Inclusion criteria were healthy adults aged between 20 and 30 years. Exclusion criteria included current neurological, cardiovascular, and metabolic diseases. All participants signed an informed consent form and were familiarized with experimental procedures. The participants were requested to avoid vigorous exercise 24 hours before visits and to avoid caffeine-containing substances and smoking 2 hours before the experiments. This study was approved by the Human Ethics Committee of University of Taipei (IRB-2019-005) and was conducted according to the Declaration of Helsinki and its later amendments.

Sample size was determined based on convenience and post hoc power analysis using dependent *t* tests carried out in G*Power [24]. A sample size of 60 participants demonstrated a 97% chance of obtaining a significant outcome measure at $P<.05$ with a moderate effect size ($d=0.50$).

Experimental Procedure

The height and weight of each participant were measured during the first visit using a portable stadiometer (Seca 213, SECA) and electrical weight scale (Xyfwt382, Teco). At the second visit, 5-minute resting HRV data were collected in a sitting position. The ECG signals with conventional lead II arrangement were set for reference, while a portable Polar HR monitor (H7, Polar Electro) was placed on the participant's chest for HR detection (Figure 1). A smartphone (PRA LX2, Huawei) with the PEP app [25] was used to record HRV signals via Bluetooth. The participants were instructed to breathe spontaneously during the HRV recording. The measurements were taken in a quiet and spacious room between 8 AM and 12 PM. Room temperature and humidity were controlled at around 25 °C and 70%-80%, respectively.

Figure 1. Illustration of the experimental setting and testing position.

HRV Recording

All participants were requested to maintain a sitting position during ECG recording. A multichannel biosignal recorder (MP160, Biopac Systems) with conventional lead II arrangement (MEC110C, Biopac Systems) was set for ECG recordings, while a telemetric HR monitoring device was used to record the resting HRV (H7, Polar Electro) via a customized smartphone app, PEP. The sampling rate of the ECG recording was set at 1000 Hz. The HRV data was exported to Google Drive and extracted to a personal laptop for data analysis. Kubios HRV Premium analysis software (Version 3.2; Kubios) was used to calculate SDNN, RMSSD, pNN50, nVLF, nLF, and nHF parameters. The SDNN, RMSSD, and pNN50 were calculated by using the standard formulas for time-domain analysis [12]. In addition, the power spectra of RR intervals were calculated by means of Fast Fourier Transformation (FFT) for frequency-domain analysis. The bands of VLF, LF, and HF ranges were set as 0-0.04 Hz, 0.04-0.15 Hz, and 0.15-0.4 Hz, respectively [12]. The normalized powers of VLF, LF, and HF were used as the autonomic indices of the participants. The formulas to calculate the normalized powers of VLF, LF, and HF bands were as follows [26-28], with nu standing for normalized unit:

$$nVLF[nu] = VLF\ (ms^2) / total\ power\ (ms^2) \times 100\ (\mathbf{1})$$

$$nLF[nu] = LF\ (ms^2) / total\ power\ (ms^2) \times 100\ (\mathbf{2})$$

$$nHF[nu] = HF\ (ms^2) / total\ power\ (ms^2) \times 100\ (\mathbf{3})$$

Strong artefact correction and smoothing priors set at 500Λ were used for HRV analysis to minimize the interference from Bluetooth transmission and the artefact resulting from physical contact between the chest strap and the skin [29,30]. The time segments of HRV recordings were divided into 0-1 minute, 1-2 minutes, 2-3 minutes, 3-4 minutes, and 4-5 minutes for ultra–short-term HRV recordings and 0-2 minutes, 0-3 minutes, 0-4 minutes, and 0-5 minutes (standard) for short-term HRV recordings.

Statistical Analysis

Statistical analyses were conducted using SPSS Statistics (Version 25.0; IBM Corp) and Microsoft Excel 2013 (Microsoft Corp). Descriptive data of the measured variables are presented as median and interquartile range (25%-75%). Magnitude of difference and agreement of HRV indices in all time segments (with the 5-minute criterion as a reference) were analyzed by using the standardized differences of variables (effect size: ES), Cohen *d*. The ES was interpreted as trivial (0.0-0.2), small (0.2-0.6), moderate (0.6-1.2), large (1.2-2.0), and very large (>2.0) [31]. In terms of validity and reliability between the ECG and PEP assessments, intraclass correlation coefficients (ICC) with a two-way random model and single measure were used to determine the relative values of reliability. The ICC values were expressed as small (0.0-0.3), moderate (0.31-0.49), large (0.50-0.69), very large (0.70-0.89), and nearly perfect (0.9-1.0) [31]. The correlation coefficient between the ECG and PEP was assessed by using the Spearman rank correlation (*r*). The level of the correlation coefficients was determined as trivial ($r<0.1$), small ($0.1<r<0.3$), moderate ($0.3<r<0.5$), high ($0.5<r<0.7$), very high ($0.7<r<0.9$), nearly perfect ($r>0.9$), and perfect ($r=1$) [31]. Lastly, a Bland-Altman plot was used to evaluate the upper and lower limits of agreement among time segments of the HRV indices as determined by the ECG and PEP [32].

Results

Standardized Differences and Limits of Agreement

The descriptive information and standardized differences of HRV indices for all time segments of the ECG and PEP measurements are presented in Tables 1 and 2. The results showed trivial ES in all time segments of the SDNN, RMSSD, and pNN50, compared to the 0-5–minute standard measurement. In contrast, a variation of ES from trivial to small effect was found in the nVLF, nLF, and nHF variables.

Table 1. Median and interquartile range (25%-75%) of time-domain and frequency-domain heart rate variability parameters in different time segments of the electrocardiogram and Pulse Express PRO measurements[a].

Device and time segment (minutes)	SDNN, ms	RMSSD, ms	pNN50, %	nVLF, nu	nLF, nu	nHF, nu
	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)
ECG						
0-5 (standard)	41.63 (30.24-53.93)	39.38 (27.73-50.56)	20.56 (6.89-35.12)	3.82 (1.80-6.85)	54.61 (41.82-65.32)	41.02 (26.17-51.47)
0-1	37.83 (28.75-55.43)	37.76 (30.40-58.16)	19.86 (6.88-40.09)	2.92 (1.30-7.62)	47.57 (33.56-62.89)	46.02 (31.52-61.12)
1-2	42.36 (30.27-60.03)	40.89 (26.92-55.40)	20.71 (6.47-37.20)	3.23 (1.80-7.10)	50.13 (37.36-71.45)	42.82 (23.30-59.14)
2-3	40.08 (29.24-55.39)	38.20 (27.85-51.59)	20.32 (4.84-36.24)	5.09 (1.99-8.19)	47.92 (33.34-63.49)	45.22 (28.46-61.41)
3-4	39.59 (30.73-54.09)	37.84 (26.30-50.93)	19.64 (4.09-32.36)	3.39 (1.20-6.57)	45.97 (31.62-65.00)	49.34 (30.12-64.34)
4-5	39.13 (28.80-48.01)	35.27 (26.70-47.15)	15.53 (5.12-33.94)	3.77 (1.74-7.87)	49.73 (27.99-66.25)	44.25 (24.85-63.73)
0-2	41.79 (29.81-58.74)	40.70 (29.05-56.91)	23.49 (6.37-35.78)	3.42 (1.61-6.89)	43.04 (31.61-61.63)	51.65 (26.64-62.59)
0-3	40.87 (30.74-56.58)	40.06 (28.23-52.88)	21.84 (5.82-34.26)	3.64 (1.36-6.67)	51.41 (38.21-67.54)	42.00 (27.84-57.41)
0-4	41.59 (30.63-57.47)	40.23 (27.66-52.04)	21.59 (7.11-35.49)	3.51 (1.91-6.62)	53.12 (43.26-66.91)	39.57 (27.73-51.41)
PEP						
0-5 (standard)	41.51 (30.33-53.85)	39.09 (27.79-50.63)	20.20 (7.08-35.16)	3.90 (1.80-6.98)	54.65 (42.63-65.55)	41.92 (26.95-51.81)
0-1	39.52 (28.65-55.03)	38.72 (31.25-57.80)	20.24 (7.91-38.26)	2.80 (1.20-7.78)	48.20 (32.93-62.48)	45.34 (32.66-60.23)
1-2	41.80 (30.20-60.71)	40.98 (27.00-55.54)	22.84 (5.10-36.78)	3.35 (1.55-6.54)	50.16 (35.69-71.22)	42.74 (24.45-59.85)
2-3	40.44 (30.51-55.29)	38.26 (27.54-51.65)	21.19 (4.82-36.46)	4.76 (1.91-8.75)	48.07 (33.45-62.41)	42.67 (27.55-62.10)
3-4	39.31 (30.53-54.32)	38.73 (26.70-51.73)	20.59 (4.95-32.33)	3.43 (1.13-6.25)	46.18 (34.18-64.19)	48.33 (28.20-62.80)
4-5	39.01 (28.76-47.35)	36.30 (26.64-47.35)	17.29 (5.11-33.81)	4.38 (1.63-7.85)	48.97 (26.83-64.78)	44.56 (24.59-64.24)
0-2	42.90 (29.78-59.09)	40.90 (29.25-57.07)	24.67 (6.62-37.26)	3.46 (1.71-7.22)	42.11 (31.50-59.15)	52.88 (26.82-64.49)
0-3	41.62 (31.14-56.75)	39.99 (28.09-53.32)	22.13 (5.89-35.72)	3.84 (1.33-6.55)	51.19 (38.00-67.06)	41.64 (28.85-57.86)
0-4	41.43 (30.74-57.72)	40.18 (27.79-52.08)	21.78 (6.92-35.22)	3.46 (1.92-6.70)	53.55 (42.65-67.67)	39.36 (29.08-51.90)

[a]ECG: electrocardiogram; ms: milliseconds; nHF: normalized high-frequency power; nLF: normalized low-frequency power; nu: normalized unit; nVLF: normalized very-low–frequency power; PEP: Pulse Express PRO; pNN50: proportion of NN50 divided by the total number of RR intervals; RMSSD: root mean square of successive R-R intervals; SDNN: standard deviation of normal R-R intervals.

Table 2. Standardized differences (95% CI) of time-domain and frequency-domain heart rate variability parameters in different time segments of the electrocardiogram and Pulse Express PRO measurements compared with the 0-5–minute standard[a].

Device and time segment (minutes)	SDNN (95% CI)	RMSSD (95% CI)	pNN50 (95% CI)	nVLF (95% CI)	nLF (95% CI)	nHF (95% CI)
ECG						
0-1	0.01 (–0.35 to 0.36)	–0.07 (–0.43 to 0.29)	–0.10 (–0.46 to 0.26)	–0.03 (–0.39 to 0.33)	0.40 (0.04 to 0.76)	–0.38 (–0.74 to –0.02)
1-2	–0.12 (–0.48 to 0.24)	–0.04 (–0.39 to 0.32)	–0.09 (–0.44 to 0.27)	0.05 (–0.31 to 0.41)	0.08 (–0.28 to 0.44)	–0.09 (–0.45 to 0.27)
2-3	0.05 (–0.31 to 0.41)	0.05 (–0.31 to 0.41)	0.01 (–0.34 to 0.37)	–0.01 (–0.61 to 0.11)	0.34 (–0.02 to 0.71)	–0.26 (–0.62 to 0.10)
3-4	0.05 (–0.31 to 0.40)	0.05 (–0.31 to 0.41)	0.04 (–0.32 to 0.40)	–0.04 (–0.40 to 0.32)	0.29 (–0.07 to 0.65)	–0.27 (–0.63 to 0.08)
4-5	0.12 (–0.23 to 0.48)	0.08 (–0.28 to 0.44)	0.11 (–0.24 to 0.47)	–0.20 (–0.56 to 0.16)	0.25 (–0.11 to 0.61)	–0.20 (–0.56 to 0.16)
0-2	–0.07 (–0.43 to 0.29)	0.06 (–0.41 to 0.30)	–0.09 (–0.45 to 0.26)	–0.02 (–0.37 to 0.34)	0.35 (0.00 to 0.72)	–0.34 (–0.70 to 0.02)
0-3	0.04 (–0.40 to 0.32)	–0.03 (–0.38 to 0.33)	–0.06 (–0.41 to 0.30)	0.14 (–0.22 to 0.49)	0.08 (–0.28 to 0.44)	–0.11 (–0.46 to 0.25)
0-4	0.03 (–0.38 to 0.33)	–0.01 (–0.37 to 0.34)	–0.03 (–0.39 to 0.33)	0.06 (–0.30 to 0.42)	0.01 (–0.35 to 0.37)	–0.02 (–0.38 to 0.34)
PEP						
0-1	0.01 (–0.35 to 0.37)	–0.07 (–0.43 to 0.29)	–0.10 (–0.46 to 0.26)	0.01 (–0.35 to 0.36)	0.38 (0.02 to 0.74)	–0.37 (–0.73 to –0.01)
1-2	–0.12 (–0.48 to 0.24)	–0.03 (–0.39 to 0.32)	–0.08 (–0.44 to 0.27)	0.05 (–0.31 to 0.41)	0.09 (–0.26 to 0.45)	–0.10 (–0.46 to 0.26)
2-3	0.05 (–0.31 to 0.41)	0.06 (–0.30 to 0.42)	0.03 (–0.33 to 0.39)	–0.25 (–0.61 to 0.10)	0.33 (–0.03 to 0.69)	–0.24 (–0.60 to 0.11)
3-4	0.04 (–0.31 to 0.40)	0.04 (–0.31 to 0.40)	0.04 (–0.32 to 0.40)	–0.07 (–0.43 to 0.29)	0.25 (–0.11 to 0.61)	–0.23 (–0.59 to 0.13)
4-5	0.12 (–0.24 to 0.48)	0.08 (–0.28 to 0.44)	0.10 (–0.26 to 0.46)	–0.23 (–0.59 to 0.13)	0.30 (–0.06 to 0.66)	–0.23 (–0.59 to 0.13)
0-2	–0.06 (–0.42 to 0.29)	–0.06 (–0.41 to 0.30)	–0.09 (–0.45 to 0.27)	–0.03 (–0.39 to 0.33)	0.38 (0.03 to 0.75)	–0.36 (–0.73 to 0.00)
0-3	–0.03 (–0.39 to 0.33)	–0.02 (–0.38 to 0.33)	–0.05 (–0.41 to 0.31)	0.13 (–0.23 to 0.49)	0.10 (–0.26 to 0.45)	–0.12 (–0.48 to 0.24)
0-4	–0.02 (–0.38 to 0.34)	–0.01 (–0.37 to 0.34)	–0.03 (–0.38 to 0.33)	0.06 (–0.30 to 0.42)	0.02 (–0.34 to 0.37)	–0.03 (–0.39 to 0.33)

[a]ECG: electrocardiogram; nHF: normalized high-frequency power; nLF: normalized low-frequency power; nVLF: normalized very-low–frequency power; PEP: Pulse Express PRO; pNN50: proportion of NN50 divided by the total number of RR intervals; RMSSD: root mean square of successive R-R intervals; SDNN: standard deviation of normal R-R intervals.

In Table 3, the Bland-Altman analysis demonstrated relatively small bias in all comparisons of the SDNN, RMSSD, pNN50, and nVLF. In contrast, a relatively small bias in the nLF and nHF variables occurred during short-term recordings of 0-3 minutes and 0-4 minutes.

Table 3. Limits of agreement (± 1.96*SD) of time-domain and frequency-domain heart rate variability parameters in different time segments of the electrocardiogram and Pulse Express PRO measurements compared with the 0-5–minute standard[a].

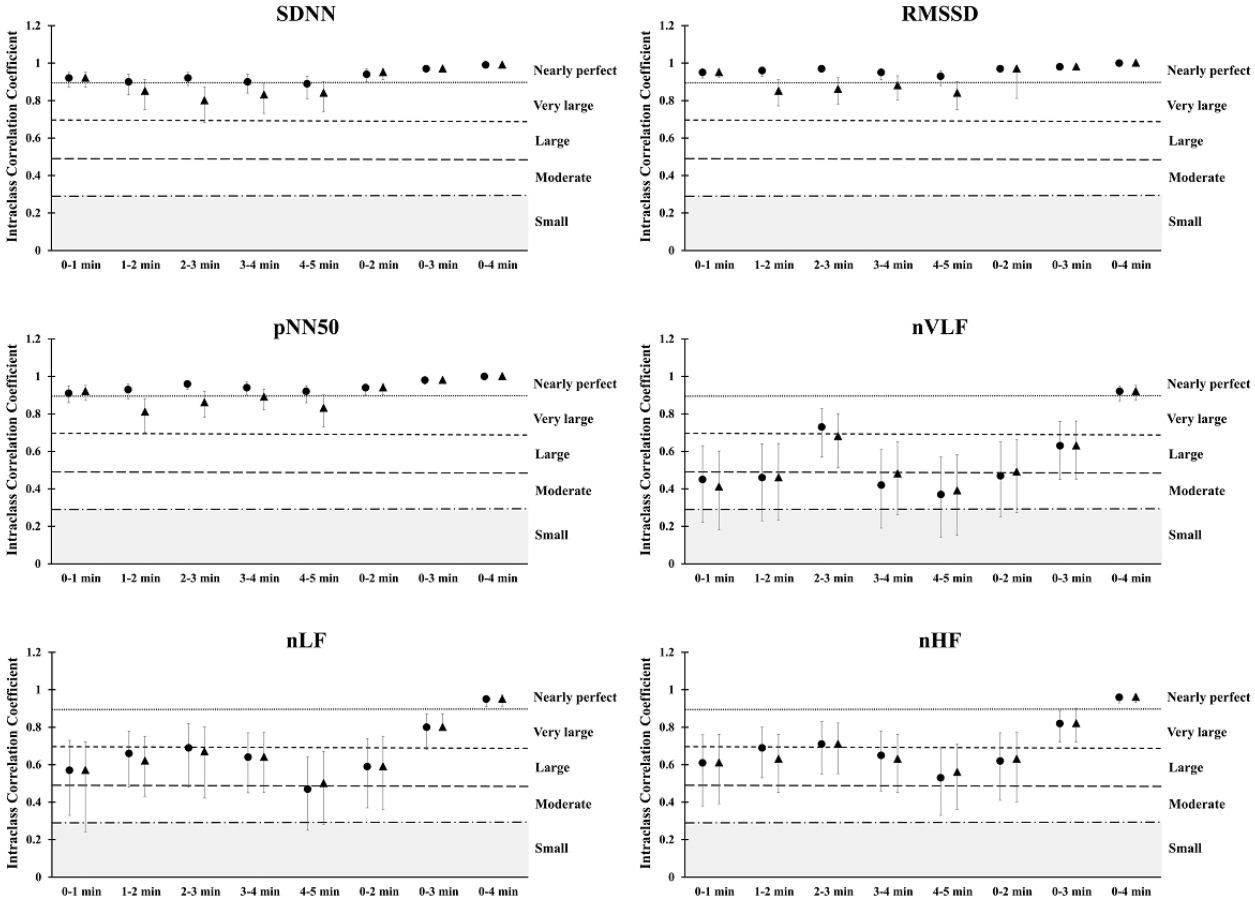
Device and time segment (minutes)	SDNN (± 1.96*SD)	RMSSD (± 1.96*SD)	pNN50 (± 1.96*SD)	nVLF (± 1.96*SD)	nLF (± 1.96*SD)	nHF (± 1.96*SD)
ECG						
0-1	0.01 (–14.23 to 14.43)	–1.40 (–13.48 to 10.69)	–1.93 (–17.10 to 13.24)	–0.15 (–9.88 to 9.58)	7.75 (–25.66 to 41.16)	–7.63 (–40.54 to 25.29)
1-2	–2.31 (–18.53 to 13.96)	–0.72 (–11.84 to 10.40)	–1.61 (–15.16 to 11.94)	0.21 (–8.49 to 8.92)	1.59 (–30.41 to 33.59)	–1.80 (–32.81 to 29.21)
2-3	0.83 (–12.63 to 14.30)	0.93 (–8.46 to 10.32)	0.26 (–9.93 to 10.44)	–1.30 (–8.59 to 6.00)	6.63 (–21.17 to 34.44)	–5.32 (–33.83 to 23.19)
3-4	0.83 (–14.72 to 16.37)	0.92 (–11.86 to 13.71)	0.70 (–11.20 to 12.60)	–0.20 (–10.39 to 10.00)	5.73 (–25.74 to 37.20)	–5.54 (–37.56 to 26.49)
4-5	2.18 (–13.70 to 18.05)	1.64 (–13.16 to 16.44)	2.04 (–11.47 to 15.56)	–1.00 (–11.78 to 9.82)	5.08 (–34.82 to 44.98)	–4.12 (–42.78 to 34.54)
0-2	–1.24 (–13.01 to 10.53)	–1.14 (–10.90 to 8.62)	–1.73 (–13.96 to 10.51)	–0.06 (–12.11 to 12.00)	7.07 (–26.77 to 40.92)	–7.03 (–40.37 to 26.32)
0-3	0.66 (–7.35 to 8.66)	0.52 (–6.28 to 7.32)	–1.01 (–6.62 to 8.64)	–0.52 (–6.85 to 5.82)	–1.50 (–24.85 to 21.85)	2.03 (–20.03 to 24.09)
0-4	–0.44 (–4.14 to 3.26)	–0.28 (–3.85 to 3.30)	–0.54 (–3.91 to 2.84)	0.24 (–2.78 to 3.26)	–0.15 (–10.98 to 11.28)	–0.38 (–11.00 to 10.23)
PEP						
0-1	0.20 (–13.41 to 13.81)	–1.35 (–13.06 to 10.36)	–1.86 (–16.16 to 12.43)	0.03 (–9.77 to 9.82)	7.32 (–26.30 to 40.94)	–7.37 (–40.44 to 25.70)
1-2	–2.24 (–22.67 to 18.91)	–0.70 (–21.94 to 20.54)	–1.60 (–24.24 to 21.04)	0.20 (–8.56 to 8.96)	1.84 (–32.13 to 35.80)	–2.04 (–36.20 to 32.12)
2-3	0.89 (–21.01 to 22.79)	1.18 (–18.49 to 20.84)	0.53 (–17.35 to 18.41)	–1.40 (–9.61 to 6.82)	6.35 (–22.36 to 35.06)	–4.94 (–34.00 to 24.11)
3-4	0.80 (–19.94 to 21.55)	0.88 (–18.35 to 20.10)	0.71 (–15.93 to 17.35)	–0.31 (–10.53 to 9.92)	4.93 (–26.76 to 36.62)	–4.61 (–37.63 to 28.42)
4-5	2.06 (–16.98 to 21.09)	1.58 (–20.36 to 23.52)	1.75 (–18.13 to 21.63)	–0.63 (–10.11 to 8.84)	5.87 (–31.94 to 43.69)	–4.74 (–41.70 to 32.23)
0-2	–1.16 (–12.68 to 11.81)	–1.12 (–10.86 to 8.62)	–1.69 (–13.91 to 10.52)	–0.99 (–10.56 to 8.57)	7.62 (–25.58 to 40.83)	–7.51 (–40.22 to 25.19)
0-3	–0.57 (–8.65 to 7.52)	–0.46 (–7.25 to 6.33)	–0.91 (–8.66 to 6.85)	–0.22 (–6.23 to 5.79)	1.78 (–21.56 to 25.12)	–2.29 (–24.52 to 19.94)
0-4	–0.38 (–4.28 to 3.52)	–0.25 (–3.74 to 3.24)	–0.45 (–3.87 to 2.97)	–0.08 (–2.41 to 2.24)	0.28 (–10.83 to 11.38)	0.51 (–11.16 to 10.14)

[a]ECG: electrocardiogram; nHF: normalized high-frequency power; nLF: normalized low-frequency power; nVLF: normalized very-low–frequency power; PEP: Pulse Express PRO; pNN50: proportion of NN50 divided by the total number of RR intervals; RMSSD: root mean square of successive R-R intervals; SDNN: standard deviation of normal R-R intervals.

Intraclass Correlation Coefficients

The results demonstrated similar outcomes for ICC values for the ECG and PEP measurements. The SDNN, RMSSD, and pNN50 ICC values were nearly perfect in all ultra–short-term and short-term records compared to the 0-5–minute standard ECG measurement (from very large to nearly perfect, 0.89-1.0). Furthermore, the time-domain variables of PEP were very large to nearly perfect for ultra–short-term recordings, except the 0-1–minute time segment (0.81-0.94). In terms of frequency-domain analysis, nearly perfect ICC values were found in the 0-4–minute time segment of the nVLF, nLF, and nHF (0.92-0.96). Very large ICC values were found in the 0-3–minute time segments for nLF and nHF (0.80-0.82). A broad range of ICC values was identified among the other comparisons (from moderate to very large, 0.37-0.71; Figure 2).

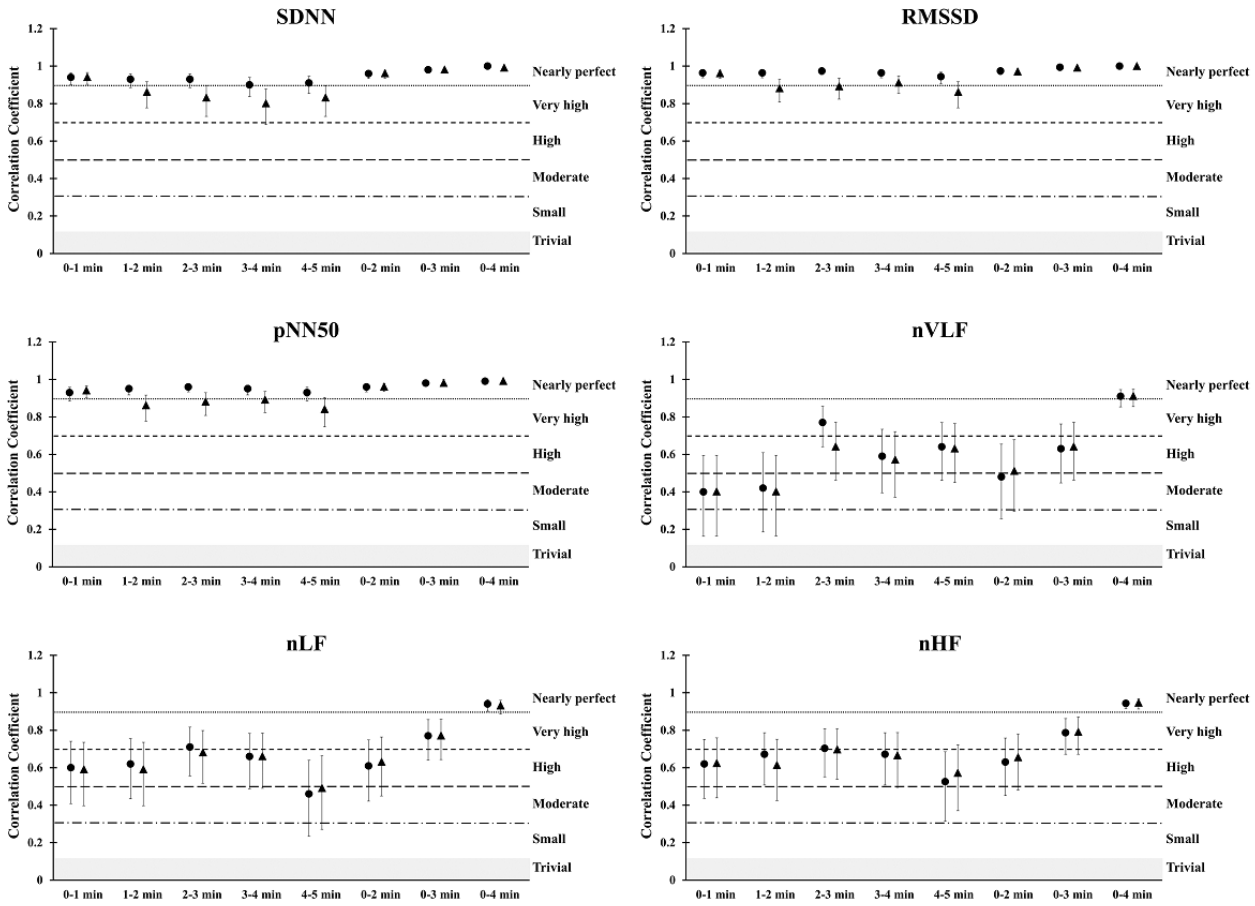
Figure 2. Intraclass correlation coefficients between the ECG and Pulse Express PRO measurements in ultra–short-term and short-term heart rate variability in time-domain and frequency-domain analyses. The grey area indicates low reliability. The black circle indicates the ECG recordings, while the black triangle indicates the Pulse Express PRO recordings. ECG: electrocardiogram.



Correlation Coefficient

Compared to the 0-5–minute standard measurement, the Spearman correlation coefficients were nearly perfect for the SDNN, RMSSD, and pNN50 variables in all time segments for the ECG measurements (0.90-1.0). Furthermore, the correlation coefficients were very large for the time-domain variables for ultra–short-term recordings using PEP (0.80-1.0), except for nearly perfect values for the 0-1–minute time segment. For frequency-domain analysis, a nearly perfect correlation coefficient was only found for 0-4–minute recordings (0.91-0.96). Furthermore, a very large correlation coefficient was found in the nLF and nHF 0-3–minute recordings (0.77-0.81). In contrast, a wide range of values was identified among the other comparisons (from moderate to very large, 0.40-0.77; Figure 3).

Figure 3. Spearmen correlation coefficients between the ECG and Pulse Express PRO measurements in ultra–short-term and short-term heart rate variability in time-domain and frequency-domain analyses. The grey area indicates trivial correlation coefficient values. The black circle indicates the ECG recordings, while the black triangle indicates the Pulse Express PRO recordings. ECG: electrocardiogram.



Bland-Altman Plots Comparing ECG and PEP Measurements

The Bland-Altman plots comparing the ECG and PEP measurements showed relatively narrow values of mean difference in all time segments (Figures 4-9). In addition, the Bland-Altman analysis found a narrow standard deviation for consecutive 2-minute recordings for SDNN, RMSSD, pNN50, and nVLF. In addition, acceptable limits of agreement were found after consecutive 3-minute recordings for nLF and nHF.

Figure 4. Bland-Altman analysis comparing the ECG and Pulse Express PRO measurements in ultra–short-term and short-term recordings of standard deviation of normal R-R intervals (SDNN). The solid line represents the mean difference and the upper and lower dashed lines represent the upper and lower limits of agreement (± 1.96*SD). ECG: electrocardiogram; PEP: Pulse Express PRO.

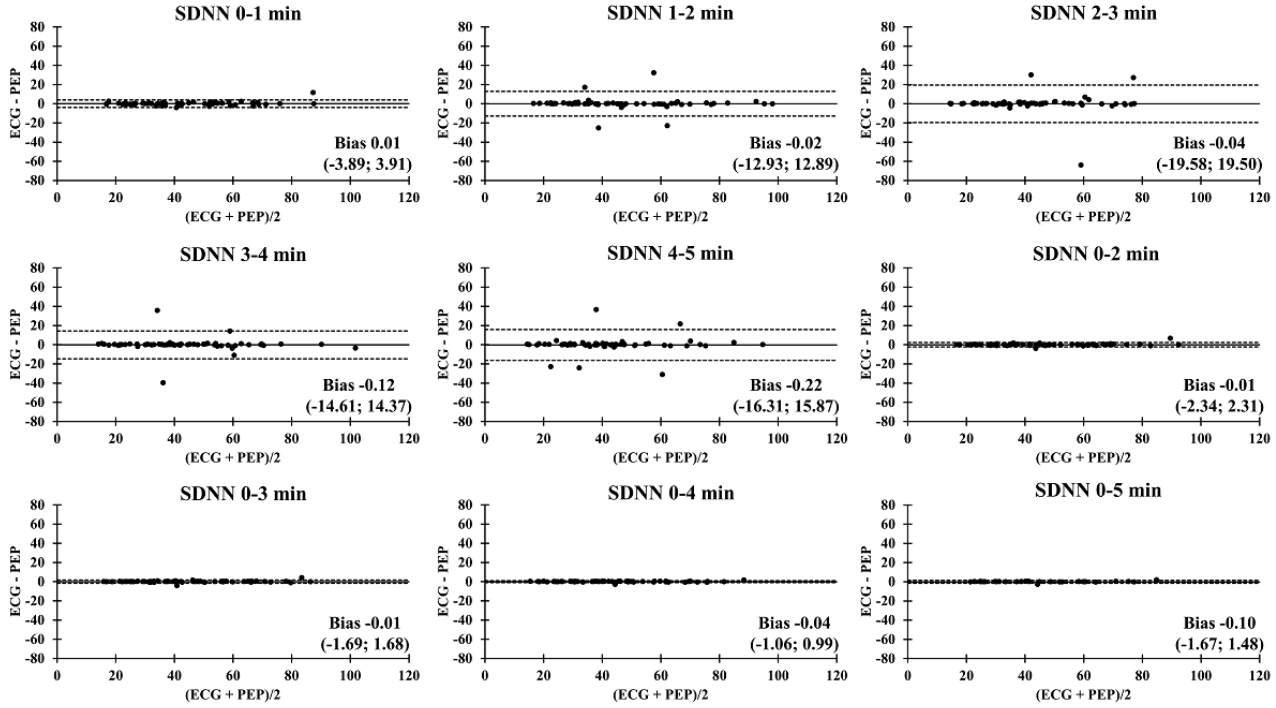


Figure 5. Bland-Altman analysis comparing the ECG and Pulse Express PRO measurements in ultra–short-term and short-term recordings of root mean square of successive R-R interval (RMSSD). The solid line represents the mean difference and the upper and lower dashed lines represent the upper and lower limits of agreement (± 1.96*SD). ECG: electrocardiogram; PEP: Pulse Express PRO.

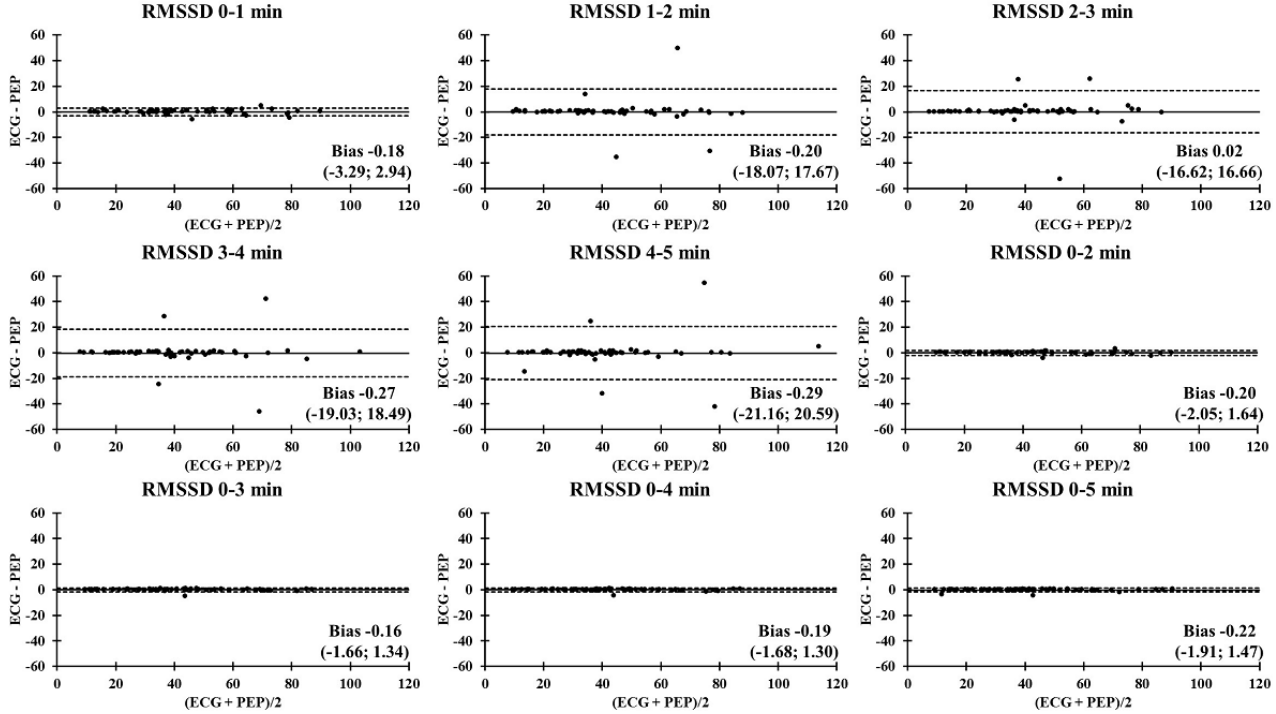


Figure 6. Bland-Altman analysis comparing the ECG and Pulse Express PRO measurements in ultra–short-term and short-term recordings of proportion of NN50 divided by the total number of RR intervals (pNN50). The solid line represents the mean difference and the upper and lower dashed lines represent the upper and lower limits of agreement (± 1.96*SD). ECG: electrocardiogram; PEP: Pulse Express PRO.

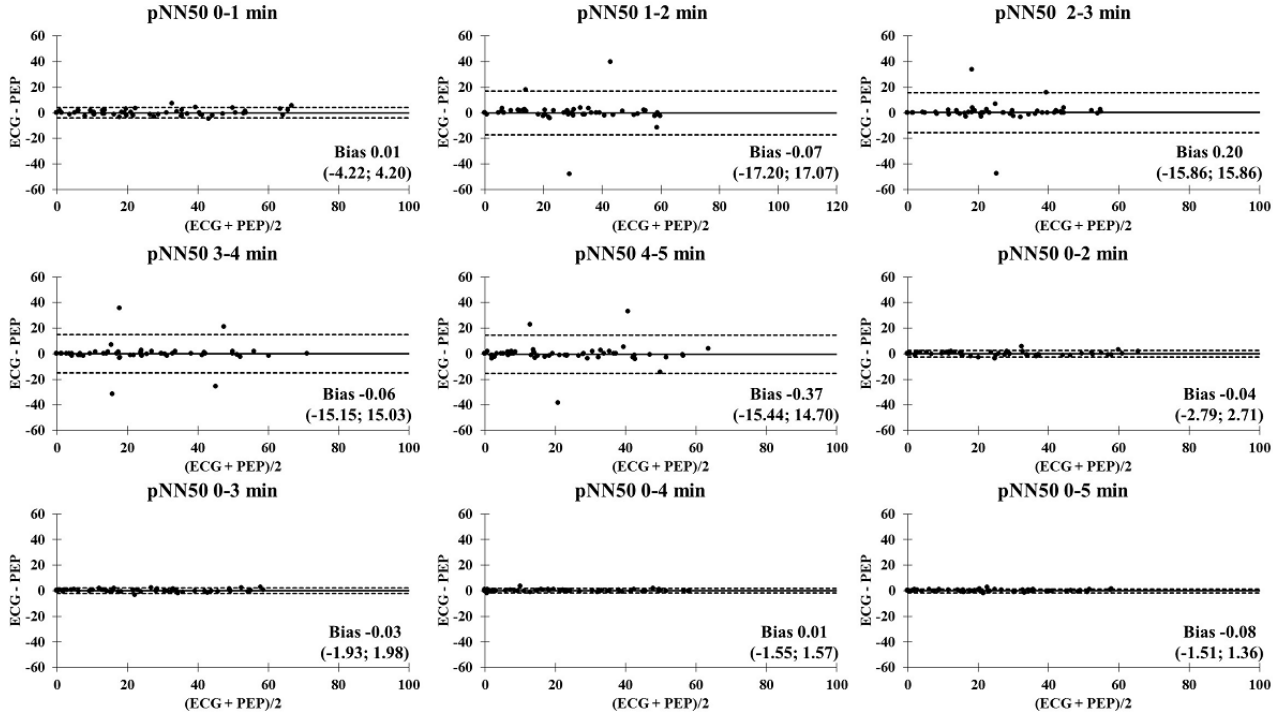


Figure 7. Bland-Altman analysis comparing the ECG and Pulse Express PRO measurements in ultra–short-term and short-term recordings of normalized very low frequency power (nVLF). The solid line represents the mean difference and the upper and lower dashed lines represent the upper and lower limits of agreement (± 1.96*SD). ECG: electrocardiogram; PEP: Pulse Express PRO.

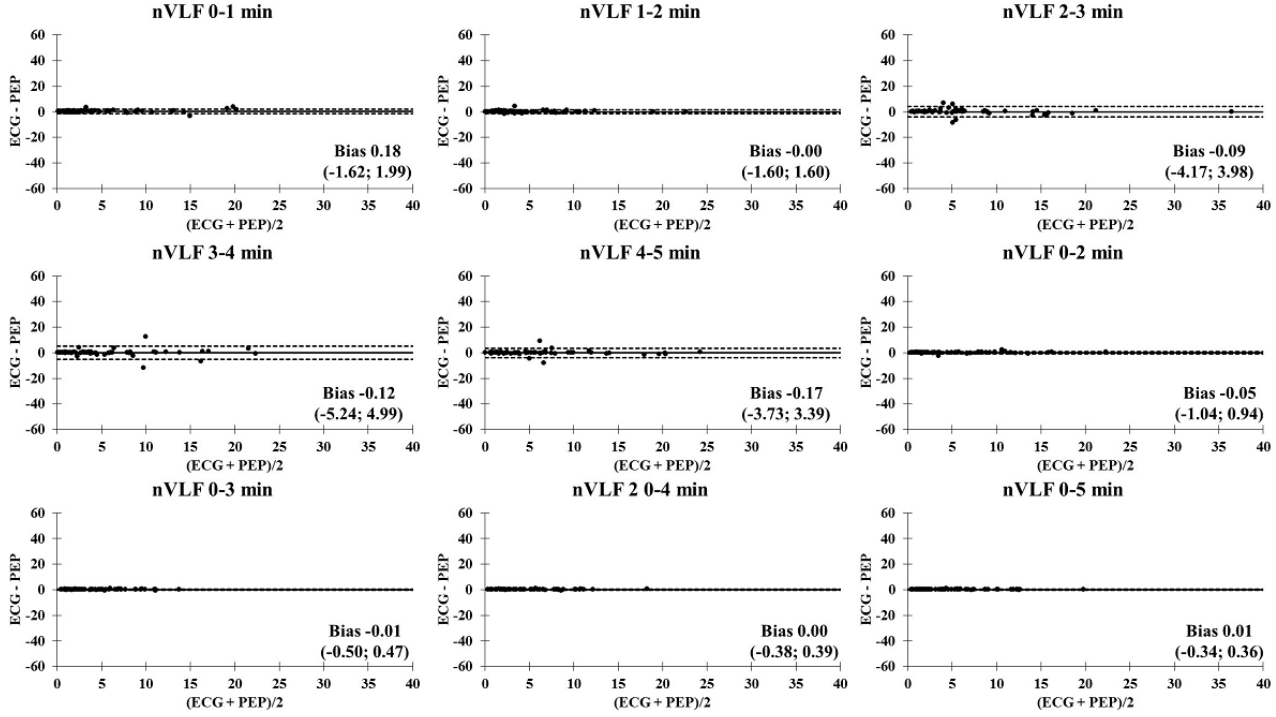


Figure 8. Bland-Altman analysis comparing the ECG and Pulse Express PRO measurements in ultra–short-term and short-term recordings of normalized low frequency power (nLF). The solid line represents the mean difference and the upper and lower dashed lines represent the upper and lower limits of agreement (± 1.96*SD). ECG: electrocardiogram; PEP: Pulse Express PRO.

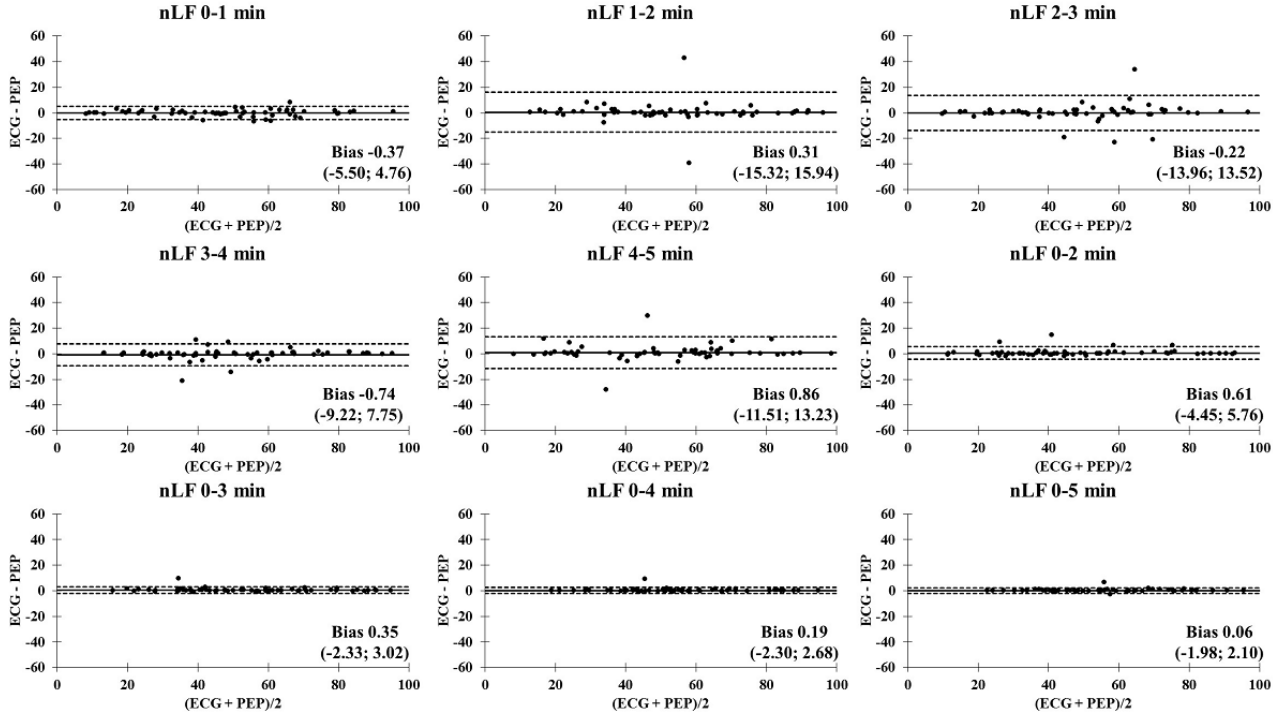
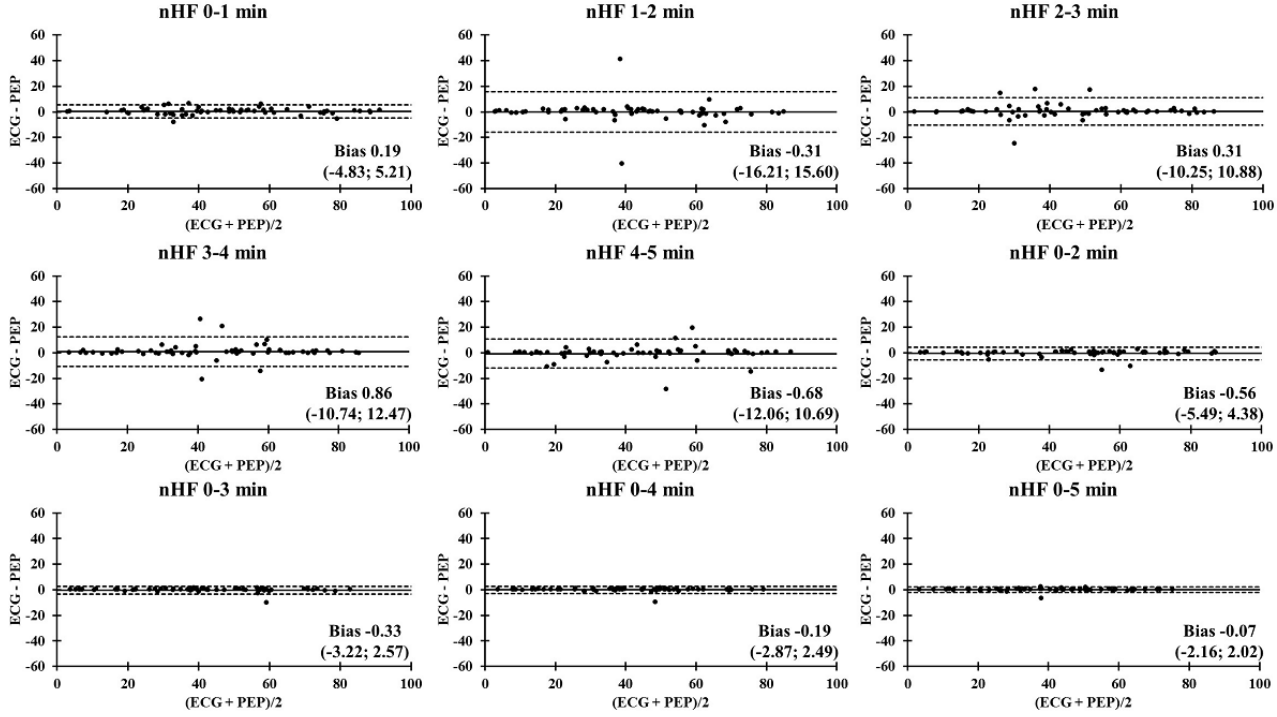


Figure 9. Bland-Altman analysis comparing the ECG and Pulse Express PRO measurements in ultra–short-term and short-term recordings of normalized high frequency power (nHF). The solid line represents the mean difference and the upper and lower dashed lines represent the upper and lower limits of agreement (± 1.96*SD). ECG: electrocardiogram; PEP: Pulse Express PRO.



Discussion

Principal Results

This study is the first to report the validity and reliability of ultra–short-term and short-term HRV via a novel smartphone app, and to compare the app with the standard ECG assessment. The limits of agreement of HRV assessments between the ECG and PEP were compared to evaluate the accuracy of measurements. The primary finding in the present study was that SDNN, RMSSD, and pNN50 parameters had very large to nearly perfect ICC and Spearman correlation coefficients in all time segments. Additionally, a large variation in ICC and Spearman correlation coefficients was found in time segments

under 2 minutes for the nVLF, nLF, and nHF parameters. The 3-minute and 4-minute nLF and nHF HRV recordings showed excellent validity and reliability and could be considered a surrogate of the standard 5-minute recording. Furthermore, with the ECG signal as a reference, the accuracy of PEP HRV recordings can be found with consecutive 1-minute recordings in the time-domain analysis (SDNN, RMSSD, and pNN50). Lastly, for the frequency-domain analysis (nLF and nHF), a recording of at least 3 minutes is required for accurate and valid PEP HRV assessment.

Time-Domain Analysis

Based on our observations, a 1-minute ultra–short-term HRV recording for the time-domain analysis revealed valid and reliable HRV features (with the 5-minute criterion as reference), despite an initial 5-minute stabilization. This indicates that the PEP app is a convenient surrogate for taking HRV measurements. It is suggested that the RMSSD is independent of respiratory sinus arrhythmia and is associated with high-frequency changes of HR modulation in response to respiratory patterns due to its strength of mathematical calculation [33]. The RMSSD has been widely accepted to evaluate cardiac-related parasympathetic activation [8,11,13,18,19,34]. Additionally, the RMSSD is recognized as a sensitive parameter to detect autonomic adaptations in response to mental stress [8,35,36] and psychophysiological strain after exercise as well as recovery status during the training period [10,37]. Long-term monitoring of resting HRV can provide valuable information to identify the chronological development of vagal-related changes related to psychometric status during sports training [38]. As demonstrated by our findings, PEP could be considered an alternative tool for short-term HRV measurements.

It is arguable that the PEP presented valid and reliable measurements in SDNN accompanied by RMSSD and pNN50 for any HRV epoch. It seems that SDNN and pNN50 are good options to integrate time-domain HRV indices. However, as the accuracy of ultra–short-term measurements of SDNN may be influenced by psychological conditions (ie, being under mental stress) [8,13], using the PEP app to facilitate 1-minute ultra–short-term HRV recordings in a quiet and relaxed manner is documented in this study.

Frequency-Domain Analysis

It is important to note that nVLF, nLF, and nHF showed trivial or small differences in association with a large variation in ICC values, correlation coefficients, and bias across all time segments compared to the standard 0-5–minute criterion. The poor validity and reliability of nVLF, nLF, and nHF in shortened epochs could be related to interindividual variations in breathing rates during measurements. Interindividual variations in breathing patterns could increase the risk of increasing HR oscillations in different time segments. Respiratory rhythm is thought of as an essential way to record frequency-domain variables such as LF and HF due to oscillations in HR responses [39]. However, breath control during resting HRV measurement does not increase accuracy and reliability during short-term recordings of frequency-domain analysis [9]. Control of respiratory frequency is not common in the general population (ie, people without appropriate respiratory training). Thus, we did not apply this instruction due to limited popularity of use.

Our findings suggest using consecutive HRV recordings of at least 3 minutes when the PEP app is used to monitor frequency-domain variables. In contrast, the minimum time requirement for HF and LF recordings has been suggested as 1 and 2 minutes, respectively [13,40]. Castaldo et al [8] showed accurate frequency-domain measurements in 1 minute for HF and 2 minutes for LF recordings after university examinations. The inconsistent findings of this study might be related to the different spectral analysis computational methods (spectrum resolution: FFT versus autoregressive) and the stabilization period prior to the HRV measurement.

Bland-Altman Analysis Comparing ECG and PEP Measurements

In an attempt to identify the agreement of biosignal measurements between the ECG and PEP, a Bland-Altman analysis was performed to compare the limit of agreement of ultra–short-term and short-term HRV recordings of the SDNN, RMSSD, pNN50, nVLF, nLF, and nHF. It is interesting to note that the PEP HRV recordings showed similar outcomes for the SDNN, RMSSD, pNN50, nVLF, nLF, and nHF measurements for all time segments, as compared to conventional lead II ECG recordings. This study revealed the accuracy and acceptance of PEP HRV recordings after consecutive 1-minute recordings in the time-domain analysis. In contrast, the degree of agreement between the ECG and PEP was relatively low for the first 3-minute assessment when frequency-domain analysis was computed. One possible explanation for less accurate measurements of frequency-domain HRV variables with shorter duration recordings may be the lack of a detrending method for processing spectral signals in the PEP app [41]. Another factor that influences measurement accuracy is related to obtaining an adequate amount of data throughout the entire measurement [42]. Lastly, acute adaptation to postural changes from standing to sitting (orthostatic stress) might be a potential mechanism to attenuate valid and reliable measurements of nLF and nHF during the 3-minute stabilization period [43,44]. Nevertheless, the PEP app is an acceptable option for HRV data collection due to its convenience and reproducibility compared to the ECG assessment.

Limitations

The first limitation of this study is that a telemetric HR sensor and a chest strap were required to detect HR responses during the PEP measurement, and that these accessories may not be commonly owned by the general population. In addition, the recording position and the HR chest strap might not be comfortable for specific populations (ie, senior adults) and clinical settings. Despite the abovementioned limitations, this is a novel study that reports the validity and accuracy of the PEP app for short-term HRV recordings.

Functional Implication

Time management is critical for professionals, including clinical practitioners and strength and conditioning coaches of elite sports teams. The PEP app is compatible with the Android operating system and can be used on low-cost smartphones. As

growing numbers of studies focus on the methodological issues related to utilizing ultra–short-term HRV recordings, the number of nonprofessionals using this free mobile app can easily be increased. We suggest that future studies should examine the use of PEP HRV assessments in the context of multidisciplinary approaches (eg, longitudinal applications in monitoring training loads, daily evaluations during competitions, and clinical evaluation).

The accuracy and reliability of the LF and HF measurements are critical to interpreting the shift of sympathovagal activities [33,45]. Excellent validity and reliability of the SDNN and RMSSD during ultra–short-term recordings indicated that the SDNN:RMSSD ratio might be appropriate to use in the first minute of PEP recording. The SDNN:RMSSD ratio is a sensitive HRV parameter that indicates autonomic adaptation in response to pathological conditions [45] and acute exercise [46]. Taking into consideration time efficiency and cross-battery assessment, our findings support the use of the SDNN:RMSSD ratio as a surrogate for the LF:HF ratio to estimate sympathovagal balance via a smartphone app.

Conclusions

In conclusion, the PEP smartphone app provides reliable and valid HRV data. It is appropriate to use the PEP app to facilitate 1-minute ultra–short-term HRV recordings during stabilization to save time when the time-domain analysis is used. Caution should be taken when the frequency-domain analysis is implemented for the interpretation of cardiac autonomic modulation. Consecutive recordings of at least 3 minutes during stabilization are suggested for accurate measurement of frequency-domain nLF and nHF indices. The use of the PEP smartphone app for ultra–short-term and short-term HRV recordings is recommended as an easy and user-friendly tool to monitor cardiac autonomic health in people with various lifestyles.

Acknowledgments

No funding was received for this study.

Authors' Contributions

YSC contributed to the study conceptualization, project administration, data collection, data analysis, methodology, and writing (including reviewing and editing) of the manuscript. WAL contributed to the study conceptualization, project administration, data collection, data analysis, and writing (including reviewing and editing) of the manuscript. JCP contributed to the study conceptualization, data analysis, supervision, and writing (including reviewing and editing) of the manuscript. CDK contributed to the study conceptualization, methodology, supervision, and writing (including reviewing and editing) of the manuscript.

Conflicts of Interest

None declared.

References

1. Proesmans T, Mortelmans C, Van Haelst R, Verbrugge F, Vandervoort P, Vaes B. Mobile Phone-Based Use of the Photoplethysmography Technique to Detect Atrial Fibrillation in Primary Care: Diagnostic Accuracy Study of the FibriCheck App. JMIR mHealth uHealth 2019 Mar 27;7(3):e12284 [FREE Full text] [doi: 10.2196/12284] [Medline: 30916656]
2. Li KHC, White FA, Tipoe T, Liu T, Wong MC, Jesuthasan A, et al. The Current State of Mobile Phone Apps for Monitoring Heart Rate, Heart Rate Variability, and Atrial Fibrillation: Narrative Review. JMIR mHealth uHealth 2019 Feb 15;7(2):e11606 [FREE Full text] [doi: 10.2196/11606] [Medline: 30767904]
3. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
4. Landreani F, Faini A, Martin-Yebra A, Morri M, Parati G, Caiani EG. Assessment of Ultra-Short Heart Variability Indices Derived by Smartphone Accelerometers for Stress Detection. Sensors (Basel) 2019 Aug 28;19(17) [FREE Full text] [doi: 10.3390/s19173729] [Medline: 31466391]
5. Landreani F, Martin-Yebra A, Casellato C, Frigo C, Pavan E, Migeotte P, et al. Beat-to-beat heart rate detection by smartphone's accelerometers: validation with ECG. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:525-528. [doi: 10.1109/EMBC.2016.7590755] [Medline: 28268385]
6. Salahuddin L, Cho J, Jeong MG, Kim D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. Conf Proc IEEE Eng Med Biol Soc 2007;2007:4656-4659. [doi: 10.1109/IEMBS.2007.4353378] [Medline: 18003044]
7. Schneider F, Martin J, Hapfelmeier A, Jordan D, Schneider G, Schulz CM. The validity of linear and non-linear heart rate metrics as workload indicators of emergency physicians. PLoS One 2017 Nov 30;12(11):e0188635 [FREE Full text] [doi: 10.1371/journal.pone.0188635] [Medline: 29190808]
8. Castaldo R, Montesinos L, Melillo P, James C, Pecchia L. Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life. BMC Med Inform Decis Mak 2019 Jan 17;19(1):12 [FREE Full text] [doi: 10.1186/s12911-019-0742-y] [Medline: 30654799]

9. Melo HM, Martins TC, Nascimento LM, Hoeller AA, Walz R, Takase E. Ultra-short heart rate variability recording reliability: The effect of controlled paced breathing. Ann Noninvasive Electrocardiol 2018 Sep 04;23(5):e12565 [FREE Full text] [doi: 10.1111/anec.12565] [Medline: 29863781]
10. Chen Y, Clemente FM, Bezerra P, Lu Y. Ultra-short-term and Short-term Heart Rate Variability Recording during Training Camps and an International Tournament in U-20 National Futsal Players. Int J Environ Res Public Health 2020 Jan 26;17(3):775 [FREE Full text] [doi: 10.3390/ijerph17030775] [Medline: 31991887]
11. Pereira L, Flatt A, Ramirez-Campillo R, Loturco I, Nakamura F. Assessing Shortened Field-Based Heart-Rate-Variability-Data Acquisition in Team-Sport Athletes. Int J Sports Physiol Perform 2016 Mar;11(2):154-158. [doi: 10.1123/ijspp.2015-0038] [Medline: 26115088]
12. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 1996 Mar;93(5):1043-1065. [doi: 10.1161/01.cir.93.5.1043]
13. Baek HJ, Cho C, Cho J, Woo J. Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. Telemed J E Health 2015 May;21(5):404-414. [doi: 10.1089/tmj.2014.0104] [Medline: 25807067]
14. Habibi M, Chahal H, Greenland P, Guallar E, Lima JA, Soliman EZ, et al. Resting Heart Rate, Short-Term Heart Rate Variability and Incident Atrial Fibrillation (from the Multi-Ethnic Study of Atherosclerosis (MESA)). Am J Cardiol 2019 Dec 01;124(11):1684-1689. [doi: 10.1016/j.amjcard.2019.08.025] [Medline: 31575421]
15. Pecchia L, Castaldo R, Montesinos L, Melillo P. Are ultra-short heart rate variability features good surrogates of short-term ones? State-of-the-art review and recommendations. Healthc Technol Lett 2018 Jun;5(3):94-100 [FREE Full text] [doi: 10.1049/htl.2017.0090] [Medline: 29922478]
16. Nussinovitch U, Cohen O, Kaminer K, Ilani J, Nussinovitch N. Evaluating reliability of ultra-short ECG indices of heart rate variability in diabetes mellitus patients. J Diabetes Complications 2012;26(5):450-453 [FREE Full text] [doi: 10.1016/j.jdiacomp.2012.05.001] [Medline: 22682758]
17. Chen X, Yao R, Yin G, Li J. Consecutive ultra-short-term heart rate variability to track dynamic changes in autonomic nervous system during and after exercise. Physiol Meas 2017 Jun 22;38(7):1384-1395. [doi: 10.1088/1361-6579/aa52b3] [Medline: 28640758]
18. Esco MR, Flatt AA. Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. J Sports Sci Med 2014 Sep;13(3):535-541 [FREE Full text] [Medline: 25177179]
19. Esco MR, Flatt AA, Nakamura FY. Agreement Between a Smartphone Pulse Sensor Application and Electrocardiography for Determining lnRMSSD. J Strength Cond Res 2017;31(2):380-385. [doi: 10.1519/jsc.0000000000001519]
20. Nakamura F, Pereira L, Cal Abad CC, Cruz I, Flatt A, Esco M, et al. Adequacy of the Ultra-Short-Term HRV to Assess Adaptive Processes in Youth Female Basketball Players. J Hum Kinet 2017 Feb;56:73-80 [FREE Full text] [doi: 10.1515/hukin-2017-0024] [Medline: 28469745]
21. Flatt AA, Esco MR. Heart rate variability stabilization in athletes: towards more convenient data acquisition. Clin Physiol Funct Imaging 2016 Sep 09;36(5):331-336. [doi: 10.1111/cpf.12233] [Medline: 25754514]
22. Flatt AA, Esco MR. Smartphone-Derived Heart-Rate Variability and Training Load in a Women's Soccer Team. Int J Sports Physiol Perform 2015 Nov;10(8):994-1000. [doi: 10.1123/ijspp.2014-0556] [Medline: 25756657]
23. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of Heart-Rate-Variability Recording With Smartphone Photoplethysmography, Polar H7 Chest Strap, and Electrocardiography. Int J Sports Physiol Perform 2017 Nov 01;12(10):1324-1328. [doi: 10.1123/ijspp.2016-0668] [Medline: 28290720]
24. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
25. Pulse Express Pro. URL: https://play.google.com/store/apps/details?id=pulseexpress.pro&hl=en_us [accessed 2020-07-15]
26. Tarvainen MP, Niskanen J, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. Comput Methods Programs Biomed 2014;113(1):210-220. [doi: 10.1016/j.cmpb.2013.07.024] [Medline: 24054542]
27. Lu W, Chen Y, Kuo C. Increased first and second pulse harmonics in Tai Chi Chuan practitioners. BMC Complement Altern Med 2016 Feb 29;16(1):87 [FREE Full text] [doi: 10.1186/s12906-016-1058-4] [Medline: 26931120]
28. Chen Y, Liao C, Lu W, Kuo C. Sympathetic enhancement in futsal players but not in football players after repeated sprint ability test. BMJ Open Sport Exerc Med 2015 Nov 27;1(1):e000049 [FREE Full text] [doi: 10.1136/bmjsem-2015-000049] [Medline: 27900135]
29. Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol 2016 Mar 26;116(3):563-571 [FREE Full text] [doi: 10.1007/s00421-015-3303-9] [Medline: 26708360]
30. Ali K, Bertoni HL. Blockage/Shadowing Polarization Measurements at 2.45GHz for Interference Evaluation between Bluetooth and IEEE 802.11 WLAN. 2001 Presented at: IEEE Antennas and Propagation Society International Symposium; 8-13 July 2001; Boston p. 376. [doi: 10.1109/aps.2001.960112]
31. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 2009 Jan;41(1):3-13. [doi: 10.1249/MSS.0b013e31818cb278] [Medline: 19092709]
32. Martin Bland J, Altman D. Statistical Methods For Assessing Agreement Between Two Methods Of Clinical Measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8]

33. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol 2014;5:73 [FREE Full text] [doi: 10.3389/fphys.2014.00073] [Medline: 24578692]
34. Flatt A, Esco M. Validity of the ithlete™ Smart Phone Application for Determining Ultra-Short-Term Heart Rate Variability. J Hum Kinet 2013 Dec 18;39:85-92 [FREE Full text] [doi: 10.2478/hukin-2013-0071] [Medline: 24511344]
35. Melo HM, Hoeller AA, Walz R, Takase E. Resting Cardiac Vagal Tone is Associated with Long-Term Frustration Level of Mental Workload: Ultra-short Term Recording Reliability. Appl Psychophysiol Biofeedback 2020 Mar 8;45(1):1-9. [doi: 10.1007/s10484-019-09445-z] [Medline: 31286301]
36. Salahuddin L, Cho J, Jeong MG, Kim D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. Conf Proc IEEE Eng Med Biol Soc 2007;2007:4656-4659. [doi: 10.1109/IEMBS.2007.4353378] [Medline: 18003044]
37. Nakamura FY, Antunes P, Nunes C, Costa JA, Esco MR, Travassos B. Heart Rate Variability Changes From Traditional vs. Ultra–Short-Term Recordings in Relation to Preseason Training Load and Performance in Futsal Players [published online ahead of print, 2018 Nov 13]. J Strength Cond Res 2020:1. [doi: 10.1519/jsc.0000000000002910]
38. Plews D, Laursen P, Kilding A, Buchheit M. Evaluating training adaptation with heart-rate measures: a methodological comparison. Int J Sports Physiol Perform 2013 Nov;8(6):688-691. [doi: 10.1123/ijspp.8.6.688] [Medline: 23479420]
39. Billman GE. The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. Front Physiol 2013;4:26 [FREE Full text] [doi: 10.3389/fphys.2013.00026] [Medline: 23431279]
40. Munoz ML, van RA, Riese H, Thio C, Oostenbroek E, Westrik I, et al. Validity of (Ultra-)Short Recordings for Heart Rate Variability Measurements. PLoS One 2015;10(9):e0138921 [FREE Full text] [doi: 10.1371/journal.pone.0138921] [Medline: 26414314]
41. Tarvainen M, Ranta-Aho PO, Karjalainen P. An advanced detrending method with application to HRV analysis. IEEE Trans Biomed Eng 2002 Feb;49(2):172-175. [doi: 10.1109/10.979357] [Medline: 12066885]
42. Plews DJ, Laursen PB, Le Meur Y, Hausswirth C, Kilding AE, Buchheit M. Monitoring training with heart rate-variability: how much compliance is needed for valid assessment? Int J Sports Physiol Perform 2014 Sep;9(5):783-790. [doi: 10.1123/ijspp.2013-0455] [Medline: 24334285]
43. Barantke M, Krauss T, Ortak J, Lieb W, Reppel M, Burgdorf C, et al. Effects of gender and aging on differential autonomic responses to orthostatic maneuvers. J Cardiovasc Electrophysiol 2008 Dec;19(12):1296-1303. [doi: 10.1111/j.1540-8167.2008.01257.x] [Medline: 18662181]
44. Young FL, Leicht AS. Short-term stability of resting heart rate variability: influence of position and gender. Appl Physiol Nutr Metab 2011 Apr;36(2):210-218. [doi: 10.1139/h10-103] [Medline: 21609282]
45. Sollers JJ, Buchanan TW, Mowrer SM, Hill LK, Thayer JF. Comparison of the ratio of the standard deviation of the R-R interval and the root mean squared successive differences (SD/rMSSD) to the low frequency-to-high frequency (LF/HF) ratio in a patient population and normal healthy controls. Biomed Sci Instrum 2007;43:158-163. [Medline: 17487074]
46. Esco MR, Williford HN, Flatt AA, Freeborn TJ, Nakamura FY. Ultra-shortened time-domain HRV parameters at rest and following exercise in athletes: an alternative to frequency computation of sympathovagal balance. Eur J Appl Physiol 2018 Jan 11;118(1):175-184. [doi: 10.1007/s00421-017-3759-x] [Medline: 29128939]

Abbreviations

ECG: electrocardiogram
ES: effect size
FFT: Fast Fourier Transformation
HR: heart rate
HRV: heart rate variability
ICC: intraclass correlation coefficient
lnRMSSD: natural logarithm of root mean square differences between adjacent normal R-R intervals
nHF: normalized high frequency power
nLF: normalized low frequency power
nVLF: normalized very low frequency power
PEP: Pulse Express PRO
pNN50: the proportion of NN50 divided by the total number of RR intervals
RMSSD: root mean square differences between adjacent normal R-R intervals
SDNN: standard deviation of normal R-R intervals

Edited by G Eysenbach; submitted 17.03.20; peer-reviewed by NJ Chang, M Altini, L Becker; comments to author 11.04.20; revised version received 05.06.20; accepted 13.06.20; published 31.07.20.

Please cite as:
Chen YS, Lu WA, Pagaduan JC, Kuo CD
A Novel Smartphone App for the Measurement of Ultra–Short-Term and Short-Term Heart Rate Variability: Validity and Reliability Study
JMIR Mhealth Uhealth 2020;8(7):e18761
URL: https://mhealth.jmir.org/2020/7/e18761
doi:10.2196/18761
PMID:32735219

Original Paper

Users' Experiences of a Mobile Health Self-Management Approach for the Treatment of Cystic Fibrosis: Mixed Methods Study

Jacqueline Floch[1], Dr Ing; Thomas Vilarinho[1], MSc; Annabel Zettl[2], BA; Gema Ibanez-Sanchez[3], MSc; Joaquim Calvo-Lerma[3,4], PhD; Erlend Stav[1], Dr Scient; Peter Halland Haro[1], MSc; Asbjørn Lein Aalberg[1], MSc; Alvaro Fides-Valero[3], MSc; José Luis Bayo Montón[3], MSc

[1]SINTEF, Trondheim, Norway

[2]YOUSE GmbH, München, Germany

[3]Universitat Politècnica de València, València, Spain

[4]Instituto de Investigación Sanitaria La Fe, València, Spain

Corresponding Author:
Jacqueline Floch, Dr Ing
SINTEF
Strindveien 4
Trondheim, 7465
Norway
Phone: 47 93008536
Email: jacqueline.floch@sintef.no

Abstract

Background: Despite a large number of clinical trials aiming at evaluating the digital self-management of chronic diseases, there is little discussion about users' experiences with digital approaches. However, a good user experience is a critical factor for technology adoption. Understanding users' experiences can inform the design of approaches toward increased motivation for digital self-management.

Objective: This study aimed to evaluate the self-management of cystic fibrosis (CF) with a focus on gastrointestinal concerns and the care of young patients. Following a user-centered design approach, we developed a self-management app for patients and parents and a web tool for health care professionals (HCPs). To evaluate the proposed solutions, a 6-month clinical trial was conducted in 6 European CF competence centers. This paper analyzes the user acceptance of the technology and the benefits and disadvantages perceived by the trial participants.

Methods: A mixed methods approach was applied. Data were collected through 41 semistructured qualitative interviews of patients, parents, and HCPs involved in the clinical trial. In addition, data were collected through questionnaires embedded in the self-management app.

Results: Support for enzyme dose calculation and nutrition management was found to be particularly useful. Patients and parents rapidly strengthened their knowledge about the treatment and increased their self-efficacy. Reported benefits include reduced occurrence of symptoms and enhanced quality of life. Patients and parents had different skills, requiring follow-up by HCPs in an introductory phase. HCPs valued obtaining precise information about the patients, allowing for more personalized advice. However, the tight follow-up of several patients led to an increased workload. Over time, as patient self-efficacy increased, patient motivation for using the app decreased and the quality of the reported data was reduced.

Conclusions: Self-management enfolds a collaboration between patients and HCPs. To be successful, a self-management approach should be accepted by both parties. Through understanding behaviors and experiences, this study defines recommendations for a complex case—the demanding treatment of CF. We identify target patient groups and situations for which the app is most beneficial and suggest focusing on these rather than motivating for regular app usage over a long time. We also advise the personalized supervision of patients during the introduction of the approach. Finally, we propose to develop guidance for HCPs to facilitate changes in practice. As personalization and technology literacy are factors found to influence the acceptance of digital self-management of other chronic diseases, it is relevant to consider the proposed recommendations beyond the case of CF.

(JMIR Mhealth Uhealth 2020;8(7):e15896) doi:10.2196/15896

KEYWORDS

mobile health; mHealth; self-management; user experience; user acceptance; mixed methods study; cystic fibrosis; pediatrics

Introduction

Background

Chronic disease self-management enfolds the idea that patients in collaboration with health care professionals (HCPs) gain knowledge about the disease and carry out part of the treatment themselves [1]. Several reviews have observed a positive effect of digital self-management approaches [2-4]. Despite several clinical trials evaluating the digital self-management of chronic diseases, there is little discussion about users' experiences with digital approaches. A good user experience, for example, making the technology useful, easy to use, and efficient, is a critical factor for technology adoption [5]. Understanding this experience can inform the design of approaches toward motivation for digital self-management. Both the viewpoints of patients and HCPs are essential for a successful collaboration.

The scope of our research is the self-management of cystic fibrosis (CF), a congenital, chronic disorder affecting the digestive and respiratory systems, resulting in malnutrition and respiratory infections [6]. Daily treatment is demanding, including physiotherapy, physical exercise, an adapted rich diet, and dosage of pancreatic enzyme supplements. Many tend not to adhere to the whole therapy [7]. Therefore, digital support that facilitates understanding of the treatment and motivates adherence is relevant but challenging.

All countries involved in this study have specialized CF centers [8]. The patients meet for consultation at least every 3 months. Between consultations, patients can contact their centers if needed. A thorough control is performed yearly, requiring a written collection of food records. Although the CF centers function well and patients express satisfaction in the services, our earlier research indicates the readiness for digital self-management in the CF care [9].

MyCyFAPP Vision

Our research is part of the European Union–funded research project MyCyFAPP [10], aiming to increase patients' knowledge regarding their treatment, facilitating adherence, and supporting teenagers' implications. Most CF patients have to follow a pancreatic enzyme replacement therapy, where enzymes are taken with each meal to help digest food. Wrong doses can cause malnutrition and gastrointestinal (GI) problems [11]. Before MyCyFAPP, there were no knowledge or tools to adjust the dose of enzymes. Rather, patients were recommended a fixed dose for each meal. A key novel component developed in MyCyFAPP is an algorithm for enzyme dose calculation [12].

Digital Self-Management in MyCyFAPP

The digital support developed in MyCyFAPP includes a self-management app targeting parents of young children with CF and teenagers with CF and a professional web tool (PWT) targeting HCPs (Multimedia Appendix 1).

The app is available in the languages of the participating countries (ie, Dutch, Spanish, Italian, and Portuguese) and English. Its main features are as follows:

1. Calculation of a personalized enzyme dose depending on meal composition
2. Follow-up of food intake
3. Food recording as a basis for both enzyme dose calculation and nutrition management
4. Access to recommended country-specific dishes to correct specific nutritional imbalances [13]
5. Health diary for recording mood and GI symptoms
6. Educational handbook about the disease and the treatment with focus on nutrition.

The PWT provides an overview of the patients' progress, mainly based on the recorded data using the self-management app. In addition, information is gathered during consultations, for example, weight and height. The monitored parameters are those included in the CF nutritional guidelines: nutrient intake, enzyme dose, and nutritional status [14]. HCPs can register health information, set nutritional goals, and send messages to patients.

Research on User Experience in Mobile Health

A recent review of qualitative studies about patients' perceptions and experiences of mobile health (mHealth) apps identified 38 scientific articles (2013-2018) related to app evaluation [15]. Most apps deal with chronic diseases and provide either access to information, communication with HCPs, peer support, or self-monitoring. No study has addressed the experiences of CF self-management. Unfortunately, the review did not assess whether evaluations were performed in controlled settings, whether HCPs were users of the technology, or the experiences of HCPs. Overall, the review finds that mHealth has great potential to engage and empower patients. Personalization, technology literacy, intrusiveness, information validity, security, and privacy appear to influence the acceptance of solutions, overlapping with identified issues in our initial research [9]. The design of MyCyFAPP digital solutions considers all these issues, as further discussed.

A motivation for that review was that mHealth apps seem to be underused after download, and app adherence is a major concern in consumer apps for health monitoring [16]. The review suggests a constant stimulation of patients to accommodate changing patients' requirements. We argue that the purpose of mHealth should be treatment adherence rather than app adherence, as apps are no longer needed when goals are achieved.

Purpose of the Study

This study is the last step in the information and communication technology (ICT) research conducted in MyCyFAPP. Clinical research applies the digital approach and evaluates its impact on quality of life [17]. Complementarily, the ICT perspective aims at developing an approach that best suits the needs of patients and HCPs and, from the evaluation of the approach, at deriving recommendations for CF care digital self-management,

also filling the gap of knowledge about experiences from the digital self-management of chronic diseases.

Adopting a *design and creation* research strategy [18,19] and applying a user-centered design approach [20], our research followed an incremental process enabling to gradually understand problems and to improve solutions. Subsequent to the extraction of requirements [9], paper prototypes and mockups were cocreated with potential users [21]. Then, an initial software prototype was developed and tested in a midterm evaluation. Finally, an improved and comprehensive prototype was developed and evaluated in a 6-month clinical trial. This final evaluation addresses 2 research questions, the first related to user acceptance and the second to impact:

1. How do the software features and the context of use affect the user experience of the proposed digital self-management approach?
2. What are the perceived benefits and disadvantages of using this digital self-management approach?

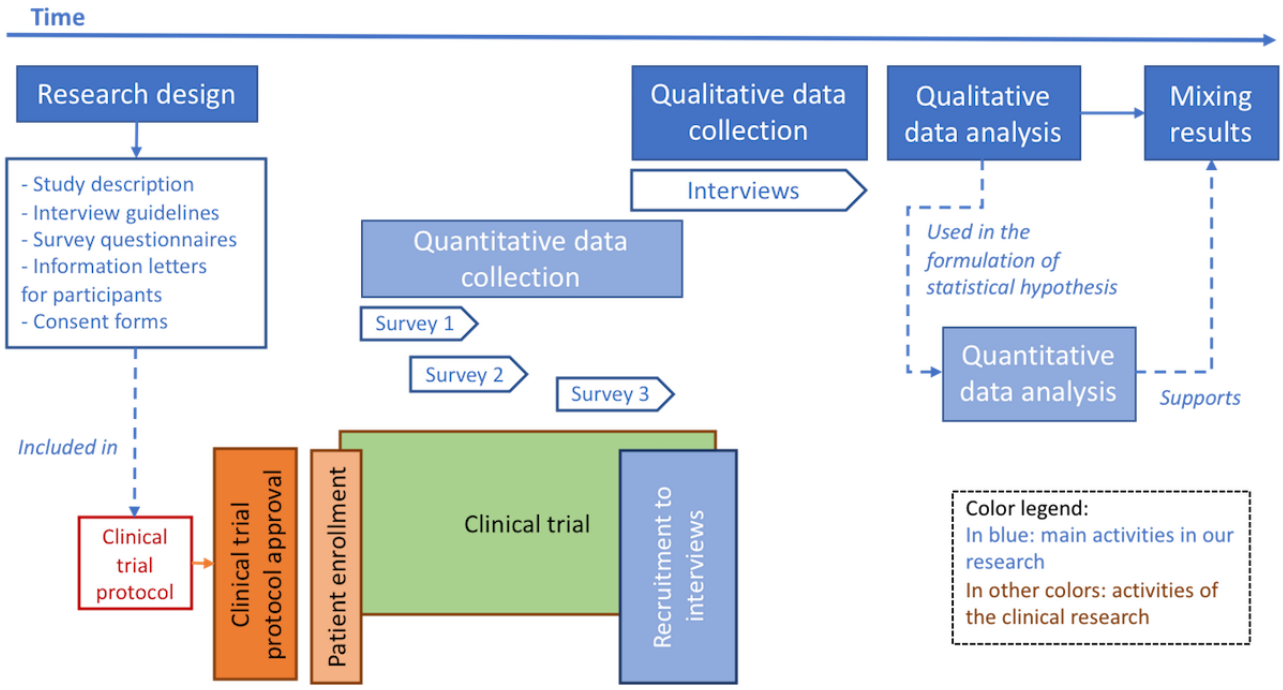
We consider user experience broadly and investigate technical and nontechnical aspects that enhance or degrade the user experience. Understanding user acceptance and the impact of the approach is cornerstone for designing appropriate digital approaches. This represents the objective of this study.

Methods

Overall Approach

Our study applied a mixed methods approach: an embedded design [22]. A flowchart depicting the design and implementation of the study is shown in Figure 1. As it had to be included in the clinical trial protocol, the study design was fixed [22]; methods were predetermined at the start of the research process, including constraints to avoid bias to the clinical study. The ICT research orientation was mainly qualitative, aiming at an in-depth understanding of system usage and experience. As usage over 6 months could not be observed, an interview approach was selected. Recruitment to interviews is often challenging and may affect results in unforeseen ways [23]. On the basis of earlier experiences with recruitment [9], convenience sampling was applied [24]. To reduce sampling limitations, 2 measures augment the findings: interviews of app users were complemented with interviews of HCPs about user feedback, and quantitative surveys were embedded in the self-management app to increase the credibility of qualitative findings [25]. Data collection was spread over time due to different start times in participating centers and spread patient enrollment. The collections of qualitative and quantitative data were independent. Qualitative analysis was performed before the quantitative analysis. The results were combined during the final step of the research process.

Figure 1. Flowchart of the research process.



MyCyFAPP Clinical Trial

A total of 171 patients from 6 CF centers in 5 countries (ie, Belgium, Italy, Netherlands, Portugal, and Spain) were recruited in the clinical trial, with most of them in a stable clinical status and with little or mild GI symptoms. A total of 154 patients completed the trial. Dropout causes were the addition of another task in the treatment or unexpected high enzyme dosage recommendations.

Patients and parents of young patients were asked to register symptoms and food intake, calculate the dose of enzyme, and consult the educational handbook. Each center appointed a dietitian for follow-up. Using PWT, dietitians monitored recorded data to check symptoms and diet quality and to compare enzyme intake with the dose recommended by the app.

They sent messages to the patients' app, including personalized advice and references to the educational handbook.

The clinical trial protocol established a minimum utilization of the app and the PWT. At months 1 and 6, food records had to be provided during 3 consecutive days, and symptoms were registered at least once a week. Otherwise, trial participants had to report at least three days per week and whenever needed (eg, new recipe). Using PWT, dietitians had to check patients and send feedback at least once a week. Depending on the dietitians' criteria and patients' needs, variations in the follow-up were allowed, such as the number of messages about educational content or the setting of nutritional goals.

Qualitative Data Collection and Analysis

Due to the explorative nature of the research, individual semistructured interviews with open-ended questions were conducted, allowing interviewees to express their viewpoints freely [26]. We conducted 3 types of interviews (Multimedia Appendix 2):

1. With teens with CF and parents of children or teens with CF, covering motivation, usage, experienced impact, and future expectations
2. With HCPs about the self-management app, covering personal opinions and feedback from app users
3. With HCPs about PWT, covering usage and experienced impact on the workflow.

Due to the geographical distribution, interviews were conducted on the web (video call). Interviews lasted 45 to 60 min. Native language was spoken with Spanish and Portuguese trial participants; otherwise, English was spoken with the constraint that participants were fluent in English. Interviews were recorded, transcribed, and coded using the Dedoose tool provided by SocioCultural Research Consultants, LLC (Manhattan Beach, California). Data analysis was performed in an inductive manner following Klein and Myers's framework [27]. The researchers performed a first independent round of thematic analysis, resulting in a number of initial topics [28]. Then, a refined set of topics was iteratively created through researchers' collaboration.

Table 1 shows the target groups and informants across the CF centers. There were 41 interviews conducted involving 19 patients and parents and 12 HCPs; 10 HCPs participated in 2 interviews about the app and the PWT. The goal of including equal numbers per target group in all countries, thus equally covering countries that differ in the organization of CF health services and cultural backgrounds, could not be achieved.

Table 1. Overview of the target groups and informants across cystic fibrosis centers and countries.

Cystic fibrosis center	Self-management app: target group					Professional web tool: profession		Total informants
	Teens	Parents of young children	Parents of teens using the app	Dietician	Doctor	Dietician	Doctor	
Rotterdam	0	1	1	1	1	(1)[a]	(1)[a]	4 (6)[a]
Leuven	2	1+(1)[b]	(1)[b]	1	1	1+(1)[a]	0	7 (8)[a]
Valencia	2	1	1	1	1	(1)[a]	(1)[a]	6 (8)[a]
Madrid	2	1	1	1	1	(1)[a]	(1)[a]	6 (8)[a]
Milan	0	1	0	1	0	(1)[a]	0	2 (3)[a]
Lisbon	2	1	1	1	1	(1)[a]	(1)[a]	6 (8)[a]
Total	8	6+(1)[b]	5	6	5	1+(6)[a]	(4)[a]	31 (41)[a]

[a]Some health care professionals participated in 2 interviews, about the app and about the professional web tool. The second is registered in parenthesis.

[b]Parents of 2 children with cystic fibrosis, 1 young child and 1 teen using the app; only 1 interview was conducted.

Quantitative Data Collection and Analysis

Questionnaires, developed in collaboration with clinical partners, were embedded in the self-management app. Standard methods, such as the system usability scale [29] or the technical acceptance model [30], were found to be too complex. Other concerns were a potential influence on participants and an additional burden. Thus, questionnaires were specifically developed for the study (Multimedia Appendix 2). They are kept short and inspired by standards. A psychologist from a CF center checked the wording.

Questionnaires were presented to users at 1 week, 1 month, and 5 months after starting using the app. The first questionnaire focused on usability and expectations and the others on user acceptance. To reduce intrusiveness and allow users to answer at any time, questionnaires were presented in a dedicated notification area rather than in pop-ups. Answering the questionnaires was not mandatory. The answers were anonymized.

Table 2 shows the response rates for each survey. As answers do not disclose the respondents' center, we group centers by app language. Table 2 depicts high response rates for survey 1 (142/171, 83.0%) and survey 2 (120/171, 70.2% to 120/154, 77.9%) and decreasing response rates for survey 3 (84/154, 54.5%). Some factors identified through qualitative analysis may influence participation: knowledge acquisition made the app less relevant after a while, the trial duration undermined

motivation, and the clinical status of participants was stable. In line with the possible ceiling effect of app usage, there were no significant differences in survey 2 answers between respondents to all questionnaires and respondents to only 2 questionnaires (all analyses ranged from t_{118}=0.043 to t_{118}=1.827 and with p values ranging from P=.06 to P=.98). Therefore, we do not use survey 3 in inferential statistics.

In the *Results* section, we present items on an ordinal Likert scale (% answered on given categories). However, inferential statistics assume the applicability of an interval scale.

Table 2. Response rates for the surveys.

Cystic fibrosis center	Clinical trial participants			Survey 1		Survey 2		Survey 3	
	Number of recruited participants	Number of dropouts	Number of participants who completed the survey, n (%)	Number of answers	Users, %	Number of answers	Estimated users[a], %	Number of answers	Users, %
Rotterdam	17	2	15 (88.2)	53[b]	93.0[b]	46[b]	80.7-90.2[b]	37[b]	72.5[b]
Leuven	40	4	36 (90.0)	—[b]	—[b]	—[b]	—[b]	—[b]	—[b]
Valencia	25	0	25 (100.0)	46[c]	80.7[c]	36[c]	63.2-65.5[c]	20[c]	36.4[c]
Madrid	32	2	30 (93.8)	—[c]	—[c]	—[c]	—[c]	—[c]	—[c]
Milan	26	6	20 (76.9)	19	73.1	17	65.4-85.0	12	60.0
Lisbon	31	3	28 (90.3)	24	77.4	21	67.7-75.0	15	53.6
Total	171	17	154 (90)	142	83.0	120	70.2-77.9	84	54.5

[a]Some dropouts answered survey 2. The rate pertains to recruited users and users who completed the survey.

[b]Data from Rotterdam and Leuven are merged (language: Dutch).

[c]Data from Valencia and Madrid are merged (language: Spanish).

Results

User Experience With the Self-Management App

Outline for the Presentation of Results

This section presents the results of the qualitative analysis of interviews about the app and the quantitative analysis of surveys. Results from both analyses were found to be consistent. To avoid repetition, we aggregate the results and structure them according to the app features rather than to the research questions. First, we present descriptive statistics showing trends. Then, for each feature, based on the qualitative analysis, we address user acceptance (ie, how features were used and what could improve the user experience) and impact (ie, perceived benefits and disadvantages). Some inferential statistics for testing hypotheses derived from the qualitative analysis are described.

We use the terms *participants* for patients and parents involved in the clinical trial, *informants* for those involved in the interviews, and *respondents* for those answering the surveys. Illustrative quotes provided in Tables 3-7 are referred by *Id* in the text. As CF is a rare disease, we randomly number the CF centers to not expose the identity of informants.

Table 3. Illustrative quotes related to using the self-management app: varying motivation to use the app.

Id	Concern	Interviewee	Quote
M1	Motivation: health improvement	Mother of 6-year-old and 10-year-old children, C4	“Getting a little bit of weight. A lot less poo. You can see the nutrition. And you can see it helps. That is a big motivation. And you know that the enzymes are important. And you have to take the right dose. So that is a big motivation.”
M2	Motivation: self-efficacy	Parents of a 6-year-old child, C2	Father: “It gives us an answer. Those doses of Creon are rather useful. And surprising.” Mother: “It feels that we can now do something ourselves. It is not something we do because the hospital asked us. It is something we do because we want to help our daughter to improve her condition.”
M3	Lack of motivation: no patience, enough knowledge	Mother of an 11-year-old child, C3	“My daughter started looking at the app and said she had no patience for that and would not use it at school. [...] She does not want to bother with it as she already knows the enzyme dosage and do it herself alone, which is great.”
M4	Decreasing motivation: long trial	Dietician, C2	“In the beginning, most of them use the app a lot. Every month it became less, and less and less. For some, 6 months was really a long time.”
M5	Lack of motivation: limited time, break of routine	Dietician, C4	“We see that when they are going to school, they have like a structure and everything is going better, but when the summer starts, they are not that motivated anymore. Especially when they are going on holidays, when they are travelling, then they really don’t like using the app. Because it is time consuming and they don’t want to think of it.”

Table 4. Illustrative quotes related to using the self-management app: enzyme dose calculation.

Id	Concern	Interviewee	Quote
E1	Increased knowledge	Mother of a 5-year-old child, C4	"Before we just looked at the calories. Every 100 calories, we gave 1 "10 000" pills. But we learnt that, for the Creon, it is not as straightforward. For mayonnaise, we gave too little and now we know. After using the app, he gained much more weight."
E2	Feeling confident	Mother of a 5-year-old child, C5	"At the beginning, I was really afraid because the doses were very high and I was scared to give this to my son, but then we decided to try. We were checking what we had entered to the app in the case it was very fatty. I felt comfortable because everything in the app is well measured. [...] Yes, I felt comfortable and there were no side effects."
E3	Developing best practices	Mother of a 7-year-old teen, C1	"Sometimes I gave her X pills for a specific meal. But later on, she doesn't want to finish up the dish. So, I wonder, what can I do? I learnt that I can give her some walnuts. It is a highly caloric food that she likes. Eating three walnuts is very easy for her, and it increases a lot the amount of Creon. I have learnt this thanks to the app. Or with the olives, it is the same."
E4	Perceived benefits: gaining weight, better digestion	Dietician, C3	"Some increased weight. There was one patient that told us that finally he knows what normal faeces are. He said he had thought his faeces were normal, and finally they weren't."
E5	Perceived benefit: child accountability	Mother of a 10-year-old teen, C4	"She is a teenager. Sometimes, she forgot to take her enzyme, but now she always takes it. We do not have to tell her..."
E6	Disadvantage: extra workload, but fewer symptoms	Parents of a 6-year-old child, C2	Father: "Yes, the app comes with extra workload. You have to do extra steps to calculate. But in the end, it makes it worthwhile to do it." Mother: "Yes, we see when she goes to the toilets, it is much better. That's worthy."
E7	Behaviors in response to the recommended dose	Doctor, C2	"It was different. There were people who did not follow the advice and did just use their own usual dosage. And there were parents who did use the advice of the app. But, in general, if there was a big difference, they used to take their own dose."
E8	Need for HCP[a] support: wrong food recording	Dietician, C2	"Because sometimes they told "We got such a weird advice. I did everything right in the app, but we have to take 8 Creon for one piece of fish." And then it turned out that they did something wrong, and if I filled it out again, it was only like 2 Creon."

[a]HCP: health care professional.

Table 5. Illustrative quotes related to using the self-management app: nutrition management.

Id	Concern	Interviewee	Quote
N1	Adjusting diet	10-year-old teen, C4	"I looked at the calories often because then I could see how many calories I needed, and how many calories I was below the goal. [...] I drink more milk now for the breakfast."
N2	Getting more disciplined	Mother of a 5-year-old child, C5	"The app helps you to be more disciplined with the food, and controlling nutrients makes you follow a more balanced diet. It has helped us to make his meals more balanced."
N3	Nutritional goals as a game: educating young teens	Parents of an 11-year-old child, C6	"He eats less than necessary, and we put a lot of effort in pressing him to eat what is recommended by the dietician. So, the app helps our son to understand why we put a big pressure on him to eat. Sometimes it is very difficult to make him understand how important the amount of calories is. I would say it is like a game, you add your meals at the end of the day, you see the amount of calories and the distance to the goal. It was very useful in letting him understand he needs to eat more."
N4	Need for HCP[a] support: not reaching goals	Father of a 3-year-old child, C2	"We saw that it was quite difficult sometimes to get this right amount of energy, fat or things like that. [...] We checked with the nutritionist. She told us that it was OK with the things that we are giving him right now."
N5	Disadvantages: obsession with goals, need for individual tailoring	Dietician, C6	"Yes, they liked it, but, in some cases, they were obsessed about the goals. Patients who have a good nutritionals status, like this information. They are concerned about the nutritional status. If the patients have problems with nutrition and see every day that they can't reach the nutritional goals, it is bad. It is important to adjust the goals to every patient."

[a]HCP: health care professional.

Table 6. Illustrative quotes related to using the self-management app: food recording.

Id	Concern	Interviewee	Quote
F1	Search: difficult in the beginning	Dietician, C4	"What I heard was that sometimes food products were difficult to find. The search function was not optimal. Especially in the beginning, it is quite hard and time consuming to know how to fill in everything and how to find the food. Once they were used to it and set their standards it was quite easy to register."
F2	Technology literacy level	Mother of a 5-year-old child, C4	"It could be more user-friendly because I know if I would give that app to my parents... They are over 60. They would not know what to do... For us, it is OK because we find our way... But for older people, it would be too technical and too hard."
F3	Food preparation, estimating quantities	Mother of a 15-year-old teen, C1	"The tedious part was the food recording. For example, there was an issue with the oil. When you created a dish and you indicated a certain amount of oil, then you could get a very high dose of enzymes, but if you changed the amount of oil the recommended dose could be reduced a lot. Also, it is difficult to estimate the amount of oil in fried products. Depending on the amount you indicated you could get too high doses."
F4	Best practices: creating own dishes	Father of a 3-year-old child, C2	"You can create the meal in your app, so you select it the next time. We use that. Quite often. When the same meal comes back after 2 or 3 weeks, that was very pleasant to use."

Table 7. Illustrative quotes related to using the self-management app: other features (health diary, educational handbook, and messages from health care professionals).

Id	Concern	Interviewee	Quote
O1	Health diary: more efficient consultation	Mother of an 11-year-old child, C3	"I actually found it useful. We try to register if our daughter goes to the bathroom or not, every day. Or if she has belly pain. [...] I think I used this most so that the doctor would know what happens with my daughter and so that I would not forget anything. Also the consultation was faster."
O2	Health diary: a tool for reflection, increased well-being	Mother of an 11-year-old child, C3	"This part where you are asked if you feel well, happy, ... I ask my daughter in the evening if we can sit together and reflect about the day. It is just a click, and she finds it fun. It is the fastest part of all, the part she liked the most. Also, this makes us talk about school, the part I liked most [...] In these 5 minutes at the end of the day, we sat together and she told if she was happy, if she had had pain."
O3	Educational handbook: better than the internet	15-year-old teen, C3	"And also the handbook. It helped because nowadays there is not much information which is specific in the Internet. Such as to explain the enzymes, how to improve things, which sports to practice. This you can't find in the web easily."
O4	Educational handbook: rather use the internet	Parents of a 6-year-old child, C2	"If we want to look up something, we will look for it on the Internet or call the hospital. There are so many other means to find information."
O5	Educational handbook: useful to explain cystic fibrosis	Mother of 6-year-old and 10-year-old children, C4	"We read it. We like it a lot. [about 10yo child] She read it to and understood. She used the text from the app, but she made her own text and presented it to the class. And everybody understood."
O6	Messages: personalized motivational messages	Mother of 6-year-old and 10-year-old children, C4	"The dietician sent us messages. "Very good. You do a good job." My daughter was sick for a few weeks, and she [the dietician] sent some solutions: "Get a bit of pudding and milk and yogurt." "You have to drink that and eat that. You are doing fine.""
O7	Messages: encouraging messages, but generic	Father of a 3-year-old child, C2	""Good job, thanks for filling in, you are doing good." Yes, it was very positive. It was good to receive, but we thought it was some kind of computer. (laugh) To be honest. Sometimes it was the same messages."
O8	Messages: lack support for messaging to HCPs[a]	Dietician, C1	"Some patients lacked the possibility to send messages to us through the app and PWT, which is something I would probably not include. It would be fine if the patients used it with moderation. Otherwise, something good can become something terrible."

[a]HCP: health care professional.

Trends in Usability, Expectations, and Experiences

Table 8 depicts the usability results and expectations (survey 1) and the perceived user experience (surveys 2 and 3).

After 1 week, most respondents had positive experiences using the app. In total, 83.0% agreed or strongly agreed that the app is easy to learn and 70.4% agreed or strongly agreed that the app is easy to use. Respondents indicated very high expectations for the enzyme dose calculation (88.0%) and over half of them for nutrition management and understanding of the treatment. After 1 and 5 months, the perceived value slightly decreased compared with initial expectations and remained high for enzyme dose calculation (72.5% and 72.6%). The results are similar for months 1 and 5, except for motivation that further decreases. This decrease might be explained by similar factors as the decrease in participation rate (*Methods* section), that is, the rapid acquisition of knowledge, the trial duration, and the stable clinical status of participants. Throughout the trial, results

related to enjoying using the app are stable (63.8%, 65.8%, and 63.1%).

The surveys included a multiple-selection question to identify features that were difficult to use (survey 1) and suggestions for enhancement (surveys 2 and 3). In survey 1, Respondents reported features related to food recording are most particularly difficult to use: MyCyFAPP dishes (reported by 21.9%), food diary (13.5%), and enzyme dose calculation (11.3%). Furthermore, around nine of ten of the suggestions (free text in the surveys) relate to food registration and almost half of those to additional food items. Suggestions are otherwise diverse and include support for medicine registration and app access from different phones allowing both parents to use their own phones.

Table 8. Percentage of respondents who agree or strongly agree with the claim and those who disagree or strongly disagree.

Claim	Week 1		Month 1		Month 5	
	Respondents who agree or strongly agree, %	Respondents who disagree or strongly disagree, %	Respondents who agree or strongly agree, %	Respondents who disagree or strongly disagree, %	Respondents who agree or strongly agree, %	Respondents who disagree or strongly disagree, %
I find the app easy to use.	70.4	11.3	Not asked	Not asked	Not asked	Not asked
I find learning to use the app easy.	83.0	5.0	Not asked	Not asked	Not asked	Not asked
I enjoy using the app.	63.8	10.6	65.8	6.7	63.1	8.3
The app will help (helps) me to find the right enzyme dose.	88.0	3.5	72.5	10.0	72.6	4.8
The app will help (helps) me to follow good eating habits.	Not asked	Not asked	54.2	9.2	53.6	10.7
The app will help (helps) me to understand the treatment.	66.2	7.7	55.0	10.0	53.6	9.5
The app will motivate (motivates) me to follow my treatment.	63.8	13.5	59.2	12.5	48.8	8.3

Variation Between App Users and Varying Motivations

Confirming our earlier research [9], a high degree of individuality in the manifestation of the disease and in patients is observed that affects needs, motivation, and perceived experience. The main influencing factors are health condition, knowledge about the disease and familiarity with the treatment, user behavior and personality (eg, compliance, structure, precision, curiosity, and reflection), and patient age.

Informants show various motivations for using the app, including expected health benefits, adjustment of enzyme dosage, tight follow-up by HCPs, and contribution to research (M1 and M2). Conversely, good health, experience and knowledge, personality, and age may reduce motivation (M3). Although HCPs expected patients with few GI symptoms to get less value from the app, most informants reported increased knowledge. On the basis of surveys 1 and 2, we tested whether positive anticipation after 1 week was connected to actual satisfaction after 1 month. Tests of correlation (Spearman R) show a moderately significant positive relationship between positive expectation and reported satisfaction (ρ=0.22-0.472; P<.001), meaning that respondents with positive expectations were more likely to report satisfaction. Although correlations are significant, they are moderate in size, meaning that a large amount of variation in satisfaction is related to factors other than positive anticipation.

HCPs described decreasing motivation along the trial and explained that increased knowledge and ability to self-manage and trial duration undermine motivation (M4). During vacation time, less structured than school time, patients were also less motivated (M5). Low motivation has a negative impact on the quality of records and, thus, their value for HCPs.

Enzyme Dose Calculation: The Most Appealing Feature

Enzyme dose calculation is the most innovative feature in MyCyFAPP as it supports the transition from learning about enzyme dosage through trial-and-error to an evidence-based method. Calculation support was indeed the feature most used and perceived as most useful. All informants from different target groups with more or less experience benefited from the feature. An analysis of variance test with Bonferroni post hoc tests showed significant differences between the reported satisfaction of the app features ($F_{3.476}$=3.151; P=.03), with enzyme dosage being the highest (significantly higher than that for eating habits).

Beyond an accurate estimation of the dose, some informants learned about the dependency between enzyme and fat and some about the relationship between enzyme and symptoms (E1, E2, and E3). Reported benefits include improved health (in terms of gaining weight/height and reduced symptoms), increased confidence and self-efficacy, and child compliance to the treatment (E4, E5, and E6). The calculation is an additional task to the demanding treatment, but for most informants, benefits

counterbalance this drawback. The informants acquired rapid knowledge about dosage and, due to stable diets, found calculation less useful after a few weeks. However, new types of food and diet changes during growth make calculation still relevant.

Despite an overall good experience, the calculation raised confusion when the recommended dose differed significantly from the usual dose. HCPs' support was sometimes needed. Informants from Northern Europe reported changes in enzyme distribution throughout the day but similar total amounts for the whole day. In Spain and Portugal, some informants were recommended much higher doses than usual, and thus expected. Dosage differences between Northern Europe and Southern Europe were earlier found in a previous study conducted in MyCyFAPP [31]. HCPs recommended patients to adjust high doses to one-half or one-third, but informants handled recommendations differently, either giving it a try, adjusting to an intermediate dose, or ignoring it (E7). Although HCPs could set an individual correction factor for enzyme dose calculation, few made use of it due to lack of experience with the approach. Another related concern is the incorrect use of the app. Enzyme dose calculation requires food recording, and this was sometimes not done correctly (E8).

Interviews of HCPs in Italy indicate that the recommended doses were lower than usual, leading to a lack of motivation for using the app. In contrast to Spain and Portugal, no recommendation was given to adjust the dosage. On the basis of survey 2, we tested for differences between respondents in Italy and other countries (Table 9). Survey responses indicate that most respondents in Italy reported less benefit than others from using enzyme dose calculation (t_{118}=2.216; *P*=.04) after 1 month. In addition, Italian respondents scored lower on motivation (t_{118}=2.392; *P*=.02) and understanding of the treatment (t_{118}=2.211; *P*=.03), but not on enjoying using the app or following good eating habits. The findings should be interpreted cautiously because of a low number of respondents in Italy.

Table 9. Comparison of responses after 1 month between Italy and other countries.

Item and country	n	Value, mean (SD)	*P* value	*t* test (*df*=118)
The app will help (helps) me to find the right enzyme dose			.04	2.216[a]
Others	103	4 (0.89		
Italy	17	3.29 (1.26)		
The app will help (helps) me to follow good eating habits			.28	1.082
Others	103	3.61 (0.92)		
Italy	17	3.35 (0.86)		
The app will motivate (motivates) me to follow my treatment			.02	2.392
Others	103	3.73 (0.96)		
Italy	17	3.12 (1.05)		
I enjoy using the app			.38	−0.884[a]
Others	103	3.78 (0.97)		
Italy	17	3.94 (0.66)		
The app will help (helps) me to understand the treatment			.03	2.211
Others	103	3.70 (0.89)		
Italy	17	3.18 (0.95)		

[a]Indicates a significant Levene test for equality of variance.

Nutritional Management Toward a Balanced Diet

The follow-up of food intake was also perceived as a useful feature. Informants reported increased knowledge about nutrition and increased awareness about food intake, often leading to diet changes (N1 and N2). Few informants explicitly reported benefits from using the feature, except parents of small children who feel they were understanding and doing well, leading to less pressure on children (N3). However, interviews indicate that good nutrition has a high status. Several teenagers proudly reported that they adapted their diet. Acknowledgment of diet changes is sometimes needed from HCPs (N4).

On the downside, some participants found nutritional goals difficult to reach and experienced them as negative (N5). HCPs could set individual goals, but this was not done by all (preset goals depending on age were then used). In some centers, HCPs chose not to set goals for young patients, but parents felt that this decision was not well communicated.

Food Recording Needs to Be Simplified

Enzyme dose calculation and nutrition management require food intake recording. Many participants experienced recording difficult the first time they used the app, and, later on, they become a little efficient (F1 and F2). A combination of complex naming of food products and limited search functionality made the retrieval of products difficult. Beyond technical issues, other

challenges included estimation of ingredients' weights, specification of cooking method, and lack of knowledge about dish composition (F3). Most suggestions for app enhancements relate to food recording, for example, product barcode scanning and voice-based recording. Poor usability has led to incorrect recording and incorrect enzyme dose calculation. A dietitian explained that she was worried that some of the patients would not use the app properly, leading to incorrect enzyme intake. Indeed, it occurred in some cases (E8).

Functions facilitating recording were used by few respondents (F4). Training materials developed for HCPs to introduce the app use breakfast to illustrate the creation of *own dishes*. Most respondents created their breakfast, but no other dishes. In addition, some lacked usual dishes, despite the recommended country-specific dishes. These dishes were used by few, and informants reported that they failed to retrieve the dish composition in the app or lacked support for adapting dishes to personal habits.

Other Features Were Used to a Lesser Extent

Most informants only used the health diary at the start of the trial when asked to record bowel movement. Some reported that the diary was useful during the consultations for a more precise description of symptoms and that it helped them to reflect on their health status (O1 and O2). The need for a health diary decreased along the study as GI symptoms were reduced. Informants were willing to share data from health and food diaries, expecting these data will provide better insight to HCPs. Opinions about the educational handbook vary. Some informants perceived the information to be better explained and more trustable than the information available on the internet, whereas others did not find it novel (O3 and O4). For many, the handbook was not a central feature. However, teens who need to learn more about CF or wish to explain the disease to peers may benefit from the information (O5).

The digital approach has an impact on the interaction between patients and HCPs. Informants appreciated receiving motivational messages, although some of them sometimes found messages a bit annoying (O6 and O7). Some lacked support for sending messages to HCPs, and, conversely, some HCPs had wished feedback on their messages (O8). Instead, contact with HCPs was taken by phone, email, or WhatsApp.

Health Care Professionals' Experience With the Professional Web Tool

The following results are based on interviews with HCPs. Dietitians appointed for follow-up during the trial worked intensely with the PWT. Most doctors used PWT less but still experienced it in actual settings and used it themselves. This section focuses on the reported benefits and disadvantages of PWT. Single features are not discussed specifically. In general, HCPs, most of them involved in the design, found PWT easy to use and useful. Preferences regarding features differed, but food records were highlighted as very useful by most of them, especially by dietitians. Illustrative quotes are provided in Tables 10-12.

Table 10. Illustrative quotes related to using the professional web tool: perceived benefits.

Id	Concern	Interviewee	Quote
B1	Patient information quality	Dietician, C4	"That the patient can record in the app and we can see like what they have eaten so that we have a very close follow-up of what the patient eats and also when he/she reports the Creon dosage we can see if it corresponds to the theoretical Creon dosage and also that we can see symptoms. If the patient report symptoms, like nausea or diarrhoea, we have close follow-up. We can see it and we can contact patients if we see abnormalities."
B2	Identifying habits	Dietician, C1	"[Now, we] Really know what they eat. With the [paper food] records we used [before the app], it was harder to interpret. And, now it is more accessible. It is useful too, to see if they change eating habits depending on whether it's weekend, midweek, holidays; because this with the [paper] records is a bit difficult to see."
B3	Patient information quality	Dietician, C6	"In some case that I have some doubts about what the patient tells me, I control and know the two information. It was helpful to control what they have told you yes, in some case it is important to have feedback from the PWT [...] No, I don't think we have more information, maybe we have more correct information."
B4	Personalized advice	Dietician, C1	*Do you think you are giving a better service?*[a] Much better of course, especially personalized, which I think is what they most have noticed. Make it something for them, in real time and according to what they eat.
B5	Saving time: when contacting patients	Dietician, C2	"Sometimes it takes a lot of time because you call them in the morning, you get their voicemail, you leave a message in the voicemail, [...]. And in the afternoon, they have not called you back, so you call again. And every time before you call them you do check their dossier for how they are doing. So, you are actually reading on a patient again every time before you call them. And then they don't pick up, I waisted another 5 minutes. It would be easy to communicate with the web tool for patients that really need that closer follow-up [...] And now I could see before the consultation already what they were eating, so it took me less time during the consultation to ask about that. Because I already knew a bit what their eating habits are."
B6	Saving time: from paper records to digital records	Doctor, C3	*So during the trial you did not use the diet paper questionnaires you were used to?*[a] "At the beginning yes, for comparing. But after, we did not use it, because we had the applications. It is much faster. Better data and faster."
B7	Saving time: reduced workload before and during consultation, no longer need for manual calculation	Dietician, C1	*Do you think that the web tool helps you do your daily job?*[a] "Yes, it makes it easier, more enjoyable, I save time. [...] The app does our function a bit. For example, it took half an hour to explain the dietary record from the last 3 months ago. Now the app tells it, but in real time. The content of the app is very useful because if the food record shows that the patient has not an adequate nutritional intake, he is redirected to the corresponding chapter in the educational content. And he does not search other sources, that is also important. Time is saved before consultation because the graphics show what is happening and it is easy to identify why. And then during the consultation, because you have already explained [the patient] with the app what he is doing wrong or good. [...] So above all, the nutritional control [is helpful], and for me the most important, the calculation of the nutritional intake, because up to now it was done using a manual spreadsheet. So I save time."
B8	Saving time in yearly control	Doctor, C2	*Would it be useful for the yearly control?*[a] "I think that would be very practical, yes. We now use food records on paper. If we had them electronically and could calculate automatically, that would be a good application."
B9	Closer relationship, dialogue	Doctor, C2	"But if you really see the hard data, that is something you can really share with them, like "look what is happening", you can start a conversation about it, it is not that you want to blame them but more "let's see what is happening" and can we think together about the solution on how you can improve your compliance [...], but as a tool for a clinician it is great and it can really improve your practice and can give more information to have like a real useful conversation and to find more in partnership with parents and patients."
B10	Tighter follow-up, closer relationship	Dietician, C2	*Do you know the patient better?*[b] "Not really, it's just that you follow them more and are in more contact... Maybe that's because they are participating in the study, that you are helping them with things, you reach out to them to ask about the app, how it is going. That's the kind of stuff that makes the connection closer. I don't really think that the connection is closer just because they use the app. Some of the patients I did not speak with a lot, the app was fine, they did not have questions."

[a]Text in italics are questions to health care professionals.

Table 11. Illustrative quotes related to using the professional web tool: perceived disadvantages.

Id	Concern	Interviewee	Quote
D1	Useful with patients' information	Doctor, C4	"[...] so, if they [patients] don't fill it in, the clinician won't look at it [...]; *Do you see an impact if the app was used in real practice?*[a] I think that it is only the case if patients often fill information. If not frequently used, then it is also difficult to say what the impact is on their daily life and whether it will reflect on compliance or whether you could use it for symptom control or looking for causes of abdominal pain."
D2	Patients will not use app regularly	Doctor, C2	*What do you mean with "it has already a lot of impact in the daily life"?*[a] "Like the normal CF therapy, taking pills, doing nebulisation, doing physiotherapy, they have to do sports, think about Creon, so they have to think about collecting their medication in time at the pharmacy, if it is warm, they have to take salt supplements, ... Well it goes on and on and on. It is quite an organisation already. And, if you have more children or more children with CF, that is even harder to handle. If you then ask the parents "well, you just have to fill in the app once or twice a day", it adds to all the other things that you need to do and look at yourself. You really have to think about how it impacts on the daily life of family, sometimes it is a real fulltime job. There are parents who stop working because of taking care of the children."
D3	A close follow-up is time consuming	Dietician, C2	*Do you feel that the tool requires a lot of effort if you have to follow-up tightly?*[a] "Yes. It does take a lot of time and now it is just 17 patients. But if all of our patients were using it which is like 150 patients. I could not send 150 patients a message twice a week."
D4	A close follow-up is time consuming, extra effort for clinicians	Doctor, C2	*If you had these shared data, would you look at them between consultations or just during the consultation?*[a] "Maybe, if it was very easy to access, it would be quite useful, I think. *Also, between consultations?* [...] I think I would not look at the data when the patients are not coming for consultation. We don't have time to do that; [...] it takes effort for clinicians to log in into the system and to look up all the data. [...] We have the luxury to have 20 min per patient, but still this is quite short, and we have to administrate everything and also talk to the parents, find a plan and execute your plan as well."

[a]Text in italics are questions to health care professionals.

Table 12. Illustrative quotes related to using the professional web tool: interest in future use.

Id	Concern	Interviewee	Quote
I1	Future frequency of use—daily	Dietician, C4	*Would you like to use in the future?*[a] "Yes, off course, I really like it [...] I would like to use the PWT on a daily basis."
I2	Future frequency of use—monthly	Dietician, C2	"But I would not use it twice a week to send messages. *You would not follow either if they are registering things?*[a] Not as much as I did during the trial now. Maybe once a month or before they come to the hospital."
I3	Gaining weight, newly diagnosed	Dietician II, C4	"I would check it like on a weekly basis and also when I know when some patients have trouble gaining weight that I will follow them closer, or when there are new diagnosis that I also can follow them closer like when they are home, I would use it more for the kids who have problems or if parents ask me to check."

[a]Text in italics are questions to health care professionals.

Positive Impact on Health Care Services and Workload

The main benefit HCPs reported is that PWT allows them to obtain more information about patients, for example, regular information about symptoms, eating habits, and enzyme intake. This information facilitates a closer follow-up because the more information HCPs have access to, the easier it is to compare it with other insights about patients and to interpret it. Correlations between eating habits, medicine intake, and symptoms can be detected, and HCPs can react quickly when noticing adverse symptom development (B1). Furthermore, HCPs highlighted that continuous monitoring provides a better overview of patients' behaviors. Within regular visits every 3 months, HCPs only get a small glimpse or a very general overview of patients' eating habits and health status. PWT allows a fuller picture of the course of the disease and the patients' management of it (B2).

HCPs also reported that information communicated by patients is more reliable and accurate. Normally, patients do not record data systematically. They forget details or get information mixed up. Empowering patients to record events at the time they occur, HCPs felt that the data they receive are more precise and better reflect the reality of patients' status (B3). Having more and more precise information during the trial, some HCPs reported that they were able to give better and more personalized advice (B4).

Not only do HCPs think that they can deliver a better service using PWT but they also see a positive effect on their workload. Using PWT, the time needed for consultations with patients can be reduced. One HCP reported that she was able to get in contact

with patients more efficiently (B5); others reported that information can be analyzed quicker than when using paper food records (B6). In particular, dietitians looked at the registered information beforehand and spent less time on asking for symptoms or eating behaviors (B7). PWT can also help to make work easier. For instance, dietitians collect and assess food records on a regular basis. This process becomes easier with the digital approach (B8).

Furthermore, some HCPs experienced that PWT facilitated new positive interactions with patients. On the basis of the registered information, an HCP reported that it was possible to start a more objective dialogue with patients and to involve patients more intensely in their treatment (B9). Several HCPs reported that relationships with patients became closer. However, they were unsure if this was only due to the frequent interactions required by the clinical trial protocol or due to the messaging support in PWT (B10). Nonetheless, HCPs perceived this closer relationship to be positive. Not all HCPs noticed a change in the interaction with patients. Some HCPs reported very good relationships with patients, with or without the approach.

Drawbacks Related to Available Information and Workload

The HCPs also reported about challenges and disadvantages they experienced using PWT. One notable challenge is that the usefulness of tools depends strongly on the data entered by patients. If patients do not record data or record incomplete or incorrect data, PWT is of little or no use (D1). HCPs were worried that it can be a burden for patients and parents to record data and did not expect all patients to fill in information, at least over a longer period (D2).

Some HCPs also reported an additional effort with a close follow-up using PWT. In the clinical trial, with resources specifically allocated for this task, HCPs were able to invest extra effort, but they did not expect to be able to continue such a close follow-up in their regular practice for a higher number of patients (D3 and D4). This may seem contradictory to the earlier described reduced workload. If we take into account that HCPs were asked to communicate with patients much more often than they were used to during the trial, PWT itself does not necessarily mean more work for HCPs, but the tool offers new means of interaction that, depending on the self-management approach, can imply more effort for HCPs than current procedures.

Potential Future Use of Professional Web Tool

Most HCPs were interested in using PWT in the future, though they have very different ideas regarding the frequency of use. Some think of using the tool daily, and others are only interested in using the tool on a weekly or monthly basis (I1 and I2). As a close follow-up of all patients would be time consuming, some HCPs suggested using PWT for specific patient groups or using specific functionalities. This would reduce their own and their patients' efforts. Specific patient groups would be those who generally need a closer follow-up or those who need tight support for a limited period, for example, newly diagnosed children and their parents and patients currently not feeling well (I3).

Discussion

Principal Findings

Our results show a positive experience of the proposed self-management approach. They indicate that challenges relate more to human factors, context of use, and lack of experience with digital self-management than to technology. Confirming previous findings, technology literacy level [32-34], positive reinforcement [35], and contextual factors [36] influence the user experience. Patients get less eager to use the app when they have reached their goals [16,36].

Adopting a user-centered design approach, we expected the app and PWT to be easy to use and useful. They mostly are, with the exception of food recording. Feature acceptance varies, reflecting the diversity of needs that we and others previously discussed [9,15]. Enzyme dose calculation was the most useful. In addition, users benefited from nutrition management. Patients and parents increased their awareness about nutrition, and HCPs collected more reliable patient data. Patients and parents were able to rapidly increase self-efficacy. They adapted enzyme intake and nutritional habits when needed, and some reported enhanced quality of life. For HCPs, rich and reliable data enabled personalized follow-up. Although tight patient follow-up may require additional workload, messaging and digital food recording save time.

Similar to other studies, our study shows the importance of trust in information (eg, educational handbook) [37,38]. In contrast to other studies, privacy was not a concern in our study [39,40]. Patients who already used to tell HCPs about their behaviors were willing to share data and expected them to be useful for follow-up.

Recommendations

On the basis of these findings, we propose recommendations for the successful adoption of digital self-management. As there is an overlap between our findings and those from previous research, it is relevant to also consider these recommendations for self-management of other chronic diseases in addition to CF.

Identifying Relevant Use Cases

Motivating patients to use health apps over a long period is a recurrent topic in mHealth. The purpose of self-management is, however, to support the acquisition of self-efficacy skills [41], not to motivate for using an app. Despite the relatively good clinical status of trial participants, many benefited rapidly from using the app. Over time, as participants acquired new skills, motivation tended to decrease, leading to poorer registered data. Rather than seeking motivational approaches to keep using the app, we propose to identify relevant *use cases*, that is, patient groups and situations where patients are most likely to benefit from using the app. From a professional viewpoint, use cases would reduce the workload.

Obviously, the worsening of symptoms and the occurrence of new symptoms are a relevant use case, as symptom descriptions and food intake provide HCPs a basis for diagnosis and recommendations. HCPs suggested focusing on learning cases,

such as parents of young patients and adolescents. The latter case gave positive results for juvenile arthritis [42]. As we observed that patients with few symptoms can benefit from using the app, using the app a few days before the regular consultations or in connection with the annual check [8] is also a relevant use case.

Personalizing and Supervising

The need for personalization is often discussed in connection to technology [43-46]. Our study also identifies the need to personalize the introduction of technology. Technology literacy level and capacity to learn, seek new insights, and reflect have an impact on the adoption of solutions. Supervision by HCPs is, therefore, essential when introducing digital self-management and should be tailored to patients' skills. Our study also indicates the importance of a regular follow-up by HCPs for motivation.

Advice should be provided to patients during consultations or between consultations using messaging. For example, advice may relate to dealing with unusual enzyme doses to explain details that are easy to overlook for correct calculation, for example, fat registration and food amount. Some patients need support to use the app efficiently and effectively, for example, creating dishes. Some need guidance for exploring features and understanding their rationale, for example, relating symptoms to food and enzyme intake or retrieving hints about nutritional goals in the educational handbook.

Patients should be gradually introduced to the app features. In our study, all features were presented to patients and parents at the start of the trial. The app is comprehensive, and we observed that some users were not aware of all features. A gradual introduction can be addressed technically. The ultimate goal is to reduce the need for follow-up by HCPs to support patients in countries where health care services are limited. The features could be activated gradually, and patients provided hints. For example, the app could detect an unexpectedly high number of calories in food recording and notify the user.

Developing Health Care Professionals' Skills

As digital self-management is a new practice, HCPs need to develop new skills. Although HCPs were trained to teach patients to interact with the app, our results indicate that tutorials should cover guidelines about the process, for example, gradual introduction to features and advice for app exploration. Best practices developed by parents should be included in tutorials, for example, dealing with a child who eats less than expected or speeding up food recording.

Beyond technical aspects, HCPs need advice for collaborating with patients. Most importantly, HCPs should agree with patients about the frequency and type of follow-up. They should also clarify what parts of self-management are relevant. Similar to the prescription of a medication, app usage should be adapted to patients. For example, nutritional goals may be stressful for some individuals and may be irrelevant for young patients. Dietitians were found to be more or less skilled in writing motivational messages. Practices can be collected and shared between HCPs. The availability of predefined messages in PWT should be reconsidered. Automatic generic health messages may appear as depersonalized and weakened relationships with HCPs [47].

User-Centered Design in the Development of Content

Trial participants reported the complexity of product names in the app. Although software development adopted a user-centered design approach, food databases were developed by dietitians without involving users. In addition, food databases in native languages were not ready for testing at midterm evaluation. As digital solutions are increasingly being developed in clinical settings, user-centered design is a software engineering practice that HCPs should become familiar with.

Future Research

Digital food recording is generally complex, with a trade-off between usability and accuracy. Technology for food image recognition was not yet accurate enough when we developed the app [48]. Recent solutions still seem too inaccurate for the purpose of enzyme calculation.

The approach to enzyme dose calculation necessitates refinement for optimal support. Guidelines for setting the individual correction factor as well as for specifying the cooking method are needed. As the enzyme dose mainly depends on the amount and type of fat, simplification of food recording should be investigated. Simplification may, however, act to the detriment of an increased knowledge about nutrition.

Strengths and Limitations

For practical and economic reasons, interviews were conducted on the web. Video calls were set up to reduce the *distance* between the interviewee and interviewer. In some countries, English was used rather than the native language. Using English was preferred over involving an interpreter because of the explorative approach of interviews. No language barrier was experienced during the interviews.

The clinical trial context does not truly represent the regular clinical context. The tight follow-up by HCPs and the commitment of trial participants to contribute to research may have influenced the app and PWT usage. Still, the research allowed us to identify usability shortcomings and factors relevant in clinical contexts. Deployment in a real context is needed to strengthen the insights about organizational issues.

Although the clinical protocol specified requirements on app usage, it left trial participants some freedom, leading to differences in the way the app was used. HCPs also had different expectations for the approach and were more or less acquainted with mobile technology. HCPs used PWT more or less often, and messaging was performed in different ways. This influences the support provided to patients. Such differences would, however, also apply in a regular clinical context. Most importantly, extracting practices that lead to positive outcomes and building guidelines upon them are needed.

Acknowledgments

The authors of this paper, on behalf of the MyCyFAPP consortium, acknowledge the European Union and the Horizon 2020 Research and Innovation Framework Programme for funding the project (ref. 643806). The authors would like to thank all project partners for their collaboration during participant recruitment and project management. Without the dedication of participants in terms of time, effort, and valuable input, this publication would not have been possible. The authors would like to thank all the participants who contributed to this work.

Authors' Contributions

JF, TV, AZ and GIS designed the research study, conducted interviews (respectively in Netherlands and Belgium, Portugal, Italy and Spain), and collaborated on the thematic analysis. JF summarized the findings for the self-management app (patients and caregivers), TV for the app (HPCs), and AZ for PWT. ALA, JF and TV analysed the quantitative data. JCL developed the algorithm for enzyme dose calculation. ES led the SW development. ES, PHH, AFV and JLBM designed the system architecture and implemented the SW. GIS coordinated the helpdesk during the clinical trial and ES, PHH, AFV and JLBM maintained the SW. JF coordinated the writing of the manuscript, described previous research and methods, and developed the discussion. JF, TV, AZ, GIS and JCL collaborated on writing other parts. All authors contributed to the review of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Features and screenshots.
[PDF File (Adobe PDF File), 2411 KB - mhealth_v8i7e15896_app1.pdf]

Multimedia Appendix 2
Questionnaires and interview guidelines.
[PDF File (Adobe PDF File), 143 KB - mhealth_v8i7e15896_app2.pdf]

References

1. Rijken M, Jones M, Heijmans M, Dixon A. Supporting self-management. In: Nolte E, McKee M, editors. Caring for People With Chronic Conditions: A Health System Perspective. New York, USA: Open University Press; 2008:114-142.
2. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
3. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
4. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
5. Venkatesh V, Morris MG, Davis GN, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
6. Cystic Fibrosis Europe. URL: https://www.cf-europe.eu/ [accessed 2020-05-27]
7. Dodd ME, Webb AK. Understanding non-compliance with treatment in adults with cystic fibrosis. J R Soc Med 2000;93(Suppl 38):2-8 [FREE Full text] [Medline: 10911812]
8. Conway S, Balfour-Lynn IM, de Rijcke K, Drevinek P, Foweraker J, Havermans T, et al. European cystic fibrosis society standards of care: framework for the cystic fibrosis centre. J Cyst Fibros 2014 May;13(Suppl 1):S3-22 [FREE Full text] [doi: 10.1016/j.jcf.2014.03.009] [Medline: 24856776]
9. Floch J, Zettl A, Fricke L, Weisser T, Grut L, Vilarinho T, et al. User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach. JMIR Mhealth Uhealth 2018 May 8;6(5):e113 [FREE Full text] [doi: 10.2196/mhealth.8236] [Medline: 29739742]
10. Calvo-Lerma J, Martinez-Jimenez CP, Lázaro-Ramos JP, Andrés A, Crespo-Escobar P, Stav E, MyCyFAPP. Innovative approach for self-management and social welfare of children with cystic fibrosis in Europe: development, validation and implementation of an mhealth tool (MyCyFAPP). BMJ Open 2017 Mar 16;7(3):e014931 [FREE Full text] [doi: 10.1136/bmjopen-2016-014931] [Medline: 28302638]
11. Borowitz D, Gelfond D, Maguiness K, Heubi JE, Ramsey B. Maximal daily dose of pancreatic enzyme replacement therapy in infants with cystic fibrosis: a reconsideration. J Cyst Fibros 2013 Dec;12(6):784-785 [FREE Full text] [doi: 10.1016/j.jcf.2013.05.011] [Medline: 23809508]

12. Calvo-Lerma J, Fornés-Ferrer V, Peinado I, Heredia A, Ribes-Koninckx C, Andrés A. A first approach for an evidence-based in vitro digestion method to adjust pancreatic enzyme replacement therapy in cystic fibrosis. PLoS One 2019;14(2):e0212459 [FREE Full text] [doi: 10.1371/journal.pone.0212459] [Medline: 30794618]
13. Calvo-Lerma J, Hulst J, Boon M, Martins T, Ruperto M, Colombo C, et al. The relative contribution of food groups to macronutrient intake in children with cystic fibrosis: a european multicenter assessment. J Acad Nutr Diet 2019 Aug;119(8):1305-1319. [doi: 10.1016/j.jand.2019.01.003] [Medline: 30862484]
14. Turck D, Braegger CP, Colombo C, Declercq D, Morton A, Pancheva R, et al. ESPEN-ESPGHAN-ECFS guidelines on nutrition care for infants, children, and adults with cystic fibrosis. Clin Nutr 2016 Jun;35(3):557-577. [doi: 10.1016/j.clnu.2016.03.004] [Medline: 27068495]
15. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mhealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
16. 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]
17. Boon M, Calvo-Lerma J, Claes I, Havermans T, Asseiceira I, Bulfamante A, et al. Use of a mobile application for self-management of pancreatic enzyme replacement therapy is associated with improved gastro-intestinal related quality of life in children with cystic fibrosis. J Cyst Fibros 2020 Apr 22:- epub ahead of print. [doi: 10.1016/j.jcf.2020.04.001] [Medline: 32335023]
18. Oates B. Researching Information Systems and Computing. Thousand Oaks, CA: Sage Publications; 2019.
19. Hevner AR, March ST, Park J. Design science in information systems research. MIS Q 2004;28(1):75. [doi: 10.2307/25148625]
20. -. ISO 9241-210:2010 Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems. International Organization for Standardization. 2010. URL: https://www.iso.org/standard/52075.html [accessed 2020-06-05]
21. Vilarinho T, Floch J, Stav E. Co-Designing a mHealth Application for Self-Management of Cystic Fibrosis. In: Proceedings of the IFIP Conference on Human-Computer Interaction. 2017 Presented at: INTERACT'17; September 25-29, 2017; Bombay, India. [doi: 10.1007/978-3-319-67687-6_1]
22. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage Publications; 2019.
23. Kristensen G, Ravn MN. The voices heard and the voices silenced: recruitment processes in qualitative interview studies. Qual Res 2015 Feb 9;15(6):722-737. [doi: 10.1177/1468794114567496]
24. Etikan I. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Sci 2016;5(1):1. [doi: 10.11648/j.ajtas.20160501.11]
25. Bryman A. Integrating quantitative and qualitative research: how is it done? Qual Res 2016 Aug 15;6(1):97-113. [doi: 10.1177/1468794106058877]
26. Flick U. An Introduction to Qualitative Research. Thousand Oaks, CA: Sage Publications; 2019.
27. Klein HK, Myers MD. A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Q 1999 Mar;23(1):67. [doi: 10.2307/249410]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Barnum CM. Usability Testing Essentials: Ready, Set...Test!. New York, USA: Elsevier; 2019.
30. Lee Y, Kozar KA, Larsen KR. The technology acceptance model: past, present, and future. Commun Assoc Inf Syst 2003;12:-. [doi: 10.17705/1cais.01250]
31. Calvo-Lerma J, Hulst JM, Asseiceira I, Claes I, Garriga M, Colombo C, MyCyFAPP Project. Nutritional status, nutrient intake and use of enzyme supplements in paediatric patients with cystic fibrosis; a European multicentre study with reference to current guidelines. J Cyst Fibros 2017 Jul;16(4):510-518 [FREE Full text] [doi: 10.1016/j.jcf.2017.03.005] [Medline: 28320633]
32. Smaradottir B, Gerdes M, Fensli R, Martinez S. Usability evaluation of a COPD remote monitoring application. Stud Health Technol Inform 2015;210:845-849. [Medline: 25991274]
33. Johnson KB, Patterson BL, Ho Y, Chen Q, Nian H, Davison CL, et al. The feasibility of text reminders to improve medication adherence in adolescents with asthma. J Am Med Inform Assoc 2016 May;23(3):449-455 [FREE Full text] [doi: 10.1093/jamia/ocv158] [Medline: 26661717]
34. Granger D, Vandelanotte C, Duncan MJ, Alley S, Schoeppe S, Short C, et al. Is preference for mHealth intervention delivery platform associated with delivery platform familiarity? BMC Public Health 2016 Jul 22;16:619 [FREE Full text] [doi: 10.1186/s12889-016-3316-2] [Medline: 27450240]
35. Cook KA, Modena BD, Simon RA. Improvement in asthma control using a minimally burdensome and proactive smartphone application. J Allergy Clin Immunol Pract 2016 Jul;4(4):730-7.e1. [doi: 10.1016/j.jaip.2016.03.005] [Medline: 27107690]
36. 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 8;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]

37. Riis A, Hjelmager DM, Vinther LD, Rathleff MS, Hartvigsen J, Jensen MB. Preferences for web-based information material for low back pain: qualitative interview study on people consulting a general practitioner. JMIR Rehabil Assist Technol 2018 Apr 2;5(1):e7 [FREE Full text] [doi: 10.2196/rehab.8841] [Medline: 29610108]
38. 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]
39. Giunti G, Kool J, Romero OR, Zubiete ED. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 9;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
40. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mhealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
41. Aujoulat I, d'Hoore W, Deccache A. Patient empowerment in theory and practice: polysemy or cacophony? Patient Educ Couns 2007 Apr;66(1):13-20. [doi: 10.1016/j.pec.2006.09.008] [Medline: 17084059]
42. Waite-Jones JM, Majeed-Ariss R, Smith J, Stones SR, van Rooyen V, Swallow V. Young people's, parents', and professionals' views on required components of mobile apps to support self-management of juvenile arthritis: qualitative study. JMIR Mhealth Uhealth 2018 Jan 19;6(1):e25 [FREE Full text] [doi: 10.2196/mhealth.9179] [Medline: 29351898]
43. Lubberding S, van Uden-Kraan CF, te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. Improving access to supportive cancer care through an ehealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015 May;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
44. Simons L, Valentine AZ, Falconer CJ, Groom M, Daley D, Craven MP, et al. Developing mhealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs. JMIR Mhealth Uhealth 2016 Mar 23;4(1):e31 [FREE Full text] [doi: 10.2196/mhealth.5009] [Medline: 27009498]
45. Velu AV, van Beukering MD, Schaafsma FG, Frings-Dresen MH, Mol BW, van der Post JA, et al. Barriers and facilitators for the use of a medical mobile app to prevent work-related risks in pregnancy: a qualitative analysis. JMIR Res Protoc 2017 Aug 22;6(8):e163 [FREE Full text] [doi: 10.2196/resprot.7224] [Medline: 28830851]
46. Switsers L, Dauwe A, Vanhoudt A, van Dyck H, Lombaerts K, Oldenburg J. Users' perspectives on mhealth self-management of bipolar disorder: qualitative focus group study. JMIR Mhealth Uhealth 2018 May 2;6(5):e108 [FREE Full text] [doi: 10.2196/mhealth.9529] [Medline: 29720363]
47. Al Dahdah M. Health at her fingertips: development, gender and empowering mobile technologies. Gender Technol Dev 2017 Oct 25;21(1-2):135-151. [doi: 10.1080/09718524.2017.1385701]
48. Consolvo S. Designing for healthy lifestyles: design considerations for mobile technologies to encourage consumer health and wellness. FNT Human–Comput Interact 2012;6(3–4):167-315. [doi: 10.1561/1100000040]

Abbreviations

CF: cystic fibrosis
GI: gastrointestinal
HCP: health care professional
ICT: information and communication technology
mHealth: mobile health
PWT: professional web tool

Edited by G Eysenbach; submitted 16.08.19; peer-reviewed by K Stasiak, H Thabrew, A Sarradon-Eck; comments to author 03.10.19; revised version received 28.11.19; accepted 28.04.20; published 08.07.20.

Please cite as:
Floch J, Vilarinho T, Zettl A, Ibanez-Sanchez G, Calvo-Lerma J, Stav E, Haro PH, Aalberg AL, Fides-Valero A, Bayo Montón JL Users' Experiences of a Mobile Health Self-Management Approach for the Treatment of Cystic Fibrosis: Mixed Methods Study JMIR Mhealth Uhealth 2020;8(7):e15896
URL: https://mhealth.jmir.org/2020/7/e15896
doi:10.2196/15896
PMID:32673237

Original Paper

Drivers of Mobile Health Acceptance and Use From the Patient Perspective: Survey Study and Quantitative Model Development

Tânia Salgado[1*], MSc; Jorge Tavares[1*], PhD; Tiago Oliveira[1*], PhD

NOVA Information Management School, Universidade Nova de Lisboa, Lisbon, Portugal

[*]all authors contributed equally

Corresponding Author:
Jorge Tavares, PhD
NOVA Information Management School
Universidade Nova de Lisboa
Campus de Campolide
Lisbon, 1070-312 Lisboa
Portugal
Phone: 351 213828610
Email: d2012072@novaims.unl.pt

Abstract

Background: Mobile health (mHealth) has potential to play a significant role in realizing a reversal of the current paradigm in health care toward a more patient-centric and more collaborative system to improve the outcomes obtained along with the quality and sustainability of health care systems.

Objective: The aim of this study was to explore and understand individual mHealth acceptance drivers between two groups of users: those with chronic health conditions and those without.

Methods: The extended unified theory of acceptance and usage of technology (UTAUT2) was enhanced with a new health-related framework: behavior intention to recommend and new mediation effects. We applied partial least squares (PLS) causal modeling to test the research model.

Results: We obtained 322 valid responses through an online questionnaire. The drivers of behavior intention with statistical significance were performance expectancy (β=.29, *P*<.001), habit (β=.39, *P*<.001), and personal empowerment (β=.18, *P*=.01). The precursors of use behavior were habit (β= .47, *P*<.001) and personal empowerment (β=.17, *P*=.01). Behavior intention to recommend was significantly influenced by behavior intention (β=.58, *P*<.001) and personal empowerment (β=.26, *P*<.001). The model explained 66% of the total variance in behavior intention, 54% of the variance in use behavior, and 70% of the variance in behavior intention to recommend.

Conclusions: Our study demonstrates a significant role of personal empowerment, as a second-order construct, in the mHealth acceptance context. The presence of a chronic health condition predicates an impact on acceptance of this technology.

(JMIR Mhealth Uhealth 2020;8(7):e17588) doi:10.2196/17588

KEYWORDS

digital health; mHealth; UTAUT2; health management; patient empowerment

Introduction

Context

The focus of health policymakers worldwide is to change health care models from disease treatment to disease prevention, necessitating a shift to more patient-centric and more collaborative actions, and information technology is one of the paths highlighted to best achieve this goal [1-5]. In this study, we focused on the use of mobile health (mHealth), specifically on smartphones, as a health management platform. In Portugal, smartphone penetration was reported to reach up to 74.9% of the population in December 2017 [6]. The field of mHealth is in a state of rapid expansion, from a global rate of 36% use in 2016 to 46% in 2018, and nearly half (48%) of all health care consumers were using mobile/tablet apps compared to only 16% reported in 2014 [7].

The main limitations of previously published research on mHealth include underpowered pilot data [8] in specific groups of patients or with a particular app [1], or a focus on only health care professionals [9,10]. In this study, we developed a new

research model to explore and better understand individual mHealth acceptance drivers, mainly to determine how unique drivers related to health care such as patient empowerment can influence the adoption of mHealth among health care consumers with and without a chronic condition.

Theoretical Background

mHealth is a subset of the larger field of electronic health (eHealth), which was originally defined under the term "unwired e-med" [11]. Many definitions have arisen since then, and in the present work, we follow the World Health Organization definition of mHealth as the use of mobile devices such as mobile phones, patient monitoring devices, personal digital assistants, and wireless devices for medical and public health practice [12].

In the last two decades, mainly after the emergence of mobile phones followed by smartphones, several researchers have been studying electronic and mobile technologies as possible solutions to address health care challenges. Relevant studies performed in the most recent years are summarized in Table 1.

Table 1. Electronic health (eHealth) adoption models.

Theory	Dependent variable	Findings	Reference
TAM[a], HBM[b] and TPB[c]	Health consumers' behavior intention of using HIT[d]	-PT[e], PU[f], and PEOU[g] significantly affected health consumers' attitude and behavior intention. -Health consumers' health status, health belief and concerns, subjective norm, HIT characteristics, and HIT self-efficacy had a substantial indirect impact on attitude and behavioral intention through the mediators of PT, PU, and PEOU.	[13]
TAM and TPB	Internet use for health purposes	- Positive influence of perceived health risk and health consciousness on health-related internet use - Perceived health risk positively affects health-related internet use - Health consciousness has a significant positive effect on health-related internet use - Health consciousness contributes to health behavior adoption	[14]
TAM2, Dual-factor model, HBM	Patient's acceptance of smartphone health technology for chronic disease management	- PU of the app was positively influenced by the perceived health threat, relationship with doctor, and PEOU, but negatively affected by resistance to change - Usage experience and self-efficacy positively influenced patients' PEOU - Behavior intention was influenced by enablers of PU and perceived health threat, an inhibitor of resistance to change - Intention of use had a significant weak relationship with actual use	[1]
UTAUT2[h]	EHR[i] portals	- Understanding of the adoption of EHR portals is improved through the use of consumer adoption-specific constructs	[15]
Uses and gratification theory	eHealth usage by sociodemographic factors	- Female gender is a consistent predictor of eHealth usage - Age is primarily influential for health-information seeking	[16]
TAM, UTAUT2	Consumer acceptance of wearable self-tracking devices	- Positive influence of trust, perceived esthetics, personal innovativeness, perceived support of health, perceived support of fitness, and perceived support of well-being on consumer acceptance of wearables	[17]
DMISSM[j]	Evaluation of trust, security beliefs, and privacy of HIT as determinants of health care outcomes	- Increased privacy concerns reduce the frequency of patient access to health records use, positive attitudes toward HIT, and perceptions of patient care quality - Belief in the effectiveness of information security increases the frequency of patient access to health records and a positive attitude toward HIT - Trust in health information had a positive association with attitudes toward HIT and perceived patient care quality	[18]
Personal empowerment	Internet use behavior as a source of information	- Health information seeking is analyzed under three perspectives: professional logic, consumer logic, and community logic	[19]
DIT[k]	Factors impacting patient acceptance and use of consumer eHealth innovations	Health care providers need to consider and address patient characteristics, their social system, and preferences on communication channels, as well as the attributes of the innovation to guarantee its success.	[20]

[a]TAM: technology acceptance model.

[b]HBM: health belief model.

[c]TPB: theory of planned behavior.

[d]HIT: health information technology.

[e]PT: perceived threat.

[f]PU: perceived usefulness.

[g]PEOU: perceived ease of use.

[h]UTAUT2: extended unified theory of acceptance and usage of technology.

[i]EHR: electronic health record.

[j]DMISSM: DeLone and McLean information systems success model.

[k]DIT: diffusion of innovation theory.

The technology acceptance model (TAM) is the most widely used research model in this field, which is commonly combined with other models or with extensions to help explain behavior intention or use behavior. An example of this approach is a study that combined the health belief model (HBM) and TAM to explore the influence of perceived health risk and health consciousness on health-related internet use [14]. In addition, Kim and Park [13] combined the TAM, HBM, and theory of planned behavior to describe health consumers' behavior intention of using health information technologies. Dou et al [1] used a combination of TAM2, the dual-factor model, and HBM to study patients' acceptance of smartphone health technology for chronic disease management, among others.

Extended Unified Theory of Acceptance and Usage of Technology

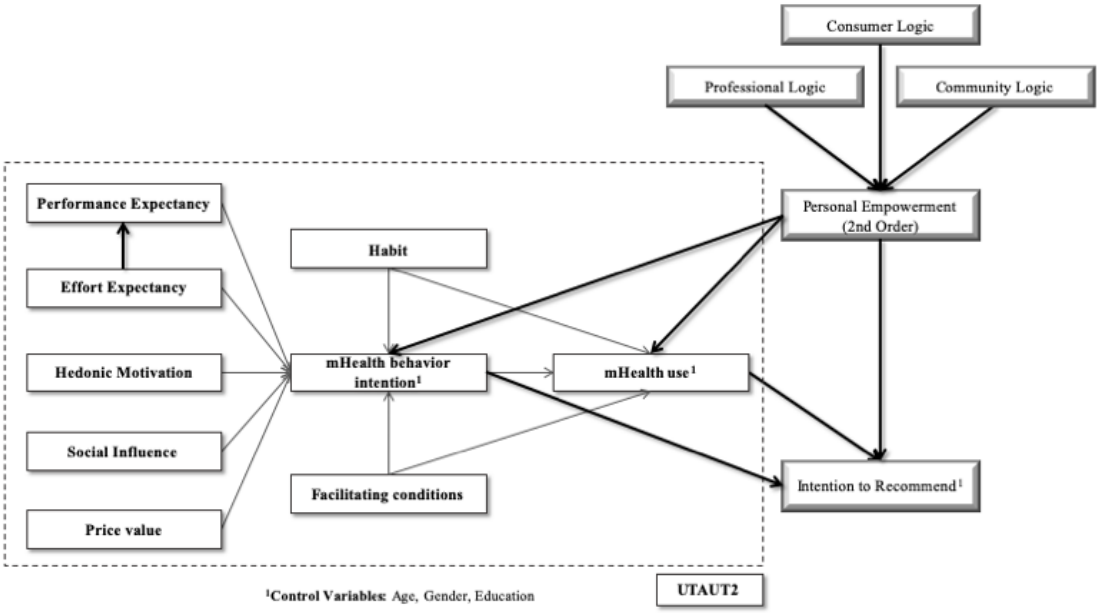
The extended version of the unified theory of acceptance and usage of technology (UTAUT), UTAUT2, is a tailor-made model of acceptance and use of technology. UTAUT2 was adapted from the UTAUT to a more consumer-centered context, with the main differences including the introduction of three new constructs, hedonic motivation, price value, and habit, with moderators of age, gender, and experience. Hedonic motivation and price value explain behavior intention, and habit explains both behavior intention and use behavior. Compared to UTAUT, the extensions proposed in UTAUT2 resulted in a substantial improvement in the variance explained in behavior intention (56% to 74%) and technology use (40% to 52%) [21].

Research Model and Hypotheses

UTAUT2 has already shown good results when applied to the health care context, with the addition of specific health-related constructs [22-24]. Concerning specific health care adoption models, the aim is to achieve an R^2 value of at least 50% [21,25], except when the model deals with constructs related to sensitive topics such as confidentiality or patient behavior in which an R^2 of 20% for the critical dependent variables can be regarded as acceptable [25,26]. If the model uses an existing adoption theory, extended by new constructs or relationships related to health care, these should be able to produce statistically significant results [21,25].

The need for change in health care models related to the scarceness of human resources and to improve health outcomes emphasizes the need to empower patients to manage their own health [27], and mHealth is considered as one of the privileged means to achieve this goal [28]. Accordingly, and due to the lack of studies using a personal/patient empowerment construct (ie, a construct associated with other drivers of information technology adoption), we here propose a model that combines UTAUT2 and personal empowerment as a second-order construct following the work of Lemire [19], which is outlined in Figure 1.

Figure 1. Research model. Original relationships from unified theory of acceptance and usage of technology (UTAUT2) are in black; new relationships from the extended model are in grey.



The hypotheses of this study were developed using the rationale of Venkatesh et al [21] with some adjustments. First, we dropped experience, as our model will be applied at a single time point. Age, gender, and education were used as control variables. In addition, the presence of chronic health conditions was added considering our intention to analyze the results between groups with and without these conditions We established the following research hypotheses based on the following constructs.

Performance Expectancy

Performance expectancy is defined as the consumers' projections of benefits provided by the use of technology and is a good predictor of behavior intention, including in the eHealth context [22].

> *H1(a): Performance expectancy will positively influence behavior intention.*
>
> *H1(b): Performance expectancy will positively influence behavior intention in the chronic condition group, and there will be a significantly greater effect when compared to that of the healthy group.*

Effort Expectancy

Effort expectancy is associated with how easy and simple it seems to be to use a particular technology. Earlier research on the usability of eHealth showed conflicting results, which was

likely related to differences in study populations; nevertheless, in the present work, we have followed the UTAUT2 rationale. Assuming that consumers who perceive mHealth to be more useful and easier to use would have a higher intention to use it [13], we also add a new mediation effect hypothesis:

H2(a): Effort expectancy will positively influence behavior intention.

H2(b): Effort expectancy will mediate the influence of performance expectancy on behavior intention.

Hedonic Motivation

Hedonic motivation is defined by the fun or pleasure derived from using technology [21]. A previous study using health and fitness apps demonstrated that hedonic motivation drives behavior intention [29].

H3: Hedonic motivation will positively influence behavior intention.

Social Influence

Social influence is the degree of consumers' perception that people who are significant to them believe they should use technology. This effect has also been verified in the case of eHealth [15].

H4: Social influence will positively influence behavior intention.

Facilitating Conditions

The construct of facilitating conditions is defined as consumer awareness of the existing support to use technology. Earlier studies also suggested that if patients with a chronic condition have the needed resources and support available, they are more likely to use eHealth technologies [30].

H5(a): Facilitating conditions will positively influence behavior intention.

H5(b): Facilitating conditions will positively influence use behavior.

H5(c): Facilitating conditions will positively influence behavior intention in the chronic conditions group, and there will be a significant increase when compared to the healthy group.

H5(d): Facilitating conditions will positively influence use behavior in the chronic conditions group, and there will be a significant increase when compared to the healthy group.

Price Value

Price value refers to the advantages obtained from a technology considering the costs of using it [21]. The use of mHealth technologies not only offers an easy way to reach health services (eg, scheduling appointments, examination results) but also provides a privileged way to continually monitor patients' parameters, allowing for better follow up with less traveling to health care institutions.

H6: Price value will positively influence behavior intention.

Habit

Habit refers to the automatic nature of a behavior response resulting from learning. As a result of prior experiences, habit has been demonstrated as a good predictor of the adoption of different technologies [21]. Therefore, we have tested this aspect for mHealth adoption.

H7(a): Habit will positively influence behavior intention.

H7(b): Habit will positively influence use behavior.

Behavior Intention

According to previous research on eHealth, the act of using eHealth tools is preceded by the behavior intention to use them [24]; therefore, we applied this concept to the mHealth context.

H8(a): Behavior intention will positively influence use behavior.

Considering the previous hypotheses, and according to earlier research about technology adoption in the health care information technology context indicating that intention to use a technology strongly influences its recommendation to others [22,31], we developed the following hypotheses for testing in the mHealth context.

H8(b): Behavior intention will positively influence the intention to recommend mHealth technologies to others.

H9: Use behavior will positively influence the intention to recommend mHealth technologies to others.

Personal Empowerment

Personal empowerment is the process and outcome through which individuals gain self-confidence and self-efficacy to actively participate in their own health care and ultimately exercise power over decision making concerning their treatment [19]. Therefore, the behavior intention and use of mHealth technologies will be positively influenced by personal empowerment, since it allows for a more active role in health management. We used personal empowerment as a second-order construct. This construct is based on three different forms of logic: professional logic, the process of empowerment in which individuals acquire expert knowledge and put it into practice so that they can act effectively in their personal health; consumer logic, the process of personal affirmation to make decisions based on one's own judgment and resources; and community logic, the dynamics of inclusion in action and social change initiatives developed from a sense of community and participation. Following the rationale of H9, we also hypothesized that behavior intention to recommend technology is influenced by personal empowerment in the same way.

H10(a): Personal empowerment will positively influence behavior intention.

H10(b): Personal empowerment will positively influence mHealth use.

H10(c): Personal empowerment will positively influence behavior intention to use.

H10(d): Personal empowerment will positively influence behavior intention in the chronic conditions group, which will be significantly higher when compared to that of the healthy group.

H10(e): Personal empowerment will positively influence mHealth use in the chronic conditions group, which will be significantly higher when compared to that of the healthy group.

H10(f): Personal empowerment will positively influence behavior intention in the chronic conditions group, which will be significantly higher when compared to that of the healthy group.

Methods

Measurement

All of the measurement items for each of the constructs described above were adapted from Venkatesh et al [21], Tavares and Oliveira [22,24], and Lemire et al [19] with minor modifications to adapt to the mHealth context. The items are described in detail in Multimedia Appendix 1. The questionnaire was developed in English and then translated to Portuguese, which was validated by two translators fluent in both languages. To guarantee that the questionnaire did not lose its original meaning, a back-translation was made to English by a third translator with no previous knowledge of the original questionnaire, which was then compared with the original [32]. The scale items were measured on a 7-point Likert scale from "strongly disagree" (1) to "strongly agree" (7). An exception was made for the use behavior construct, which was measured on a different scale ranging from "never" (1) to "whenever I need" (7) for most of the items following the study of Tavares and Oliveira [24]. Some sociodemographic questions were also added to characterize the study sample. Age was measured in years and gender was coded as a dummy variable (0 or 1), with women represented by 0. The presence of a chronic health condition was also coded as a dummy variable (0 or 1), with its absence represented by 0. Each respondent's education level was assessed by 5 different layers (1, middle school or lower; 2, high school; 3, bachelor degree or postgraduate; 4, master degree; 5, doctorate). The questionnaire started with a brief introduction explaining the mHealth concept (Multimedia Appendix 1) to ensure that all respondents had prior knowledge and contact with mHealth technologies.

Data Collection

To validate the questions and the scales of the survey, a pilot survey was conducted, from which we received 40 responses, assuring that all of the items were reliable and valid. These data were not included in the main survey. According to previous literature on health information technologies, the users of mHealth are mostly younger and have higher education levels [16]; therefore, one of the targets selected for our sample was educational institutions. In addition, considering the goal of analyzing the impact of having chronic health conditions on the use of mHealth technologies, we requested the collaboration of a set of national patients' associations to diffuse our study among their associates and allow us to reach a sample of individuals specifically with these characteristics. By the end of January 2019, an email with a collaboration request and the survey hyperlink was sent to 6 educational institutions in Lisbon and 30 national chronic patients' associations. The request included the study purpose and a statement that anonymity and confidentiality of the information collected were assured, and that by following the hyperlink, they would authorize the use of the data for academic purposes. After this first approach, we received 118 responses. Reminders were sent at 3 and 6 weeks after sending the first email to improve the response rate. Following the reminders, we obtained a total of 322 respondents, including 209 from educational institutions and 113 from the patients' associations. We verified the common method bias through Harman's one-factor test [33] that attests that the total variance for every single factor is always less than 50%. Using Lindell and Whitney's [34] approach, we found a maximum shared variance of 6.5% with other variables after adding a theoretically irrelevant marker variable in the research model, which can be considered low [35]. Therefore, using two independent approaches, we demonstrated that common method bias should not be an issue.

In our total sample, approximately 71% (229/322) of the respondents were women, the average age was 40 years, and 77% were university graduates or higher, which is in line with the literature [20,36]. The patients with chronic health conditions were older on average than the healthy group, and there were also differences in education level between the two groups. The basic characteristics of our sample are shown in Table 2 in more detail.

Table 2. Sample characteristics.

Characteristic	Total (N=322)	Chronic conditions group (n=124)	Healthy group (n=198)
Age (years), mean (SD)	40.03 (13.70)	47.66 (14.30)	35.26 (10.89)
Gender, n (%)			
Male	93 (28.9)	34 (27.4)	59 (29.8)
Female	229 (71.1)	90 (72.6)	139 (70.2)
Education level, n (%)			
Middle school or lower	11 (3.4)	8 (6.5)	3 (1.5)
High school	61 (18.9)	37 (29.8)	24 (12.1)
Bachelor or postgraduate	163 (50.6)	54 (43.6)	109 (55.1)
Master	72 (22.4)	17 (13.7)	55 (27.8)
Doctorate	15 (4.7)	8 (6.5)	7 (3.5)

Data Analysis

We used a structural equation model (SEM)-partial least square (PLS) approach to analyze the data obtained. Smart PLS v3.28 software [37] was used to analyze the relationships defined by the model. PLS is a causal modeling approach and a powerful multivariate analysis technique that enables analysis of the complexity of the model and to test the validity of the theory using empirical models [15,38]. The rationale for choosing this approach was the high model complexity (many constructs and many indicators), the incorporation of formative measured constructs as part of the structural model, and the fact that the PLS-SEM method is oriented to explain the variance of the research model and to detect statistically significant constructs [25,39].

Results

Measurement Model

Since we had both reflective and formative indicators, we applied different measures to assess the reliability and validity of the measurement model. For reflective indicators, we initially evaluated the internal consistency reliability through Cronbach alpha and composite reliability. Table 3 shows that both measures were above .70, demonstrating internal consistency [25]. Validity was examined by the convergent validity and discriminant validity metrics. Convergent validity is assured when each item has outer loadings above 0.70 and when each construct's average variance extracted is 0.50 or higher. As shown in Table 3, both of these criteria were met. Discriminant validity represents the extent to which a construct is empirically distinct from other constructs [40], which was evaluated with three methods: analysis of crossloadings (the outer loading of an indicator should be higher than all its crossloadings), as verified in Table 3; the Fornell and Larcker criterion, which states that the square root of the average variance extracted for each construct should be greater than its higher correlation with others constructs, as confirmed in Table 4; and the Heterotrait-monotrait ratio, which is an estimate of the true correlation between constructs and should be below 0.90 [25], as confirmed in Multimedia Appendix 2.

Table 3. Total sample loadings and crossloadings.

Constructs	PE[a]	EE[b]	SI[c]	HT[d]	FC[e]	HM[f]	PV[g]	PEM[h]	PEM-PL[i]	PEM-CL[j]	PEM-CCL[k]	BI[l]	BIR[m]	CA[n]	CR[o]	AVE[p]
PE														.90	0.94	0.84
PE1	*0.87*[q]	0.37	0.44	0.52	0.31	0.40	0.23	0.52	0.51	0.50	0.40	0.56	0.55			
PE2	*0.93*	0.51	0.39	0.54	0.47	0.49	0.39	0.54	0.55	0.52	0.39	0.61	0.58			
PE3	*0.94*	0.48	0.38	0.52	0.41	0.42	0.37	0.55	0.59	0.52	0.39	0.63	0.59			
EE														.93	0.95	0.82
EE1	0.43	*0.92*	0.12	0.37	0.62	0.41	0.37	0.31	0.32	0.32	0.21	0.32	0.30			
EE2	0.51	*0.93*	0.27	0.50	0.66	0.51	0.43	0.39	0.39	0.37	0.30	0.42	0.38			
EE3	0.41	*0.87*	0.22	0.42	0.61	0.48	0.45	0.39	0.39	0.34	0.32	0.34	0.35			
EE4	0.44	*0.91*	0.13	0.41	0.64	0.43	0.41	0.34	0.36	0.33	0.24	0.38	0.30			
SI														.97	0.98	0.94
SI1	0.43	0.21	*0.97*	0.58	0.29	0.40	0.27	0.55	0.52	0.47	0.48	0.47	0.48			
SI2	0.43	0.18	*0.97*	0.59	0.26	0.44	0.28	0.54	0.54	0.44	0.48	0.51	0.52			
SI3	0.40	0.21	*0.96*	0.59	0.28	0.43	0.31	0.55	0.56	0.45	0.47	0.51	0.52			
HT														.87	0.91	0.73
HT1	0.52	0.46	0.53	*0.92*	0.45	0.59	0.41	0.53	0.54	0.46	0.43	0.64	0.56			
HT2	0.31	0.21	0.49	*0.74*	0.23	0.51	0.28	0.47	0.42	0.40	0.45	0.44	0.42			
HT3	0.61	0.39	0.55	*0.83*	0.40	0.57	0.44	0.62	0.64	0.53	0.49	0.71	0.68			
HT4	0.50	0.51	0.51	*0.92*	0.51	0.60	0.45	0.57	0.58	0.50	0.47	0.68	0.61			
FC														.86	0.91	0.71
FC1	0.34	0.59	0.22	0.38	*0.86*	0.40	0.49	0.40	0.41	0.39	0.29	0.38	0.37			
FC2	0.40	0.70	0.25	0.44	*0.88*	0.45	0.50	0.38	0.39	0.37	0.27	0.43	0.36			
FC3	0.44	0.64	0.25	0.45	*0.91*	0.52	0.57	0.43	0.44	0.42	0.32	0.43	0.36			
FC4	0.29	0.39	0.26	0.31	*0.71*	0.41	0.52	0.40	0.43	0.33	0.33	0.33	0.30			
HM														.90	0.94	0.84
HM1	0.47	0.50	0.40	0.62	0.52	*0.95*	0.47	0.57	0.53	0.48	0.52	0.54	0.52			
HM2	0.52	0.53	0.43	0.70	0.54	*0.93*	0.53	0.60	0.59	0.50	0.53	0.61	0.55			
HM3	0.29	0.31	0.36	0.46	0.36	*0.87*	0.40	0.45	0.40	0.35	0.45	0.42	0.39			
PV														.92	0.95	0.87
PV1	0.31	0.40	0.21	0.36	0.57	0.42	*0.92*	0.47	0.44	0.45	0.38	0.39	0.37			
PV2	0.35	0.43	0.28	0.46	0.58	0.47	*0.96*	0.50	0.49	0.47	0.40	0.50	0.47			
PV3	0.35	0.45	0.33	0.48	0.57	0.54	*0.92*	0.55	0.53	0.51	0.45	0.49	0.49			
PEM-PL														.92	0.94	0.76
PEM-PL1	0.52	0.39	0.56	0.59	0.45	0.51	0.52	0.82	*0.91*	0.67	0.63	0.57	0.61			
PEM-PL2	0.54	0.37	0.51	0.62	0.45	0.52	0.51	0.82	*0.91*	0.68	0.63	0.59	0.61			
PEM-PL3	0.54	0.33	0.51	0.54	0.40	0.50	0.44	0.85	*0.93*	0.68	0.67	0.56	0.60			
PEM-PL4	0.58	0.40	0.47	0.58	0.46	0.51	0.43	0.81	*0.89*	0.66	0.63	0.61	0.62			
PEM-PL5	0.46	0.27	0.39	0.46	0.38	0.42	0.38	0.67	*0.72*	0.58	0.52	0.47	0.47			
PEM-CL														.90	0.93	0.77
PEM-CL1	0.56	0.39	0.45	0.54	0.43	0.45	0.43	0.80	0.73	*0.85*	0.61	0.57	0.59			
PEM-CL2	0.47	0.28	0.37	0.43	0.35	0.38	0.45	0.74	0.57	*0.89*	0.60	0.47	0.49			
PEM-CL3	0.50	0.34	0.45	0.49	0.40	0.46	0.49	0.81	0.66	*0.92*	0.67	0.54	0.58			

Constructs	PE[a]	EE[b]	SI[c]	HT[d]	FC[e]	HM[f]	PV[g]	PEM[h]	PEM-PL[i]	PEM-CL[j]	PEM-CCL[k]	BI[l]	BIR[m]	CA[n]	CR[o]	AVE[p]
PEM-CL4	0.43	0.31	0.38	0.49	0.39	0.45	0.44	0.80	0.67	*0.85*	0.69	0.51	0.55			
PEM-CCL														.93	0.95	0.78
PEM-CCL1	0.37	0.24	0.43	0.47	0.29	0.48	0.36	0.78	0.60	0.63	*0.87*	0.49	0.51			
PEM-CCL2	0.37	0.24	0.38	0.43	0.27	0.45	0.38	0.78	0.60	0.63	*0.90*	0.45	0.49			
PEM-CCL3	0.38	0.31	0.45	0.51	0.38	0.53	0.41	0.81	0.64	0.63	*0.91*	0.56	0.55			
PEM-CCL4	0.46	0.30	0.48	0.53	0.34	0.52	0.43	0.85	0.71	0.67	*0.90*	0.60	0.59			
PEM-CCL5	0.31	0.19	0.43	0.41	0.28	0.44	0.35	0.75	0.57	0.64	*0.83*	0.45	0.47			
BI														.95	0.97	0.90
BI1	0.65	0.38	0.44	0.66	0.42	0.51	0.46	0.63	0.62	0.56	0.53	*0.95*	0.76			
BI2	0.58	0.34	0.53	0.71	0.40	0.56	0.43	0.63	0.59	0.54	0.56	*0.94*	0.77			
BI3	0.66	0.44	0.50	0.71	0.50	0.59	0.52	0.66	0.63	0.60	0.56	*0.97*	0.78			
BIR														.93	0.95	0.82
BIR1	0.59	0.33	0.53	0.65	0.36	0.53	0.48	0.67	0.63	0.61	0.58	0.77	*0.93*			
BIR2	0.53	0.36	0.32	0.50	0.37	0.40	0.44	0.54	0.54	0.50	0.41	0.64	*0.84*			
BIR3	0.58	0.33	0.51	0.65	0.38	0.52	0.42	0.65	0.62	0.59	0.57	0.77	*0.95*			
BIR4	0.57	0.32	0.51	0.61	0.38	0.51	0.40	0.66	0.63	0.58	0.58	0.73	*0.90*			

[a]PE: performance expectation.

[b]EE: effort expectancy.

[c]SI: social influence.

[d]HT: habit.

[e]FC: facilitation conditions.

[f]HM: hedonic motivation.

[g]PV: price value.

[h]PEM: personal empowerment (second order).

[i]PEM-PL: personal empowerment-professional logic.

[j]PEM-CL: personal empowerment-consumer logic.

[k]PEM-CCL: personal empowerment-community logic.

[l]BI: behavior intention.

[m]BIR: behavior intention to recommend.

[n]CA: Cronbach alpha.

[o]CR: composite reliability.

[p]AVE: average variance extracted.

[q]Numbers in italics indicate loadings of the indicators for their own constructs.

Table 4. Total sample descriptive statistics, correlations, and square root of average variance extracted.

Variable	Mean (SD)	PE[a]	EE[b]	SI[c]	HT[d]	FC[e]	HM[f]	PV[g]	PEM[h]	PEM-PL[i]	PEM-CL[j]	PEM-CCL[k]	BI[l]	BIR[m]	Age	Education	Gender	PCHC[n]
PE	4.90 (1.63)	0.91[o]																
EE	4.90 (1.63)	0.50	0.91[o]															
SI	5.41 (1.34)	0.44	0.21	0.97[o]														
HT	5.18 (1.55)	0.58	0.47	0.60	0.85[o]													
FC	4.49 (1.58)	0.44	0.70	0.28	0.47	0.84[o]												
HM	3.84 (1.89)	0.48	0.51	0.44	0.67	0.53	0.91[o]											
PV	5.23 (1.36)	0.37	0.46	0.30	0.47	0.62	0.51	0.93[o]										
PEM	4.47 (1.54)	0.59	0.40	0.56	0.64	0.48	0.60	0.54	0.79[o]									
PEM-PL	4.21 (1.57)	0.61	0.40	0.56	0.64	0.49	0.56	0.53	0.91	0.87[o]								
PEM-CL	4.44 (1.54)	0.56	0.38	0.47	0.56	0.45	0.49	0.51	0.90	0.75	0.88[o]							
PEM-CCL	4.75 (1.46)	0.43	0.29	0.49	0.53	0.35	0.55	0.44	0.90	0.71	0.73	0.88[o]						
BI	4.34 (1.40)	0.66	0.41	0.51	0.73	0.47	0.58	0.50	0.67	0.64	0.60	0.58	0.95[o]					
BIR	4.04 (1.63)	0.63	0.37	0.52	0.67	0.41	0.54	0.48	0.70	0.67	0.63	0.60	0.81	0.90[o]				
Age	40.04 (13.6)	–0.08	–0.29	0.24	0.05	–0.10	0.03	–0.11	0.04	0.09	–0.05	0.04	–0.01	0.00	1.00			
Education	N/A[p]	0.06	0.20	–0.10	–0.03	0.25	0.08	0.11	0.02	0.03	0.06	–0.03	0.05	0.08	–0.12	1.00		
Gender	N/A	0.20	0.03	0.15	0.12	0.10	0.13	0.06	0.22	0.22	0.21	0.16	0.17	0.18	0.13	–0.02	1.00	
PCHC	N/A	–0.04	–0.11	0.15	0.11	–0.02	0.11	–0.06	0.04	0.11	–0.08	0.04	0.03	0.05	0.44	–0.20	–0.03	1.00

[a]PE: performance expectation.

[b]EE: effort expectancy.

[c]SI: social influence.

[d]HT: habit.

[e]FC: facilitation conditions.

[f]HM: hedonic motivation.

[g]PV: price value.

[h]PEM: personal empowerment (second order).

[i]PEM-PL: personal empowerment-professional logic.

[j]PEM-CL: personal empowerment-consumer logic.

[k]PEM-CCL: personal empowerment-community logic.

[l]BI: behavior intention.

[m]BIR: behavior intention to recommend.

[n]PCHC: presence of chronic health condition.

[o]Square root of average variance extracted.

[p]N/A: not applicable.

Use behavior is formed by 10 formative indicators, and its assessment involves specific quality criteria. No collinearity issues were detected in the total model with a variance inflation factor (VIF) below 5 for all indicators. Besides not all the indicators' weights complying with the criteria of being statistically significant, their outer loadings were all higher than 0.5, with some exceptions. Nevertheless, since all of the outer loadings were statistically significant, we retained all of the indicators in the model (see Table 5).

Table 5. Formative indicators for quality criteria.

Indicator	VIF[a]	Loading	Weight	Loading *P* value	Weight *P* value
UB[b]f1: What is your actual frequency of use of mHealth[c] to collect biometric data for medical follow-up?	2.10	0.74	0.21	<.001	.007
UBf2: What is your actual frequency of use of mHealth to collect biometric data related to well-being (fitness apps)?	1.68	0.84	0.45	<.001	<.001
UBf3: What is your actual frequency of use of mHealth to access a patient portal (eg, manage appointments, results of clinical analysis, application for online prescription)?	1.94	0.77	0.34	<.001	<.001
UBf4: What is your actual frequency of use of mHealth to monitor therapeutic compliance/adhesion (prescribed drugs/medicine intake follow up)?	2.01	0.53	–0.06	<.001	.43
UBf5: What is your actual frequency of use of mHealth for scientific observational studies (eg, medicine, app, or innovative treatment trial)?	1.86	0.52	0.07	<.001	.45
UBf6: What is your actual frequency of use of mHealth for health information research?	1.80	0.64	0.11	<.001	.21
UBf7: What is your actual frequency of use of mHealth for clinical screening and counselling?	2.04	0.64	0.17	<.001	.04
UBf8: What is your actual frequency of use of mHealth for making remote medical consultations/appointments?	1.90	0.42	0.08	<.001	.34
UBf9: What is your actual frequency of use of mHealth to request home medical consultation?	1.64	0.31	–0.14	<.001	.05
UBf10: What is your actual frequency of use of mHealth to participate in peer support groups or online communities of patients?	1.81	0.46	0.07	<.001	.48

[a]VIF: variance inflation factor.

[b]UB: use behavior.

[c]mHealth: mobile health.

Personal empowerment is designed as a reflective formative–type higher-order construct [25,41]. We assessed its multicollinearity according to the VIF, which indicated no collinearity issues as the VIF varied from 2.44 to 2.76 (ie, <5), and all of the weights were statistically significant and positive (Table 6).

Table 6. Measurement model evaluation for the higher-order formative constructs personal empowerment.

Constructs	VIF[a]	Weight	*P* value
Personal Empowerment - Community Logic	2.44	0.38	<.001
Personal Empowerment - Consumer Logic	2.76	0.32	<.001
Personal Empowerment - Professional Logic	2.61	0.41	<.001

[a]VIF: variance inflation factor.

These same assessments were also applied separately to the two groups under analysis (with and without a chronic health condition), as shown in Multimedia Appendix 3. Considering the results, we concluded that all of the constructs were suitable to test the conceptual model.

Structural Model

Before assessing the structural model, we first tested the multicollinearity of all constructs based on the VIF. All VIF values were below the threshold of 5, ranging from 1.00 to 2.73, indicating the absence of multicollinearity among the variables.

The structural model path significance levels were estimated using bootstrap resampling with 5000 iterations to achieve the maximum possible consistency in the results. The R^2 value was used to assess the structural model. The total model explained 66% of the variance in behavior intention, 54% of the variance in use behavior, and 70% of the variance in behavior intention to recommend.

We performed a PLS multigroup analysis to analyze the two groups in our sample. However, the results obtained were not globally statistically significant. Nevertheless, analysis of the two groups independently demonstrated some significant differences with comparison. Behavior intention explained a higher percentage of the variance in the model for the patients with chronic health conditions than for the healthy group (74% vs 65%), whereas higher variance for use behavior and behavioral intention to recommend was found for the healthy group than for the chronic conditions group (63% vs 51% and 75% vs 62%, respectively). Table 7 summarizes the detailed structural model results (R^2, path coefficients significance, and significance between groups).

Figure 2 and Figure 3 show the structural model results of the total model and for each group, respectively.

Table 7. Structural model results.

Dependent/independent variables	β				*t* value				R^2 (%)		
	Total	CHP[a]	WCHC[b]	CHCP – WCHC	Total	CHCP	WCHC	CHCP – WCHC	Total	CHCP	WCHC
BI[c]									66	74	65
PE[d]	.29**	.46**	.22**	.19	5.13**	5.20**	3.13**	1.68			
EE[e]	–.11	–.15	–.05	.15	1.76	1.43	0.71	1.07			
SI[f]	.02	–.10	.11	.04	0.38	1.79	1.36	0.43			
HT[g]	.39**	.35**	.37**	.12	5.68**	3.23**	4.20**	0.84			
FC[h]	.04	.16	.00	.07	0.56	1.42	0.06	0.46			
HM[i]	.05	–.10	.10	.16	0.87	1.01	1.51	1.25			
PV[j]	.10	.11	.12	.11	1.74	1.26	1.69	1.04			
PEM[k]	.18*	.23*	.11	.02	2.55*	2.47*	1.10	0.14			
Age	–.04	–.03	–.03	.00	1.06	0.54	0.70	0.00			
Gender	.02	.00	.03	.04	0.64	0.09	0.79	0.55			
Education	.04	.02	.02	.00	1.06	0.27	0.36	0.02			
UB[1]									54	51	63
HT	.47**	.26	.54**	.23	6.16**	1.67	6.09**	1.40			
FC	.04	–.01	.09	.14	0.68	0.13	1.44	1.17			
PEM	.17**	.19	.16*	.01	2.62**	1.44	2.19	0.04			
BI	.15	.27	.09	.23	1.83	1.76	0.96	1.38			
Age	–.04	–.20*	.07	.26	0.72	2.44*	1.26	2.87**			
Gender	.02	.06	.00	.07	0.47	0.81	0.06	0.74			
Education	.10*	.10	.05	.05	2.33*	1.12	0.92	0.56			
BIR[m]									70	62	75
PEM	.26**	.34**	.23**	.09	4.67**	3.49**	3.56**	0.85			
BI	.58**	.42**	.65**	.11	9.17**	4.19**	8.26**	0.82			
UB	.08	.09	.06	.15	1.56	0.43	0.99	1.43			
Age	–.01	.00	–.03	.03	0.35	0.02	0.83	0.40			
Gender	.01	.01	.00	.01	0.37	0.17	0.06	0.10			
Education	.04	.10	.03	.07	1.38	1.73	0.78	1.08			

[a]CHCP: patients with chronic health conditions.

[b]WCHC: without chronic health conditions.

[c]BI: behavior intention.

[d]PE: performance expectation.

[e]EE: effort expectancy.

[f]SI: social influence.

[g]HT: habit.

[h]FC: facilitation conditions.

[i]HM: hedonic motivation.

[j]PV: price value.

[k]PEM: personal empowerment (second order).

[l]UB: use behavior.

[m]BIR: behavior intention to recommend.
** *P*<.01, **P*<.05; df (bootstrap)=4999.

Figure 2. Results of the structural model for the total sample. UTAUT: unified theory of acceptance and usage of technology; mHealth: mobile health.

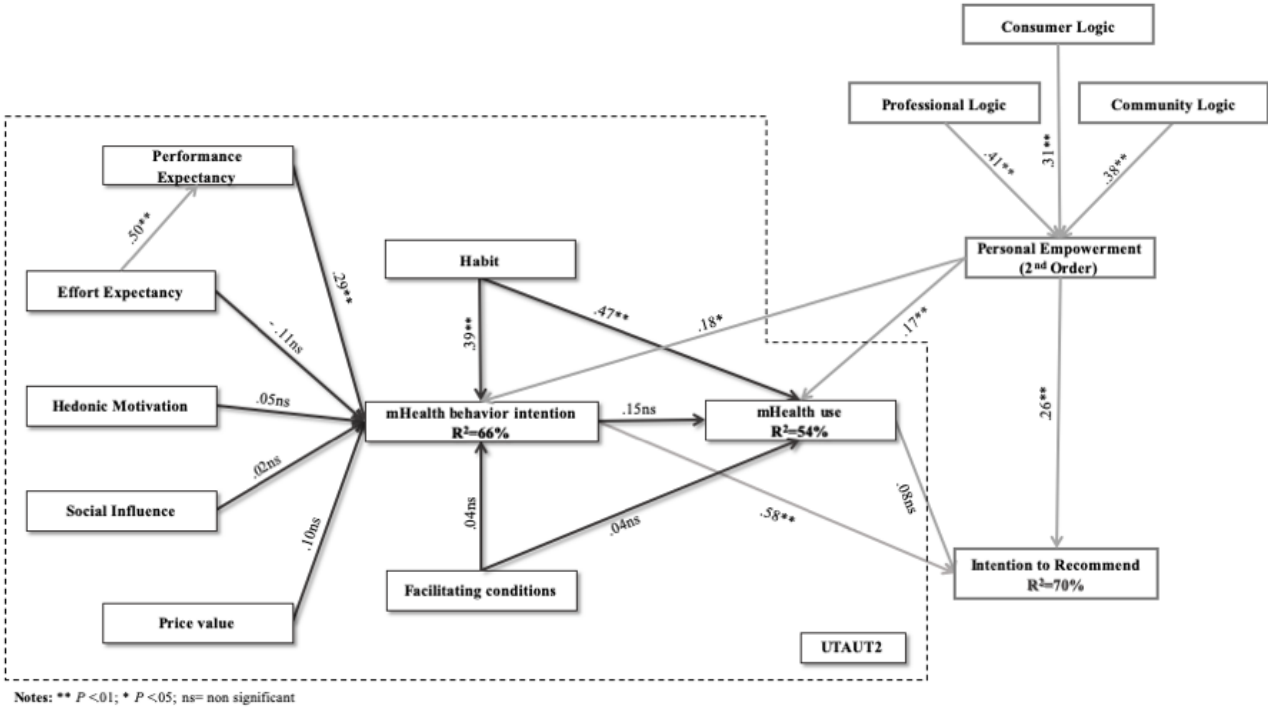
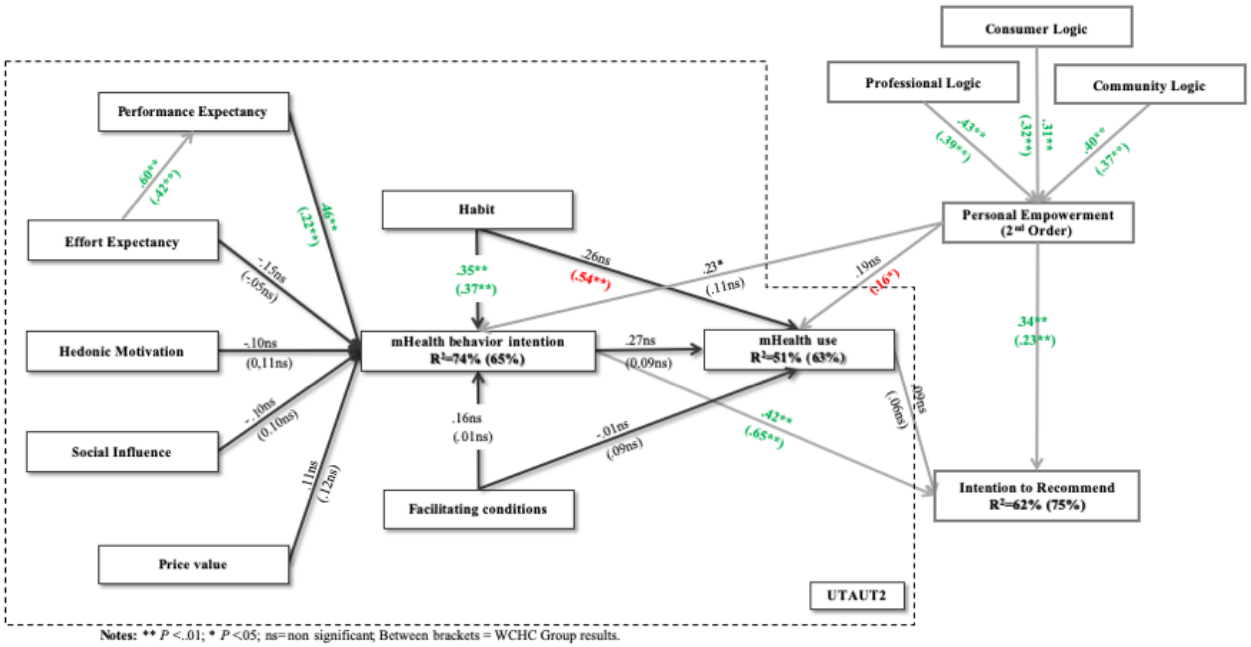


Figure 3. Structural model results for the chronic health conditions patients (CHCP) group and the without chronic health conditions (WCHC) group (in parentheses). Statistically significant relationships observed for both groups are presented in green. Significant relationships only in the WCHC group are presented in red. Results with no significant relationship in either group are presented in black with ns.



Discussion

Principal Findings

The results of our study support development of the UTAUT2 model with constructs that are more specific to the health domain. Most of the constructs of UTAUT2 did not show statistical significance in the sample analyzed in this study. In contrast, an important role of personal empowerment was revealed for the intention to use, use behavior, and behavior intention to recommend mHealth. Further, the addition of

behavior intention to recommend mHealth had a substantial contribution to explaining the mHealth phenomenon with a high R^2 value. The mediation effect of effort expectancy on performance expectancy influenced behavior intention, demonstrating a relevant role to effort expectancy in the understanding of mHealth. However, when focusing on the group of patients with chronic health conditions, the results were not as meaningful as expected, disclosing significance of the independent variables only on behavior intention. These results lead to an important concern that patients with chronic conditions, who would more likely receive benefits from the use of mobile technology to manage their health condition, not only still do not use it as often as they could (mean use behavior 3.32, SD 0.84, range 1-7) but also their intention to use is not a predecessor of use.

Theoretical Implications

According to the results summarized in Table 8, performance expectancy and habit were predictors of behavior intention in the total model and across the two groups analyzed (with and without chronic conditions) with a positive statistically significant impact, supporting H1(a) and 7(a). Previous studies addressing the eHealth context reported similar results [15].

The impact of performance expectancy on behavioral intention in the chronic conditions group was highly significant (β=.46, P<.001), confirming our rationale that the expectancy of the benefits obtained with the use of mHealth technology to manage health strongly influences the intention of using it for patients with chronic conditions. Nevertheless, the hypothesized difference between groups related to performance expectancy was not verified, thereby not supporting H1(b).

The predictor effect of effort expectancy has shown contradictory results in previous studies [42,43]. In our sample, the influence of effort expectancy on behavioral intention was not statistically significant, thereby not supporting H2(a). However, the proposed mediation effect of effort expectancy for the influence of perceived effort on behavioral intention was statistically significant, revealing a full mediation effect, since only the indirect effect of effort expectancy was significant [25], which supports H2(b).

The personal empowerment construct that we included in the model showed its importance to the study with a positive statistically significant impact on behavior intention in the total model and in the chronic conditions group. For use behavior, this construct revealed a statistically significant positive impact in the full sample and in the healthy group, and also had a statistically significant positive impact in all three analyses when related to behavioral intention to recommend, thereby corroborating H10(a), H10(b), and H10(c). Nevertheless, the expected difference between the chronic conditions and healthy groups was not confirmed, thereby not supporting H10(d), H10(e), and H10(f). This may also explain why hedonic motivation and social influence were not found to have significant effects (H3 and H4 not supported). Since personal empowerment, as a second-order construct, has a personal/consumer and a community logic, inclusion of these concepts can better capture a model considering the health context.

Our results do not support the influence of facilitation conditions on behavioral intention or use behavior, thereby not supporting H5(a) and H5(b), and the difference expected between groups was also not verified, thereby not supporting H5(c) and H5(d). This element suggests that the individuals in our sample consider that the resources or knowledge to use mHealth are not an issue, which is likely related to the current natural increasing access to a mobile phone and mobile internet. According to the 2017 ANACOM mobile services report [6], the number of mobile broadband users in Portugal reached 7.2 million, representing an increase of 9.5% from the prior year.

Price value also did not show a significant effect in our sample, thereby not supporting H6. The absence of influence could be related to the Portuguese national health service universal coverage concept, with health care services tending to be free, but also with the fact that eHealth/mHealth models such as teleconsultation are not yet very common in Portugal. Another reason may be that benefits other than the price of the technology are not being perceived by end users.

Table 8. Summary of findings regarding the hypotheses.

Hypothesis	Path	β	*t* value	Result
1 (a)	PE[a] to BI[b]	.29	5.13**	Supported
1 (b)	(PE_{CHCP}[c] to BI_{CHCP}) - (PE_{WCHC}[d] to BI_{WCHC})	.19	1.68	Not supported
2(a)	EE[e] to BI	–.11	1.76	Not supported
2(b)	EE to PE to BI	.27	4.47**	Supported
3	HM[f] to BI	.05	0.87	Not supported
4	SI[g] to BI	.02	0.38	Not supported
5 (a)	FC[h] to BI	.04	0.56	Not supported
5(b)	FC to UB[i]	.04	0.68	Not supported
5(c)	(FC_{CHCP} to BI_{CHCP}) - (FC_{WCHC} to BI_{WCHC})	.07	0.46	Not supported
5(d)	(FC_{CHCP} to UB_{CHCP}) - (FC_{WCHC} to UB_{WCHC})	.14	1.17	Not supported
6	PV[j] to BI	.10	1.74	Not supported
7(a)	HT[k] to BI	.39	5.68**	Supported
7(b)	HT to UB	.47	6.16**	Supported
8(a)	BI to UB	.15	1.83	Not supported
8(b)	BI to BIR[l]	.58	9.17**	Supported
9	UB to BIR	.08	1.56	Not supported
10(a)	PEM[m] to BI	.18	2.55*	Supported
10(b)	PEM to UB	.17	2.62**	Supported
10(c)	PEM to BIR	.26	4.67**	Supported
10(d)	(PEM_{CHCP} to BI_{CHCP}) - (PEM_{WCHC} to BI_{WCHC})	.02	0.14	Not supported
10(e)	(PEM_{CHCP} to UB_{CHCP}) - (PEM_{WCHC} to UB_{WCHC})	.01	0.04	Not supported
10(f)	(PEM_{CHCP} to BIR_{CHCP}) - (PEM_{WCHC} to BIR_{WCHC})	.09	0.85	Not supported

[a]PE: performance expectation.
[b]BI: behavior intention.
[c]CHCP: patients with chronic conditions.
[d]WCHC: without chronic conditions.
[e]EE: effort expectancy.
[f]HM: hedonic motivation.
[g]SI: social influence.
[h]FC: facilitation conditions.
[i]UB: use behavior.
[j]PV: price value.
[k]HT: habit.
[l]BIR: behavior intention to recommend.
[m]PEM: personal empowerment.
** *P*<.01, **P*<.05; df (bootstrap)=4999.

In contrast to the UTAUT2 rationale, our results do not attest to the influence of behavioral intention on use behavior, thereby not supporting H8(a). The same conclusion was reached in a study by Lim et al [44] on acceptance of using mobile phones to seek health information by women in Singapore, which raises the concern that behavior intentions are not translated in actual use, but only on the intention to recommend, thereby supporting H8(b). This aspect is even more relevant when we consider only the patients with chronic conditions in which none of the independent variables showed statistical significance on use behavior.

Our control variables were globally not statistically significant, with the exception of a significant impact of age in the chronic

conditions group. This outcome shows that younger individuals tend to use mHealth more than older individuals [45]. In addition, in the total sample, higher education levels were associated with a higher level of use of this technology, which is in accordance with expectations and in line with previous research [16]. Gender was not statistically significant for any group or dependent variable.

Overall, our results reinforce the importance of performance expectancy and habit as drivers of technology acceptance, which is aligned with previous studies, and specifically in mHealth acceptance. Moreover, we found a significant effect of personal empowerment, demonstrating the beneficial effect of adding this aspect to UTAUT2 when analyzing the acceptance of mobile technology in a health management context.

Health Policy Implications

The results of our study show that personal empowerment is a key driver of mHealth. This suggests that governments worldwide should focus more on patient-centric policies, with more direct communication to the patients to promote their empowerment as this will drive the adoption of mHealth. Personal empowerment has been studied in many contexts and with different levels of significance [46]. Our results show how consumers truly value this aspect from the eHealth context. The anticipated scarcity of human health care resources in the near future emphasizes how critical it is to involve the patient in the management of their own health, leading to a change of the health care paradigm to a more predictive, personalized, preventive, and participatory approach [47], which has been promoted by emerging technologies. Owing to its proximity, accessibility, and increasing diffusion, mHealth is one of the privileged paths to foster this transformation, increasing not only the interactivity between patients and health care providers but also the engagement of individuals, underlining their active and determinant role as key actors in their health management cycle to help toward democratizing health care [48]. Currently, mHealth users are diverse, being not only patients but also general consumers who aim to be more informed and diligent in every aspect of their life, particularly with respect to health-related issues. Therefore, this aspect must be considered as a primary driver to the integration of this technology as part of health care systems.

Effort expectancy did not show significant predictive power as a direct influencer of behavioral intention, which is in opposition to the results of a UTAUT2 study and other previous studies [43]. Nevertheless, effort expectancy had a strong mediator effect on performance expectancy, indicating the need to provide more information to the public about mHealth. Such clarification refers to its applications and ease of use, and to reinforce its role as a pivotal tool to ease the interaction and proximity between patients and health care providers in such a way that promotes its integration in health care management.

Our research shows that mHealth technologies still have low usage in Portugal, as demonstrated by the use behavior descriptive statistics. Therefore, targeting early adopters and its continuance of use can be one of the main strategies adopted to increase the diffusion of mHealth technologies by fostering word-of-mouth recommendation. This was attested by the significant role of habit and the good results of behavioral intention to recommend. This outcome shows that governments and health care institutions should realize that the current users of mHealth also play a role in the diffusion of mHealth tools. Therefore, this aspect should be considered as part of the promotion strategy that governments and health care institutions implement to increase the adoption of mHealth tools.

Experts, clinical/managerial staff, and health care providers should also be aware that behavior intention is only a proxy for measuring technology acceptance [49]. More specifically, a positive behavioral intention does not always translate into the actual use of technology, as demonstrated in the present study. The ease to operate a technology, health and eHealth literacy level, belief in the importance of having an active role in one's own health management, and even the relationship with health care providers [50] also have to be considered.

Another important topic to take in account is that the highest risk group (ie, patients that are older and have chronic health conditions [51]) demonstrated a lower use of mHealth. The majority of the features showed higher usage rates in the younger group (see Table 9) as expected; however, we would like to highlight that older people (aged ≥55 years) mentioned that they use mHealth more than the younger group to request home medical appointments (UBf9) and to obtain information about scientific studies and new treatments (UBf5). Clearly, these two aspects show that older people with chronic diseases are concerned about their health. This is intuitive as a chronic condition gets worse with age [51], and patients will therefore seek information about treatments with increasing age, which also impacts their mobility so that remote appointments are a good solution [51]. From a standpoint of health policy, by addressing these topics to the older population in a more effective manner, such as by promoting and providing support for the use of remote medical consultations, the problem of lack of mobility for the older population with chronic conditions can be solved while also allowing for more effective management of health care resources in the future that will become increasingly scarce to manage the growing increase of chronic diseases related with aging [3]. For example, with the COVID-19 outbreak, it is even more critical to make sure that older people with chronic health conditions remain at home in isolation [52,53]. Nevertheless, this population still needs to be in contact with health care providers; therefore, communication and remote home medical appointments are excellent options to keep these high-risk people safe via mHealth.

Table 9. Median use behavior (UB) formative indicators in patients with chronic conditions according to age.

UB formative indicators	Younger patients (<55 years; n=44), median	Older patients (≥55 years; n=80), median
UBf1: What is your actual frequency of use of mHealth to collect biometric data for medical follow-up?	4.00	3.00
UBf2: What is your actual frequency of use of mHealth to collect biometric data related to well-being (fitness apps)?	4.00	3.00
UBf3: What is your actual frequency of use of mHealth to access a patient portal (eg, manage appointments, results of clinical analysis, application for online prescription)?	6.00	5.00
UBf4: What is your actual frequency of use of mHealth to monitor therapeutic compliance/adhesion (prescribed drugs/medicine intake follow up)?	4.00	3.00
UBf5: What is your actual frequency of use of mHealth for scientific observational studies (eg, medicine, app, or innovative treatment trial)?	2.00	3.00
UBf6: What is your actual frequency of use of mHealth for health information research?	5.00	5.00
UBf7: What is your actual frequency of use of mHealth for clinical screening and counselling?	3.50	3.50
UBf8: What is your actual frequency of use of mHealth for making remote medical consultations/appointments?	1.00	1.00
UBf9: What is your actual frequency of use of mHealth to request home medical consultation?	1.00	1.50
UBf10: What is your actual frequency of use of mHealth to participate in peer support groups or online communities of patients?	2.00	2.00

Limitations and Future Research

One of the limitations of this study was the use of a convenience sampling approach, which was limited to educational institutions and patients' associations, thereby restraining the extrapolation of the results for the general population. The cross-sectional design of the study, which did not allow to capture changes over time, and the geographical circumscription are additional limitations. Namely, mHealth is still at an early use stage in Portugal, and the national health system, besides being in a transition process, is still very "paternalistic," meaning there is a highly dependent relationship between health care professionals and patients, which follows a model of care that is strongly focused on institutions such as hospitals [54]. Therefore, future studies may benefit from a cross-country comparison, ideally including countries at different stages of mHealth adoption so as to explore better paths to increase the use of mHealth. In this context, it would be very interesting to develop a longitudinal study and add to the research experience information, at least in a self-reported manner, and the continuance of use constructs.

Additionally, regarding the extension of UTAUT2, application of the moderation effects of the original model, such as experience considering that habit had a significant impact on the dependent variables, or even the exploration of new effects, could improve the explanatory power of the model. Our research goal focused on understanding whether having a chronic condition/disability would influence the acceptance drivers of mHealth. However, we only collected this information as a binary coded variable, without details on the type or severity. Therefore, future researchers could develop this perspective with more detail. Another important topic in future studies would be to increase the sample size, particularly in the group of patients with chronic diseases, because age (as a control variable) had a significant effect in use for this group. Patients with chronic diseases have a higher risk at older age [3]. A relevant sample size (eg, ≥100 individuals per age group) [25] could allow for conducting a subgroup analysis between older and younger patients with chronic conditions.

Conclusion

mHealth technologies are suggested as one of the privileged means to address emerging problems in health care, including the need to improve access to health services, regardless of time and place [27]. Governments worldwide are concerned about the increased prevalence of chronic diseases and age-related diseases with regard to the potentially insufficient human health care resources available in the future to deal with this new situation. Our study showed that promoting the engagement of patients in their own self-care via mHealth can be a viable solution for this problem [27]. In addition, mHealth can present a viable solution for high-risk patients such as older patients

with chronic diseases that need to be in isolation to protect them from infectious diseases like COVID-19 [52,53]. These patients can be reached by health care providers remotely using mHealth. The primary goal of this study, using a novel theoretical model, was to explore and understand individual mHealth acceptance drivers, and to further explore if having a chronic health condition influences these drivers. The tested model incorporates an extension of UTAUT2 with personal empowerment and behavior intention to recommend constructs, and established a new mediation effect between perceived effort and behavior intention through effort expectancy.

The study was conducted in Portugal with a sample of 322 individuals recruited from educational institutions and patients' associations, and the model explained 66% of the variance of behavior intention, 54% of use behavior, and 70% of behavioral intention to recommend. Performance expectancy and habit emerged as predictors of behavior intention in the total model and across the two groups analyzed (those with and without chronic health conditions). Effort expectancy had a significant effect on the influence of perceived effort to behavior intention, revealing the impact of this construct. In the chronic conditions group, the percentage of behavior intention variance explained by the model was higher than that obtained in the healthy group (74% vs 65%), whereas the percentages for use behavior and behavior intention to recommend were higher in the healthy group than in the chronic conditions group (63% vs 51% and 75% vs 62%, respectively). The personal empowerment construct had a significant effect on behavior intention in the total model and for the chronic conditions group. Personal empowerment had a positive impact on use behavior in the total sample and the healthy group, and had a significant effect on behavior intention to recommend for the total sample as well as in each of the two groups.

Overall, our findings show that by using constructs that are specifically health-related, namely personal empowerment as a second-order construct, we achieved a model that could offer a better explanation of mHealth acceptance drivers. With this study, we advance the perspective of technology acceptance at the individual/patient level, thereby reinforcing the existing knowledge and highlighting the need for further research to develop more evidence-based theories in this field.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire items.
[DOCX File , 34 KB - mhealth_v8i7e17588_app1.docx]

Multimedia Appendix 2
Total sample heterotrait-monotrait ratio (HTMT).
[PDF File (Adobe PDF File), 42 KB - mhealth_v8i7e17588_app2.pdf]

Multimedia Appendix 3
Group analysis tables.
[PDF File (Adobe PDF File), 317 KB - mhealth_v8i7e17588_app3.pdf]

References

1. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' Acceptance of Smartphone Health Technology for Chronic Disease Management: A Theoretical Model and Empirical Test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
2. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A Framework and Toolkit of Strategies for the Development of More Effective Digital Interventions to Change Health Behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
3. Alpay LL, Henkemans OB, Otten W, Rövekamp TAJM, Dumay AC. E-health applications and services for patient empowerment: directions for best practices in The Netherlands. Telemed J E Health 2010 Sep;16(7):787-791. [doi: 10.1089/tmj.2009.0156] [Medline: 20815745]
4. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
5. Mathur M, Vamadevan A, Reddy K. E-Health in Emerging Economies. Soc Policies 2018;2:235-244. [doi: 10.7389/90596]
6. Estatísticas Serviço Móvel Q4 2017. ANACOM. 2018. URL: https://www.anacom.pt/streaming/STM_4T2017.pdf?contentId=1433890&field=ATTACHED_FILE [accessed 2018-09-18]
7. Consumer Survey on Digital Health Summary. Accenture. 2018. URL: https://www.accenture.com/us-en/insight-new-2018-consumer-survey-digital-health [accessed 2018-10-28]
8. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]

9. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
10. Kim S, Lee K, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
11. Laxminarayan S, Istepanian RS. Unwired E-Med: the next generation of wireless and internet telemedicine systems. IEEE Trans Inf Technol Biomed 2000 Sep;4(3):189-193. [doi: 10.1109/titb.2000.5956074] [Medline: 11026588]
12. World Health Organization. Global diffusion of eHealth: Making universal health coverage achievable. Geneva, Switzerland: Global Observatory for eHealth; 2016.
13. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
14. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
15. Tavares J, Goulão A, Oliveira T. Electronic Health Record Portals adoption: Empirical model based on UTAUT2. Inform Health Soc Care 2018 Mar 16;43(2):109-125. [doi: 10.1080/17538157.2017.1363759] [Medline: 29035646]
16. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
17. Pfeiffer J, Von EM, Urbach N, Buchwald A. Quantify-Me: Consumer acceptance of wearable self-tracking devices. : Association for Information Systems AIS Electronic Library (AISeL); 2016 Presented at: 2016 European Conference on Information Systems (ECIS 2016); June 12-15, 2016; Istanbul, Turkey.
18. Kisekka V, Giboney JS. The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes. J Med Internet Res 2018 Apr 11;20(4):e107 [FREE Full text] [doi: 10.2196/jmir.9014] [Medline: 29643052]
19. Lemire M, Sicotte C, Paré G. Internet use and the logics of personal empowerment in health. Health Policy 2008 Oct;88(1):130-140. [doi: 10.1016/j.healthpol.2008.03.006] [Medline: 18436330]
20. Zhang X, Yu P, Yan J, Ton AM Spil I. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15:71 [FREE Full text] [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
21. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quart 2012;36(1):157. [doi: 10.2307/41410412]
22. Tavares J, Oliveira T. New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
23. Tavares J, Oliveira T. Electronic Health Record Patient Portal Adoption by Health Care Consumers: An Acceptance Model and Survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
24. Tavares J, Oliveira T. Electronic Health Record Portal Adoption: a cross country analysis. BMC Med Inform Decis Mak 2017 Jul 05;17(1):97 [FREE Full text] [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
25. Hair J, Hult G, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Second edition. Los Angeles: Sage; 2017.
26. Angst C, Agarwal R. Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. MIS Quart 2009;33(2):339-370. [doi: 10.2307/20650295]
27. Silva BMC, Rodrigues JJPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
28. Bradway M, Årsand E, Grøttland A. Mobile Health: empowering patients and driving change. Trends Endocrinol Metab 2015 Mar;26(3):114-117. [doi: 10.1016/j.tem.2015.01.001] [Medline: 25727462]
29. Yuan S, Ma W, Kanthawala S, Peng W. Keep Using My Health Apps: Discover Users' Perception of Health and Fitness Apps with the UTAUT2 Model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
30. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
31. Oliveira T, Thomas M, Baptista G, Campos F. Mobile payment: Understanding the determinants of customer adoption and intention to recommend the technology. Comput Hum Behav 2016 Aug;61:404-414. [doi: 10.1016/j.chb.2016.03.030]
32. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for TranslationCultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]

33. Podsakoff P, MacKenzie S, Lee J, Podsakoff N. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
34. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]
35. Johnson RE, Rosen CC, Djurdjevic E. Assessing the impact of common method variance on higher order multidimensional constructs. J Appl Psychol 2011 Jul;96(4):744-761. [doi: 10.1037/a0021504] [Medline: 21142343]
36. Or CKL, Karsh B. A Systematic Review of Patient Acceptance of Consumer Health Information Technology. J Am Med Inform Assoc 2009 Jul 01;16(4):550-560. [doi: 10.1197/jamia.m2888]
37. Ringle C, Wende M, Beck JM. SmartPLS 3. Boenningstedt: SmartPLS GmbH; 2015. URL: https://www.smartpls.com [accessed 2018-01-23]
38. Beran TN, Violato C. Structural equation modeling in medical research: a primer. BMC Res Notes 2010 Oct 22;3:267 [FREE Full text] [doi: 10.1186/1756-0500-3-267] [Medline: 20969789]
39. Hair Jr JF, Sarstedt M, Hopkins L, Kuppelwieser VG. Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research. Eur Bus Rev 2014 Mar 04;26(2):106-121. [doi: 10.1108/EBR-10-2013-0128]
40. Sarstedt M, Ringle C, Smith D, Reams R, Hair J. Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. J Fam Bus Strateg 2014 Mar;5(1):105-115. [doi: 10.1016/j.jfbs.2014.01.002]
41. Becker J, Klein K, Wetzels M. Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models. Long Range Plann 2012 Oct;45(5-6):359-394. [doi: 10.1016/j.lrp.2012.10.001]
42. Apolinário-Hagen J, Menzel M, Hennemann S, Salewski C. Acceptance of Mobile Health Apps for Disease Management Among People With Multiple Sclerosis: Web-Based Survey Study. JMIR Form Res 2018 Dec 12;2(2):e11977 [FREE Full text] [doi: 10.2196/11977] [Medline: 30684408]
43. Khechine H, Lakhal S, Ndjambou P. A meta-analysis of the UTAUT model: Eleven years later. Can J Adm Sci 2016 Jun 06;33(2):138-152. [doi: 10.1002/cjas.1381]
44. Lim S, Xue L, Yen CC, Chang L, Chan HC, Tai BC, et al. A study on Singaporean women's acceptance of using mobile phones to seek health information. Int J Med Inform 2011 Dec;80(12):e189-e202. [doi: 10.1016/j.ijmedinf.2011.08.007] [Medline: 21956003]
45. Bol N, Helberger N, Weert J. Differences in mobile health app use: A source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
46. Lemire M. What can be expected of information and communication technologies in terms of patient empowerment in health? J Health Organ Manag 2010 May 25;24(2):167-181. [doi: 10.1108/14777261011047336] [Medline: 20698192]
47. Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. Biotechnol J 2012 Aug 20;7(8):992-1001 [FREE Full text] [doi: 10.1002/biot.201100306] [Medline: 22815171]
48. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015 Jan 08;6:2013. [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
49. Bhattacherjee A, Sanford C. The intention–behaviour gap in technology usage: the moderating role of attitude strength. Behav Inf Technol 2009 Jul;28(4):389-401. [doi: 10.1080/01449290802121230]
50. De Rosis S, Barsanti S. Patient satisfaction, e-health and the evolution of the patient-general practitioner relationship: Evidence from an Italian survey. Health Policy 2016 Nov;120(11):1279-1292 [FREE Full text] [doi: 10.1016/j.healthpol.2016.09.012] [Medline: 27836231]
51. Quinaz Romana G, Kislaya I, Salvador M, Cunha Gonçalves S, Nunes B, Dias C. Multimorbidity in Portugal: Results from The First National Health Examination Survey. Acta Med Port 2019 Feb 01;32(1):30-37 [FREE Full text] [doi: 10.20344/amp.11227] [Medline: 30753801]
52. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]
53. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
54. Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the regions on enabling the digital transformation of health and care in the Digital Single Market;. EUR-Lex. 2018. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A233%3AFIN [accessed 2018-09-23]

Abbreviations

eHealth: electronic health
HBM: health belief model
mHealth: mobile health
PLS: partial least squares

SEM: structural equation model
TAM: technology acceptance model
UTAUT: Unified theory of acceptance and usage of technology
UTAUT2: extended unified theory of acceptance and usage of technology
VIF: variance inflation factor

Edited by G Eysenbach; submitted 23.12.19; peer-reviewed by S Jones, N Deng; comments to author 28.01.20; revised version received 24.03.20; accepted 30.03.20; published 09.07.20.

Please cite as:
Salgado T, Tavares J, Oliveira T
Drivers of Mobile Health Acceptance and Use From the Patient Perspective: Survey Study and Quantitative Model Development
JMIR Mhealth Uhealth 2020;8(7):e17588
URL: https://mhealth.jmir.org/2020/7/e17588
doi:10.2196/17588
PMID:32673249

Original Paper

Perceptions of Patients Regarding Mobile Health Interventions for the Management of Chronic Obstructive Pulmonary Disease: Mixed Methods Study

Meshari F Alwashmi[1], PhD; Beverly Fitzpatrick[1], PhD; Jamie Farrell[1], MD, FRCPC; John-Michael Gamble[2], PhD; Erin Davis[1], PharmD; Hai Van Nguyen[1], PhD; Gerard Farrell[1], MD; John Hawboldt[1], PharmD

[1]Memorial University of Newfoundland and Labrador, St. John's, NL, Canada

[2]School of Pharmacy, Faculty of Science, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Meshari F Alwashmi, PhD
Memorial University of Newfoundland and Labrador
230 Elizabeth Avenue
St. John's, NL, A1C 5S7
Canada
Phone: 1 7096910728
Email: mfa720@mun.ca

Abstract

Background: Using a mobile health (mHealth) intervention consisting of a smartphone and compatible medical device has the potential to enhance chronic obstructive pulmonary disease (COPD) treatment outcomes while mitigating health care costs.

Objective: This study aims to describe the demographics, use, and access to smartphones of patients with COPD. It also aims to explore and develop an understanding of potential facilitators and barriers that might influence patients using mHealth interventions for COPD management.

Methods: This was an explanatory, sequential mixed methods study. Patients who attended respirology clinics completed a questionnaire on technology access and use. We conducted semistructured individual interviews with the patients. Interview topics included the following: demographics, mHealth use, perceptions toward challenges of mHealth adoption, factors facilitating mHealth adoption, and preferences regarding features of mHealth interventions for COPD management.

Results: A total of 100 adults completed the survey but 22 participants were excluded because they were not diagnosed with COPD. Of these, 10 patients with COPD participated in the interview. The quantitative component revealed that many patients with COPD owned a mobile phone, but only about one-fourth of the participants (18/77, 23%) owned a smartphone. The likelihood of owning a smartphone was not associated with age, sex, marital status, or geographical location, but patients with high educational status were more likely to own a smartphone. The qualitative component found that patients with COPD, in general, had a positive attitude toward mHealth adoption for COPD management, but several facilitators and barriers were identified. The main facilitators of mHealth adoption are possible health benefits for patients, ease of use, educating patients, and credibility. Alternatively, the barriers to adoption are technical issues, lack of awareness, potential limited uptake from older adults, privacy and confidentiality issues, finances, and lack of interest in mHealth

Conclusions: It is important to understand the perceptions of patients with COPD regarding the adoption of innovative mHealth interventions for COPD management. This study identifies some potential facilitators and barriers that may inform the successful development and implementation of mHealth interventions for COPD management.

(JMIR Mhealth Uhealth 2020;8(7):e17409) doi:10.2196/17409

KEYWORDS

mhealth; COPD; health technology; smartphone; mobile phone

Introduction

Although chronic obstructive pulmonary disease (COPD) is a preventable and treatable disease, it is currently the third leading cause of death worldwide [1,2]. According to the Canadian Institute for Health Information, COPD now accounts for the highest rate of hospital admission and readmission among major chronic illnesses in Canada [3]. The Conference Board of Canada has stated that the combined direct and indirect costs of COPD will increase from just under Can $4 billion (US $2.93 billion) in 2010 to Can $9.5 billion (US $6.96 billion) by 2030, an increase of 140% [4]. These costs include direct costs (including drugs, hospitals, and physicians) and indirect costs (including long-term disability losses and mortality). Advances in technology have the potential to enhance both pharmacological and nonpharmacological interventions for COPD management.

This surge in computing power and mobile connectivity has led to the inception of mobile health (mHealth), which can transform clinical research and health care [5]. mHealth is defined by the National Institutes of Health as the "use of mobile and wireless devices to improve health outcomes, healthcare services, and health research." An mHealth intervention could also include the use of a medical device that is compatible with a smartphone. Previous research suggests that mHealth interventions may benefit patients with many chronic health conditions, including COPD [6]. Some of these benefits include improving the knowledge about COPD, increasing physical activity, and reducing exacerbations [7-9]. Alwashmi et al [6] noted that the current literature on the role of smartphones in reducing COPD exacerbations is limited, but they suggest that smartphone interventions may reduce COPD exacerbations. To potentially enhance the adoption and outcome of mHealth interventions, key users should be involved in the development of these interventions.

The International Organization for Standardization (ISO) 9241-210 defines human-centered design (HCD) as "an approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques" [10]. The ISO uses the term HCD instead of user-centered design to "address impacts on a number of stakeholders, not just those typically considered as users" [10]. However, in practice, these terms are often used synonymously.

Many researchers use mHealth to assist in the management of chronic diseases; nevertheless, gaps still exist regarding the development process of these mHealth interventions [11]. Testing mHealth interventions with patients has revealed preferences and concerns unique to the tested population [12-14]. Developing a COPD mHealth intervention with insights from patients will potentially improve the process and outcome of mHealth interventions.

Patient perspectives toward using mHealth for COPD management are relatively unexplored [11,15]. A recent meta-analysis on the remote monitoring of patients with COPD concluded that some interventions may prove to be promising in changing clinical outcomes in the future, but there are still large gaps in the evidence base [16]. Noah et al [16] suggested that adding a qualitative component would give researchers insight into which elements best engage and motivate patients and health care professionals (HCPs).

To improve the success of mHealth interventions that target patients with COPD, we included patients in the development process. The lessons learned will bridge the knowledge gap between barriers and facilitators for mHealth uptake in COPD management. Key lessons learned will be offered as a guide for research and technology developers who are developing mHealth interventions for COPD management [12,14,17]. Furthermore, before implementing mHealth interventions, it is important to understand the use of and access to mHealth.

Methods

Purpose

This study aimed to describe the demographics, use, and access to smartphones of patients with COPD. It also aimed to assess whether demographic factors predict engagement with mHealth and to explore and develop an understanding of potential facilitators and barriers that might influence patients using mHealth interventions for COPD management.

Study Design

First, participants completed a questionnaire about technology use and access. The findings from the questionnaire were used to develop the sample and questions for the qualitative phase [18]. A subset of participants were interviewed to provide deeper insights into technology access and use among patients with COPD. In this study, the quantitative component was fully completed before the design and implementation of the qualitative component. The results of the quantitative component were used to define the qualitative sample.

Ethical Considerations

Ethical approval for this study was obtained from the Newfoundland and Labrador Health Research Ethics Authority (HREB-2017-194). Before agreeing to participate, all subjects were informed about the nature of the research project, possible risks and benefits, and their rights as research subjects. All participants in the interview completed a written consent form and were given a copy.

Quantitative Phase

The quantitative phase aimed to describe the demographics, use, and access to smartphones among patients with COPD. It also aimed to assess whether demographic factors predicted engagement with mHealth. The results were used to define the qualitative sample and create the questions required for the qualitative phase.

Recruitment and Study Setting

Participants were recruited during routine visits to their respirologists at outpatient respirology clinics in St John's, Newfoundland and Labrador, Canada. There are 3 clinics in the city. Participants were eligible for the study if they met the following inclusion criteria:

1. A COPD diagnosis (self-report).
2. Aged ≥30 years at study enrollment.
3. Ability to answer questionnaires in English.
4. Ability to provide informed consent.

Patients who attended respirology clinics received a consent cover letter for research. The cover letter included a questionnaire about technology access and use (Multimedia Appendix 1). After reading the cover letter, interested participants completed the questionnaire and submitted it in a locked box (red box) available at the respirology clinic.

The last section of the questionnaire included a question about the participant's interest in participating in an interview regarding the same topic. Interested participants provided their contact information and placed it in another box (blue box) located at the respirology clinic.

We separated the contact information from the questionnaire to ensure that the questionnaire was kept anonymous. Each questionnaire had a unique identifier. Only the primary investigator was able to link the questionnaire with the contact information. The sample size (n=77) is comparable to that reported by Granger et al [19] and Kayyali et al [20] who assessed the use of technology among patients. Patients did not receive remuneration for completing the questionnaire.

Data Collection

The questionnaire was adapted from Ramirez et al [21] to assess how patients use various types of mobile technology and to what extent they use mHealth. In addition, the questionnaire was used to assess the need for and interest in using mobile health technology, such as mobile phone apps and social media, to help manage health. Similar to Ramirez et al [21] the questionnaire was not validated due to the lack of standardized instruments regarding the use and access to mHealth; however, the questions were consistent with previously published literature in the field.

The questionnaire also included demographic information such as age, sex, race, ethnicity, primary language spoken, annual household income, and education level. In addition, participants were asked about their mobile phone ownership and if their mobile phone had internet capabilities. They were also asked if they used the internet on their mobile phones to learn about their health. Afterward, the participants were asked about their knowledge of mobile phone apps, if they used such apps, and if they were currently using any mobile health apps. In addition, participants identified individuals who they might rely on to use mobile phones and/or mobile apps for them (eg, partner, child, friend).

Statistical Analysis

A database of the questionnaire results was created using unique nonidentifying numbers. The information was password-protected. Before conducting the analysis, data were cleaned, coded, and entered into SPSS version 25.0 (IBM Corporation). Unclear or incomplete survey items were flagged for queries. These were brought to the attention of the research team, then each item was discussed, and a decision concerning its eligibility and entry was made.

Baseline characteristics of participants were summarized with percentages for categorical variables and means and standard deviations for continuous variables. Crude and adjusted odds ratios were measured using univariate and multivariable logistic regression analyses to determine if smartphone ownership was independently associated with age (30-64 years/65 years or older), sex (female/male), marital status (in a relationship/not in a relationship), education level (less than high school/high school/more than high school), and geographical location (rural area/small population center/medium or large population center). All these variables have a known association with smartphone ownership and have been previously reported in the literature [22]. All statistical tests were performed with an alpha significance level of .05.

Qualitative Phase

We used a descriptive qualitative research design grounded in pragmatism [23,24]. Using a qualitative methodology allowed us to achieve an in-depth, contextualized picture of how a diverse sample of patients with COPD think and feel about the possibilities and challenges of using mHealth. This has pragmatic value, as mHealth is an emerging option for delivering health care.

Recruitment and Study Setting

Once all the questionnaires were collected and analyzed, we only contacted participants who agreed to participate in the interview. On the bases of the demographic information collected from the questionnaire, a purposeful sampling strategy was used to identify key informants who could provide rich and diverse interview data. We also used a criterion-based selection [24] so that we could categorize participant characteristics such as age, familiarity with mHealth, smartphone ownership, and years living with COPD.

After interviewing 7 to 8 patients, we reached saturation, as we were not gathering new information. However, we continued interviewing until 10 patients were interviewed to strengthen the validity of inferences [25]. Our final sample size was comparable with similar qualitative studies [15,26].

The study was conducted in St John's, Newfoundland and Labrador, Canada. We conducted interviews at the Memorial University of Newfoundland and others at the participants' homes. Participants were recruited from April to August 2018. After completing the interviews, participants were offered a gift card (Can $30 [US $1.3]).

Data Collection

We conducted individual semistructured interviews to gain an understanding of the lived experiences of patients with COPD in relation to using mHealth [27,28]. Using semistructured interviews allowed the interviewer to begin with a broad question to direct the focus of the interview and then to provide an opportunity for patients to bring forth their thoughts and feelings about the phenomena they thought were important [27,28]. The interview prompts are available in Multimedia Appendix 2. If participants identified that they had not used mHealth, they were asked questions pertaining to why they had not used mHealth (barriers). However, we did not ask them

about facilitators because we did not think they had the experience needed to answer these questions. To facilitate discussions, the interviews were conversational in nature, and items were not asked verbatim or in the order presented. As the study progressed, emerging issues were explored with subsequent participants to refine the themes.

The interview questions and prompts were informed by findings from the literature and input from the authors, who have diverse backgrounds including in mHealth, pharmacy, nursing, medicine, respirology, family medicine, education, and qualitative research. They were also informed by interviewing HCPs regarding the use of mHealth in COPD management [29]. Finally, the interview questions were informed by the results of the quantitative phase.

The interviews were recorded to enable transparent and accurate transcriptions. Interview lengths ranged from 20 to 40 min. Topics included demographics, mHealth usage, perceptions toward challenges of mHealth adoption, factors facilitating mHealth adoption, and preferences regarding features of the mHealth intervention for COPD management. Owing to the large amount of data, preferences regarding features of the mHealth intervention will be published in another article. Data consisted of about 4 hours of interview time with approximately 100 pages of transcription.

Data Analysis

The interviews were transcribed verbatim and compared against the digital recordings to ensure the accuracy of the content. Identifying information (names) was removed to ensure anonymity. We used NVivo (version 12; QSR International) to organize the data and examine the words, including frequency counts, as in classical content analysis [30]. All data were analyzed, but we only coded the data that were relevant for answering the research questions, as recommended by Saldana [31], Wolcott [32], and Yin [33]. An audit trail was created to keep track of all analytic decisions [34].

After using NVivo, we used first cycle coding that was both structural and holistic [31], meaning that we used the interview prompts and the literature to guide some of the coding. One researcher analyzed the transcripts and developed a set of themes and subthemes and then obtained input from a second researcher. In the second cycle of coding, the 2 researchers independently coded the data using pattern coding to develop themes [31]. They then discussed commonalities and differences in their coding and theme development until consensus was reached. Qualitative data analysis is mainly inductive as the researcher does not approach the data with a pre-existing hypothesis that is meant to be supported or refuted. Instead, the researcher reads and rereads the data, using specific information to generate more general ideas that may take the form of themes or explanations. For this study, we had already analyzed data from health care providers [29]. We used a deductive approach to look for patient data that would either support or disconfirm the HCP results from the point of view of the patients. In addition, inductive analysis was also used to search for new ideas not mentioned by the HCPs. The iterative process continued as these analyses were conducted to find commonalities, differences, and new patterns in thinking.

Mixed Methods Integration

In addition to the integration at the study design level, we implemented integration at the methods, interpretation, and reporting levels [35]. We implemented integration at the methods level in 2 ways: connecting and building. Connecting occurs when a researcher links one type of data to another type of data through sampling [35]. In this study, we used data from the questionnaire to purposefully sample participants for follow-up interviews. Specifically, we were able to categorize participant characteristics such as age, familiarity with mHealth, smartphone ownership, and years living with COPD. Obtaining information from a diverse group may generate a more complete picture, reveal patterns that would otherwise go unnoticed, and may also help identify novel relationships between variables and concepts [18]. Building occurs when one database informs the data collection approach of the other [35]. In this study, we used the questionnaire responses to develop some aspects of the interview guide. This allowed us to gain further insights regarding mHealth use. For example, knowing if patients used mHealth in the past was helpful in understanding the facilitators of mHealth adoption.

Furthermore, we implemented integration at the interpretation and reporting levels. We used both integrations through narratives and the use of a joint display [35,36]. Integration through a narrative occurs when a researcher describes the quantitative and qualitative findings in a single report or series of reports [35]. In this paper, we described the findings in a single report and used the contiguous approach, in which qualitative and quantitative findings are reported in different sections. Finally, we used a joint display to provide a structure to discuss the integrated analysis [36].

Results

A Cross-Sectional Survey of Patients

Demographics

A total of 100 adults completed the survey from January to November 2018. Only 77 participants reported that they were diagnosed with COPD and were included in the analysis. Table 1 provides an overview of the demographic and health information of the sample. Table 1 indicates that most participants were aged 55 years or older, and 61.5% had an annual income of less than $40,000. About 70% had earned at least a high school diploma, and there was a mixture of rural and urban participants. Participants had a range of comorbidities, and almost 65% were taking 5 or more medications.

Table 1. Participant demographics and health information (N=77).

Variables	Values, n (%)
Age (years; n=77)	
30-34	2 (2.6)
45-54	3 (3.9)
55-64	15 (19.5)
65 or older	57 (74)
Sex (n=74)	
Female	44 (59.5)
Male	30 (40.5)
Marital status (n=74)	
Married	44 (59.5)
Common law	6 (8.1)
Single (never married)	6 (8.1)
Widowed, separated, or divorced	18 (24.3)
Income (Can $; n=52)	
Under 20,000	14 (26.9)
20,000-39,000	18 (34.6)
40,000-59,000	6 (11.5)
60,000-79,000	4 (7.7)
80,000-150,000	8 (15.4)
Over 150,000	2 (3.8)
Employment (n=68)	
Employed full time	6 (8.8)
Employed part time	2 (2.9)
Self-employed	2 (2.9)
Retired	52 (76.5)
Unemployed	6 (8.8)
Education level (n=68)	
Less than high school	14 (20.6)
High school equivalency (GED)	9 (13.2)
High school	25 (36.8)
College/trade	10 (14.7)
Bachelor’s degree	5 (7.4)
Master’s degree	4 (5.9)
PhD/MD/JD	1 (1.5)
Population size (n=71)	
Rural area (with a population less than 1000)	18 (25.4)
Small population center (with a population between 1000 and 29,999)	23 (32.4)
Medium population center (with a population between 30,000 and 99,999)	6 (8.5)
Large urban population center (with a population of 100,000 or more)	24 (33.8)
Self-reported comorbidities[a] (n=65)	
Cancer	18 (28.1)
Diabetes	15 (23.4)

Variables	Values, n (%)
Heart disease	14 (21.9)
Skeletal or muscular disease	12 (18.8)
Kidney disease	4 (6.3)
Mental health issues	2 (3.1)
Medication intake (n=68)	
None	2 (2.9)
1-2	10 (14.7)
3-4	12 (17.6)
4-6	16 (23.5)
More than 6	28 (41.2)

[a]Some patients reported several comorbidities.

mHealth Technology Ownership

Table 2 illustrates the findings regarding mHealth technology ownership. A total of 73% (56/77) participants owned a mobile phone, but only about one-fourth of the participants, 23% (18/77), owned a smartphone. The number of iPad or tablet owners was slightly higher than that of smartphones at 33% (25/77). In terms of the availability of a smartphone in the household, 27% (21/77) participants stated that a person in the household owns a smartphone. Although only 23% (18/77) participants reported having a smartphone, 29% (22/77) participants were able to access the internet via their phone, suggesting a higher number of smartphones than initially reported. About a third of participants 35% (27/77) owned a blood pressure monitor, and 22% (17/77) participants owned a glucometer.

Logistic regression was performed to ascertain the association of age, sex, marital status, education level, and geographical location with the likelihood that participants owned a smartphone. Owing to missing observations, the true sample used in the regression was 65 of 77.

We measured crude odds ratios to determine whether smartphone ownership was independently associated with the occurrence of predictor variables. The likelihood of owning a smartphone was reduced in participants earning less than a high school diploma (crude odds ratio [cOR] 0.11, 95% CI 0.02-0.64; *P*=.01) and participants earning a high school diploma (cOR 0.14, 95% CI 0.04-0.5; *P*=.002) compared with participants who received education beyond high school.

The logistic regression model was statistically significant, χ^2_7=15.8, *P*=.01. The model explained 30.4% (Nagelkerke R^2) of the variance in smartphone ownership and correctly classified 77% of cases. Sensitivity was 50%, and specificity was 89%. Of the 5 predictor variables, only 2 were statistically significant, and both were related to education level (as shown in Table 3). The likelihood of owning a smartphone was reduced in participants earning less than a high school diploma (adjusted odds ratio [aOR] 0.12, 95% CI 0.017-0.86; *P*=.03) and participants earning a high school diploma (aOR 0.13, 95% CI 0.03-0.54; *P*=.005) compared with participants who received education beyond high school.

Table 2. Mobile health technology ownership (N=77).

Mobile health technology ownership	Values, n (%)
Mobile phone	56 (72.7)
Smartphone	18 (23.4)
iPad	25 (32.5)
Availability of a smartphone in the household	21 (27.3)
Internet access through a mobile phone	22 (28.6)
Spirometer/peak flow meter	4 (5.2)
Glucometer	17 (22.0)
Blood pressure monitor	27 (35.1)
Heart rate monitor	10 (13.0)
Accelerometer/activity counter	3 (3.9)
Scale	15 (19.5)
Thermometer	18 (23.4)

Table 3. Logistic regression predicting the likelihood of smartphone ownership.

Variables	Adjusted odds ratio (95% CI)	*P* value	Crude odds ratio (95% CI)	*P* value
Age (years)				
30-64	2.24 (0.57-8.77)	.25	1.97 (0.65-6.04)	.23
65 or older	Reference	N/A[a]	Reference	N/A
Sex				
Female	2.10 (0.54-8.19)	.29	1.03 (0.36-2.94)	.95
Male	Reference	N/A	Reference	N/A
Marital status				
In a relationship[b]	2.36 (0.494-11.29)	.28	1.63 (0.51-5.17)	.41
Not in a relationship[c]	Reference	N/A	Reference	N/A
Education level				
Less than high school	0.12 (0.02-0.86)	*.03* [d]	0.11 (0.02-0.64)	*.01* [c]
High school[e]	0.13 (0.03-0.54)	*.005* [d]	0.14 (0.04-0.50)	*.002* [c]
More than high school	Reference	N/A	Reference	N/A
Population size				
Rural area	0.50 (0.09-2.76)	.43	0.35 (0.08-1.47)	.15
Small population center	0.46 (0.11-1.92)	.29	0.61 (0.19-2.01)	.42
A medium population center or a large population center	Reference	N/A	Reference	N/A

[a]N/A: not applicable.

[b]In a relationship includes being married or in common law.

[c]Not in a relationship includes being single, widowed, separated, or divorced.

[d]Significance level <.05.

[e]High school includes General Educational Development.

mHealth Technology Use

Table 4 highlights the findings related to mHealth technology use. Only a third of the participants 29% (20/77) understood the term “app.” Of these, 50% (10/20) used apps. Among app users, only 3/10 participants used health apps, and 7/10 participants were interested in using a health app. Six app users said they would be comfortable allowing their family members to access their health information, and 7 said they would be comfortable with their HCP accessing their health care information. The most common social media platform used was Facebook 38% (29/77). Among users of social media, 45% (13/29) reported using social media at least once a day.

Participants completed questions about their concerns regarding mHealth adoption. The first question was about concerns regarding smartphones. The following 3 options were chosen by participants: cost of smartphones 24.5% (21/77), reducing face-to-face interactions 20.4% (10/77), and not easy to use 18.4% (9/77). Of the participants who used apps (n=10), the following concerns about app use were chosen: worried about personal information disclosure (n=6), extra fees to use the app (n=3), apps use a lot of data (n=3), apps are not easy to use (n=1), and I do not know if they are effective (n=1). No participants chose “taking too much time to use” or “not recommended by a health care provider” as a concern arising from using apps.

Table 4. Mobile health technology use.

Variables	Values, n (%)
Understood the term "app" (n=77)[a]	20 (26)
Use apps (n=20)[b]	10 (50)
Use health apps (n=10)[c]	3 (30)
Interested in using health apps (n=10)	7 (70)
Comfortable allowing a family member to access health information (n=10)	6 (60)
Comfortable allowing a health care provider to access health information (n=10)	7 (70)
Social media use (n=77)	
Facebook	29 (38)
Twitter	2 (3)
Instagram	2 (3)
Snapchat	1 (1)
Interested in using social media to share health experience (n=77)	9 (12)
Frequency of social media use (n=29)[d]	
Never	1 (3.4)
A few times a month	5 (17.2)
A few times a week	3 (10.3)
About once a day	7 (24.1)
More than once a day	13 (44.8)

[a]Total study population.

[b]Sample population that understood the term "app."

[c]Sample population that uses apps.

[d]Sample population that uses social media.

Semistructured Interviews With Patients

We developed themes under 2 categories: facilitators and barriers that would influence the feasibility and use of mHealth. We have included details and examples to illustrate patients' thoughts and beliefs. These findings expand on the barriers and facilitators reported previously by health care providers who treat patients with COPD [29]. Textbox 1 highlights the facilitators and barriers reported by patients.

Textbox 1. The facilitators and barriers to mobile health adoption.

Facilitators

1. There are possible health benefits for patients
2. The software needs to be easy to use
3. Patients need to be educated on the use of mobile health (mHealth)
4. The credibility of mHealth needs to be evident

Barriers

1. There are technical issues with mHealth
2. Lack of awareness is a challenge
3. There may be limited uptake from the elderly
4. There are possible financial barriers
5. There may be privacy and confidentiality concerns
6. There was little interest in using an mHealth intervention

Demographics

A diverse group of 10 patients with COPD participated in face-to-face interviews. The mean age was 67.6 (SD 7.58) years, and the range was from 51 to 80 years. There were 4 females and 6 males. Participants stated that the mean number of years living with COPD was 8.4 (SD 4.45), and the range was from 3 to 15 years.

Most patients expressed interest in using mHealth to assist in the management of their COPD. In terms of smartphone ownership, 6 participants owned a smartphone, 2 owned a mobile phone, and 2 did not have either. Participants used their phones for different purposes, including communication, managing finances, gaming, and browsing the internet. One patient stated "I use it for everything. I never thought I'd see the day where I was dependant on my phone." On the other hand, another participant stated that she did not use her smartphone beyond making phone calls: "I have a phone, but I am not smart enough to use it." Some participants used it to monitor their physical activity: "I have a health thing on it and I look at it every once in a while just to see how many steps I've done that day because it...Improve health activity." One patient used his smartphone to "get pollen reports and anything that will trigger a COPD attack...to do research on nutraceutical products or on COPD-related matters." Four participants were enrolled in an mHealth intervention to manage their COPD.

Facilitators

Patients reported 4 facilitators, which are discussed below.

1. There Are Possible Health Benefits for Patients

Participants agreed that mHealth has the potential to provide health benefits to patients. One patient who used a fitness tracker remarked:

> *I think it's better for me to track what I do every day. It is going to make me feel better...That would help me a lot with my pains...I know how my day was and I know how I feel like.*

Another patient who kept track of his vitals, weight, and medication intake stated that:

> *if I had been monitored, I might not have this broken arm...That should've been picked up on. I mean, I got the records there and you look back on it, I can easily look back on them now and say you had water retention, your resting heart rate was way too high, and your blood pressure was low. There's something wrong.*

Two patients described that mHealth could provide a sense of security and reduce hospitalization:

> *...it felt good to have that, you know, that security there at least that in those four months that I had it. So there was no guessing because you don't know if you should or you shouldn't go to a hospital.*

2. The Software Needs to Be Easy to Use

Usability was highlighted by many patients as an important factor in increasing the uptake of mHealth. One patient cautioned that he might not use an intervention if it was not easy to use:

> *I'm not getting into something that's going to fill up my day ferreting around. But if it's something that I've got to look at for five or ten minutes, I'm okay with that.*

Participants who were enrolled in an mHealth intervention mentioned that there was "no trouble setting up. It's all there, so all you had to do is turn it on."

3. Patients Need to Be Educated on the Use of mHealth

Patients learned how to use mHealth interventions via different sources. The majority asked their family members for assistance "...I'll go to my 14-year-old who is generationally more apt to be able to teach me a new technology." This was also mentioned by another patient who used a fitness tracker: "My nieces. They buy it for me and they set it up for me and everything." Other patients taught themselves about mHealth interventions: "If I can read it, I can learn it...I generally research it myself" or used the library: "Also the libraries here will help you in any programs." One patient said that her "...own care worker helps with it, I don't know how to do it." It was also recommended by some that technical support staff be available as a resource for patients to call when they needed technical help.

4. The Credibility of mHealth Needs to be Evident

Some patients thought that the credibility of mHealth needs to be made evident. Some patients stated that they would use an mHealth intervention if it was recommended by their HCP: "if he told me that, I probably would try to do it." This was reiterated by a patient who regularly monitored his COPD: "Doctor xx was the one that said to me...If you're going to cope with this and keep it under control, you're going to have to learn how to look after yourself."

Barriers

Patients reported 6 barriers to mHealth adoption, which are discussed below.

1. There Are Technical Issues With mHealth

A few patients expressed that they did not have the technical expertise to use mHealth. One patient expressed his concern: "I wouldn't know how to turn a computer on. I'm not very good... You know, I never grew up with computers, but I have seen on." This was reiterated by another patient: "It just looked way too complicated to download the app so I didn't because I'm technology averse." For patients who used mHealth, technical issues included limited cellular and Wi-Fi connections: "we travel out to the cabin every weekend and the cabin's out in central Newfoundland out in Terra Nova and now it's getting better now because these phone services getting better out there." Another issue was moving and setting up the mHealth intervention components, such as a blood pressure monitor or a scale, when traveling: "...it's not a problem. But when you go on vacation, sometimes you got to take this along with you and set it up somewhere."

2. Lack of Awareness is a Challenge

Many patients indicated that a lack of awareness is a barrier to mHealth adoption. One patient expressed this concern: "I'm not aware of everything that's out there, but people need to be more aware of their COPD and know more about it." In addition, a family member, who accompanied her mother in the interview, stated the following about her mother who has COPD:

> *I think it's more she probably don't know what her phone can do, right. But if you say to her okay we're going to start using this, this is going to be useful and it's going to be beneficial I'm sure she'd be game for it.*

However, patients who participated in an mHealth intervention stated that their HCP recommended mHealth interventions when they were in the hospital or attended a community health event. Another patient mentioned that "it was advertised and I called in about it and they got in contact with me, set me up with it."

3. There May Be Limited Uptake by the Elderly

A few patients mentioned that they face issues in adopting technology because of their age. One patient stated, "I'm not generationally born into technology that is prevalent and considered a norm of the day." This was also mentioned by another: "I try but I'm a little bit nervous, sometimes I'll ask my daughter or someone else around because we didn't grow up with the phones as you do today."

4. There Are Possible Financial Barriers

A few patients said they cannot afford the costs associated with mHealth, such as the cost of a smartphone. One patient expressed it this way: "there's no way I'll pay that money." Expenses incurred through the use of data were also discussed: "it's a bit expensive for like I've got no data right now because it's all extra." In addition to individual patient costs, one patient raised the concern of additional costs required by HCPs: "doctors are not going to do that without a fee." In addition, some mHealth programs are limited to a certain period, which may increase costs.

5. There May Be Privacy and Confidentiality Concerns

One patient thought privacy and confidentiality could be a barrier to mHealth adoption. She was open to sharing the results with an HCP but not to a family member: "I don't want to make them worry (her family) because I told them nothing about my cancer... I don't like to worry my family." The rest of the patients were willing to share the results with their family members and HCPs: "I don't care who sees it...They can put it in the Evening Telegram, doesn't bother me." This was echoed by another patient: "they (his family) watch over my shoulder like nothing else now."

6. There Was Little Interest in Using an mHealth Intervention

Some patients indicated that a lack of interest in mHealth is a major barrier to its adoption. One patient expressed this concern: "I do try to help myself but when it comes to using the phone and that stuff and the computer, it's not for me." This was also mentioned by another patient with limited technology experience: "I know on smartphones you can dial, you can play a game and some they can even watch movies probably, but I got no interest." Another patient mentioned that they are too busy to include an mHealth intervention in their routine: "So my day is pretty filled with different things. So remembering is a problem. Sometimes I just say: To hell with it and I am not going to do it. But remembering is probably the biggest thing."

When patients use mHealth, they may share the results with their HCP. This occurs if the mHealth intervention is not integrated in the health care system. One patient posited that some HCPs are not interested in reviewing the records brought by patients: "And my sheet, he didn't even look at it...That's, that's depressing."

Discussion

Principal Findings

We conducted an explanatory, sequential mixed methods study with patients and identified their perceptions regarding the use of mHealth for COPD management. The quantitative component revealed that over 70% of patients owned a mobile phone, but only about a quarter of the participants (18/77, 23.4%) owned a smartphone. The likelihood of owning a smartphone was not associated with age, sex, marital status, or geographical location. However, patients with a high educational status were more likely to own a smartphone. The qualitative component found that patients, in general, had a positive attitude toward mHealth adoption for COPD management, but several facilitators and barriers were identified. It is important to promote facilitators and address the barriers to optimize the successful implementation of mHealth interventions.

Using a mixed methods approach allowed us to produce a diverse sample of patients with COPD. The quantitative and qualitative components complemented each other to improve the validity of inferences and expand on why participants answered quantitative questions in a certain way. For example, the number of participants who stated that they had a smartphone (n=18) was lower than the number of participants who accessed the internet through their mobile phone (n=22), suggesting a lack of understanding of what a smartphone is. This was further explored during the interviews. Although some participants owned a smartphone, their use was limited to making phone calls and taking pictures. On the other hand, some participants did not own a smartphone, but they were able to enroll in an mHealth intervention and complete the program. As explained by one patient, "Eastern Health sets you up with everything. It's so different, there's nothing to it, it's just hit the button, use your device and it's so easy to use." This finding highlights the need for education and confidence building among patients with COPD.

We created a joint display (Table 5) to clarify some of the barriers that were reported quantitatively. The table is organized by survey items. It merges the related quantitative results and provides typical comments from patients.

Table 5. Joint display of barriers to mobile health adoption.

Quantitative results: variables	Values, n (%)	Qualitative results: exemplar quotes	Interpretation of mixed methods findings
Concerns regarding smartphones (N=77[a])			
Cost of smartphones	21 (27.3)	• "I can't afford one (smartphone)" • "I've got no data right now because it's all extra" • "we can always get one (smartphone)"	Costs include the cost of a smartphone and the data to enable its functionalities. However, some patients could afford to get a smartphone, or it could be provided by the health care system
Not easy to use	9 (11.7)	• "I wouldn't know how to turn a computer on. I'm not very good..." • "There's nothing to it, it's just hit the button, use your device and it's so easy to use"	Although some participants owned a smartphone, their use was limited to making phone calls and taking pictures. On the other hand, some participants did not own a smartphone, but they were able to enroll in a mobile health intervention and complete the program. This finding highlights the need for education and confidence building among patients with COPD[b]
Concerns regarding app use (N=10[c])			
Worried about personal information disclosure	6 (60)	• "I don't want to make them worry because I told them nothing about my cancer...I just told my sister a week before I had my surgery...I don't like to worry my family" • "I don't care who sees it...They can put it in the Evening Telegram, doesn't bother me"	There was inconclusive evidence regarding confidentiality. Patients should have a choice in what to share and who should have access to their health information

[a]The total study population.

[b]COPD: chronic obstructive pulmonary disease.

[c]The sample population that uses apps.

Comparison With Prior Work

Our study provides a meaningful contribution to the literature, as few prior studies have specifically examined the use of mHealth among patients with COPD. It is important to note that the published literature on mHealth access and use was focused on general and largely healthy populations, with little attention to individuals with chronic illnesses, such as COPD [37].

A study investigated smartphone ownership among the general public and reported a high smartphone adoption rate of 76% [20]. Ramirez et al investigated smartphone ownership in primary care clinics and found a high adoption rate of 76%; however, 96% of the participants were younger than 65 years. Among patients with COPD from several locations in the United States, we found lower smartphone adoption rates (47%), which was double the rate reported in our sample [38]. The qualitative data expand on this finding by clarifying that some smartphone owners use their device in ways similar to mobile phone owners. Knowing how to use the features of a smartphone, such as using apps, is necessary for using the mHealth intervention. This finding highlights the need for education and confidence building among some smartphone users to help them use their devices for COPD management.

Our logistic regression results support the claim that a lower level of education is associated with limited access to mobile devices [39-41]. Other researchers also found that the individuals more likely to use health apps tended to be younger and have higher incomes [38-41]. Our findings echo concerns about the relationship between mHealth access and health disparities [38]. Limiting mHealth interventions to smartphone or tablet owners could disproportionately benefit highly educated and wealthy individuals.

We also investigated the association between technology use and geographical location. Although we did not find differences between smartphone ownership among urban and rural patients with COPD, a report suggests that individuals living in rural areas are less likely to have smartphones than individuals not living in rural areas [42]. Previous studies suggest that among patients with COPD, living in rural areas was associated with a worse health status [43,44]. The authors suggest that the higher prevalence of COPD in rural areas could be linked to an increased proportion of older residents, a shortage of HCPs, the underuse of spirometry and pulmonary rehabilitation, and problems with access to medical care [43,44]. Limited access to smartphones may further exacerbate health disparities for rural patients [38].

Similar to Kayyali et al [20], the majority of the participants were open to data-sharing options with an HCP through mHealth apps. Nevertheless, Kayyali et al [20] stated that data sharing can be ineffective if the participant is not honest or if data sharing is overused.

Some of the findings presented in this study confirm previously reported findings in the context of mHealth for COPD management. Our findings are in agreement with those of Vorrink et al [45], who stressed the importance of training patients and HCPs on the proper use of mHealth. As suggested by Korpershoek et al [15], we confirm that the expected benefits of using mHealth contribute to the success of mHealth uptake, and our study provides additional insight regarding these perceptions, such as on the ease of use, educating patients about

mHealth, and the importance of credibility. Our findings echo some and add to the barriers reported by Krebs and Duncan [39], which include technical difficulties and a lack of interest. We included quotes from participants to expand on these insights.

In comparison with the facilitators reported by HCPs, patients had 4 parallel facilitators: there are possible health benefits for patients, the software needs to be easy to use, patients need to be educated on the use of mHealth, and the credibility of mHealth should be evident [29]. The only facilitator that was not mentioned by patients is that mHealth should reduce the cost to the health care system. On the other hand, patients had 5 parallel barriers with the HCPs: there are technical issues with mHealth, lack of awareness is a challenge, there may be limited uptake from the elderly, there are possible financial barriers, and there may be privacy and confidentiality concerns [29]. The possibility of mHealth limiting the personal connection between HCPs and patients was not mentioned as a barrier by patients. Furthermore, one new barrier emerged from interviewing patients; there was little interest in using mHealth interventions. Our findings and the limited literature on this matter emphasize the need for further research into the use of mHealth in COPD management.

Strengths and Limitations

This study has several strengths. First, we used a mixed methods approach to produce a diverse sample of participants. This human-centered approach ensures that the needs and challenges of a diverse group of patients can be considered before developing an mHealth intervention. Second, some patients had experience in using mHealth interventions to manage their COPD, which further increases the richness of the data. Third, all the interviews were conducted in a similar manner to ensure consistency during data collection and analysis. Finally, we recruited patients with COPD from outpatient respirology clinics. This has led to the capture of a well-characterized cohort of individuals with COPD.

There were also several limitations. First, the number of patients who completed the survey was relatively small. However, all efforts were made to recruit as many participants as possible and facilitate the completion of the survey. Owing to the small sample size, regression coefficients may have been imprecisely estimated. However, the age of the sample was reflective of a representative sample of patients with COPD in Canada who were recruited from a respirology clinic [46]. Furthermore, the main objective of the survey was to obtain a diverse sample of patients with COPD for the qualitative phase. Second, not all patients had experience using mHealth. Thus, the perceptions of these participants were not based on actual interventions with patients. Third, conducting focus groups with some of the participants following the individual interviews could have yielded richer information, as participants would have been given the opportunity to compare their thoughts and confirm or expand upon each other's ideas. Fourth, there were no questions in the quantitative survey about facilitators of mHealth uptake. Including this topic would have been beneficial, as it could have been expanded upon when conducting the interviews.

Implications for Practice and Future Research

The findings of this study may help various stakeholders who are planning to use mHealth interventions for COPD management. It is important to consider the low rate of smartphone use among patients when implementing an mHealth intervention for COPD management. Some lessons learned include the importance of raising awareness among patients regarding the potential of mHealth interventions in COPD management. Family members could play a significant role in raising awareness as well as in teaching patients with COPD about mHealth. The findings also emphasize the importance of developing a user-friendly mHealth intervention. This could reduce the time and resources required to teach patients about the mHealth intervention. The lack of an internet connection could limit access to mHealth interventions. This should be taken into consideration when measuring access to health resources in rural communities. Some of the barriers and facilitators have the potential to be applied to other chronic diseases. For example, these findings could be beneficial for stakeholders who plan to develop a mHealth intervention for heart failure or diabetes.

mHealth is particularly important in geographical locations with a relatively large proportion of rural residents such as Newfoundland and Labrador. Of the Atlantic provinces, NL has the highest proportion of its population (60%) living in rural areas [47]. mHealth may enhance care provider access throughout sparsely populated rural areas. Newfoundland and Labrador has a substantial remote and rural population; therefore, our results may be more applicable to rural areas.

Future studies would benefit from conducting focus groups with some of the participants following individual interviews. Focus groups could yield rich information, as participants would be given the opportunity to compare their thoughts and confirm or expand upon each other's ideas. After developing a user-centered mHealth intervention, the authors recommend using a mixed methods framework for usability testing [48]. Additional trials will be required to provide data regarding the efficacy and cost-effectiveness of mHealth interventions in COPD management.

Conclusions

It is important to understand access to mHealth among patients with COPD and their perceptions regarding the adoption of mHealth for COPD management. Despite the rise in smartphone adoption, the rate of adoption among patients with COPD remains to be low. Additionally, it is important to consider that owning a smartphone does not mean that one has the ability to use it for mHealth. This finding highlights the need for education and confidence building among some smartphone users to be able to use their devices for COPD management. This study identifies some potential facilitators and barriers that may inform the successful development and implementation of mHealth interventions for COPD management. We recommend that those who develop mHealth interventions for COPD should consider the facilitators and barriers highlighted in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient questionnaire.
[DOCX File , 21 KB - mhealth_v8i7e17409_app1.docx]

Multimedia Appendix 2
Patient interview prompts.
[DOCX File , 18 KB - mhealth_v8i7e17409_app2.docx]

References

1. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. Am J Respir Crit Care Med 2017 Mar;195(5):557-582. [doi: 10.1164/rccm.201701-0218pp]
2. The Top 10 Causes of Death. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2020-06-23]
3. secure.cihi.ca. CIHI: Canadian Institute for Health Information. 2008. URL: https://secure.cihi.ca/free_products/HealthIndicators2008_ENGweb.pdf [accessed 2019-03-07] [WebCite Cache ID 76gVKOWKJ]
4. Najafzadeh M, Marra CA, Lynd LD, Sadatsafavi M, FitzGerald JM, McManus B, et al. Future impact of various interventions on the burden of COPD in Canada: a dynamic population model. PLoS One 2012;7(10):e46746 [FREE Full text] [doi: 10.1371/journal.pone.0046746] [Medline: 23071626]
5. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
6. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble J, Abu Ashour W. The effect of smartphone interventions on patients with chronic obstructive pulmonary disease exacerbations: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 Sep 1;4(3):e105 [FREE Full text] [doi: 10.2196/mhealth.5921] [Medline: 27589898]
7. Joe J, Demiris G. Older adults and mobile phones for health: a review. J Biomed Inform 2013 Oct;46(5):947-954 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.008] [Medline: 23810858]
8. Juen J, Cheng Q, Schatz B. A natural walking monitor for pulmonary patients using mobile phones. IEEE J Biomed Health Inform 2015 Jul;19(4):1399-1405. [doi: 10.1109/JBHI.2015.2427511] [Medline: 25935052]
9. Zhang J, Song YL, Bai CX. MIOTIC study: a prospective, multicenter, randomized study to evaluate the long-term efficacy of mobile phone-based internet of things in the management of patients with stable COPD. Int J Chron Obstruct Pulmon Dis 2013;8:433-438 [FREE Full text] [doi: 10.2147/COPD.S50205] [Medline: 24082784]
10. Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems. International Organization for Standardization. 2010. URL: https://www.iso.org/standard/52075.html [accessed 2020-06-30]
11. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
12. Nelson LA, Mayberry LS, Wallston K, Kripalani S, Bergner EM, Osborn CY. Development and usability of REACH: a tailored theory-based text messaging intervention for disadvantaged adults with type 2 diabetes. JMIR Hum Factors 2016 Sep 8;3(2):e23 [FREE Full text] [doi: 10.2196/humanfactors.6029] [Medline: 27609738]
13. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
14. Gray CS, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
15. Korpershoek YJ, Vervoort SC, Trappenburg JC, Schuurmans MJ. Perceptions of patients with chronic obstructive pulmonary disease and their health care providers towards using mHealth for self-management of exacerbations: a qualitative study. BMC Health Serv Res 2018 Oct 4;18(1):757 [FREE Full text] [doi: 10.1186/s12913-018-3545-4] [Medline: 30286761]
16. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Erratum: author correction: impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digit Med 2018;1:17 [FREE Full text] [doi: 10.1038/s41746-018-0027-3] [Medline: 31305588]

17. Alshahrani A. Self-management and diabetes mellitus mhealth, a glance at the present and a glimpse into the future. Stud Health Technol Inform 2019 Jul 4;262:352-355. [doi: 10.3233/SHTI190091] [Medline: 31349340]
18. Curry L. In: Nunez-Smith M, editor. Mixed Methods in Health Sciences Research: A Practical Primer. Volume 1. Thousand Oaks, CA: Sage Publications; 2015.
19. Granger CL, Wijayarathna R, Suh E, Arbane G, Denehy L, Murphy P, et al. Uptake of telehealth implementation for COPD patients in a high-poverty, inner-city environment: a survey. Chron Respir Dis 2018 Feb;15(1):81-84 [FREE Full text] [doi: 10.1177/1479972317707653] [Medline: 28513197]
20. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and use of mhealth apps: a study from England. Pharmacy (Basel) 2017 Jun 14;5(2):33 [FREE Full text] [doi: 10.3390/pharmacy5020033] [Medline: 28970445]
21. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
22. Mobile Fact Sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-06-24]
23. Merriam S, Tisdell E. Qualitative Research. Fourth Edition. New York, USA: Jossey-Bass; 2015.
24. Patton M. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2015.
25. Maxwell JA. Qualitative Research Design. Third Edition. Thousand Oaks, CA: Sage Publications; 2013.
26. Barken TL, Thygesen E, Söderhamn U. Unlocking the limitations: living with chronic obstructive pulmonary disease and receiving care through telemedicine-a phenomenological study. J Clin Nurs 2018 Jan;27(1-2):132-142. [doi: 10.1111/jocn.13857] [Medline: 28425194]
27. Brinkman S, Kvale S. InterViews: Learning the Craft of Qualitative Research Interviewing. Third Edition. Thousand Oaks, CA: Sage Publications; 2015:-.
28. Brooks FB, Rubin HJ, Rubin IS. Qualitative Interviewing: The Art of Hearing Data. Thousand Oaks, CA: Sage Publications; 1996.
29. Alwashmi MF, Fitzpatrick B, Davis E, Gamble J, Farrell J, Hawboldt J. Perceptions of health care providers regarding a mobile health intervention to manage chronic obstructive pulmonary disease: qualitative study. JMIR Mhealth Uhealth 2019 Jun 10;7(6):e13950 [FREE Full text] [doi: 10.2196/13950] [Medline: 31199330]
30. Leech NL, Onwuegbuzie AJ. Qualitative data analysis: a compendium of techniques and a framework for selection for school psychology research and beyond. School Psychol Q 2008 Dec;23(4):587-604. [doi: 10.1037/1045-3830.23.4.587]
31. Saldaña J. The Coding Manual for Qualitative Researchers. Thousand Oaks, CA: Sage Publications; 2016.
32. Wolcott H. Writing Up Qualitative Research. Thousand Oaks, CA: Sage Publications; 2008.
33. Yin KR. Qualitative Research from Start to Finish. Second Edition. New York, USA: GuilFord Press; 2011.
34. Guest G, MacQueen K, Namey E. Applied Thematic Analysis. Thousand Oaks, CA: Sage Publications; 2012.
35. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
36. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
37. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
38. Witry M, Comellas A, Simmering J, Polgreen P. The association between technology use and health status in a chronic obstructive pulmonary disease cohort: multi-method study. J Med Internet Res 2018 Apr 2;20(4):e125 [FREE Full text] [doi: 10.2196/jmir.9382] [Medline: 29610113]
39. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
40. Fox S, Duggan M. Mobile Health 2012. Pew Research Center. 2012. URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2015-04-23]
41. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht U. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
42. Perrin A. Digital Gap Between Rural and Nonrural America Persists. Pew Research Center. 2017. URL: https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/ [accessed 2020-06-30]
43. Abrams TE, Vaughan-Sarrazin M, Fan VS, Kaboli PJ. Geographic isolation and the risk for chronic obstructive pulmonary disease-related mortality: a cohort study. Ann Intern Med 2011 Jul 19;155(2):80-86. [doi: 10.7326/0003-4819-155-2-201107190-00003] [Medline: 21768581]
44. Jackson BE, Coultas DB, Suzuki S, Singh KP, Bae S. Rural-urban disparities in quality of life among patients with COPD. J Rural Health 2013 Aug;29(Suppl 1):s62-s69 [FREE Full text] [doi: 10.1111/jrh.12005] [Medline: 23944281]

45. Vorrink S, Huisman C, Kort H, Troosters T, Lammers J. Perceptions of patients with chronic obstructive pulmonary disease and their physiotherapists regarding the use of an ehealth intervention. JMIR Hum Factors 2017 Sep 19;4(3):e20 [FREE Full text] [doi: 10.2196/humanfactors.7196] [Medline: 28928110]
46. Safka KA, Wald J, Wang H, McIvor L, McIvor A. GOLD stage and treatment in COPD: a 500 patient point prevalence study. Chronic Obstr Pulm Dis 2016 Dec 22;4(1):45-55 [FREE Full text] [doi: 10.15326/jcopdf.4.1.2016.0126] [Medline: 28848910]
47. Simms A, Greenwood R. Newfoundland and Labrador's Place in Canada. Government of Newfoundland and Labrador. 2015. URL: https://www.gov.nl.ca/publicat/royalcomm/research/Pollara.pdf [accessed 2020-06-30]
48. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]

Abbreviations

aOR: adjusted odds ratio
COPD: chronic obstructive pulmonary disease
cOR: crude odds ratio
HCD: human-centered design
HCP: health care professional
ISO: International Organization for Standardization
mHealth: mobile health

Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by S Lundell, S Chen; comments to author 10.02.20; revised version received 10.05.20; accepted 14.06.20; published 23.07.20.

Please cite as:
Alwashmi MF, Fitzpatrick B, Farrell J, Gamble JM, Davis E, Nguyen HV, Farrell G, Hawboldt J
Perceptions of Patients Regarding Mobile Health Interventions for the Management of Chronic Obstructive Pulmonary Disease: Mixed Methods Study
JMIR Mhealth Uhealth 2020;8(7):e17409
URL: http://mhealth.jmir.org/2020/7/e17409/
doi:10.2196/17409
PMID:32706697

Original Paper

User-Dependent Usability and Feasibility of a Swallowing Training mHealth App for Older Adults: Mixed Methods Pilot Study

HyangHee Kim[1], PhD, CCC-SLP; Sang-Ho Lee[2], BA; Nam-Bin Cho[2], BA; Heecheon You[3], PhD; Teukgyu Choi[4]; Jinwon Kim[3], MSc

[1]Graduate Program in Speech-Language Pathology, Department and Research Institute of Rehabilitation Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Graduate Program in Speech-Language Pathology, Yonsei University, Seoul, Republic of Korea

[3]Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Republic of Korea

[4]Humanopia, Pohang, Republic of Korea

Corresponding Author:
HyangHee Kim, PhD, CCC-SLP
Graduate Program in Speech-Language Pathology
Department and Research Institute of Rehabilitation Medicine
Yonsei University College of Medicine
11-12 Yondaedongmun-gil, Seodaemun-gu
Seoul, 03721
Republic of Korea
Phone: 82 2 2228 3900
Email: h.kim@yonsei.ac.kr

Abstract

Background: Swallowing difficulties (ie, dysphagia) are common among older adults, with a 13% to 54% prevalence. Adequate interventions to improve the swallowing function of older adults would reduce morbidity and enhance health-related quality of life outcomes. Mobile health (mHealth) apps may help alleviate dysphagia symptoms by providing programs that maximize the intensity and frequency of training without requiring high costs or regular clinic visits.

Objective: The aim of this pilot study was to assess the usability of swallowing training apps by quantitatively and qualitatively evaluating older adults' self-reported data, taking into consideration their educational levels and exposure to mobile technology. We conducted surveys and brief interviews while the participants used a swallowing intervention app we developed. We subsequently identified and resolved individual-specific usability issues to improve future implementation of the app protocol for older persons with swallowing difficulties.

Methods: A total of 11 participants (10 women, 91%; mean age 75.7 years, SD 3.93) from two district-run senior welfare centers took part in this study. The participants were divided into a high-potential group and a low-potential group based on their total number of years of education and smart device usage. To investigate the usability of the app twice (ie, in the second week of the intervention and the postintervention stage), we used mixed methods consisting of both quantitative approaches, namely the System Usability Scale (SUS) and modified Computer Self-Efficacy Scale (mCSES) surveys, and qualitative approaches (ie, interviews).

Results: The quantitative results of the SUS and mCSES surveys revealed that the high-potential group was more inclined to adopt and learn new technology than the low-potential group. Specifically, within the high-potential group, a Wilcoxon signed-rank test indicated that the postintervention mCSES scores (median 65.50) were significantly higher than those in the second week of intervention (median 54.00; z=–2.023, P=.04). Additionally, the usability scores in the low-potential group were within the "marginal acceptability" range even after completion of an 8-week intervention program. Qualitative analyses via semi-structured interviews yielded promising outcomes regarding app acceptability, training program utilization, emotional responses, and learning experience.

Conclusions: To the best of the authors' knowledge, this usability and feasibility study is the first report of a swallowing training app designed to improve the swallowing function of older adults. Future research should consider several issues, such as user characteristics, pretraining education, and the intensity and innate characteristics of the intervention program.

(JMIR Mhealth Uhealth 2020;8(7):e19585) doi:10.2196/19585

KEYWORDS

older adults; dysphagia; swallowing; mHealth; thematic analysis; usability; apps; education; experience; sociodemographic

Introduction

Condition-focused health issues, such as swallowing difficulties (ie, dysphagia), are prevalent among older adults. Approximately 13% to 54% of older adults are reported to have swallowing difficulties, depending on their age, underlying diseases, and care level [1-3]. In a study on the prevalence of swallowing difficulties among older Koreans, the age- and sex-standardized prevalence of dysphagia in Koreans aged 65 years or older was 98/415 participants (23.6%) [4]. Major complaints resulting from dysphagia include coughing or choking [2,5], spillage of food [5], and difficulty swallowing hard foods [6]. Because dysphagia in older adults increases their risk of malnutrition, dehydration, weight loss, and aspiration pneumonia [7], it is imperative to detect dysphagia symptoms early and provide adequate preventative efforts to reduce morbidity and improve health-related quality of life outcomes.

Rehabilitative interventions include swallowing-focused therapies such as the Mendelsohn maneuver [8,9] and effortful swallowing [10]. Indirect methods to improve swallowing also exist, such as the effortful pitch glide (EPG) exercise [11] and tongue rotation exercise [12]. These training exercises are usually conducted via interactions between swallowing clinicians and patients during therapy sessions at various health care facilities. However, to induce experience-dependent neural plasticity, repetition and intensity of training are key factors in motor behaviors [13] and swallowing regimes [14]. While goal-setting training usually requires regular clinic visits, a few studies [15,16] have reported methods of maximizing effective therapeutic parameters for better outcomes in terms of duration, frequency, and intensity of training.

However, older adults may experience barriers that make regular clinic visits a challenge. Several studies have reported that the following factors are associated with the difficulty faced by older adults when visiting clinics: expenses, distance, transportation, time, and greater difficulties associated with advanced age [17-19]. For example, one of the reasons for unmet health care needs among older Korean adults is economic adversity [20], which prevents these older adults from attending regular and intensive therapeutic training. Accordingly, mobile health (mHealth) technologies have tremendous potential to bring health care into the digital age by providing more efficient health care for people with limited resources—especially older adults, who may be less capable of visiting clinics on their own.

mHealth technology can be utilized to monitor, control, or deliver training exercises. Pervasive health monitoring or control via mHealth apps has been developed for various medical conditions, such as diabetes, hypertension, and headache [21,22]; some of these apps have proved to be successful [23]. However, self-exercises incorporating rehabilitation expand far beyond simply monitoring and controlling, and they may require a higher degree of patient adherence. It has been reported that nonadherence to home exercise in rehabilitation programs is as high as 50%; thus, more rigorous mHealth exercise systems that make use of coaching, self-monitoring, and education are needed to achieve the desired goals [24].

Some apps are commercially available for improving swallowing function; however, very few of these apps make use of swallowing research [25,26]. mHealth apps may benefit individuals who are capable of achieving the maximum intensity and frequency of training programs at home but cannot afford high health care costs. Through these apps, individuals may be able to alleviate their dysphagia symptoms. However, the use of mHealth apps to improve swallowing function is not commonly practiced among older adults. Although older adults can benefit from mHealth to address both general and specific health issues [27], they may encounter usability issues due to the complexity and unfamiliarity of the apps [28]. These issues may be more substantial among older adults with limited formal education and little previous experience of high technology. Thus, it is reasonable to presume that use of and receptivity to technology varies depending on these sociodemographic characteristics [29].

Thus, in this pilot study, we assessed the usability and feasibility of a swallowing training app by quantitatively and qualitatively evaluating self-reported data from older adults, taking into consideration their educational levels and exposure to mobile technology. Specifically, we conducted surveys and brief interviews while the study participants used a swallowing intervention app we developed. We thereby identified and resolved individual-specific usability and feasibility issues to improve future implementation of the app protocol for older adults with swallowing difficulties. In the current study, the term “usability” was defined as the degree to which the participants used the mHealth app as intended by the researchers, and “feasibility” was defined as the degree to which the app effectively assisted older adults in improving their swallowing difficulties [30].

Methods

Apps and Hardware

The 365 Healthy Swallowing Coach app was developed for the intervention program. The app was designed with two fundamental characteristics. First, it was simple to use irrespective of the user’s familiarity with mobile apps because the target app users were older adults. Second, it allowed users to effectively complete the swallowing exercises without a clinician physically present because older adults are generally unfamiliar with technology and often face physical and cognitive barriers [31].

We were particularly concerned with how well the users would navigate the app, reach the training page, and follow the training protocol. If the app required users to remember a complex sequence of actions, older adults might experience difficulties

in navigating its functions [32]. Figure 1 provides an overview of the 365 Healthy Swallowing Coach app. In the first step, the user entered a user ID and password to log in. After logging in, the user was directed to the main page, Swallowing Training.

The Swallowing Training tab in the navigation menu enabled the users to access the Training Instruction, Training of the Day, and Training Record menus. The Training Instruction menu contained information regarding the training protocol, which was presented via animations and demonstration videos. Upon entering the Training of the Day menu, the user was shown three large tabs labeled Morning Training, Afternoon Training, and Evening Training. Each tab contained three exercises aimed at improving swallowing function: effortful prolonged swallow (EPS) [8-10], effortful pitch glide (EPG) [11], and effortful tongue rotation (ETR). [12] In our training protocol, users were required to complete 2 sets of 10 repetitions of each exercise in the morning, afternoon, and evening sessions (resulting in 60 repetitions of each exercise per day) on 5 days of the week of their choice. The app additionally included feedback system options, such as video demonstrations and a mirror function, real-time graphing, and audiovisual instructions, to monitor and correct the user's performance [33].

The Training Record screen (Figure 2) enables users to keep track of the extent to which they have completed their exercises on any given day. On the left, the level of completion of each exercise in each session is indicated by bars and percentages; on the right is a calendar in which users can choose a day to review. The overall rate of completeness is marked on the calendar by a circle around each specific date. The Training of the Day screen also provides feedback on the user's progress through a horizontal bar that "fills in" green as each training session is started and completed. When each training session is completed, the training data are automatically recorded through the automatic data-logging system and then saved in the database. The saved data can be extracted as comma-separated value (.csv) files when desired.

In addition, because limitations associated with aging may influence a user's interactions with apps [34], the current app used a design that was specialized for older adults [35]. The buttons on the app contained both icons and text for easier viewing; these buttons were structured so that clicking anywhere on them performed the target action. In addition, to prevent users from pressing unwanted buttons by mistake, the sizes of the buttons and the spaces between them were adjusted appropriately and the color of the buttons was differentiated from that of the background. In particular, the buttons on the edges of the app were placed so that there was sufficient space between them and the soft buttons (ie, the Back and Recent Apps buttons) to prevent users from inadvertently pressing the latter.

Figure 1. Overview of the 365 Healthy Swallowing Coach app. EPG: effortful pitch glide; EPS: effortful prolonged swallow; ETR: effortful tongue rotation.

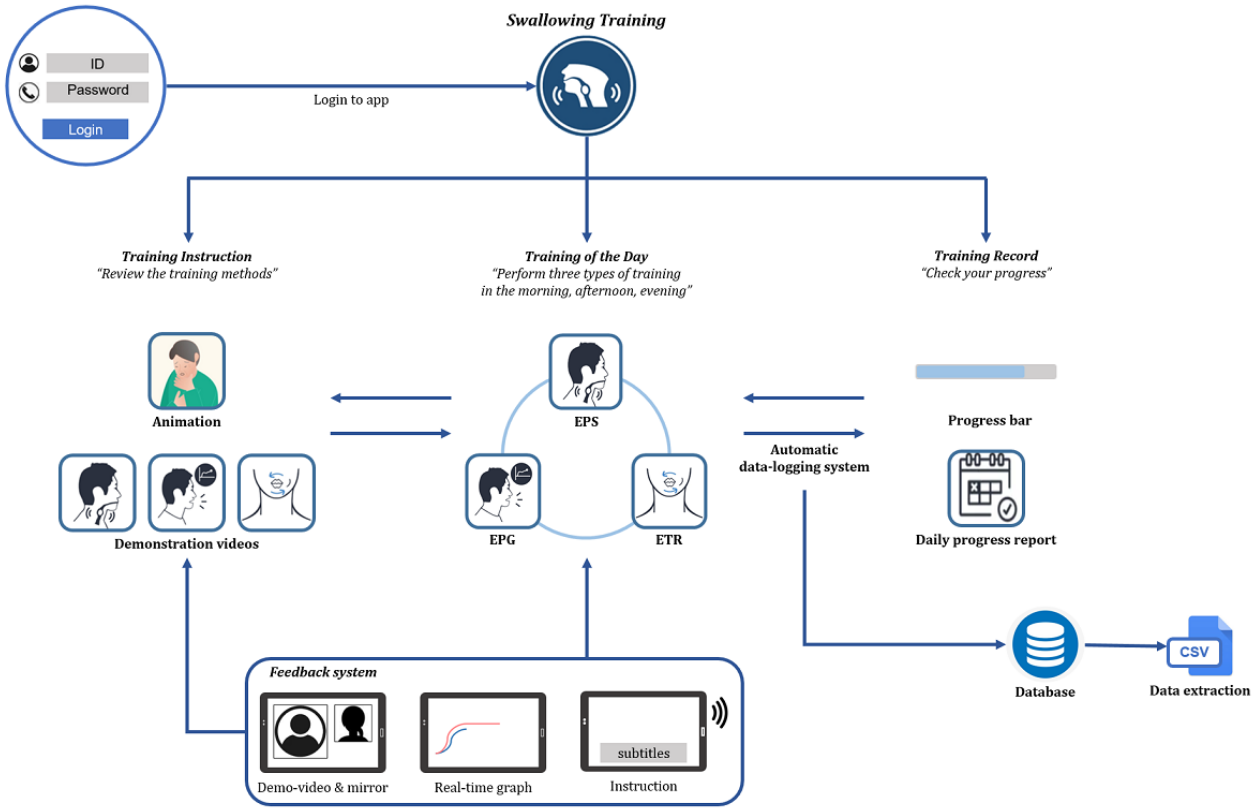
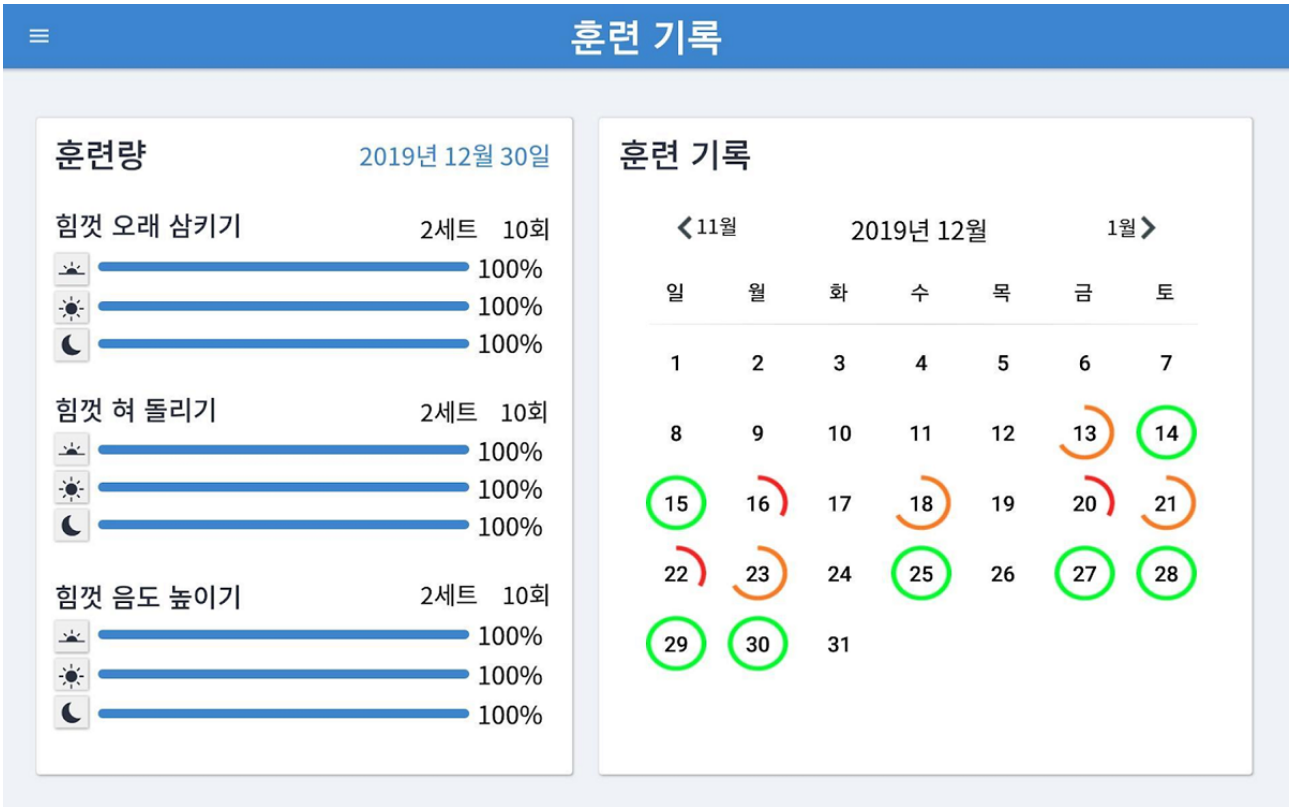


Figure 2. Screenshot of the Training Record screen of the 365 Healthy Swallowing Coach app.



The prototype of the 365 Healthy Swallowing Coach app was specially designed to be compatible with a specific tablet PC model to minimize unforeseen bugs or errors after the launch of the intervention program. The app was built for the Galaxy Tab A model no. SM-P580 (Samsung Corporation) with a 1.6 GHz octa-core processor running the Android Oreo (version 8.1.0) operating system. The tablet has an internal storage capacity of 32 GB and 3 GB of RAM. The dimensions of the tablet are 254.3×164.2×8.2 millimeters, with a weight of 554 grams. The diagonal of the display dimensions is 255.4 mm, with a screen resolution of 1920×1200 pixels.

Participants

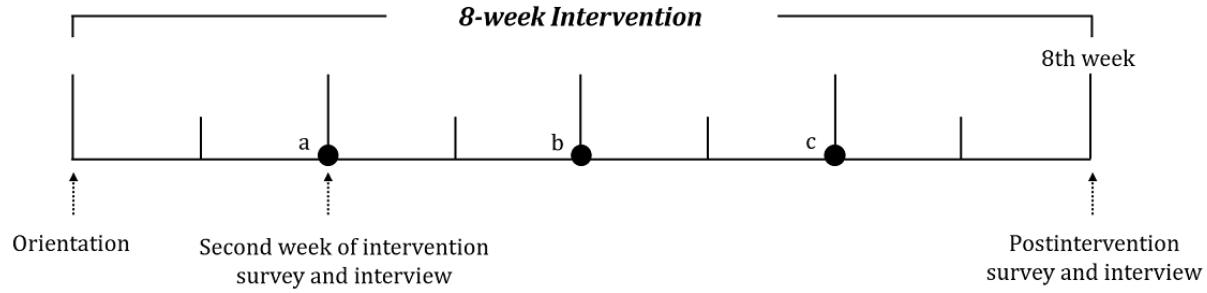
The participants were recruited from two district-run welfare centers between November 11, 2019 and November 20, 2019. An advertisement for the study was placed in the centers. The individuals voluntarily presented themselves to research staff and were provided with a brief face-to-face introduction to the intervention program (eg, requirements for participation and training methods) by the authors HK, SHL, and NBC. This study was approved by the Institutional Review Board of the Yonsei University Health System (No. 4-2019-0888) and complied with the Helsinki Declaration of 1974 as revised in 2000. Written consent was obtained from all participants in the study.

Several inclusion criteria were established for the study. Participants were required to be older than 65 years; report swallowing difficulties (eg, experience aspirations more frequently than in the past or feel as if something is stuck in the throat); meet age and education level criteria for the Korean version of the Mini-Mental State Examination (K-MMSE) [36]; and have normal vision, hearing, and motor function of the upper limbs so they could use a tablet PC for the swallowing intervention. Specific exclusion criteria were also established. Individuals who had experienced neurological disorders and who required non-oral feeding (eg, nasogastric and percutaneous endoscopic gastrostomy tubes) were excluded.

mHealth Intervention

The swallowing intervention in this study consisted of an 8-week program (Figure 3). The participants were asked to engage in three different types of training (ie, EPS, EPG, and ETR) 20 times during a single session. The program consisted of 120 sessions (5 days a week, 3 times a day) in total. The participants also took part in four face-to-face meetings with the researchers. The first meeting (ie, orientation), which was held prior to commencement of the intervention, was designed to guide participants to learn how to use the 365 Healthy Swallowing Coach app and the tablet PC. Subsequently, meetings were held every other week to monitor the training progress and adherence of the participants to the program.

Figure 3. Timeline of the usability study. a,b,c: biweekly face-to-face meeting.



Procedure

After written informed consent was obtained, demographic and socioeconomic status information was collected from the participants. Following this, they were asked if they had ever owned a mobile smart device, and if so, how long they had used or had been using it. Participants who reported having owned a mobile smart device were also asked if they had ever used a health-related app on their devices.

As shown in the usability study timeline (Figure 3), two short surveys were conducted in the second week of intervention and postintervention stages: the System Usability Scale (SUS) [37] and the modified Computer Self-Efficacy Scale (mCSES) [38], respectively. The SUS was conducted to gather information on the participants' subjective judgments of the usability of the app, and the mCSES was conducted to evaluate their perceived confidence in coping with new technology. We translated the two scales into Korean with written email permission from the original authors for academic/nonprofit research use.

In addition, in the second week of the intervention and in the postintervention phase, semistructured interviews were conducted, and the results were used as the key source of information for our thematic analysis. The second week interview took place two weeks after the initiation of intervention program. The postintervention interview was conducted upon completion of the program. Both interviews were digitally recorded and transcribed verbatim. To thematically analyze the transcribed text, the participants' statements were categorized into themes and subthemes depending on the context in which the responses were elicited.

System Usability Scale

The SUS [37] is a widely used 10-item usability measurement scale. Several studies have demonstrated that the SUS is valid (ie, face, concurrent, and convergent validity) and reliable and that it can be used on a broad range of participants [39-41]. The scale was described by its creator as "a quick and dirty usability scale." It is widely considered to be an effective usability assessment tool that can accommodate a variety of user interfaces [39]. The original SUS items are presented in Multimedia Appendix 1.

Responses to the SUS are rated on a 5-point scale from 1 to 5, with 1 corresponding to "strongly disagree" and 5 corresponding to "strongly agree." The scoring system used to determine the overall SUS score is based on alternating positive items (nos. 1, 3, 5, 7, and 9) and negative items (nos. 2, 4, 6, 8, and 10). Responses to the odd-numbered items are subtracted by 1 and summed; responses to the even-numbered items are subtracted from 5 and summed. The sum of each subscore multiplied by 2.5 yields the total SUS score [42]. An example of the survey is presented in Multimedia Appendix 2.

Modified Computer Self-Efficacy Scale

The original 10-item Computer Self-Efficacy Scale (CSES) was developed for workplace professionals by Compeau and Higgins [43] in 1995. Previous studies that used this or other questionnaires to assess perceived competence in using various types of technology have shown that higher self-efficacy is a crucial contributor to the likelihood of acceptance of the new technology. Laver et al [38] modified the original CSES (the mCSES) and tested its validity and reliability to determine whether it could be adapted to clinical rehabilitation settings for older and disabled populations. The results revealed high internal consistency, reliability, construct validity, and acceptance of the mCSES. The mCSES appeared to be suitable for our purpose of measuring the participants' perceived confidence in coping with new technology. We made slight changes to wording of the mCSES to make it specific to mobile apps, such as changing the words "product" and "technology" to "app." The mCSES items are presented in Multimedia Appendix 1.

The mCSES items are rated on a 10-point scale, ranging from 1 to 10, with 1 corresponding to "not confident at all" and 10 corresponding to "completely confident." To calculate the total mCSES score, the point values of all 10 items are summed. The maximum score is 100 points [38].

Interviews

The interviews were conducted by the authors HK, SHL, and NBC at 2 weeks and 8 weeks after the commencement of the intervention. At the time of the interviews, the first author (HK, female, PhD) was a university professor of a graduate program in speech-language pathology (SLP), with certificates in SLP (CCC-SLP, BC-ANCDS) and previous experience of conducting a qualitative study and a thematic analysis [44]. SHL and NBC (both male) were graduate students who conducted interviews under the supervision of the first author after interview training.

The items included in the first interview, which took place in the second week of the intervention program, were as follows: how adept participants believed they were at using the app; whether they experienced any difficulties when using the app,

and how they coped with these issues; what participants' opinions were regarding the strengths and weaknesses of the app; and whether the participants would like the app to be improved in any way. Depending on their responses, the participants were asked additional follow-up questions.

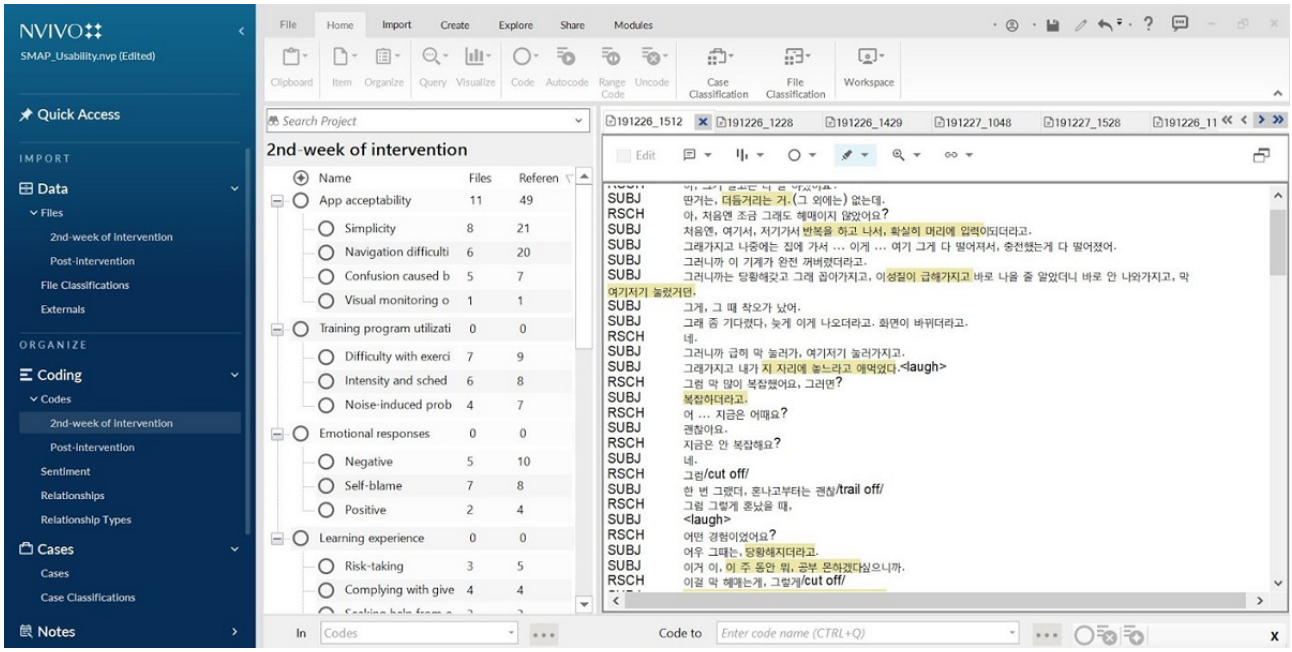
The second interview was conducted on the postintervention assessment day (ie, after 8 weeks of using the app). The second interview was identical to the first. The topic guide for the second week of the intervention and the postintervention is presented in Multimedia Appendix 3. The interviews were digitally recorded using an ICD-UX560F device (Sony Corporation) and transcribed verbatim for thematic analysis. The average duration of the interviews was 369 seconds (SD 141).

Data Coding and Thematic Analysis

To identify key issues regarding the participants' mHealth app usage, thematic analysis was used. We derived four major themes (ie, app acceptability, training program utilization, emotional responses, and learning experience) and multiple subthemes from the transcripts. The authors selected the aforementioned themes on the basis that the themes reflected patterns of responses in the interview data and were relevant to our research question [45], namely, "What factors affect the usability of swallowing training app for older adults?" The details of the themes are shown in Multimedia Appendix 4.

Data coding for the thematic analysis was conducted using NVivo software version 1.0 (QSR International) [46], which was used to code themes for the transcripts. Specifically, the first and second authors (HK and SHL) reviewed the interview data and selected specific phrases that were later assigned (using NVivo) to themes and subthemes according to their content. The selection of phrases and theme assignments were cross-checked by the three authors (HK, SHL, NBC). When disagreement occurred, the theme assigned by most of the authors was assigned to the utterance. The number of responses that corresponded to each theme was automatically calculated by the software. An example of the use of NVivo 1.0 is shown in Figure 4.

Figure 4. Example of the use of NVivo 1.0.



Results

Participants' Characteristics

Eleven older adults participated in the study; 10 (91%) were female, and the mean age was 75.7 years (SD 3.93, range 67-83). This number met the optimal sample size requirement for a usability assessment [47]. All 11 participants successfully completed the 8-week intervention program. As shown in Table 1, the mean number of years of formal education was 9.9 (SD 3.36, range 5-16). The mean score of the K-MMSE was 28.2 (SD 1.26, range 24-30). Among the 11 participants, 8 (73%) reported currently owning or having previously owned a mobile smart device (ie, a smartphone), with the mean number of years of usage being 3.5 (SD 2.31, range 0-9). Among these 8 participants, only 2 (25%) had used a health-related app, such as a pedometer app (ie, a step-counter app), which can track a person's number of steps and distance walked. The participants were divided into two groups based on their number of years of education and history of smart device usage (>10 years and ≤10 years, respectively): a high-potential group (n=6, mean age 74.3 years, SD 5.47; mean education 12.8 years, SD 3.00) and a low-potential group (n=5, mean age 77.4 years, SD 4.22; mean education 6.4 years, SD 1.52). There was statistical significance between the groups with regard to number of years of education (Mann-Whitney u=20.50, P=.004).

Table 1. Demographic characteristics of the participants (N=11).

Group and participants	Sex	Age (years; mean 75.7, SD 3.93)	Education (years; mean 9.9, SD 3.36)	K-MMSE score[a] (points; mean 28.2, SD 1.26)	Smart device usage (years; mean 3.5, SD 2.31)	Health app use
High-potential group (n=6)						
No. 1	F	77	12	27	3	No
No. 2	F	73	12	28	9	No
No. 3	F	83	9	24	5	No
No. 4	F	71	16	29	2	Yes
No. 5	M	75	16	28	8	Yes
No. 6	F	67	12	30	4	No
Mean (SD)	N/A[b]	74.3 (5.47)	12.8 (3.00)	27.7 (1.92)	5.2 (3.05)	N/A
Low-potential group (n=5)						
No. 7	F	78	6	29	0	N/A
No. 8	F	82	6	29	4	No
No. 9	F	71	9	29	0	N/A
No. 10	F	80	6	30	0	N/A
No. 11	F	76	5	27	3	No
Mean (SD)	N/A	77.4 (4.22)	6.4 (1.52)	28.8 (1.10)	1.4 (1.95)	N/A

[a]K-MMSE: Korean Mini-Mental State Examination, maximum score=30 points.

[b]N/A: not applicable.

SUS Scores

As shown in Tables 2 and 3, within the high-potential group, a Wilcoxon signed-rank test indicated that the postintervention SUS scores (median 71.25) were not significantly different from scores obtained from the survey in the second week of intervention (median 73.75; z=–0.271, P=.79). Likewise, within the low-potential group, the postintervention SUS scores (median 70.00) were not significantly different from those in the second week of the intervention (median 55.00; z=–1.761, P=.08).

Table 2. Scores of the System Usability Scale (SUS) from the second-week and postintervention surveys for the high-potential and low-potential groups (N=11).

Group and survey	Mean (SD)	Minimum	Maximum	Median
High-potential group (n=6)				
Second week of intervention	70.83 (8.75)	60.0	80.0	73.75
Postintervention	72.08 (7.31)	65.0	82.5	71.25
Low-potential group (n=5)				
Second week of intervention	56.00 (6.51)	50.0	65.0	55.00
Postintervention	67.50 (7.28)	57.5	75.0	70.00

Table 3. Ranks and test statistics for the System Usability Scale (SUS) scores from the second-week and postintervention surveys for the high-potential and low-potential groups (N=11).

Group and rank	n	Rank (postintervention – second week)		Test statistics[a] (postintervention – second week)	
		Mean rank	Sum of ranks	*z* value	*P* value
High-potential group (n=6)				–0.271[b]	.79
Negative rank[c]	2	3.25	6.50		
Positive rank[d]	3	2.83	8.50		
Tied[e]	1	N/A[f]	N/A		
Low-potential group (n=5)				1.761[b]	.08
Negative rank	1	1.00	1.00		
Positive rank	4	3.50	14.00		
Tied	0	N/A	N/A		

[a]Wilcoxon signed-rank test.

[b]Based on negative ranks.

[c]Postintervention score<score at second week of intervention.

[d]Postintervention score>score at second week of intervention.

[e]Postintervention score=score at second week of intervention.

[f]N/A: not applicable.

mCSES Scores

As indicated in Tables 4 and 5, within the high-potential group, a Wilcoxon signed-rank test indicated that the postintervention mCSES scores (median 65.50) were significantly higher than those obtained during the second week of the intervention (median 54.00; z=–2.023, P=.04). In contrast, within the low-potential group, the postintervention mCSES scores (median 66.00) were not significantly different from those obtained during the second week of the intervention (median 48.00; z=–1.761, P=.08).

Table 4. Scores of the modified Computer Self-Efficacy Scale (mCSES) from the second-week and postintervention surveys for the high-potential and low-potential groups (N=11).

Group and survey	Mean (SD)	Minimum	Maximum	Median
High-potential group (n=6)				
Second week of intervention	54.00 (17.28)	35.0	85.0	54.0
Postintervention	65.00 (19.43)	38.0	96.0	65.5
Low-potential group (n=5)				
Second week of intervention	44.44 (11.58)	29.0	57.0	48.0
Postintervention	59.20 (14.92)	40.0	75.0	66.0

Table 5. Ranks and test statistics for the Computer Self-Efficacy Scale (mCSES) scores from the second-week and postintervention surveys for the high-potential and low-potential groups (N=11).

Group and rank	n	Rank (postintervention – second week)		Test statistics[a] (postintervention – second week)	
		Mean rank	Sum of ranks	*z* value	*P* value
High-potential group (n=6)				–2.023[b]	.04
Negative rank[c]	0	.00	.00		
Positive rank[d]	5	3.00	15.00		
Tied[e]	1	N/A[f]	N/A		
Low-potential group (n=5)				1.761[b]	.08
Negative rank	1[c]	1.00	1.00		
Positive rank	4[d]	3.50	14.00		
Tied	0[e]	N/A	N/A		

[a]Wilcoxon signed-rank test.

[b]Based on negative ranks.

[c]Postintervention score<score at second week of intervention.

[d]Postintervention score>score at second week of intervention.

[e]Postintervention score=score at second week of intervention.

[f]N/A: not applicable.

Thematic Analysis of Interview Content

As shown in Table 6 and Multimedia Appendix 5, we devised and defined themes to incorporate all content extracted from the interviews while minimizing the conceptual overlap among key themes. From the first and second interviews, we extracted 107 and 132 responses, respectively. Multimedia Appendix 5 (participants' interview responses with 4 themes and 15 subthemes) presents the number of responses categorized within each theme and subtheme and the contrast in data between the high- and low-potential groups. We categorized the responses into four main themes; hereafter, the numbers in parentheses indicate the number of responses categorized into each theme in the first and second interviews, respectively. The themes selected were important app quality indicators according to the experiences of the participants: app acceptability (49; 51), training program utilization (24; 16), emotional responses (22; 37), and learning experience (12; 28). The theme of app acceptability, which pertained to different aspects of the app, was divided into four subthemes: simplicity (21; 17), navigation difficulties (20; 7), confusion caused by session selection (7; 17), and visual monitoring of exercise progress (1; 10).

Reports of navigation difficulties substantially decreased between the first and second interviews, possibly due to increased app experience. In contrast, the confusion caused by the session selection increased. This indicates that the design of the app and the location of the session selection buttons may have induced confusion among the older adult users when they selected the corresponding training sessions. Responses regarding the visual monitoring theme increased significantly between interviews, suggesting that the visual monitoring system of the app facilitated the self-monitoring process for the participants.

The theme of training program utilization encompassed the use of the training protocol and was divided into three subthemes: difficulty with exercises (9; 3), intensity and scheduling of the training protocol (8; 11), and noise-induced problems (7; 2). The participants reported experiencing less difficulty with exercises and noise-induced problems during the second interview than the first, which indicates that as the training program progressed, they were able to adapt to each of the training methods and cope with the noise-related issues mentioned during the first interview. In contrast, responses related to the intensity and scheduling of the training protocol were more frequent in the second interview. This suggests that the participants found that the intensity of the training protocol became more burdensome as they proceeded with the program.

The theme of emotional responses evoked by the use of the app was divided into three subthemes: negative responses (10; 9), self-blame (8; 17), and positive responses (4; 11). The incidence of self-blame and positive responses dramatically increased during the second interview, whereas the number of negative responses remained similar between interviews.

Finally, the theme of learning experiences was composed of five subthemes: risk-taking (5; 2), complying with given instructions (4; 5), seeking help from others (3; 5), progress in using the app (0; 12), and forgetfulness (0; 4). During the second interview, the users reported more incidences of complying with the given instructions and seeking help from others, while the incidence of risk-taking decreased. The three subthemes represent the way in which the older adults coped with issues regarding their experience with the app; the results suggest that the participants chose to adhere to the given instructions and ask for help instead of taking risks to solve problems as the program proceeded. The users also reported a significant

increase in progress using the app, which indicates that the users were becoming skilled in app use.

More responses were generated by members of the high-potential group (69; 78) than by members of the low-potential group (38; 54). In addition, the high-potential group reported better app acceptability in the second interview, as indicated by fewer navigation difficulties (13; 2) and greater perceived usefulness of visual monitoring of training progress (0; 9). In contrast, the low-potential group reported increased confusion caused by session selection, with an increase from 1 response in the first interview to 8 responses in the second interview. This suggests that at the time of the first interview, the low-potential group did not realize that they were not appropriately selecting the training sessions, and they only began to perceive the app as confusing once they realized that training session selection was something they needed to consider when following the exercise routines.

Regarding training program utilization, both groups reported less difficulty with exercises and fewer noise-induced problems in the second interview. In the high-potential group, the number of related responses decreased from 6 to 2, whereas in the low-potential group, the number of related responses decreased from 3 to 1. Interestingly, more high-potential participants reported the intensity and scheduling of the training program as being laborious in the second interview than in the first, as evidenced by an increase in the number of related comments from 4 to 11. Participants in the low-potential group expressed more self-blaming responses in the second interview than in the first (3; 9), although positive emotional responses also increased in frequency (2; 6). Finally, both groups demonstrated progress in app use competency after the 8-week intervention program.

Table 6. List of the 4 themes and 15 subthemes with examples of related statements.

Themes and subthemes	Example statements
App acceptability	
Simplicity	*"Nothing was particularly complicated, everything showed up once I launched [the app]."*
Navigation difficulties	*"I kept pushing this and that button but it still didn't work."*
Confusion caused by session selection	*"Even when I did all the sets, it doesn't say that I completed them. Why is that?"*
Visual monitoring of exercise progress	*"Because of the green indicators, I was able to see what I missed."*
Training program utilization	
Difficulty with exercises	*"The one where you have to say 'eee' was the most difficult . "*
Intensity and scheduling of the training protocol	*"Doing this three times a day is too much given my daily schedule."*
Noise-induced problems	*"I couldn't do [the pitch glide exercise] because I was concerned with alarming my neighbors."*
Emotional responses	
Negative	*"When [the app] didn't work as I wanted it to, I got annoyed."*
Self-blame	*"Old people like me need time to register things in the head."*
Positive	*"I was happy to learn something new."*
Learning experience	
Risk-taking	*"I touched the buttons here and there."*
Complying with given instructions	*"I was scared of getting lost, so I just stuck to what you taught me."*
Seeking help from others	*"I asked my grandson to help me."*
Progress in using the app	*"After figuring things out, it could not have been easier."*
Forgetfulness	*"When I go back home, I forget how."*

Discussion

Principal Findings

mHealth is being increasingly recognized as an effective approach to deliver health care in a more accessible and cost-effective manner [48]. To the best of the authors' knowledge, this usability and feasibility study is the first research on the use of swallowing training apps to improve the swallowing function of older adults. In this study, the home-based 365 Healthy Swallowing Coach app was used for 8 weeks by older adults. As a result, several issues emerged regarding the results of the analysis of the self-reported data.

First, user characteristics, such as the duration of formal education and smart device usage, were key variables related to usability, as reflected in the survey scores and interview data. Among the participants, we identified low-potential and high-potential groups based on their total years of education

and smart device usage. In the second week of the intervention period, the low-potential group demonstrated a mean SUS score that fell within the "low marginal acceptability" range [39]. Even after completion of the training program, the mean SUS score of the low-potential group did not reach the "acceptable" category, falling within the "high marginal acceptability" range. In contrast, mean SUS score in the high-potential group fell within the "acceptable" range in both the second-week and postintervention surveys. Without previous experience using swallowing apps, the high-potential group reported the app to be promising in terms of feasibility and user satisfaction.

It was also interesting to note that regardless of their educational attainment and exposure to smart device usage, by the second week of the intervention program, both groups were unlikely to adopt a new app unless they were offered help and assistance with using it. However, as evidenced by the mCSES scores of the second survey, the high-potential group later gained confidence in using the new app; meanwhile, the low-potential group lacked confidence in adopting a new app even after completion of the 8-week intervention program. Thus, the discomfort with technology felt by older adults with diverse degrees of education and smart device usage should be approached differently to reduce socioeconomic disparity. For example, more extensive education or orientations regarding the use of the app would facilitate the execution of the intervention program among older adults, especially those with lower education levels [49]. A study has shown that even young older adults (ie, those over 50 years of age) had difficulties with certain aspects of smartphone usage after training, even though they demonstrated significant improvement in competency [50]. This suggests that there is a need for individually tailored education programs that consider user characteristics, such as smartphone proficiency and education.

Second, as both groups continued to learn via the app, the learning experiences accumulated and resulted in mixed learning outcomes. In this respect, we gained valuable insights from the qualitative data collected in the two interviews. For example, more participants claimed to have benefited from the self-monitoring features of the app in the second interview versus the first, which is consistent with previously suggested strategies for increasing adherence to home exercise [24]. In addition, the frequency of comments related to app navigation difficulties decreased. However, the frequency of self-directed blame for negative experiences with the app also increased, especially among individuals in the low-potential group. In previous research [51], "old age" or "aging" in older adults was cited as a main cause of functional limitation by 20% of the 230 participants. Identifying specific sources of aging-related issues in individual users, such as visual and hearing impairment as well as memory decline [18], would reduce self-blame and negative reactions to app use. Moreover, some older adults are passionate about acquiring new forms of technological skills that can aid them in maintaining their independence as well as their quality of life [29]. This is similar to a previous study in which it was reported that more than 80% of community-dwelling older adults expressed interest in using health technology [52].

Third, it is of the utmost importance to find the optimal balance among training intensity, frequency, duration, and level of adherence to maximize the benefits of app use among older adults. This balance is necessary because a discrepancy between the users' perceived symptom severity and training intensity may harm their overall adherence [53]. Self-management and support while using apps without the presence of clinicians requires thorough scheduling and compliance with the schedule. Some participants in our study expressed dissatisfaction with the frequency of training; they claimed that three sessions a day (for a total of six sets of exercises per day) was excessive given that they had other plans or performed other activities during the day as well. Because patient compliance and adherence varies depending on the disease type and the participants' perception of disease severity [53], the participants may have felt that the program was excessively time-consuming because they had a lower degree of swallowing difficulty and thus may have felt less motivation [54]. The optimal number (ie, frequency) and intensity of training sessions needed for older adults to induce experience-dependent neural plasticity [13,14] is a major concern when planning therapy programs. One study reported that the implementation of intervention programs in a "distributed" manner (6 hours/week, 8 weeks) resulted in significantly greater improvements compared to an "intensive" manner (16 hours per week, 3 weeks) for patients with aphasia [15]. The 365 Healthy Swallowing Coach app requires 7.5 hours per week, which is comparable to "distributed" therapy. Because the intensity needed for "intensive" swallowing exercise protocols varies [55], further research is needed to elucidate the relationship between the amount of training and therapeutic effects.

Fourth, the innate characteristics of an intervention program can create difficulties in its actual usage. For example, almost all of the reported difficulties in training program utilization referred to the effortful pitch glide (EPG) program, as high-pitched voice production during the program is a source of noise; this could be problematic, especially during the evening. Currently, no training technique has been established as an appropriate alternative to EPG. However, if there were options that could replace EPG, choosing the most appropriate option for home app use would be beneficial to increase the feasibility of the app. It is generally agreed that unlike treatment in clinical settings, which allow for patient-focused approaches, home therapies necessitate considering inspections of other complex factors, including the patient's home environment and social context [56]. Therefore, until a substitute for EPG is found, a home environment assessment could be performed to determine whether to recommend a particular exercise.

Limitations and Future Directions

Although it is the first usability and feasibility study of a swallowing training program for older adults, this study is not without limitations. First, all participants of this study are residents of Seoul, which is the largest metropolis in South Korea. Thus, further study is warranted to determine the extent to which the findings will be applicable to older adults from more rural regions.

Second, this study included only individuals who were 65 years of age or older (range 67-83 years). The inclusion of a broad age range of participants, such as the "oldest-old" (ie, people who are aged at least 85 years) [57], who are more susceptible to swallowing difficulties and have less experience with high technology, would reveal additional usability and feasibility issues with at-home training apps.

Further, this study only included participants with mild swallowing difficulties, which also limits the generalizability of the findings. Investigations of patients at increased risk or with a chronic course of dysphagia and with special intervention needs would fill major research gaps.

Conclusions

Quantitative measurements may highlight the characteristics of different groups of users rather than indicate actual usability and feasibility problems. However, combined with qualitative data, we acquired insights into problems that older adults face when using mHealth technology as well as how a swallowing training app can be tailored to such persons. Qualitative investigation revealed that to provide an effective mHealth app to treat swallowing problems in older adults, the users' years of formal education and smart device usage must be taken into consideration even during the training phase of app use. Additionally, despite some early difficulties as they became familiar with using the app, more participants expressed comfort with app usage later in the intervention, highlighting the potential of mHealth apps for older adults. However, self-blaming behavior may be a limitation when testing usability with older adults as evaluators. We also note the importance of our users' home environment, intensity and adherence to the training program, and the nature of the training exercises in the usability and feasibility of our app.

Acknowledgments

This research was supported by a National Research Foundation of Korea (NRF) Grant funded by the Ministry of Science and ICT for convergent research in Development program over Science and Technology Liberal Arts (NRF-2017M3C1B6070665, NRF-2017M3C1B6070527). We would like to express our sincere gratitude to all the participants for their time and effort in participating in this research. We would also like to acknowledge the contributions of two senior community centers for their support in recruiting participants. We also wish to thank our graduate students at the Graduate Program in Speech-Language Pathology of Yonsei University, Ms Kyung Min Kim and Ms Minji Kang, for their assistance with editing the manuscript.

Authors' Contributions

All authors made substantial contributions to this study. HK contributed to the app content and design; study concept and design; data acquisition, analysis, and interpretation; drafting and revision of the manuscript; and revision of the manuscript after the first peer review. SHL contributed to the data acquisition and analysis and to the drafting of the manuscript. NBC contributed to the app design, data acquisition and analysis, drafting of the manuscript, and revision of the manuscript after the first peer review. HY, TC, and JK contributed to the design and development of the app and the revision of the manuscript. All authors certify that they reviewed and approved the final version of the manuscript being submitted. The article is the authors' original work and is not under consideration for publication elsewhere.

Conflicts of Interest

None declared.

Multimedia Appendix 1
System Usability Scale (SUS) and modified Computer Self-Efficacy Scale (mCSES).
[PDF File (Adobe PDF File), 344 KB - mhealth_v8i7e19585_app1.pdf]

Multimedia Appendix 2
An example of the System Usability Scale (SUS) response sheet.
[PDF File (Adobe PDF File), 224 KB - mhealth_v8i7e19585_app2.pdf]

Multimedia Appendix 3
Interview topic guide.
[PDF File (Adobe PDF File), 33 KB - mhealth_v8i7e19585_app3.pdf]

Multimedia Appendix 4
Codes for thematic analysis.
[PDF File (Adobe PDF File), 39 KB - mhealth_v8i7e19585_app4.pdf]

Multimedia Appendix 5

Participants' interview responses with 4 themes and 15 subthemes.
[PDF File (Adobe PDF File), 75 KB - mhealth_v8i7e19585_app5.pdf]

References

1. Adkins C, Takakura W, Spiegel BMR, Lu M, Vera-Llonch M, Williams J, et al. Prevalence and Characteristics of Dysphagia Based on a Population-Based Survey. Clin Gastroenterol Hepatol 2019 Oct 24 [FREE Full text] [doi: 10.1016/j.cgh.2019.10.029] [Medline: 31669055]
2. Igarashi K, Kikutani T, Tamura F. Survey of suspected dysphagia prevalence in home-dwelling older people using the 10-Item Eating Assessment Tool (EAT-10). PLoS One 2019;14(1):e0211040 [FREE Full text] [doi: 10.1371/journal.pone.0211040] [Medline: 30673750]
3. Kawashima K, Motohashi Y, Fujishima I. Prevalence of dysphagia among community-dwelling elderly individuals as estimated using a questionnaire for dysphagia screening. Dysphagia 2004;19(4):266-271. [doi: 10.1007/s00455-004-0013-6] [Medline: 15667063]
4. Yang EJ, Kim KW, Lim J, Paik N. Relationship between dysphagia and mild cognitive impairment in a community-based elderly cohort: the Korean longitudinal study on health and aging. J Am Geriatr Soc 2014 Jan;62(1):40-46. [doi: 10.1111/jgs.12606] [Medline: 25180377]
5. Kim H, Lee H, Pyo HY, Kim JW, Choi SH, Choi H, et al. Concurrent Validity of the Swallowing Monitoring & Assessment Protocol for the Elderly. Commun Sci Disord 2019 Jun 30;24(2):507-517. [doi: 10.12963/csd.19599]
6. Nimmons D, Michou E, Jones M, Pendleton N, Horan M, Hamdy S. A Longitudinal Study of Symptoms of Oropharyngeal Dysphagia in an Elderly Community-Dwelling Population. Dysphagia 2016 Aug;31(4):560-566 [FREE Full text] [doi: 10.1007/s00455-016-9715-9] [Medline: 27307155]
7. Wirth R, Dziewas R, Beck AM, Clavé P, Hamdy S, Heppner HJ, et al. Oropharyngeal dysphagia in older persons - from pathophysiology to adequate intervention: a review and summary of an international expert meeting. Clin Interv Aging 2016;11:189-208 [FREE Full text] [doi: 10.2147/CIA.S97481] [Medline: 26966356]
8. McCullough GH, Kamarunas E, Mann GC, Schmidley JW, Robbins JA, Crary MA. Effects of Mendelsohn Maneuver on Measures of Swallowing Duration Post Stroke. Top Stroke Rehabil 2015 Jan 05;19(3):234-243. [doi: 10.1310/tsr1903-234]
9. Mendelsohn MS, McConnel FM. Function in the pharyngoesophageal segment. Laryngoscope 1987 Apr;97(4):483-489. [doi: 10.1288/00005537-198704000-00014] [Medline: 3561135]
10. Jang HJ, Leigh JH, Seo HG, Han TR, Oh BM. Effortful swallow enhances vertical hyolaryngeal movement and prolongs duration after maximal excursion. J Oral Rehabil 2015 Oct;42(10):765-773. [doi: 10.1111/joor.12312] [Medline: 26013277]
11. Miloro KV, Pearson WG, Langmore SE. Effortful pitch glide: a potential new exercise evaluated by dynamic MRI. J Speech Lang Hear Res 2014 Aug;57(4):1243-1250 [FREE Full text] [doi: 10.1044/2014_JSLHR-S-13-0168] [Medline: 24686494]
12. Arakawa I, Koide K, Takahashi M, Mizuhashi F. Effect of the tongue rotation exercise training on the oral functions in normal adults - Part 1 investigation of tongue pressure and labial closure strength. J Oral Rehabil 2015 Jun;42(6):407-413. [doi: 10.1111/joor.12271] [Medline: 25640889]
13. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008 Feb;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
14. Robbins J, Butler SG, Daniels SK, Diez Gross R, Langmore S, Lazarus CL, et al. Swallowing and dysphagia rehabilitation: translating principles of neural plasticity into clinically oriented evidence. J Speech Lang Hear Res 2008 Feb;51(1):S276-S300. [doi: 10.1044/1092-4388(2008/021)] [Medline: 18230851]
15. Dignam J, Copland D, McKinnon E, Burfein P, O'Brien K, Farrell A, et al. Intensive Versus Distributed Aphasia Therapy: A Nonrandomized, Parallel-Group, Dosage-Controlled Study. Stroke 2015 Aug;46(8):2206-2211. [doi: 10.1161/STROKEAHA.115.009522] [Medline: 26106114]
16. O'Brien SR, Zhang N. Association Between Therapy Intensity and Discharge Outcomes in Aged Medicare Skilled Nursing Facilities Admissions. Arch Phys Med Rehabil 2018 Jan;99(1):107-115 [FREE Full text] [doi: 10.1016/j.apmr.2017.07.012] [Medline: 28860096]
17. Fitzpatrick AL, Powe NR, Cooper LS, Ives DG, Robbins JA. Barriers to health care access among the elderly and who perceives them. Am J Public Health 2004 Oct;94(10):1788-1794. [doi: 10.2105/ajph.94.10.1788] [Medline: 15451751]
18. Kim H, Park E, Lee S, Kim M, Park EJ, Hong S. Self-Management of Chronic Diseases Among Older Korean Adults: An mHealth Training, Protocol, and Feasibility Study. JMIR Mhealth Uhealth 2018 Jul 29;6(6):e147 [FREE Full text] [doi: 10.2196/mhealth.9988] [Medline: 29959109]
19. Murata C, Yamada T, Chen C, Ojima T, Hirai H, Kondo K. Barriers to health care among the elderly in Japan. Int J Environ Res Public Health 2010 Apr;7(4):1330-1341 [FREE Full text] [doi: 10.3390/ijerph7041330] [Medline: 20617033]
20. Kim Y, Lee J, Moon Y, Kim KJ, Lee K, Choi J, et al. Unmet healthcare needs of elderly people in Korea. BMC Geriatr 2018 Apr 20;18(1):98 [FREE Full text] [doi: 10.1186/s12877-018-0786-3] [Medline: 29678164]
21. Lewinski AA, Patel UD, Diamantidis CJ, Oakes M, Baloch K, Crowley MJ, et al. Addressing Diabetes and Poorly Controlled Hypertension: Pragmatic mHealth Self-Management Intervention. J Med Internet Res 2019 Apr 09;21(4):e12541 [FREE Full text] [doi: 10.2196/12541] [Medline: 30964439]

22. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
23. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
24. Argent R, Daly A, Caulfield B. Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence? JMIR Mhealth Uhealth 2018 Mar 01;6(3):e47 [FREE Full text] [doi: 10.2196/mhealth.8518] [Medline: 29496655]
25. Constantinescu G, Loewen I, King B, Brodt C, Hodgetts W, Rieger J. Designing a Mobile Health App for Patients With Dysphagia Following Head and Neck Cancer: A Qualitative Study. JMIR Rehabil Assist Technol 2017 Mar 24;4(1):e3 [FREE Full text] [doi: 10.2196/rehab.6319] [Medline: 28582245]
26. Constantinescu G, Kuffel K, King B, Hodgetts W, Rieger J. Usability testing of an mHealth device for swallowing therapy in head and neck cancer survivors. Health Informatics J 2019 Dec;25(4):1373-1382. [doi: 10.1177/1460458218766574] [Medline: 29618274]
27. Morey SA, Stuck RE, Chong AW, Barg-Walkow LH, Mitzner TL, Rogers WA. Mobile Health Apps: Improving Usability for Older Adult Users. Ergon Des 2019 Apr 01;27(4):4-13. [doi: 10.1177/1064804619840731]
28. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review protocol. BMJ Open 2020 Mar 16;10(3):e033870 [FREE Full text] [doi: 10.1136/bmjopen-2019-033870] [Medline: 32184309]
29. Heinz M, Martin P, Margrett JA, Yearns M, Franke W, Yang H, et al. Perceptions of technology among older adults. J Gerontol Nurs 2013 Jan;39(1):42-51. [doi: 10.3928/00989134-20121204-04] [Medline: 23244061]
30. Monitoring and evaluating digital health interventions. World Health Organization. 2016 Dec 13. URL: https://www.who.int/reproductivehealth/publications/mhealth/digital-health-interventions/en/ [accessed 2020-06-04]
31. Lee C, Coughlin JF. PERSPECTIVE: Older Adults' Adoption of Technology: An Integrated Approach to Identifying Determinants and Barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
32. Wildenbos GA, Jaspers MWM, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: Using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Apr;124:68-77. [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
33. Völter C, Schirmer C, Hinsen D, Roeber M, Dazert S, Bilda K. Therapist-Guided Telerehabilitation for Adult Cochlear Implant Users: Developmental and Feasibility Study. JMIR Rehabil Assist Technol 2020 May 28;7(1):e15843 [FREE Full text] [doi: 10.2196/15843] [Medline: 32255434]
34. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging 2018 Nov 20;1(2):e11569 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]
35. Barros ACD, Leitão R, Ribeiro J. Design and Evaluation of a Mobile User Interface for Older Adults: Navigation, Interaction and Visual Design Recommendations. Procedia Computer Science 2014;27:369-378. [doi: 10.1016/j.procs.2014.02.041]
36. Kang Y. A Normative Study of the Korean-Mini Mental State Examination (K-MMSE) in the Elderly. Korean J Psychol 2006 Jun;25:1-12.
37. Brooke J. SUS - a quick and dirty usability scale. In: Jordan W, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. London: CRC Press; Jun 11, 1996:189-194.
38. Laver K, George S, Ratcliffe J, Crotty M. Measuring technology self efficacy: reliability and construct validity of a modified computer self efficacy scale in a clinical rehabilitation setting. Disabil Rehabil 2012;34(3):220-227. [doi: 10.3109/09638288.2011.593682] [Medline: 21958357]
39. Bangor A, Kortum P, Miller J. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
40. Mol M, van Schaik A, Dozeman E, Ruwaard J, Vis C, Ebert DD, et al. Dimensionality of the system usability scale among professionals using internet-based interventions for depression: a confirmatory factor analysis. BMC Psychiatry 2020 May 12;20(1):218 [FREE Full text] [doi: 10.1186/s12888-020-02627-8] [Medline: 32398111]
41. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.
42. System Usability Scale (SUS) Plus. usabiliTEST. URL: https://www.usabilitest.com/system-usability-scale [accessed 2020-06-01]
43. Compeau DR, Higgins CA. Computer Self-Efficacy: Development of a Measure and Initial Test. Manag Inf Syst Q 1995 Jun;19(2):189 [FREE Full text] [doi: 10.2307/249688]
44. Kim S, Kim H, Nam S, Kim H. A Thematic Analysis of Speech Self-Awareness in the Community-Dwelling Elderly. Commun Sci Disord 2020 Mar 31;25(1):156-168. [doi: 10.12963/csd.19585]
45. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. NVivo. Discover the Power of NVivo Today. URL: https://www.qsrinternational.com/nvivo/trial [accessed 2020-06-01]

47. Hwang W, Salvendy G. Number of people required for usability evaluation. Commun. ACM 2010 May;53(5):130-133. [doi: 10.1145/1735223.1735255]
48. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
49. Gatti FM, Brivio E, Galimberti C. "The future is ours too": A training process to enable the learning perception and increase self-efficacy in the use of tablets in the elderly. Educational Gerontology 2017 Jan 09;43(4):209-224. [doi: 10.1080/03601277.2017.1279952]
50. Lu S, Wen T, Chang P. The Study of Smartphone Usage Competency Assessment and Training for the Elderly. Stud Health Technol Inform 2017;245:161-165. [Medline: 29295073]
51. Williamson JD, Fried LP. Characterization of older adults who attribute functional decrements to "old age". J Am Geriatr Soc 1996 Dec;44(12):1429-1434. [doi: 10.1111/j.1532-5415.1996.tb04066.x] [Medline: 8951311]
52. Wang J, Du Y, Coleman D, Peck M, Myneni S, Kang H, et al. Mobile and Connected Health Technology Needs for Older Adults Aging in Place: Cross-Sectional Survey Study. JMIR Aging 2019 May 15;2(1):e13864 [FREE Full text] [doi: 10.2196/13864] [Medline: 31518283]
53. DiMatteo MR, Haskard KB, Williams SL. Health beliefs, disease severity, and patient adherence: a meta-analysis. Med Care 2007 Jun;45(6):521-528. [doi: 10.1097/MLR.0b013e318032937e] [Medline: 17515779]
54. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: A scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
55. Balou M, Herzberg EG, Kamelhar D, Molfenter SM. An intensive swallowing exercise protocol for improving swallowing physiology in older adults with radiographically confirmed dysphagia. Clin Interv Aging 2019;14:283-288 [FREE Full text] [doi: 10.2147/CIA.S194723] [Medline: 30804667]
56. Chen Y, Abel KT, Janecek JT, Chen Y, Zheng K, Cramer SC. Home-based technologies for stroke rehabilitation: A systematic review. Int J Med Inform 2019 Mar;123:11-22 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.12.001] [Medline: 30654899]
57. Mehra S, Visser B, Cila N, van den Helder J, Engelbert RH, Weijs PJ, et al. Supporting Older Adults in Exercising With a Tablet: A Usability Study. JMIR Hum Factors 2019 Feb 01;6(1):e11598. [doi: 10.2196/11598]

Abbreviations

CSES: computer self-efficacy scale
csv: comma-separated value
EPG: effortful pitch glide
EPS: effortful prolonged swallow
ETR: effortful tongue rotation
K-MMSE: Korean Mini-Mental State Examination
mCSES: modified Computer Self-Efficacy Scale
mHealth: mobile health
SLP: speech-language pathology
SUS: System Usability Scale

Edited by G Eysenbach; submitted 02.05.20; peer-reviewed by C Cheng, T Alessa, R Argent; comments to author 25.05.20; revised version received 11.06.20; accepted 14.06.20; published 27.07.20.

Please cite as:
Kim H, Lee SH, Cho NB, You H, Choi T, Kim J
User-Dependent Usability and Feasibility of a Swallowing Training mHealth App for Older Adults: Mixed Methods Pilot Study
JMIR Mhealth Uhealth 2020;8(7):e19585
URL: http://mhealth.jmir.org/2020/7/e19585/
doi:10.2196/19585
PMID:32663161

Original Paper

The Cedar Project - Mobile Phone Use and Acceptability of Mobile Health Among Young Indigenous People Who Have Used Drugs in British Columbia, Canada: Mixed Methods Exploratory Study

Kate Jongbloed[1], PhD; Margo E Pearce[1], PhD; Vicky Thomas[2,3]; Richa Sharma[1], BA, MA; Sherri Pooyak[4,5], MSW; Lou Demerais[6,7]; Richard T Lester[8], MD, FRCPC; Martin T Schechter[1], OBC, MD, PhD, FRSC, FCAHS; Patricia M Spittal[1], PhD; The Cedar Project Partnership[2,7]

[1]School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[2]The Cedar Project, Prince George, BC, Canada

[3]Wuikinuxv Nation, Prince George, BC, Canada

[4]Aboriginal HIV/AIDS Community-Based Research Collaborative Centre, Victoria, BC, Canada

[5]Cree, Victoria, BC, Canada

[6]Cree & Métis, Surrey, BC, Canada

[7]The Cedar Project, Vancouver, BC, Canada

[8]Division of Infectious Diseases, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Patricia M Spittal, PhD
School of Population and Public Health
Faculty of Medicine
University of British Columbia
2206 East Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 604 875 2345 ext 5944
Email: spittal@sm.hivnet.ubc.ca

Abstract

Background: Indigenous leaders continue to be concerned about high rates of HIV and barriers to HIV treatment among young Indigenous people involved in substance use. Growing evidence suggests that using mobile phones for health (mHealth) may be a powerful way to support connection with health services, including HIV prevention and treatment.

Objective: This study examined the patterns of mobile phone ownership and use among young Indigenous people who have used drugs living with or vulnerable to HIV and explored the acceptability of mHealth to support access to health care in this population.

Methods: The Cedar Project is a cohort study involving young Indigenous people who have used drugs in Vancouver and Prince George, British Columbia. This mixed methods exploratory study involved 131 Cedar Project participants enrolled in our WelTel mHealth program. At enrollment, participants completed a questionnaire related to mobile phone use and interest in mHealth. Data were linked to Cedar Project questionnaires and serodata. We present comparative statistics (quantitative) and results of a rapid thematic analysis (qualitative) related to mobile phone patterns and interest in receiving mHealth.

Results: Less than half of the participants (59/130; 45.4%) reported owning a phone. Among those with a phone, the majority owned a smartphone (46/59; 78%). Most participants with a phone reported having an unlimited texting plan (39/55; 71%), using the internet on their phone (44/59; 75%), and texting daily (44/55; 80%). A majority reported that using a mobile phone for health would be invaluable (120/130; 92.3%). There were no differences in mHealth acceptance between participants who owned a phone and those who did not (P>.99). All but one participant living with HIV felt using a mobile phone would be helpful for their health, while a small proportion of HIV-negative participants remained unsure (1.9% vs 11.7%; P=.047). In response to open-ended questions asking why using a mobile phone may be helpful for health, participants identified a diverse set of anticipated benefits: (1) connection for emotional, mental, and spiritual support, (2) connection to family, (3) staying in touch and/or being

reachable, (4) overcoming current barriers to phone use, (5) convenience, privacy, and safety, and (6) access to health care and emergency services.

Conclusions: We observed high acceptance and interest in using mobile phone technology for health despite low rates of personal mobile phone connectivity among young Indigenous people who have used drugs living with and vulnerable to HIV in British Columbia, Canada. Mobile phones were viewed as a way to support connections and relationships that are seen as critical to health and well-being among young Indigenous people in this study. Findings may be useful for health care providers preparing to scale up mHealth programs to support HIV prevention and treatment in this population.

(JMIR Mhealth Uhealth 2020;8(7):e16783) doi:10.2196/16783

KEYWORDS

Indigenous; mobile health; mHealth; text messaging; substance use; HIV/AIDS

Introduction

Researchers have started to understand mobile phones as a necessity comparable to other utilities and argue that gaps in access may re-enforce or exacerbate other disparities, including those related to health [1,2]. However, marginalized groups may be excluded from access due to structural inequalities (including health, economic, and gender disparities) that create a "digital divide," or differential access to digital technologies [3]. Mobile phones have been identified as a critical resource for individuals who require regular contact with health care providers, social services, and social support [2,4,5]. Conversely, interruptions in phone connectivity may disrupt important health-related communication networks, such as contact with health care providers, social services, and social support [2].

Health care providers have begun to embrace the potential of widespread mobile phone usage by offering mobile health (mHealth) programs: mobile phone-based interventions that aim to improve health outcomes among clients experiencing a variety of health conditions. These mHealth interventions utilize mobile phone functions such as calling, texting, and/or smartphone apps, and may be used to provide reminders, information, or support. mHealth initiatives utilizing text messaging have been found to be successful in supporting engagement in health care for people living with HIV [6-14]. Such initiatives can facilitate real-time problem solving between patients and health care providers when medication, health, or other issues arise; remind clients to take medications, attend appointments, and take care of their health; and demonstrate that "somebody cares" [7,15,16]. More recently, mHealth initiatives have aimed to address substance use as well as optimize and expand treatment for substance use and HIV [17-22]. Notably, there is a paucity of evidence of mobile phone use for health and mHealth interventions for Indigenous people living with or at risk of HIV, including those using drugs [10].

Indigenous scholars and leaders continue to be concerned about high rates of HIV and other harm among Indigenous young people who use drugs resulting from the ongoing impacts of colonization [23,24]. Substance use, poverty, barriers to health care access, and limited mobile phone ownership may be mutually reinforcing, leading to further marginalization from care and services. Despite the potential to connect young Indigenous people who have used drugs to health services, little is known about mobile phone access and acceptability of mHealth programs for this key population. This study (1) examines patterns of mobile phone ownership and use among young Indigenous people who have used illicit drugs in British Columbia and (2) explores the acceptability of mHealth for this population.

Methods

Study Design and Setting

This is a mixed methods cross-sectional analysis involving young Indigenous people who have used drugs participating in the Cedar Project WelTel mHealth Study. The study took place in inner-city settings of two cities in British Columbia, Canada. Vancouver is a large city in southern British Columbia on Coast Salish territory. In 2016, nearly 14,000 Indigenous people lived in Vancouver, accounting for 2.2% of the population [25]. Prince George is a forestry and mining town in British Columbia's Northern Interior on the traditional territory of the Lheidli T'enneh people. Just over 11,000 Indigenous people lived in Prince George in 2016, accounting for 15.4% of the population [26]. Both cities are home to large "away from home" Indigenous populations with people from nations and territories across the province and country.

Participants

The Cedar Project is a cohort involving 782 young Indigenous people who have used drugs in Vancouver and Prince George, British Columbia. The term Indigenous is used as Cedar Project participants represent many of the diverse First Nations, Inuit, and Métis communities across Canada and often live far away from their home communities. Methods have previously been described in detail [27]. Briefly, participants were recruited through health care providers, street outreach, and word of mouth. Initial recruitment took place between 2003 and 2007 and reopened in 2011. Participants were eligible if they self-identified as Indigenous, were between 14 and 30 years of age, and had smoked or injected illicit drugs (other than cannabis) in the month before enrollment. Drug use was confirmed using saliva screens (ORAL-screen, Avitar Onsite Diagnostics). Follow-up interviews were carried out every 6 months, and blood samples were collected for HIV and hepatitis C antibody tests.

In September 2014, the Cedar Project WelTel mHealth Study was initiated. The mHealth program consisted of a structured mobile phone initiative to connect young Indigenous people who have used drugs with Cedar Case Managers in a

community-based setting. It included a package of supports, including a mobile phone and cellular plan, alongside weekly two-way text messaging and support from Cedar Case Managers. Each Monday at noon, a text message saying, "how's it going?" was automatically sent to participants through the WelTel mHealth platform. Cedar Case Managers responded to all participants and followed up with participants who replied with a specific problem or need. On Wednesday, those who had not replied received an additional text saying, "Haven't heard from you, are you ok?" On Thursday or Friday, Case Managers attempted to call all remaining participants who had not responded to the text message. The program was offered between September 2014 and January 2016. Of the 60 HIV-positive participants in the Cedar Project Blanket Case Management study, 52 (88.3%) agreed to participate in the mHealth study. In addition, with the aim of recruiting 94 HIV-negative Cedar Project participants, 131 were randomly selected to be invited, of whom 79 (78.7%) agreed to join. Thus, 131 Cedar Project participants were enrolled in the mHealth study and provided a mobile phone and plan, weekly text messaging, and connection to a Case Manager. They also continued with their regular visits to the main Cedar Project cohort study.

Data Sources

At enrollment into the mHealth study, participants completed a short questionnaire on mobile phone use, which is the focus of the analysis presented here. Close- and open-ended questions were related to interest in and concerns about using mobile phones and text messaging for health. Data were linked with questionnaires and serodata collected every 6 months as part of the main Cedar cohort. Time-independent variables were obtained from baseline questionnaires and time-dependent variables were collected from the follow-up visit that occurred closest to (but <30 days after) the mHealth baseline visit.

Analytical Approach

We conducted descriptive analyses related to phone ownership and patterns, and participants' interest and concerns using mobile phones for health. Differences in characteristics and acceptability of mHealth by phone ownership and HIV status were compared using the Chi-square and Fisher exact tests (dichotomous variables), and *t* tests and Mann-Whitney-Wilcoxon Test (continuous variables). All *P* values are 2-sided. Analyses were performed using R version 3.5.0 (R Foundation for Statistical Computing) [28].

Short responses (1-40 words) to open-ended questions were recorded verbatim. Using a rapid qualitative analysis approach [29], two authors independently read and reread responses to identify recurring themes. Emerging themes were discussed, defined, and a coding manual was created. Responses were sorted into categories using the NVivo software Version 10 [30]. Representative responses were chosen to highlight the themes.

Ethical Considerations

The Cedar Project follows the guidelines provided in the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans—Chapter Nine: Research involving the First Nations Inuit and Métis Peoples of Canada [31] and adheres to the principles of Ownership, Control, Access, and Possession [32]. The Cedar Project Partnership, an independent body of Indigenous Elders, leaders, and health experts, governed all aspects of this study. The study was approved by the University of British Columbia/Providence Health Care Research Ethics Board. All participants gave both verbal and written consent, and it was emphasized that deciding not to participate in this substudy would not affect continued involvement with Cedar or support from staff.

Results

Baseline Characteristics

More than half (81/131, 61.8%) of mHealth participants were women, and half of mHealth participants lived in Prince George (65/131, 49.6%; Table 1). Approximately half (63/130, 48.6%) had a parent who attended residential school, and substantial proportions had been apprehended from their parents (100/131, 76.3%) and/or experienced childhood sexual abuse (70/123, 56.9%). Among parents, 53.5% (61/114) reported that they had ever had a child apprehended. Fewer participants reported connection to Indigenous cultures either in the past or present, including having a traditional language spoken at home (54/130, 41.5%), speaking a traditional language (53/130, 40.8%), often/always speaking a traditional language today (4/130, 3.1%); participating in ceremony (30/130, 23.1%); recently accessing traditional food (63/130, 48.5%); and living by traditional culture (22/127, 17.3%). Recent involvement in sex work was reported by 16.9% (15/89) participants, and recent injection drug use was reported by 44.2% (57/129) participants. Overall, 60.3% (79/131) and 40.5% (53/131) participants were living with hepatitis C and HIV, respectively.

Table 1. Characteristics of young Indigenous people who have used drugs enrolled in the Cedar Project WelTel mHealth Program by phone ownership.

Characteristic	Total (n=131)		No phone (n=71)		Own a phone (n=59)		*P*[a] value
	N	n (%)	N	n (%)	N	n (%)	
Demographics							
Age (years), median (IQR)	131	33 (30-36)	71	34 (31-36)	59	33 (29-36)	.26
Sex (female)	131	81 (61.8)	71	41 (57.7)	59	39 (66.1)	.43
Location (Prince George)	131	65 (49.6)	71	29 (40.8)	59	35 (59.3)	.05
In a relationship	129	33 (25.6)	70	19 (27.1)	58	14 (24.1)	.85
Sexual identity (LGBTQ[b])	131	23 (17.6)	71	7 (9.9)	59	16 (27.1)	*.02*
Education (Did not graduate high school)	129	106 (82.2)	69	58 (84.1)	49	47 (79.7)	.68
Recent[c] homelessness	131	34 (26.0)	71	17 (23.9)	59	17 (28.8)	.67
Recent housing instability	126	56 (44.4)	69	32 (46.4)	56	24 (42.9)	.83
Recent incarceration	128	19 (14.8)	69	13 (18.8)	58	6 (10.3)	.28
Cultural Connection & Resilience							
Traditional language spoken often at home growing up	130	54 (41.5)	71	34 (47.9)	58	19 (32.8)	.12
Speak traditional language (*yes or a bit*)	130	53 (40.8)	71	32 (45.1)	58	20 (34.5)	.30
Often/always speak traditional language today	130	4 (3.1)	71	2 (2.8)	58	1 (1.7)	>.99
Ever participated in traditional ceremonies	130	30 (23.1)	71	12 (16.9)	58	17 (29.3)	.14
Often or always live by traditional culture	127	22 (17.3)	70	12 (17.1)	56	9 (16.1)	>.99
Recent access to traditional food	130	63 (48.5)	71	30 (42.3)	58	32 (55.2)	.20
Resilience, mean (SD)	122	63.37 (21.35)	65	61.37 (20.05)	57	65.19 (22.75)	.33
Trauma							
Either parent at residential school	130	63 (48.5)	71	39 (54.9)	58	24 (41.4)	.18
Apprehended from biological parents	131	100 (76.3)	71	54 (76.1)	59	46 (78.0)	.96
Childhood sexual abuse (≤13)	123	70 (56.9)	66	30 (45.5)	56	39 (69.6)	*.01*
Ever attempted suicide	129	44 (34.1)	69	22 (31.9)	59	21 (35.6)	.80
Ever had a child apprehended[d]	114	61 (53.5)	61	28 (45.9)	52	33 (63.5)	.09
Sexual vulnerability							
Recent sex work[e]	15/89	16.9%	44	8 (18.2)	44	7 (15.9)	>.99
Recent sexual assault	5/129	3.9%	70	2 (2.9)	58	3 (5.2)	.66
Substance use							
Recent injection drug use	129	57 (44.2)	69	34 (49.3)	59	23 (39.0)	.32
Ever overdosed	130	52 (40.0)	71	29 (40.8)	58	23 (39.7)	>.99
Recent alcohol/drug treatment	130	51 (39.2)	71	29 (40.8)	58	22 (37.9)	.88
Current methadone treatment[f]	66	38 (57.6)	38	25 (65.8)	28	13 (46.4)	.19
Ever tried to quit drugs/alcohol	129	109 (84.5)	70	55 (78.6)	58	53 (91.4)	.08
Health outcomes							
HIV infection	131	53 (40.4)	71	35 (49.3)	59	18 (30.5)	*.046*
Hepatitis C virus infection	131	79 (60.3)	71	46 (64.8)	59	33 (55.9)	.40
Psychological distress, mean (SD)	131	1.00 (0.84)	71	1.04 (0.90)	59	0.98 (0.77)	.71
Recent hospitalization	131	15 (11.5)	71	5 (7.0)	59	10 (16.9)	.10

[a]*P* values indicated in italics are statistically significant.

[b]LGBTQ refers to lesbian, gay, bisexual, transgender, queer or questioning sexual identities.

[c]Recent refers to the 6-month period prior to the interview.

[d]Among a subset of people who were parents (n=114).

[e]Among a subset of people who said yes to having sex in the past 6 months (n=91).

[f]Among a subset of people who said yes to ever being on methadone (n=66).

Patterns of Mobile Phone Use

Slightly less than half (59/130, 45.4%) of the participants reported owning a phone at baseline. Among those, the majority owned a smartphone (46/59, 78%), had an unlimited texting plan (39/55, 71%), used the internet on their phone (44/59, 75%), and texted daily (44/55, 80%; Table 2). No differences in the patterns of mobile phone use were observed between men and women. Those who identified as lesbian, gay, bisexual, transgender, queer or questioning (LGBTQ), or reported experiencing childhood sexual abuse, were more likely at enrollment to own a phone, while people living with HIV were less likely to own one (Table 1).

Table 2. Baseline mobile phone use patterns among young Indigenous people who have used drugs who reported owning a phone at enrollment into the Cedar Project WelTel Mobile Health Study (N=59).

Mobile phone use pattern	Value, n (%)
Type of phone (n=59)	
Basic	13 (22)
Smart	46 (78)
Texting plan (n=59)	
Yes	53 (89)
No	6 (10)
Type of text plan (n=55)	
Pay as you text	12 (21)
Unlimited	39 (71)
Limited	2 (3)
Unsure	2 (3)
Access internet on phone (n=59)	
Yes	44 (75)
No	15 (25)
Frequency of texting (n=55)	
Never	3 (5)
Rarely (1× per month)	2 (3)
Occasionally (1× per week)	0 (0)
Frequently (2-3× per week)	6 (10)
Very frequently (daily)	44 (80)

Mobile Health Acceptance

Participants were asked whether they felt using a mobile phone would be helpful for health care and if they had any concerns using text messaging for their health (Table 3). A majority reported that using a mobile phone for health would be invaluable (120/130, 92.3%). There were no differences in mHealth acceptance among participants who owned a phone and those who did not. All but one participant living with HIV felt using a mobile phone would be helpful for their health, while some HIV-negative participants remained unsure (1.9% vs 11.7%; *P*=.047). No differences in concerns using text messaging for health were observed between those living versus not living with HIV.

Table 3. Self-reported mHealth acceptance stratified by phone ownership and HIV status (N=130).

Mobile health acceptance	Total (n=130), n (%)	Phone ownership			HIV status		
		Yes (n=59), n (%)	No (n=71), n (%)	*P* value	HIV+ (n=53), n (%)	HIV- (n=77), n (%)	*P* value
Do you think using a cell phone would help with your health care and be helpful to you?							**.047**
Yes	120 (92.3)	55 (93.2)	65 (91.5)	>.99	52 (98.1)	68 (88.3)	
No	0 (0)	0 (0)	0 (0)		0 (0	0 (0)	
Unsure	10 (7.7)	4 (6.8)	6 (8.5)		1 (1.9)	9 (11.7)	
Do you have any concerns about using text messaging for your health care?							**.45**
Yes	4 (3.1)	2 (3.4)	2 (2.8)	>.99	1 (1.9)	3 (3.9)	
No	125 (96.2)	57 (96.6)	68 (95.8)		51 (96.2)	74 (96.1)	
Unsure	1 (0.7)	0 (0)	1 (1.4)		1 (1.9)	0 (0)	

Benefits of Phone Use for Health

The survey included an open-ended question asking why using a mobile phone may be helpful for health. Nearly all participants (127/130, 97.7%) responded, suggesting a diverse set of anticipated benefits. Analysis of these responses revealed 6 themes: (1) connection for emotional, mental, and spiritual support, (2) connection to family, (3) staying in touch and/or being reachable, (4) overcoming current barriers to phone use, (5) convenience, privacy, and safety, and (6) access to health care and emergency services.

Connection for Support

Many participants (29/127, 22.8%) anticipated that a mobile phone would enable them to reach out or be contacted for emotional, mental, and spiritual support. Supporters included health professionals, "workers," spouses, friends, and family. One participant explained,

> *I am a social person…I constantly stay in contact with people so I don't get depressed.* [Female (F), Prince George (PG)]

Another said,

> *Having someone to talk to or someone to call. That helps.* [F, Vancouver (Van)]

A third said,

> *Communicating, letting people know how you are, what mood you're in…if they know you're in a bad mood they might call you.* [M, Van]

Having support reducing or abstaining from substance use was how 4.7% (6/127) participants anticipated using or currently used their phone for support. Participants felt that they could use the phone to call their 12-step/recovery sponsor, remember to take their methadone, access detox, and/or avoid a relapse. For example, one participant explained:

> *My mental health hasn't been all that good lately but having someone to talk/text with will be better than turning to the bottle or needle. I won't be so alone.* [F, Van]

Connection to Family

Connection to family was identified as an important potential health benefit by 8.7% (11/127) participants. One described how having a phone would facilitate in-person visits with family:

> *So I can get ahold of family more and probably go see them for Christmas!* [M, Van]

A total of 5.5% (7/127) participants anticipated benefits related to pregnancy and parenting, including being able to stay in touch with social workers, receiving health care leading up to and during labor, and calling for help in case of emergency.

Staying in Touch

Participants also spoke about the importance of "staying in touch" more broadly, suggesting that the benefit of a mobile phone was connection in and of itself. Some spoke about how a phone would help reduce feelings of isolation and allow them to, "keep connected with the world." One person described how,

> *If I don't have a phone I feel cut off, I get anxiety.* [F, Van]

Another said,

> *It's nice to have someone check up on you sometime.* [F, PG]

Eight (8/127, 6.3%) participants described the value of being more "reachable", as being unreachable was perceived as a source of stress for themselves and people around them. As one participant put it,

> *I don't have to stress—people can reach me. The phone is my lifeline.* [F, PG]

Overcoming Current Barriers to Phone Use

A number of participants spoke of barriers they currently faced with accessing a phone or making calls. Many who did not currently have a phone relied on borrowed phones, including those belonging to friends, in the lobby of their building, or available at local service agencies. Borrowing phones had drawbacks including time and energy it took to find a phone, time limits on phone calls, inability to receive messages or a call back, lack of privacy, and potential for stigma or disclosure. One participant explained the impact of time limits:

> *You have to get hold of people right away or it doesn't work out.* [F, Van]

Another described concerns when calling from service agencies, especially those known to serve people who use drugs:

> *If you call from [Vancouver's supervised injection facility], it has call display…then people know that you're a drug user.* [M, Van]

Convenience, Privacy, and Safety

Some participants (16/127, 12.6%) indicated that having a mobile phone would make communication more convenient, including reducing the need to seek support on foot and save you "the mileage on your shoes" [M, PG]. For 2.4% (3/127) participants, privacy was an important anticipated benefit of having their own phone. One explained that owning her own phone would provide more control over the circumstances of engagement with health care providers, whereas currently her partner was too involved:

> *I can talk to my doctor one-on-one without having [partner] follow me or know everything.* [F, PG]

In addition, some participants felt that having a phone would afford some degree of safety, including from violence in relationships and street life.

Access to Health Care and Emergency Services

More than 20% of participants expected to use the study phones, or currently used their own mobile phones, to connect directly to health care, including calling doctors, nurses, and/or counselors, arranging out-of-town medical care, making appointments, organizing rides, and receiving messages from health services. A total of 11.0% (14/127) participants also planned to use their phones to receive medication and appointment reminders, including reminder calls from health care providers or setting a reminder on the phone itself. Owning a mobile phone was also seen as a potentially important resource in the case of physical or mental health emergencies faced by the participants themselves or those around them.

Concerns About Using Texting for Health

Few participants (4/127, 3.1%) had concerns about using text messaging for health (Table 3). In total, 2.4% (3/127) participants spoke about anticipated challenges with mobile phone technology, including poor eyesight and low literacy, making it difficult to read and respond to text messages; 1.6% (2/127) participants reported fear that mobile phones may cause cancer. One participant raised issues about confidentiality, explaining that it may be hard to confirm,

> *that you're talking to the right person.* [M, Van]

However, others spoke about how confidentiality was not a concern because either they did not feel they had serious health concerns or they would take steps to protect their privacy. For example, 1 participant said,

> *No. Nobody's going to be looking at my phone. Even if they try there'll be a password.* [F, Van]

Discussion

Patterns of Mobile Phone Ownership

To our knowledge, this is the first study to report on patterns of mobile phone ownership and use among young Indigenous people who have used drugs in Canada. Less than half (45.4%) of the participants in this study owned a mobile phone. Observed phone ownership was considerably lower than rates in Canada and North America in general, and similar or lower compared with other marginalized groups, including street-involved youth (45%-63%), homeless adults (44%-78%), and people who use drugs (83%-86%) [33-38]. Many young Indigenous people have been acutely affected by colonization, including having parents and family members who were forced into residential schools, removal from family into the child welfare system, and experiences of childhood sexual abuse [39,40]. Some have turned to substance use as a way to cope with the effects of these historical and lifetime traumas [40-42]. Intersections of substance use and poverty can create barriers to connectedness, for example, by contributing to incarceration and housing transitions [43]. Lack of phone ownership must be understood within the context of colonization, which continues to impact the well-being of Indigenous people across Canada and is a key consideration for future mHealth programs [44].

Despite low phone connectivity, a majority of participants, including those living with HIV, felt that using a mobile phone for their health would be helpful. Mobile phones were viewed as a way to support social connections, which may reflect an Indigenous worldview that highlights connectivity, relationships, interconnectedness, and interdependence as critical to health and well-being [45-50]. This was demonstrated in participants' emphasis on the role of mobile phones as a way to re/connect with family, despite experiences of childhood trauma and, often, many years of separation [39-42]. Three-quarters (76%) of the participants had been taken from their parents into the child welfare system, and as a result, many lived far from their home communities. Yet, despite experiencing traumatic separations, family connections remain a powerful source of strength and resilience among young Indigenous people, as highlighted in other studies [51,52]. From a wholistic perspective of health and well-being, phones may provide an important way of connecting with family and loved ones, especially when living far from home.

The vital role of family in health and well-being has also been identified specifically among Indigenous people living with HIV [24]. Family relationships can be an important source of strength, a foundation of emotional support, and a motivator to stay healthy [53-55]. Family members, including partners, parents, and children often provide primary or first line support [55], helping to meet the logistical and emotional demands of dealing with diagnosis [56], initiating care [57], and adhering to medication [58]. Among Cedar participants living with HIV, access to mobile phones may enhance access to family networks who may provide social, material, and emotional support as they navigate complex health issues, systems, and treatment regimens [59].

Mobile phones were also seen as an opportunity to support health and well-being in the context of pregnancy and parenting. Previous Cedar research among young Indigenous mothers observed that being able to parent their children was key to participants' own wellness, while those whose children had been taken into care expressed feelings of deep regret and loss [41]. Participants in this study included pregnant mothers preparing for birth, parents of children currently in care of the state who were interested in visitation and regaining custody, and parents with custody of children who were navigating ongoing relationships with social workers as well as trying to protect their children's health and safety. Participants identified diverse ways in which being connected by phone would support carrying out their responsibilities as parents as well as nurturing their connections with children in care.

Implications of Phonelessness for Health and Wellness

Indigenous people who use drugs have frequently encountered systemic and interpersonal racism, stigma, and judgment within harm reduction and health services [23,54,60-62]. As noted by other researchers, participants in this study reflected that periods of "phonelessness," contribute to substantial burdens of time and energy required to remain in contact with services through in-person visits or borrowing a phone [2]. These efforts are not always successful and can result in missed follow-up calls, appointments, and test results [2]. Cedar participants anticipated that having a mobile phone would support access to health care, such as allowing them to speak directly with a provider (doctor, clinic, nurse, pharmacist, and counsellor), coordinate appointments, and visit logistics, or call 911 in an emergency. Participants also viewed mobile phones as potential tools for accessing emotional, mental, and spiritual support from a diverse group of care providers when experiencing crises related to unaddressed trauma, substance use, and/or mental health challenges. This may reflect a desire to access what Indigenous scholars have described as "relational care," which emphasizes connections and takes a "whole person" perspective of well-being [54].

The impact of phonelessness on overall health may be especially pronounced in the context of HIV care. Findings from our recent systematic review highlighted profound gaps in access to the HIV cascade of care among Indigenous peoples [24]. Emerging evidence also indicates that barriers to phone access may be associated with poorer health outcomes among people with HIV [38]. However, our study indicates that young Indigenous people living with HIV are especially interested in receiving mHealth support to enhance engagement with health care. Others have shown that mHealth programs may improve relationships between health care providers and patients over time [16,63], and that strong patient-provider relationships that are engaging, validating, and emphasize partnership are more likely to facilitate engagement and retention in HIV care [64]. For young Indigenous people who use drugs, mHealth programs that take a culturally safe approach, including avoiding judgment of drug use and honoring Indigenous identities, may help to strengthen relationships with health care providers and engagement in care [65,66].

A mobile phone may also be vital for the safety of young Indigenous people who use drugs, in the face of housing instability [43], police surveillance [29], transitions into injection [67], sexual assault [68], suicide [69,70], and overdose [70]. A qualitative study of 43 street-involved youth in Seattle observed that mobile phones have become intrinsic to safety [71]. Among 100 homeless people in Philadelphia, participants described how with access to a mobile phone, "help is a phone call away," especially in the context of threats to health and safety [35]. Cedar participants discussed how phones would support their safety by enhancing their privacy and enabling them to assist others by calling 911 in an emergency. Mobile phones may also be an important source of safety information for people who use drugs by allowing them to subscribe to relevant text message alerts [35]. For example, the Vancouver Community Network Street Messaging System sends emergency text alerts such as extreme weather shelter openings or "bad batch" incidents to enrolled residents [72]. Closely related to safety is the impact of not having a phone on privacy. This may be of particular concern for people living with HIV or using substances in the context of persistent stigma and criminalization. For example, Gonzales et al described how reliance on borrowed phones by 29 low-income clients of an HIV clinic created a risk of HIV disclosure, such as if the clinic returned a call or if the owner of the phone was present during the call [2]. For one of our participants who relied on her partner's phone, having a phone of her own presented an opportunity to speak to her doctor privately without her partner "knowing everything." Taken together, these findings suggest that mobile phones may provide a degree of safety and autonomy for young Indigenous people who use drugs.

Addressing Mobile Connectivity Challenges

Continuity of phone connection was often not consistently available to Cedar participants via a mobile phone and plan [2,35]. Our survey did not differentiate between having a phone versus having a phone with a phone number and airtime credit. However, when asked if they had a phone, many participants produced a handset, but explained that it was not connected except through Wi-Fi. Through informal discussions that emerged while administering this survey, we learned that mobile phone handsets are fairly accessible to participants and they reported getting phones from various sources, including second hand, as gifts, and through trades. However, substantial barriers to maintaining cellular connectivity exist, including high costs and existing debts with providers, that prevent participants from having a phone plan [33,35]. Of note, the costs of cellular services are notably higher in Canada than in other international markets, especially when the most basic packages are compared [73]. Gonzales [4] offers the concept of "technology maintenance" to highlight the time, energy, and money required to maintain phone connectivity even after a phone handset is acquired and uses the term "dependable instability" to refer to the frequent, short-term disconnections among low-income people in the United States. Many Cedar participants may experience such periods as a result of jail, housing transitions, relationship breakdowns, phone loss, lack of phone credit, or missed payments. Participants identified the resultant loss of touch with support and service networks as significant barriers

to access. In her study of 37 low-income clients attending two free health clinics in the US Midwest, Gonzales et al [2] found that short-term phonelessness contributed to lost employment, lost welfare benefits, and strains on social support networks, which she describes as critical for health. The authors argued that frequently changing mobile phone numbers may disrupt access to health services, resources, and social support [2]. Homeless adults in Philadelphia also characterized phones as being important tools to fulfill responsibilities to work, housing, and social support—all powerful social determinants of health [35,74].

In the absence of a phone number where they can be reached consistently over time, many participants use social media and messaging apps (eg, Facebook) that can be accessed via Wi-Fi on a smart phone that does not have a cellular plan. High use (90%) of Facebook and other Web-based communication tools have been observed among other marginalized groups, including street-involved youth in Denver [34], British Columbia [33], and Seattle [75]. When a phone is lost, stolen, or disconnected, these points of digital connection are not severed in the same way as when the phone number is cancelled. However, at present, it is not common practice for health care providers to connect using these alternative digital technologies, perhaps because of institutional policies limiting the use of social media [2]. Reaching clients via messaging apps other than SMS may be a useful way to stay connected despite phone disruptions. Previous research has suggested that the success of mHealth programs that rely on text messages may be a result of capitalizing on technology that is both familiar and part of regular habits and routines, rather than creating something new (such as a new mobile app) that requires behavior change [76]. Given that many participants had access to a mobile phone without a cellular plan, future programs may be interested in using Web-based messaging programs that have better continuity in the event of phone loss or missed payments.

Recommendations for Future Mobile Health Programs

Taken together, the findings discussed so far emphasize the importance of addressing ongoing phone access and connectivity within future mHealth programs involving young Indigenous people who use drugs. In addition, participants' perspectives support growing evidence that two-way, open-ended supportive mHealth interventions are more effective than those that are more narrowly focused (eg, reminders only, single health condition) [77-79]. First, participants voiced a broad view of their health, which captures physical, emotional, mental, and family wellness. This is affirmed by studies that articulate a diversity of health priorities among Indigenous people living with/affected by HIV, beyond those directly related to the virus [80,81]. Others have described a variety of strategies Indigenous people living with HIV use to stay healthy that include, but are not limited to, taking antiretroviral therapy (ART) medications [57]. Narrowly focusing on HIV/AIDS prevention and treatment outcomes, such as substance use recovery or ART adherence, may limit the potential for mHealth programs to address a person's own health priorities and goals. Further, frameworks developed by Indigenous bodies to guide mental wellness and substance use services in British Columbia and Canada call for programs that build on the principle of wholistic wellness [48,82]. Second, highly targeted approaches, such as text message reminders to take a specific medication such as ART or methadone may disclose the participants' HIV status or drug use if intercepted. Open-ended text messages, such as the "how are you?" approach taken in the original WelTel Kenya trial minimizes the possibility of disclosure by allowing participants to direct the conversation according to their comfort level [13]. In the context of ongoing stigma and criminalization of drug use and HIV in Canada and elsewhere, avoiding unwanted disclosure is essential [83,84]. However, the participants reported few concerns in this regard. It is possible that lack of concern stemmed from the fact that participants already constantly navigate privacy and disclosure related to drug use and HIV status and have established strategies they would apply in the mHealth context.

Limitations

This study has several potential limitations. The cross-sectional design limits the identification of trends or causal associations. Our sample may not be representative of all young Indigenous people who have used illicit drugs in British Columbia; however, efforts were made to ensure that a diversity of characteristics were represented, including gender, city of residence, and injection and noninjection drug use. Further, Indigenous peoples are diverse, and as this survey involved a particularly vulnerable group of young Indigenous people who have used drugs, findings cannot be generalized to young Indigenous people in general. Finally, our survey did not capture participants' use of mobile handsets that were not connected via a cellular plan. However, this has emerged as an area for future research.

Conclusions

While interest in using mobile phones for health is high among young Indigenous people who have used drugs in British Columbia, low rates of phone ownership present a barrier to engagement in mHealth. Future mHealth programs will need to take this into account, either by providing mobile phone handsets and cellular plans, by supporting texting through Web-based platforms currently utilized by their client base or innovating to reduce periods of phonelessness and/or loss of connectivity. Mobile phones were viewed as a way to support connections and relationships that are seen as critical to health and well-being among the young Indigenous people in this study. In addition, participants articulated a wholistic view of health that included physical, mental, emotional, and family well-being. As a result, open-ended mHealth initiatives for young Indigenous people who have used drugs that strengthen relationships with care providers and other social supports, enable individuals to set their own priorities for health and well-being, and take a culturally safe approach are recommended.

Acknowledgments

The authors are indebted to the study participants who continue to share their stories through The Cedar Project. The authors especially thank the Cedar Project Partnership for their conviction and for holding them accountable to the voices of young Indigenous people. To the Elders Earl Henderson (Cree and Métis) and Violet Bozoki (Lheidli T'enneh First Nation) who support our study, thank you for your continued wisdom and guidance. The authors are grateful to the past and present study staff who bring tremendous care and energy to this work. The Cedar Project receives ongoing support from the Canadian Institutes of Health Research, which has no role in the preparation of data or manuscripts.

Authors' Contributions

KJ lead study design, data collection, analysis, and manuscript composition. MP, VT, RL, MS, and PS contributed to the development of study design and instruments, interpretation of results, and provided feedback on the manuscript. MP also contributed to the analysis. VT was also involved in data collection. RS made substantial contributions to manuscript development and review. SP and LD provided mentorship and guidance at all stages in their role as Cedar Project Partners and Indigenous mentors.

Conflicts of Interest

RL has founded WelTel International mHealth Society (a nonprofit organization) and WelTel Inc (a company) to help develop and scale technologies to deliver research-based services and has an interest in both organizations. There are no other conflicts of interest to declare.

References

1. International Telecommunication Union. Geneva, Switzerland: Telecommunication Development Bureau; 2015. Measuring the Information Society Report URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf [accessed 2020-03-28]
2. Gonzales AL, Ems L, Suri VR. Cell phone disconnection disrupts access to healthcare and health resources: a technology maintenance perspective. New Media Soc 2016 Jul 10;18(8):1422-1438. [doi: 10.1177/1461444814558670]
3. van Dijk JA. The Deepening Divide: Inequality in the Information Society. Thousand Oaks, California: Sage Publications; 2005.
4. Gonzales A. The contemporary US digital divide: from initial access to technology maintenance. Inf Commun Soc 2016;19(2):234-248. [doi: 10.1080/1369118x.2015.1050438]
5. Gonzales AL. Health benefits and barriers to cell phone use in low-income urban US neighborhoods: indications of technology maintenance. Mob Media Commun 2014 Aug 19;2(3):233-248. [doi: 10.1177/2050157914530297]
6. Smillie K, van Borek N, van der Kop ML, Lukhwaro A, Li N, Karanja S, et al. Mobile health for early retention in HIV care: a qualitative study in Kenya (WelTel Retain). Afr J AIDS Res 2014;13(4):331-338 [FREE Full text] [doi: 10.2989/16085906.2014.961939] [Medline: 25555099]
7. Smillie K, van Borek N, Abaki J, Pick N, Maan EJ, Friesen K, et al. A qualitative study investigating the use of a mobile phone short message service designed to improve HIV adherence and retention in care in Canada (WelTel BC1). J Assoc Nurses AIDS Care 2014;25(6):614-625 [FREE Full text] [doi: 10.1016/j.jana.2014.01.008] [Medline: 24768442]
8. Murray MC, O'Shaughnessy S, Smillie K, van Borek N, Graham R, Maan EJ, WelTel BC1 Study Team. Health care providers' perspectives on a weekly text-messaging intervention to engage HIV-positive persons in care (WelTel BC1). AIDS Behav 2015 Oct;19(10):1875-1887. [doi: 10.1007/s10461-015-1151-6] [Medline: 26297567]
9. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
10. Jongbloed K, Parmar S, van der Kop M, Spittal PM, Lester RT. Recent evidence for emerging digital technologies to support global HIV engagement in care. Curr HIV/AIDS Rep 2015 Dec;12(4):451-461 [FREE Full text] [doi: 10.1007/s11904-015-0291-7] [Medline: 26454756]
11. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment & prevention: a systematic review of the literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
12. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
13. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, 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]
14. Mbuagbaw L, Ongolo-Zogo P, Thabane L. Investigating community ownership of a text message programme to improve adherence to antiretroviral therapy and provider-client communication: a mixed methods research protocol. BMJ Open 2013 Jun 25;3(6):e002816 [FREE Full text] [doi: 10.1136/bmjopen-2013-002816] [Medline: 23801710]

15. Mahal D, van der Kop M, Murray M, Jongbloed K, Patel A, Kazi M, et al. Mobile health: an update on BC projects that use WelTel to enhance patient care. BC Med J 2014 Mar;56(2):90-93 [FREE Full text]
16. van der Kop ML, Karanja S, Thabane L, Marra C, Chung MH, Gelmon L, et al. In-depth analysis of patient-clinician cell phone communication during the WelTel Kenya1 antiretroviral adherence trial. PLoS One 2012;7(9):e46033 [FREE Full text] [doi: 10.1371/journal.pone.0046033] [Medline: 23049928]
17. Young SD, Swendeman D, Holloway IW, Reback CJ, Kao U. Use of technology to address substance use in the context of HIV: a systematic review. Curr HIV/AIDS Rep 2015 Dec;12(4):462-471 [FREE Full text] [doi: 10.1007/s11904-015-0295-3] [Medline: 26475670]
18. Cordova D, Alers-Rojas F, Lua FM, Bauermeister J, Nurenberg R, Ovadje L, et al. The usability and acceptability of an adolescent mhealth HIV/STI and drug abuse preventive intervention in primary care. Behav Med 2018;44(1):36-47 [FREE Full text] [doi: 10.1080/08964289.2016.1189396] [Medline: 27223646]
19. Przybyla SM, Eliseo-Arras RK, Krawiec G, Gower E, Dermen K. Feasibility and acceptability of a smartphone app for daily reports of substance use and antiretroviral therapy adherence among HIV-infected adults. AIDS Res Treat 2016;2016:9510172 [FREE Full text] [doi: 10.1155/2016/9510172] [Medline: 27610243]
20. Horvath KJ, Alemu D, Danh T, Baker JV, Carrico AW. Creating effective mobile phone apps to optimize antiretroviral therapy adherence: perspectives from stimulant-using HIV-positive men who have sex with men. JMIR Mhealth Uhealth 2016 Apr 15;4(2):e48 [FREE Full text] [doi: 10.2196/mhealth.5287] [Medline: 27084049]
21. Kirk GD, Himelhoch SS, Westergaard RP, Beckwith CG. Using mobile health technology to improve HIV care for persons living with HIV and substance abuse. AIDS Res Treat 2013;2013:194613 [FREE Full text] [doi: 10.1155/2013/194613] [Medline: 24381751]
22. Cordova D, Bauermeister JA, Fessler K, Delva J, Nelson A, Nurenberg R, Youth Leadership Council. A community-engaged approach to developing an mhealth HIV/STI and drug abuse preventive intervention for primary care: a qualitative study. JMIR Mhealth Uhealth 2015 Dec 18;3(4):e106 [FREE Full text] [doi: 10.2196/mhealth.4620] [Medline: 26685288]
23. Allan B, Smylie J. Wellesley Institute. Toronto: Wellesley Institute; 2015. First Peoples, Second Class Treatment: The Role of Racism in the Health and Well-Being of Indigenous Peoples in Canada URL: https://www.wellesleyinstitute.com/wp-content/uploads/2015/02/Summary-First-Peoples-Second-Class-Treatment-Final.pdf [accessed 2020-03-28]
24. Jongbloed K, Pooyak S, Sharma R, Mackie J, Pearce ME, Laliberte N, Cedar Project Partnership. Experiences of the HIV cascade of care among indigenous peoples: a systematic review. AIDS Behav 2019 Apr;23(4):984-1003. [doi: 10.1007/s10461-018-2372-2] [Medline: 30600452]
25. Statistics Canada. Ottawa; 2018. Aboriginal Population Profile, 2016 Census. Vancouver, City [Census subdivision], British Columbia URL: https://tinyurl.com/y7uh372u [accessed 2020-03-28]
26. Statistics Canada. Ottawa; 2018. Aboriginal Population Profile, 2016 Census. Prince George, City [Census subdivision], British Columbia URL: https://tinyurl.com/ycs53nts [accessed 2020-03-28]
27. Spittal PM, Craib KJ, Teegee M, Baylis C, Christian WM, Moniruzzaman AK, Cedar Project Partnership. The Cedar project: prevalence and correlates of HIV infection among young aboriginal people who use drugs in two Canadian cities. Int J Circumpolar Health 2007 Jun;66(3):226-240. [doi: 10.3402/ijch.v66i3.18259] [Medline: 17655063]
28. R Development Core Team. The R Project for Statistical Computing. Vienna; 2019. URL: https://www.r-project.org/ [accessed 2020-03-28]
29. Pan SW, Christian CW, Pearce ME, Blair AH, Jongbloed K, Zhang H, et al. The Cedar Project: impacts of policing among young aboriginal people who use injection and non-injection drugs in British Columbia, Canada. Int J Drug Policy 2013 Sep;24(5):449-459. [doi: 10.1016/j.drugpo.2013.04.009] [Medline: 23731672]
30. QSR International. 2014. NVivo Qualitative Data Analysis Software URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2020-03-28]
31. CIHR, NSERC, SSHRC. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2018). 2014. TCPS 2 (2018) – Chapter 9: Research Involving the First Nations, Inuit and Métis Peoples of Canada URL: https://ethics.gc.ca/eng/tcps2-eptc2_2018_chapter9-chapitre9.html [accessed 2020-03-22]
32. Schnarch B. Ownership, control, access, and possession (OCAP) or self-determination applied to research. J Aborig Health 2004 Jan;1(1):80-95 [FREE Full text]
33. Selfridge M. The Homeless Hub. Victoria: University of Victoria; 2016 Jan 20. Infographic: Social Media, Technology and BC Street-Involved Youth URL: https://www.homelesshub.ca/blog/infographic-social-media-technology-and-bc-street-involved-youth [accessed 2020-03-28]
34. Harpin S, Davis J, Low H, Gilroy C. Mobile phone and social media use of homeless youth in Denver, Colorado. J Community Health Nurs 2016;33(2):90-97. [doi: 10.1080/07370016.2016.1159440] [Medline: 27074405]
35. Eyrich-Garg KM. Mobile phone technology: a new paradigm for the prevention, treatment, and research of the non-sheltered 'street' homeless? J Urban Health 2010 May;87(3):365-380 [FREE Full text] [doi: 10.1007/s11524-010-9456-2] [Medline: 20397058]
36. Asgary R, Sckell B, Alcabes A, Naderi R, Adongo P, Ogedegbe G. Perceptions, attitudes, and experience regarding mHealth among homeless persons in New York City shelters. J Health Commun 2015;20(12):1473-1480 [FREE Full text] [doi: 10.1080/10810730.2015.1033117] [Medline: 26313765]

37. Milward J, Day E, Wadsworth E, Strang J, Lynskey M. Mobile phone ownership, usage and readiness to use by patients in drug treatment. Drug Alcohol Depend 2015 Jan 1;146:111-115. [doi: 10.1016/j.drugalcdep.2014.11.001] [Medline: 25468818]
38. Genz A, Kirk G, Piggott D, Mehta SH, Linas BS, Westergaard RP. Uptake and acceptability of information and communication technology in a community-based cohort of people who inject drugs: implications for mobile health interventions. JMIR Mhealth Uhealth 2015 Jun 25;3(2):e70 [FREE Full text] [doi: 10.2196/mhealth.3437] [Medline: 26111915]
39. Cedar Project Partnership, Clarkson AF, Christian WM, Pearce ME, Jongbloed K, Caron NR, et al. The Cedar Project: negative health outcomes associated with involvement in the child welfare system among young indigenous people who use injection and non-injection drugs in two Canadian cities. Can J Public Health 2015 May 7;106(5):e265-e270 [FREE Full text] [doi: 10.17269/cjph.106.5026] [Medline: 26451986]
40. Cedar Project Partnership, Pearce ME, Christian WM, Patterson K, Norris K, Moniruzzaman A, et al. The Cedar Project: historical trauma, sexual abuse and HIV risk among young Aboriginal people who use injection and non-injection drugs in two Canadian cities. Soc Sci Med 2008 Jun;66(11):2185-2194 [FREE Full text] [doi: 10.1016/j.socscimed.2008.03.034] [Medline: 18455054]
41. Shahram SZ, Bottorff JL, Kurtz DL, Oelke ND, Thomas V, Spittal PM, Cedar Project Partnership. Understanding the life histories of pregnant-involved young aboriginal women with substance use experiences in three Canadian cities. Qual Health Res 2017 Jan;27(2):249-259. [doi: 10.1177/1049732316657812] [Medline: 27401489]
42. Shahram SZ, Bottorff JL, Oelke ND, Kurtz DL, Thomas V, Spittal PM, And For The Cedar Project Partnership. Mapping the social determinants of substance use for pregnant-involved young aboriginal women. Int J Qual Stud Health Well-being 2017 Dec;12(1):1275155 [FREE Full text] [doi: 10.1080/17482631.2016.1275155] [Medline: 28140776]
43. Jongbloed K, Thomas V, Pearce ME, Christian KW, Zhang H, Oviedo-Joekes E, Cedar Project Partnership. The Cedar Project: residential transience and HIV vulnerability among young aboriginal people who use drugs. Health Place 2015 May;33:125-131. [doi: 10.1016/j.healthplace.2015.02.008] [Medline: 25840350]
44. Greenwood M, de Leeuw S, Lindsay N, Reading C, editors. Determinants of Indigenous Peoples' Health in Canada: Beyond the Social. Toronto: Canadian Scholars' Press Inc; 2015.
45. Marshall M, Marshall A, Bartlett C. Two-eyed seeing in medicine. In: Greenwood M, de Leeuw S, Lindsay N, Reading C, editors. Determinants of Indigenous Peoples' Health in Canada: Beyond the Social. Toronto: Canadian Scholars' Press Inc; 2015.
46. McCormick RM. Healing through interdependence: the role of connecting in first nations healing practices. Can J Couns 1997;31(3):84 [FREE Full text]
47. Waller M, Okamoto S, Miles B, Hurdle D. Resiliency factors related to substance use/resistance: perceptions of native adolescents of the Southwest. J Soc Soc Welfare 2003;30(4):79-94 [FREE Full text]
48. Assembly of First Nations, Health Canada. First Nations Mental Wellness Continuum Framework: Summary Report. Ottawa: Health Canada; 2015.
49. van Uchelen CP, Davidson SF, Quressette SV, Brasfield CR, Demerais LH. What makes us strong: urban aboriginal perspectives on wellness and strength. Can J Commun Ment Health 1997;16(2):37-50. [doi: 10.7870/cjcmh-1997-0005] [Medline: 10181476]
50. Christian WM. Voice of a leader: if you truly believe children are our future - the future is now!. First Peoples Child Fam Rev 2010;5(1) [FREE Full text]
51. Pooyak S. UVic. Victoria: University of Victoria; 2009. My Life is My Ceremony: Indigenous Women of the Sex Trade Share Stories about Their Families and Their Resiliency URL: https://www.uvic.ca/hsd/socialwork/assets/docs/research/thesis-project/SPooyak%20Thesis.pdf [accessed 2020-03-28]
52. Goodkind JR, Hess JM, Gorman B, Parker DP. 'We're still in a struggle': Diné resilience, survival, historical trauma, and healing. Qual Health Res 2012 Aug;22(8):1019-1036 [FREE Full text] [doi: 10.1177/1049732312450324] [Medline: 22707344]
53. Cain R, Jackson R, Prentice T, Collins E, Mill J, Barlow K. The experience of HIV diagnosis among aboriginal people living with HIV/AIDS and depression. Qual Health Res 2013 Jun;23(6):815-824. [doi: 10.1177/1049732313482525] [Medline: 23539094]
54. Reading C, Brennan R, Masching R. Cultural concepts of care among aboriginal people living with HIV and AIDS: a study by the Canadian Aboriginal AIDS Network. Canadi J Aborig Community Based HIV/AIDS Res 2013;5:23-37 [FREE Full text]
55. Bouey PD, Druan BE. The Ahalaya case-management program for HIV-infected American Indians, Alaska Natives, and Native Hawaiians: quantitative and qualitative evaluation of impacts. Am Indian Alsk Native Ment Health Res 2000;9(2):36-52. [Medline: 11279557]
56. Mill J, Keenan L, Lambert D, Larkin K, Ward K. 'I can't deal with this alone': support for aboriginal men and women living with HIV. Canadi J Aborig Community Based HIV/AIDS Res 2008:16-29 Supplemental Edition [FREE Full text]
57. Prentice T, Mill J, Archibald CP, Sommerfeldt S, Worthington C, Jackson R, et al. Aboriginal youth experiences of accessing HIV care and treatment. J HIV AIDS Soc Serv 2011 Oct;10(4):395-413. [doi: 10.1080/15381501.2011.623903]

58. Ka'opua LS, Mueller CW. Treatment adherence among native Hawaiians living with HIV. Soc Work 2004 Jan;49(1):55-63. [doi: 10.1093/sw/49.1.55] [Medline: 14964518]
59. Duran B, Harrison M, Shurley M, Foley K, Morris P, Davidson-Stroh L, et al. Tribally-driven HIV/AIDS health services partnerships: evidence-based meets culture-centered interventions. J HIV AIDS Soc Serv 2010 May 26;9(2):110-129. [doi: 10.1080/15381501003795444]
60. Simoni JM, Sehgal S, Walters KL. Triangle of risk: urban American Indian women's sexual trauma, injection drug use, and HIV sexual risk behaviors. AIDS Behav 2004 Mar;8(1):33-45. [doi: 10.1023/b:aibe.0000017524.40093.6b] [Medline: 15146132]
61. McKenzie HA, Dell CA, Fornssler B. Understanding addictions among Indigenous people through social determinants of health frameworks and strength-based approaches: a review of the research literature from 2013 to 2016. Curr Addict Rep 2016 Sep 28;3(4):378-386. [doi: 10.1007/s40429-016-0116-9]
62. Goodman A, Fleming K, Markwick N, Morrison T, Lagimodiere L, Kerr T, Western Aboriginal Harm Reduction Society. 'They treated me like crap and I know it was because I was Native': the healthcare experiences of aboriginal peoples living in Vancouver's inner city. Soc Sci Med 2017 Apr;178:87-94 [FREE Full text] [doi: 10.1016/j.socscimed.2017.01.053] [Medline: 28214449]
63. Mbuagbaw L, Thabane L, Ongolo-Zogo P. Opening communication channels with people living with HIV using mobile phone text messaging: insights from the CAMPS trial. BMC Res Notes 2013 Apr 4;6:131 [FREE Full text] [doi: 10.1186/1756-0500-6-131] [Medline: 23557081]
64. Mallinson RK, Rajabiun S, Coleman S. The provider role in client engagement in HIV care. AIDS Patient Care STDS 2007;21(Suppl 1):S77-S84. [doi: 10.1089/apc.2007.9984] [Medline: 17563293]
65. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D, et al. Unpacking the black box: a formative research approach to the development of theory-driven, evidence-based, and culturally safe text messages in mobile health interventions. JMIR Mhealth Uhealth 2016 Jan 22;4(1):e10 [FREE Full text] [doi: 10.2196/mhealth.4994] [Medline: 26800712]
66. McCall J, Lauridsen-Hoegh P. Trauma and cultural safety: providing quality care to HIV-infected women of aboriginal descent. J Assoc Nurses AIDS Care 2014;25(1 Suppl):S70-S78. [doi: 10.1016/j.jana.2013.05.005] [Medline: 24012166]
67. Miller CL, Pearce ME, Moniruzzaman A, Thomas V, Christian W, Schechter MT, Cedar Project Partnership. The Cedar project: risk factors for transition to injection drug use among young, urban aboriginal people. Can Med Assoc J 2011 Jul 12;183(10):1147-1154 [FREE Full text] [doi: 10.1503/cmaj.101257] [Medline: 21670106]
68. Pearce ME, Blair AH, Teegee M, Pan SW, Thomas V, Zhang H, et al. The Cedar project: historical trauma and vulnerability to sexual assault among young aboriginal women who use illicit drugs in two Canadian cities. Violence Against Women 2015 Mar;21(3):313-329. [doi: 10.1177/1077801214568356] [Medline: 25648945]
69. Moniruzzaman A, Pearce ME, Patel SH, Chavoshi N, Teegee M, Adam W, Cedar Project Partnership. The Cedar project: correlates of attempted suicide among young aboriginal people who use injection and non-injection drugs in two Canadian cities. Int J Circumpolar Health 2009 Jun;68(3):261-273. [doi: 10.3402/ijch.v68i3.18338] [Medline: 19705658]
70. Jongbloed K, Pearce ME, Pooyak S, Zamar D, Thomas V, Demerais L, Cedar Project Partnership. The Cedar project: mortality among young Indigenous people who use drugs in British Columbia. Can Med Assoc J 2017 Nov 6;189(44):E1352-E1359 [FREE Full text] [doi: 10.1503/cmaj.160778] [Medline: 29109208]
71. Woelfer J, Iverson A, Hendry D, Friedman B, Gill B. Improving the Safety of Homeless Young People With Mobile Phones: Values, Form and Function. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7 - 12, 2011; Vancouver, Canada p. 1707-1716. [doi: 10.1145/1978942.1979191]
72. Vancouver Community Network. StreetMessenger. Vancouver: Telus Foundation; 2015 Jul. Street Messaging System (SMS) Feasibility Report URL: https://www.streetmessenger.io/report.pdf [accessed 2020-03-28]
73. NGL Nordicity Group Ltd. The Canadian Radio-television and Telecommunications Commission (CRTC). Ottawa, Canada; 2016 Mar 22. 2016 Price Comparison Study of Telecommunications Services in Canada and Select Foreign Jurisdictions URL: https://crtc.gc.ca/eng/publications/reports/compar/compar2016.pdf [accessed 202-03-27]
74. Loppie C, Wien F. National Collaborating Centre for Aboriginal Health. Prince George: NCCIH; 2013. Health Inequalities and the Social Determinants of Aboriginal Peoples' Health. Second Edition URL: https://www.ccnsa-nccah.ca/docs/determinants/RPT-HealthInequalities-Reading-Wien-EN.pdf [accessed 2020-03-28]
75. Woelfer J, Hendry D. Homeless Young People on Social Network Sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5 - 10, 2012; Texas, USA p. 2825-2834. [doi: 10.1145/2207676.2208686]
76. Lanham HJ, Leykum LK, Taylor BS, McCannon CJ, Lindberg C, Lester RT. How complexity science can inform scale-up and spread in health care: understanding the role of self-organization in variation across local contexts. Soc Sci Med 2013 Sep;93:194-202. [doi: 10.1016/j.socscimed.2012.05.040] [Medline: 22819737]
77. Wald DS, Butt S, Bestwick JP. 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]

78. Lester RT. Ask, don't tell - mobile phones to improve HIV care. N Engl J Med 2013 Nov 7;369(19):1867-1868. [doi: 10.1056/NEJMc1310509] [Medline: 24195570]
79. Mills EJ, Lester R, Ford N. Adherence to antiretroviral therapy: supervision or support? Lancet Infect Dis 2012 Feb;12(2):97-98. [doi: 10.1016/S1473-3099(11)70354-1] [Medline: 22281134]
80. Sazegar P, Tu D, Littlejohn D, Myran A. Perspectives of Canadian Inner City Aboriginal and Non-Aboriginal People Living with HIV and AIDS. Canadi J Aborig Community Based HIV/AIDS Res 2013;5:39-56.
81. McCall J, Browne AJ, Reimer-Kirkham S. Struggling to survive: the difficult reality of Aboriginal women living with HIV/AIDS. Qual Health Res 2009 Dec;19(12):1769-1782. [doi: 10.1177/1049732309353907] [Medline: 19949225]
82. First Nations Health Authority, BC Ministry of Health, Health Canada. First Nations Health Authority. Vancouver; 2013. A Path Forward: BC First Nations and Aboriginal People's Mental Wellness and Substance Use - 10 Year Plan URL: https://www.fnha.ca/documents/fnha_mwsu.pdf [accessed 2020-03-28]
83. Adam BD, Elliott R, Corriveau P, English K. Impacts of criminalization on the everyday lives of people living with HIV in Canada. Sex Res Soc Policy 2014;11(1):39-49. [doi: 10.1007/s13178-013-0131-8]
84. Wood E, Werb D, Kazatchkine M, Kerr T, Hankins C, Gorna R, et al. Vienna declaration: a call for evidence-based drug policies. Lancet 2010 Jul 31;376(9738):310-312. [doi: 10.1016/S0140-6736(10)60958-0] [Medline: 20650517]

Abbreviations

ART: antiretroviral therapy
mHealth: mobile health

Edited by CL Parra-Calderón; submitted 25.10.19; peer-reviewed by N Muinga, M Bendtsen; comments to author 02.01.20; revised version received 16.01.20; accepted 22.02.20; published 27.07.20.

Please cite as:
Jongbloed K, Pearce ME, Thomas V, Sharma R, Pooyak S, Demerais L, Lester RT, Schechter MT, Spittal PM, The Cedar Project Partnership
The Cedar Project - Mobile Phone Use and Acceptability of Mobile Health Among Young Indigenous People Who Have Used Drugs in British Columbia, Canada: Mixed Methods Exploratory Study
JMIR Mhealth Uhealth 2020;8(7):e16783
URL: https://mhealth.jmir.org/2020/7/e16783
doi:10.2196/16783
PMID:32716311

Original Paper

Perceived Need and Acceptability of an App to Support Activities of Daily Living in People With Cognitive Impairment and Their Carers: Pilot Survey Study

Rhoda Lai[1], BA (Hons); Maria Tensil[2], MSc; Alexander Kurz[2], MD; Nicola T Lautenschlager[1,3,4], MD; Janine Diehl-Schmid[2], MD

[1]Academic Unit for Psychiatry of Old Age, Department of Psychiatry, The University of Melbourne, Melbourne, Australia

[2]Department of Psychiatry and Psychotherapy, School of Medicine, Technical University of Munich, Munich, Germany

[3]Aged Persons Mental Health Program, North Western Mental Health, Melbourne Health, Melbourne, Australia

[4]School of Psychiatry and Clinical Neurosciences & WA Centre for Health and Ageing, University of Western Australia, Perth, Australia

Corresponding Author:
Janine Diehl-Schmid, MD
Department of Psychiatry and Psychotherapy
School of Medicine
Technical University of Munich
Ismaninger Str 22
Munich, 81675
Germany
Phone: 49 89 41406488
Email: janine.diehl-schmid@tum.de

Abstract

Background: Modern technologies, including smartphone apps, have the potential to assist people with cognitive impairment with activities of daily living, allowing them to maintain their independence and reduce carer burden. However, such tools have seen a slow rate of uptake in this population, and data on the acceptability of assistive technologies in this population are limited.

Objective: This pilot study included older adults with cognitive impairment and their carers, and explored the perceived needs for and acceptability of an app that was designed to be a simple assistive tool for activities of daily living. In particular, this study aimed to assess the acceptability of common app functions such as communication, reminder, navigation, and emergency tools in this population, and to compare patients' and carers' responses to them.

Methods: A total of 24 German participants with mild cognitive impairment or dementia and their family carers separately completed two short questionnaires. The first questionnaire asked the participants with cognitive impairment and their carers to self-rate the patients' cognitive impairment levels and affinity to technology. Following a demonstration of the app, participants rated the usability and acceptability of the app and its functions in a second questionnaire.

Results: Participants rated themselves as much less cognitively impaired than their carers did (P=.01), and insight into the level of support they received was low. The majority of the participants (19/24, 79%) and their carers (20/24, 83%) had low affinity to technology, and even after the demonstration, 63% (15/24) of the participants had low interest in using the app. A breakdown of acceptability responses by app function revealed that participants were more amenable to the reminder function, the emergency feature, and a wearable form of the app. Features that centered around carers monitoring participants' movements were reported to be less acceptable to participants.

Conclusions: This study highlights the importance of focusing on acceptability and the consumer's perceptions in the development of assistive technology for older adults with cognitive impairment. Participants showed an aversion to functions they perceived as eroding their independence, while functions that more closely aligned with independence and autonomy were perceived as more acceptable.

(JMIR Mhealth Uhealth 2020;8(7):e16928) doi:10.2196/16928

KEYWORDS

Aged; dementia; memory disorders; carers; mobile apps

Introduction

By 2050, the World Alzheimer's Report 2015 projected that over 20% of the world's population will be over the age of 60 years [1]. With this comes a projected rise in the prevalence of dementia and mild cognitive impairment (MCI) [1]. The greatest cost of dementia care has been found to be unpaid or informal care, which costs 16,473€ (US $21,011) per patient annually in Germany, in addition to productivity losses for caregivers who are employed [2]. Most informal carers of people with memory problems are their spouses, who provide help with daily activities [2].

In recent years, there has been an increased interest in how modern technology can assist with the daily activities of older adults with cognitive impairment [3,4]. Technologies that are being explored include reminder systems (such as medication prompts), tools for maintaining social contact, navigational and safety monitoring tools, and stimulation tools (for entertainment or cognitive training) [3]. There is evidence to suggest that these technologies have the potential to not only benefit people with cognitive impairment by increasing or preserving their independence and safety, but also to subsequently support carers by reducing the time and energy spent on daily caregiving activities and reducing anxiety about taking respite time [4,5].

However, technologies for people with cognitive impairment have seen a slow rate of uptake, which could be partly attributed to a lack of user-centered design and validation [6,7]. Studies have tended to focus on technological possibilities and design rather than the usability and acceptability of these technologies for older people [8], which are particularly important to evaluate for this population that is often reluctant to adopt new technologies [9]. Even in studies that have evaluated technology usability among people with cognitive impairment, carers have often been the participants [10]—few have evaluated usability or acceptability for people with memory problems [9,11,12]. Furthermore, in the studies that have included people with memory problems, their opinions have not been reported separately from their carers', which may differ significantly [9,12]. In order to be acceptable, the technology must not only be easy to use (ie, it must have high usability) but must also spark motivation in the user to adopt it in everyday life [9]. Little is known about which features people with cognitive impairment would value in assistive technology, and which would increase their motivation for using it. These insights may be key to increasing uptake rates in this population.

In this pilot study, we engaged in a preliminary exploration of the acceptability of a simple smartphone app among people with mild cognitive impairment and dementia and their family carers living in the community. The app was designed to aid the person with cognitive impairment in activities of daily living, including communicating with friends and family, navigating, and serving as a memory prompt and emergency alert system. The study aimed to obtain feedback from people with cognitive impairment and their family carers separately on their perceived need for such a support, and whether its functions would be useful and feasible for everyday use. Although we explored the usability and acceptability of each of the app's functions, the focus of this study was not on the design of this app specifically, but rather on the participants' and carers' perceived need and usability of an assistive app in general.

Methods

The App

The app used was a pilot developed in collaboration with professionals in the fields of informatics and dementia. It used a simple interface and had four key functions: (1) simplified phone calls and voice or text messages to predefined contacts; (2) a memory aid, used to save speech and text notes as reminders, such as reminders for taking medications; (3) a one-click emergency call to the predefined carer; and (4) a simplified pedestrian navigation system. The app interface was in German.

Participants and Recruitment

From an outpatient memory clinic in Germany, 24 older adults with a diagnosis of MCI or dementia and their family or informal carers were recruited (demographic data are displayed in Table 1). Patients were invited to participate in the study if (1) they had been diagnosed with MCI or mild dementia at a prior visit, and (2) if they had a close relative/carer. All patients who visited the memory clinic (irrespective of the reason, whether it was for a follow-up visit, clinical trial participation, or counseling) and fulfilled the inclusion criteria were asked to participate, and if they agreed, an appointment for the study visit was scheduled. Demographic data for the participants are displayed in Table 1. Of the 24 patients recruited, 14 (58%) of the patients lived in the same household as their carer.

Table 1. Participant demographic data.

Characteristic	Patients (N=24)	Carers (N=24)
Age (years), mean (SD, range)	74.5 (6.1, 57-84)	62.4 (16.0, 31-83)
Gender, n (%)		
Men	8 (33)	15 (62)
Women	16 (67)	9 (38)
Education, > 12 years, n (%)	8 (33)	11 (46)
MMSE[a] score, mean (SD, range)	22.4 (4.2, 10-28)	
Diagnosis, n (%)		
MCI[b]	6 (25)	N/A[c]
Dementia caused by Alzheimer's disease	16 (67)	N/A
Other neurodegenerative dementia	2 (8)	N/A
Relationship to patient, n (%)		
Spouse	N/A	7 (29)
Child	N/A	7 (29)
Other type of relative	N/A	1 (4)
Friend	N/A	2 (8)

[a]MMSE: Mini-Mental State Examination.

[b]MCI: mild cognitive impairment.

[c]N/A: not applicable.

Procedure and Measures

Patients and carers first completed a short, simple questionnaire in separate face-to-face interviews conducted in their native German. They each self-rated the patient's impairments and obstacles in daily living by responding to questions about how often appointments and medications are forgotten, how often memory aids are used, and how often they become disoriented outside the home. Responses were indicated using "rarely," "sometimes," "often," or "supported by carer" options, with the latter indicating that a task was perceived as possible only with carer support. They also each rated the patient's affinity to technology and answered questions about the patient's experience with using devices such as smartphones and computers. Demographic information was also collected in this questionnaire.

The interviewer then demonstrated the app to the patients and their carers, explaining all functions and answering any questions. Patients and carers were given the opportunity to try using the app's functions, with help from the interviewer if needed. This lasted for approximately one hour.

Finally, patients and carers rated the usability and acceptability of the app. This was assessed through a short questionnaire administered in separate face-to-face interviews about the clarity of the layout, interest in finding out more about the app, interest in trying the app, and ratings of each function's usefulness. Ratings of usability and acceptability were measured using "low," "moderate," "high," or "don't know" responses.

This project was approved by the Ethics Committee of the Medical Faculty of the Technical University of Munich (335/18S).

Data Analysis

Data analysis was performed using the Statistical Package for the Social Sciences (SPSS; IBM Corp). Descriptive analyses were performed to summarize the socio-demographic data and frequencies generated for survey answers. A Wilcoxon signed-rank test was used to analyze differences between patients' and carers' perceptions of impairments in daily living. A statistical threshold of $P<.05$ was considered statistically significant.

Results

The patients tended to rate their impairments as less severe than their carers did (see Table 2). When asked how often they had problems with aspects of their memory or concentration, carers responded that patients were affected more often than the patients did. The Wilcoxon signed-rank test revealed that perceptions differed significantly between patients and carers on every aspect of daily living that was asked about in the survey (see Table 2).

Table 2. Participants' (N=24) self-rating, carers' (N=24) rating of the participants' impairments, and *P* values for differences between their perceptions.

Question	Response				
	Rarely	Sometimes	Often	Carer supported	*P* value
How often is the patient's memory and concentration impaired?					.012[a]
Patients, n (%)	5 (21)	14 (58)	5 (21)	N/A[b]	
Carers, n (%)	2 (8)	7 (29)	15 (63)	N/A[b]	
How often does the patient forget appointments?					.001[a]
Patients, n (%)	12 (50)	6 (25)	2 (8)	4 (17)	
Carers, n (%)	5 (21)	1 (4)	3 (13)	15 (62)	
How often does the patient forget to take their medication?					.007[a]
Patients, n (%)	10 (42)	3 (12)	0 (0)	11 (46)	
Carers, n (%)	3 (13)	2 (8)	2 (8)	17 (71)	
How often does the patient get disoriented outside the home?					.028[a]
Patients, n (%)	17 (71)	3 (12)	0 (0)	4 (17)	
Carers, n (%)	11 (46)	5 (21)	2 (8)	6 (25)	

[a]A statistical threshold of $P<.05$ was considered statistically significant.

[b]N/A: not applicable.

Of the 24 carers, 20 (83%) rated their own general affinity to technology as "low." Despite this, half of the 24 carers reported that they had "high" (7/24, 29%) or "very high" (5/24, 21%) levels of preparedness for using an app, indicating their interest in using such an assistive tool. Half (12/24, 50%) of the 24 patients never used a mobile phone, and out of those who did, only 3 (13%) used a mobile phone at least once a week. Of the 24 carers, 19 (79%) reported that the patient's ability to successfully use an app would be "low," while the remaining 5 carers believed it would be "moderate." In the predemonstration interview, most patients (18/24, 75%) reported "low" interest in getting a short introduction to the app and its use, which corresponded with the 19 (19/24, 79%) carers who believed their relative's willingness and ability to use an app would be "low." These responses were given despite the fact that 10 (42%) of the 24 patients reported frequent use of aids like calendars or notebooks for everyday tasks (while 4 patients, 17%, reported using such aids "occasionally").

After the demonstration of the app, patients' willingness to try the app only changed moderately—15 (63%) of the 24 patients still reported "low" interest in trying the app despite most of them (18/24, 75%) reporting that the layout was "clear." However, their perceptions about the usefulness of each function varied (see Table 3). Patients rated the reminder and emergency functions more highly than the navigation function. Interestingly, carers were more likely to feel safe if they could see the patient's location, whereas patients did not rate the usefulness of the navigation and location function highly. Surprisingly, a large proportion of patients thought that having the app in a wearable form, such as on a smartwatch, would be useful—7 (29%) of the 24 patients reported "moderate" or higher usefulness.

Table 3. Patients' (N=24) and carers' (N=24) ratings of the usefulness of the app's functions.

Question	Response			
	Low	Moderate	High	Don't know
How likely is it that the patient would use the navigation function?				
Patients, n (%)	14 (58)	1 (4)	1 (4)	8 (33)
Carers, n (%)	16 (67)	8 (33)	0 (0)	0 (0)
How much would it increase your feeling of safety if the carer could see the patient's location?				
Patients, n (%)	16 (67)	2 (8)	2 (8)	4 (17)
Carers, n (%)	13 (54)	7 (29)	4 (17)	0 (0)
How likely is it that the patient would use the reminder function for appointments/medications?				
Patients, n (%)	9 (38)	7 (29)	2 (8)	6 (25)
Carers, n (%)	13 (54)	4 (17)	7 (29)	0 (0)
How useful would it be if the carer could update appointments in the app via the internet?				
Patients, n (%)	12 (50)	4 (17)	4 (17)	4 (17)
Carers, n (%)	15 (62)	5 (21)	4 (17)	0 (0)
How likely is it that the emergency function would be useful?				
Patients, n (%)	5 (21)	7 (29)	5 (21)	7 (29)
Carers, n (%)	10 (42)	6 (25)	8 (33)	0 (0)
How useful would it be if the app came in a wearable form, like on a smartwatch?				
Patients, n (%)	8 (33)	2 (8)	5 (21)	9 (38)
Carers, n (%)	6 (25)	11 (46)	7 (29)	0 (0)

Discussion

Principal Findings

This pilot study included several findings that will be useful in the design of apps for people with cognitive impairment, including the significant discrepancy between perceived levels of cognitive impairment between people with cognitive impairment and their carers. There was low overall interest in such an app, and specifically, patients were least interested in functions that facilitated carers with monitoring their activities. This study has begun to address a gap in the research, surveying not only carers' perceptions of assistive technology but also the perceptions of people with cognitive impairment.

In this study, there were clear discrepancies between the ratings given by the participants with cognitive impairment and their carers in every aspect of cognition. Participants with cognitive impairment most frequently responded that they rarely or only sometimes forgot appointments or became disoriented outside the home, and did not acknowledge that they were supported by their carers as often as carers reported they were. Limited insight into their own need for support in daily activities may be a barrier to the uptake of assistive technology by this population [6].

Even after receiving the demonstration of the app, almost two-thirds (15/24, 63%) of the participants with cognitive impairment had low-interest levels in trying the app, and they frequently responded with the "don't know" option when rating the app functions' usefulness, showing ambivalence toward adopting the technology. Previous research has shown that the use of technology can ignite positive feelings of mastery in older adults [12], and a review of mobile assistive technologies for people with dementia found that there was a dire need for these tools to tap into higher-level human needs, such as self-esteem and creativity [13]. However, the low-interest levels in this study point to a barrier to technology adoption. Studies have found that learning to use new technologies can call an older adult's attention to their diminishing memory and functioning [14], and it is possible that participants in this study were responding to such a recognition with denial of their need for assistive apps. It is clear from this study that simply demonstrating the functionality of assistive technologies is not enough to spark interest in their use, and that it may even turn potential consumers away if their use is perceived to be demeaning or to highlight impairments [12].

Although this study does not offer a promising likelihood of people with cognitive impairment adopting such an app, it did provide useful findings on the acceptability of the app's different functions. Most evident was the fact that people with cognitive impairment appeared to have a desire to maintain their independence. While participants with cognitive impairment were moderately open to the usefulness of the reminder function, they were less open to the idea of their carer being able to update their appointment information remotely. This desire also manifested in responses to the app's navigation function—nearly half (11/24, 46%) of the carers reported that being able to see the patient's location would at least moderately increase their feelings of safety, but about two-thirds (16/24, 67%) of the

patients reported that it would not make them feel safer. The patients were more amenable to having an emergency function handy, which suggests that they preferred carer help to be on an on-call basis rather than as an ongoing presence. This view differed from the typical view of the carers, who leaned toward being supportive of surveillance and monitoring technology [12,15]; this should be noted in future studies.

Interestingly, in a qualitative study, Hill et al [16] found that older adults perceived technology to be contributing to the erosion of independence and skills such as problem-solving in younger people; these older adults correlated the ability to live without technology dependence with self-reliance and the preservation of cognitive skills [16]. Despite 42% (10/24) of the participants with cognitive impairment in this study reporting the use of aids to assist with everyday memory tasks, it is possible that participants could have been similarly resistant to using technological aids for the same reason. Further, in another qualitative study by Arntzen et al [17], people with younger onset dementia reported that new technology had to fit well into their habitual routines to facilitate adoption. The relative similarity between a smartwatch and a regular watch may bridge the adoption of a new technology, and this is supported by the finding that 7 out of 24 patients (29%) thought it would be useful to have the app on a smartwatch, a greater proportion of patient agreement than was observed for any other app function rating. Although this is mere speculation, the need to relate assistive technology to higher-level needs in order to facilitate better technology adoption rates is apparent [13]—the need for autonomy and identity maintenance was clear in this study's population sample [17].

Limitations and Future Directions

This study had several limitations. Participants were given limited time to trial the app; had they been permitted to trial the app at home, it is possible that they could have developed greater confidence in and acceptability of using the app over time. Furthermore, we did not capture in the questionnaire whether the participants completely understood the functionality of the app (although anecdotal impressions from the interviewer were that they grasped the concept). The study's sample size was small, which limited the ability to conduct statistical analyses on the data.

However, this study captures attitudes that would be present in the decision-making moments regarding the use or purchase of an app. Future studies would benefit from involving people with cognitive impairment in all aspects of assistive technology development. While this population may find technologies acceptable when scaffolded by training and provision of devices, or in the evaluation phase, gauging interest at the outset may be vital to unpacking higher-level needs that increase technology use in this population [18].

Conclusions

Overall, this study has provided useful preliminary findings on older adults with cognitive impairment and their carers' perceptions of the usefulness of different functions that can be provided within an app to assist with everyday tasks. The need for assistive apps and technology to be tailored toward autonomy and identity maintenance to be acceptable was apparent. Aligning assistive technology to these needs may improve the uptake of technology in this population.

Acknowledgments

This work was supported by the Technical University of Munich's Institute for Advanced Study, funded by the German Excellence Initiative.

Authors' Contributions

JDS, NL, and AK conceived of and designed the project. JDS and MT collected the data. MT and RL performed the analysis. RL, JDS, and NL wrote the paper.

Conflicts of Interest

None declared.

References

1. Prince M, Wimo A, Guerchet M. World Alzheimer's Report 2015. London: Alzheimer's Disease International; 2015.
2. Michalowsky B, Thyrian JR, Eichler T, Hertel J, Wucherer D, Flessa S, et al. Economic Analysis of Formal Care, Informal Care, and Productivity Losses in Primary Care Patients who Screened Positive for Dementia in Germany. J Alzheimers Dis 2016;50(1):47-59. [doi: 10.3233/JAD-150600] [Medline: 26639964]
3. D'Onofrio G, Sancarlo D, Ricciardi F, Panza F, Seripa D, Cavallo F, et al. Information and Communication Technologies for the Activities of Daily Living in Older Patients with Dementia: A Systematic Review. J Alzheimers Dis 2017;57(3):927-935. [doi: 10.3233/JAD-161145] [Medline: 28304297]
4. Madara Marasinghe K. Assistive technologies in reducing caregiver burden among informal caregivers of older adults: a systematic review. Disabil Rehabil Assist Technol 2016;11(5):353-360. [doi: 10.3109/17483107.2015.1087061] [Medline: 26371519]

5. Brown EL, Ruggiano N, Li J, Clarke PJ, Kay ES, Hristidis V. Smartphone-Based Health Technologies for Dementia Care: Opportunities, Challenges, and Current Practices. J Appl Gerontol 2019 Jan;38(1):73-91. [doi: 10.1177/0733464817723088] [Medline: 28774215]
6. Ienca M, Fabrice J, Elger B, Caon M, Scoccia Pappagallo A, Kressig RW, et al. Intelligent Assistive Technology for Alzheimer's Disease and Other Dementias: A Systematic Review. J Alzheimers Dis 2017;56(4):1301-1340. [doi: 10.3233/JAD-161037] [Medline: 28222516]
7. Guisado-Fernández E, Giunti G, Mackey LM, Blake C, Caulfield BM. Factors Influencing the Adoption of Smart Health Technologies for People With Dementia and Their Informal Caregivers: Scoping Review and Design Framework. JMIR Aging 2019 Apr 30;2(1):e12192 [FREE Full text] [doi: 10.2196/12192] [Medline: 31518262]
8. Bächle M, Daurer S, Judt A, Mettler T. Assistive technology for independent living with dementia: Stylized facts and research gaps. Health Policy and Technology 2018 Mar;7(1):98-111. [doi: 10.1016/j.hlpt.2017.12.002]
9. Holthe T, Halvorsrud L, Karterud D, Hoel K, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: A systematic literature review. Clin Interv Aging 2018;13:863-886 [FREE Full text] [doi: 10.2147/CIA.S154717] [Medline: 29765211]
10. Sriram V, Jenkinson C, Peters M. Informal carers' experience of assistive technology use in dementia care at home: a systematic review. BMC Geriatr 2019 Jun 14;19(1):160 [FREE Full text] [doi: 10.1186/s12877-019-1169-0] [Medline: 31196003]
11. Meiland F, Innes A, Mountain G, Robinson L, van der Roest H, García-Casal JA, et al. Technologies to Support Community-Dwelling Persons With Dementia: A Position Paper on Issues Regarding Development, Usability, Effectiveness and Cost-Effectiveness, Deployment, and Ethics. JMIR Rehabil Assist Technol 2017 Jan 16;4(1):e1 [FREE Full text] [doi: 10.2196/rehab.6376] [Medline: 28582262]
12. van Boekel LC, Wouters EJM, Grimberg BM, van der Meer NJM, Luijkx KG. Perspectives of Stakeholders on Technology Use in the Care of Community-Living Older Adults with Dementia: A Systematic Literature Review. Healthcare (Basel) 2019 May 28;7(2):73 [FREE Full text] [doi: 10.3390/healthcare7020073] [Medline: 31141999]
13. Delello JA, McWhorter RR. Reducing the Digital Divide: Connecting Older Adults to iPad Technology. J Appl Gerontol 2015 Jun 17;36(1):3-28. [doi: 10.1177/0733464815589985] [Medline: 26084479]
14. Koo BM, Vizer LM. Examining Mobile Technologies to Support Older Adults With Dementia Through the Lens of Personhood and Human Needs: Scoping Review. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e15122 [FREE Full text] [doi: 10.2196/15122] [Medline: 31710305]
15. Mulvenna M, Hutton A, Coates V, Martin S, Todd S, Bond R, et al. Views of caregivers on the ethics of assistive technology used for home surveillance of people living with dementia. Neuroethics 2017;10(2):255-266 [FREE Full text] [doi: 10.1007/s12152-017-9305-z] [Medline: 28725288]
16. Hill R, Betts LR, Gardner SE. Older adults' experiences and perceptions of digital technology: (Dis)empowerment, wellbeing, and inclusion. Computers in Human Behavior 2015 Jul;48:415-423. [doi: 10.1016/j.chb.2015.01.062]
17. Arntzen C, Holthe T, Jentoft R. Tracing the successful incorporation of assistive technology into everyday life for younger people with dementia and family carers. Dementia (London) 2016 Jul;15(4):646-662. [doi: 10.1177/1471301214532263] [Medline: 24784941]
18. Suijkerbuijk S, Nap HH, Cornelisse L, IJsselsteijn WA, de Kort YAW, Minkman MMN. Active involvement of people with dementia: A systematic review of studies developing supportive technologies. J Alzheimers Dis 2019;69(4):1041-1065 [FREE Full text] [doi: 10.3233/JAD-190050] [Medline: 31156158]

Abbreviations

MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
SPSS: Statistical Package for the Social Sciences

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by L Sbaffi, C Fernández; comments to author 29.02.20; revised version received 24.04.20; accepted 26.04.20; published 31.07.20.

Please cite as:
Lai R, Tensil M, Kurz A, Lautenschlager NT, Diehl-Schmid J
Perceived Need and Acceptability of an App to Support Activities of Daily Living in People With Cognitive Impairment and Their Carers: Pilot Survey Study
JMIR Mhealth Uhealth 2020;8(7):e16928
URL: http://mhealth.jmir.org/2020/7/e16928/
doi:10.2196/16928
PMID:32735223

Original Paper

Implementing mHealth Interventions in a Resource-Constrained Setting: Case Study From Uganda

Amanda J Meyer[1,2], BA, MPH; Mari Armstrong-Hough[2,3], BA, MA, MPH, PhD; Diana Babirye[2], MSc; David Mark[2], BA; Patricia Turimumahoro[2], MBChB; Irene Ayakaka[2], MBChB; Jessica E Haberer[4], MD; Achilles Katamba[2,5], MBChB, PhD; J Lucian Davis[1,2,6,7], AB, MS, MD

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, United States

[2]Uganda Tuberculosis Implementation Research Consortium, Makerere University, Kampala, Uganda

[3]Departments of Social and Behavioral Sciences and Epidemiology, School of Global Public Health, New York University, New York, NY, United States

[4]Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States

[5]Clinical Epidemiology and Biostatistics Unit, Department of Medicine, Makerere University College of Health Sciences, Kampala, Uganda

[6]Center for Methods in Implementation and Prevention Science, Yale School of Public Health, New Haven, CT, United States

[7]Pulmonary, Critical Care and Sleep Medicine Section, Yale School of Medicine, New Haven, CT, United States

Corresponding Author:
J Lucian Davis, AB, MS, MD
Department of Epidemiology of Microbial Diseases
Yale School of Public Health
60 College Street
New Haven, CT, 06520
United States
Phone: 1 203 785 3665
Fax: 1 203 785 4782
Email: lucian.davis@yale.edu

Abstract

Background: Mobile health (mHealth) interventions are becoming more common in low-income countries. Existing research often overlooks implementation challenges associated with the design and technology requirements of mHealth interventions.

Objective: We aimed to characterize the challenges that we encountered in the implementation of a complex mHealth intervention in Uganda.

Methods: We customized a commercial mobile survey app to facilitate a two-arm household-randomized, controlled trial of home-based tuberculosis (TB) contact investigation. We incorporated digital fingerprinting for patient identification in both study arms and automated SMS messages in the intervention arm only. A local research team systematically documented challenges to implementation in biweekly site visit reports, project management reports, and minutes from biweekly conference calls. We then classified these challenges using the Consolidated Framework for Implementation Research (CFIR).

Results: We identified challenges in three principal CFIR domains: (1) intervention characteristics, (2) inner setting, and (3) characteristics of implementers. The adaptability of the app to the local setting was limited by software and hardware requirements. The complexity and logistics of implementing the intervention further hindered its adaptability. Study staff reported that community health workers (CHWs) were enthusiastic regarding the use of technology to enhance TB contact investigation during training and the initial phase of implementation. After experiencing technological failures, their trust in the technology declined along with their use of it. Finally, complex data structures impeded the development and execution of a data management plan that would allow for articulation of goals and provide timely feedback to study staff, CHWs, and participants.

Conclusions: mHealth technologies have the potential to make delivery of public health interventions more direct and efficient, but we found that a lack of adaptability, excessive complexity, loss of trust among end users, and a lack of effective feedback systems can undermine implementation, especially in low-resource settings where digital services have not yet proliferated. Implementers should anticipate and strive to avoid these barriers by investing in and adapting to local human and material resources, prioritizing feedback from end users, and optimizing data management and quality assurance procedures.

Trial Registration: Pan-African Clinical Trials Registration PACTR201509000877140; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=877

(JMIR Mhealth Uhealth 2020;8(7):e19552) doi:10.2196/19552

KEYWORDS

mHealth; implementation; tuberculosis; consolidated framework for implementation science; Uganda; framework; intervention; app

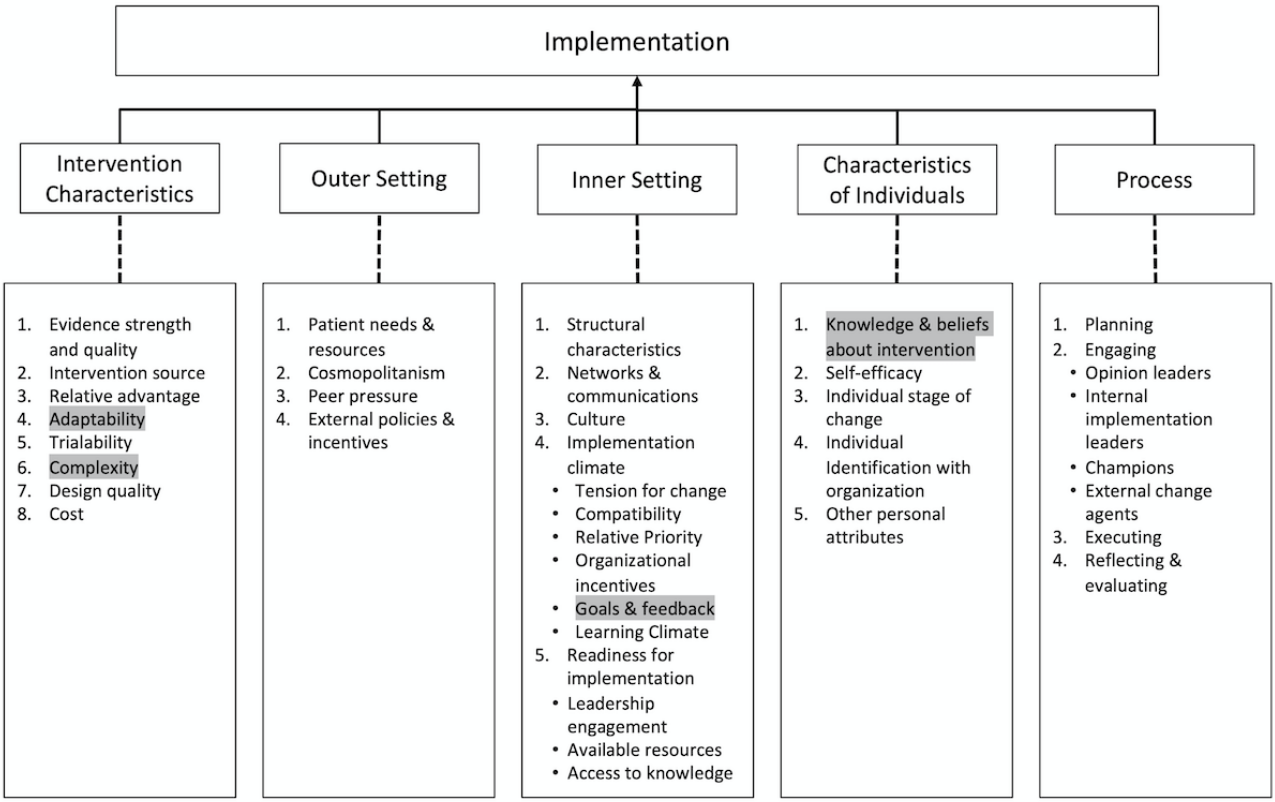
Introduction

Mobile health (mHealth) and other electronic health (eHealth) interventions are becoming more common in low-income countries as advances in technology have enabled researchers and practitioners to engineer seemingly simple solutions to complex public health problems [1,2]. In many countries in sub-Saharan Africa, access to mobile phones is nearly universal, and mHealth apps such as mobile data collection software and SMS are showing great promise for enhancing capacity in resource-constrained health systems [3-5]. For example, mHealth technologies can promote treatment adherence [6-10], enhance access to and uptake of maternal and child health care services [11,12], and improve the quality and reliability of services by connecting external quality assurance teams directly to diagnostic instruments [13]. However, mHealth studies have met with mixed success in achieving their aims in these settings: many interventions have been piloted, yet few have made it into routine practice [1,3,14,15]. In many cases, it is unclear whether this lack of success reflects the ineffective implementation of these technologies or a lack of efficacy of the interventions themselves as most pilot studies do not report on implementation outcomes.

Implementation science is an emerging field that is highly relevant for answering such questions in mHealth by using interdisciplinary approaches, including quantitative and qualitative process evaluation. The goals of such investigations may be exploratory (ie, to plan for future implementation) or explanatory (ie, to identify causes of past implementation failures). Implementation science methods may be particularly useful for evaluating complex interventions, which are characterized by the presence of multiple interacting components that are a common feature of mHealth deployments. These methods are also well-suited to showing how changes in the implementation context may influence delivery [16]. Understanding the process and context can help improve fidelity to the core elements that drive the effectiveness of an intervention, as well as the adaptability of the intervention to the local setting [17].

The Consolidated Framework for Implementation Research (CFIR) is among the most widely used tools for planning and evaluating implementation in public health and population health [18]. It draws heavily on a rich literature on innovation in health and society [19,20]. According to the CFIR, factors at multiple levels may contribute to the success or failure of implementation, including (1) the characteristics of the intervention; (2) the inner setting where implementation is occurring; (3) the outer setting (ie, the economic, political, and social context within which an organization resides); (4) the characteristics of the implementers; and, finally, (5) the processes of implementation (Figure 1). Although the CFIR and other determinant frameworks have been widely applied to characterize implementation research [21-23], the CFIR has rarely if ever been used to examine the unique challenges associated with implementing mHealth interventions. Moreover, formal evaluations of mHealth implementation outcomes are especially important in low- and middle-income countries where information technology (IT) resources are more constrained, yet very few analyses of mHealth interventions in these settings have directly examined these challenges [24,25]. We used the CFIR to describe the process of implementing a complex mHealth intervention for tuberculosis (TB) case finding in Kampala, Uganda, in the context of a randomized controlled trial that showed no effect on the primary outcome. Our objectives in this study were to characterize the key challenges to introducing this intervention as planned to use a widely established implementation framework and to use the insights gained to propose generalizable, setting-appropriate solutions to mHealth implementation challenges.

Figure 1. The Consolidated Framework for Implementation Research (CFIR), showing the five principal domains and highlighting 4 of the 39 underlying constructs/subconstructs that were identified as the most significant barriers and facilitators in the current analysis. Barriers and facilitators associated with constructs in the Outer Setting domain were not identified in this analysis. Barriers and facilitators associated with constructs/subconstructs in the Process domain have been previously described, and are not highlighted here [26,27].



Methods

Setting

Kampala, the capital city of Uganda, has approximately 3 million people. Rapid population growth in recent years has strained the capacity of the health care infrastructure to meet public health demands, and a high burden of infectious diseases compounds these problems. Given the high penetration of mobile phones and internet access, there has been interest in using new information and communication technologies (ICT) to support health infrastructure and improve patient care [4,28]. Although many mHealth and eHealth interventions have been piloted throughout Uganda, a shortage of trained ICT support personnel and a lack of public funding for digital health has limited adoption of technology into routine health practice [29].

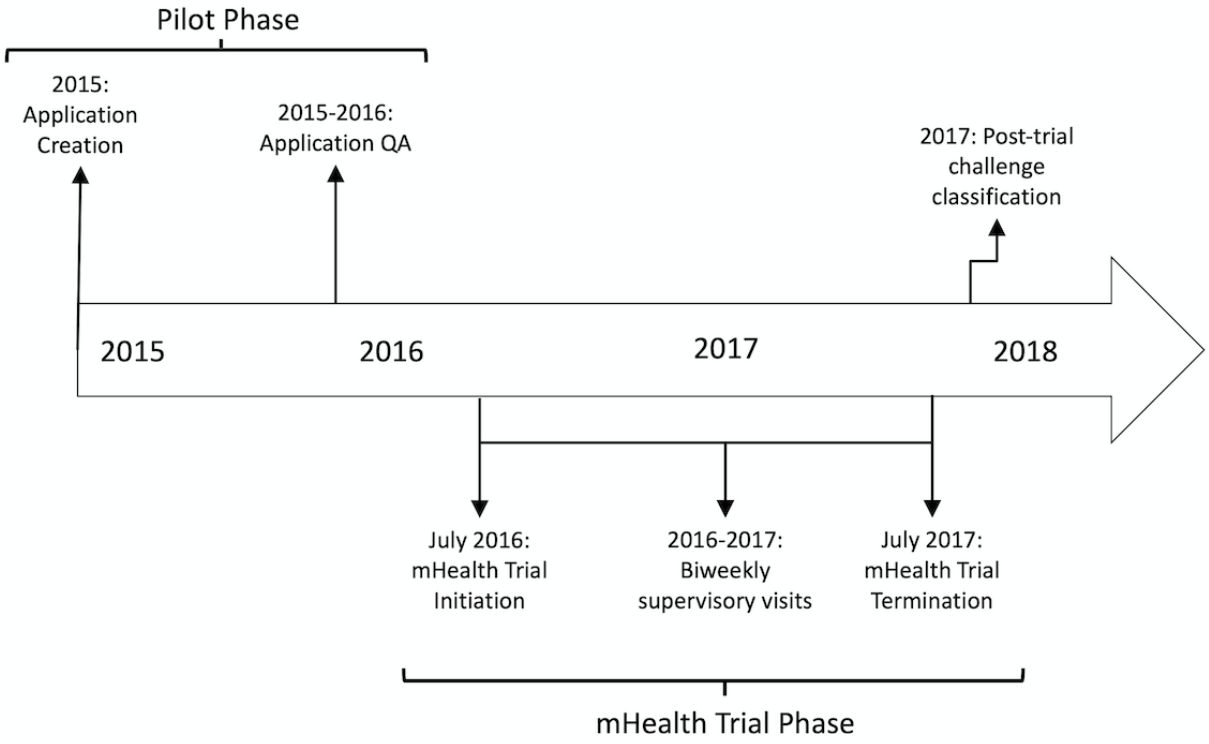
Developing the mHealth App

In 2015 (Figure 2), after a period of gathering mixed-methods data on user preferences and technology use [4], our study team, which included an experienced mHealth evaluator (JEH), an experienced mHealth implementer (DM), and commercial partners (Dimagi), created a customized mHealth app using CommCare software (Dimagi). The app was designed with two goals in mind: (1) to facilitate the implementation of household TB contact investigation using an evidence-based strategy for finding and treating close contacts with undiagnosed TB disease among household contacts of active TB patients; and (2) to evaluate this strategy in a household-randomized, controlled trial (called the "mHealth Trial," Pan-African Clinical Trials Registration #201509000877140). The app incorporated several mHealth functions, including a survey app to collect participant information; digital fingerprinting (Biometrac) to help identify and track participants between their homes and multiple clinics (given widely recognized challenges with using a name, date of birth, and address for this purpose); and automated SMS messaging to communicate TB test results and follow-up instructions. We aimed to make the app easy for community health workers (CHWs) to use as they collected data during study-related household or clinic visits. For example, we added variable logic to display relevant questions based on participant characteristics and decision-support tools to guide on-site procedures and referrals. All data could be collected with or without internet connectivity; in-built functionality enabled subsequent, automated wireless syncing with a remote cloud-based server (CommCareHQ, Dimagi) via a third-generation (3G) mobile-broadband connection (Airtel Uganda) whenever connectivity was available. Local ICT staff (DB), customized the app in collaboration with an external ICT

adviser (DM), a US-based data manager (AJM) and programmers and field officers from Dimagi.

As previously reported [30], the mHealth trial showed no improvement in the primary outcome of completing TB evaluation. Although household contacts and health workers spoke positively about the mobile app and text messaging, low rates of delivery and uptake of text-messaged instructions were observed [26,31]. Fingerprinting was also deemed acceptable, but implementation challenges resulted in inconsistent and declining use of the technology over time [27].

Figure 2. Timeline for the design and implementation of our mobile application. QA: quality assurance.



Implementation of the Mobile Health App

We piloted the app for 10 months. During this time, we developed detailed quality assurance (QA) procedures by creating test cases that reflected all possible participant characteristics. Specifically, a team of staff members systematically verified all possible survey outcomes by entering all possible responses to all questions and verifying the observed behaviors on both health worker devices and in the back-end database. We repeatedly used these test cases to confirm expected app behaviors and to localize errors in variable logic or the functioning of hardware and software. We recorded all QA results in a spreadsheet, including errors encountered or barriers to implementing changes. During this time, we also received feedback from CHWs on user preferences or errors encountered. After rigorous testing and the adoption of changes requested by CHWs, we launched the app in July 2016 with the same staff and clinics for the start of the mHealth trial. Each time a modification was proposed to the app, such as an update to survey questions or variable logic, we conducted QA in a parallel app environment to verify that the changes operated as intended and to ensure that no other parts of the app were compromised. We released new versions of the app only after completing these QA procedures and confirming the full functionality of the revised app.

Identifying Challenges to Implementation

In addition to the QA spreadsheet, study staff recorded implementation challenges in bi-weekly site visit reports, project management lists, and conference call minutes. A study coordinator (AJM) who oversaw programming of the survey app throughout the project reviewed these documents to identify key themes. We then mapped these themes using constructs and subconstructs from CFIR to classify key challenges to mHealth implementation. CFIR operationalizes these as 39 constructs/subconstructs within five previously described domains that together can be used to plan for and evaluate the success of implementation [21].

Study Approvals

The mHealth trial protocol was approved by the School of Medicine Research Ethics Committee at Makerere University

College of Health Sciences, the Uganda National Council for Science and Technology, and the Yale University Human Investigation Committee. In addition, the Uganda National TB and Leprosy Programme and the Health Directorate at the Kampala Capital City Authority approved project activities.

Results

We reviewed implementation source documents collected between July 2015 and July 2017, including site visit reports, task lists recorded in an online project management tool, and minutes from weekly team conference calls, to identify barriers to implementation. We identified challenges relevant to three of the five principal CFIR domains: (1) *intervention characteristics*, (2) *inner setting*, and (3) *characteristics of individuals* who are implementing the intervention (Figure 1). We did not identify information on *outer setting* challenges related to technology. We largely excluded the implementation process domain from this analysis, as we have previously described these factors [26,27] concerning the key construct of execution (ie, the extent to which implementation went according to plan). We found that there was low fidelity for both SMS messages and fingerprinting but high levels of acceptability and feasibility, giving insight into the overall execution of the implementation strategy, therefore allowing us to exclude this domain from this analysis.

Challenge: Adaptability and Complexity

Study staff identified limited adaptability of the mHealth app and the related hardware as a key challenge. According to CFIR, *adaptability* measures the degree to which an intervention can be altered to fit the local setting [18]. Although survey questions could be easily added or removed from the app, specialized software and hardware requirements for fingerprinting made the overall app less adaptable to the local setting. The need to have offline access to fingerprinting (eg, in the absence of internet connectivity) placed processing demands on the app that altered how participant data were collected. Offline access required custom re-coding of a sensitive section of the app that required the expertise of Dimagi engineers. The requirements for custom coding by external programmers reduced the app's flexibility and adaptability and reduced its transparency to frontline ICT managers and its flexibility for design refinements based on input from end users. Additional challenges arose because digital fingerprinting required special tablets with specific hardware and software features that were not available in Uganda and had to be procured overseas. Replacing lost or broken tablets and peripheral connector cables, therefore, proved to be time-consuming and expensive, especially as cable connectors became obsolete as technology evolved towards using wireless connections for peripheral hardware. Increased rates of hardware failure over time ultimately reduced enrollment capacity for CHWs.

The complexity of the intervention and the logistics of its implementation also hindered the adaptability of the app. Within CFIR, *complexity* refers to implementation procedures that contain multiple interacting components, the perceived difficulty of implementing the intervention, and the number of steps required to implement an intervention [18]. Altering the app was complex because to ensure sustained functionality, every change had to be (1) assessed for both perceived usefulness to the overall study and risk to the functionality of the app, (2) tested using the QA procedures, and (3) introduced and implemented at the study sites. Because the last two steps were especially laborious, we managed all changes in batches, an approach that sometimes delayed updates by up to 4 months and limited our ability to iterate upon and adapt the app to local needs rapidly.

Similarly, creating appropriate logic for the automated SMS system proved surprisingly challenging and burdensome. Ultimately, over 50 new variables had to be created and continuously updated within the app as new clinical data became available during the evaluation process, all to make sure that the app correctly assigned participants to the appropriate end states required for them to receive timely and accurate messages. When we attempted to create test users to verify every possible pathway to every possible end state, this process added to the complexity of implementation, and the enormous number of possible pathways finally made this infeasible. The complex logic ultimately contributed to errors within the app and to mistimed SMS messages that were therefore never initiated, as previously described [26].

Challenge: Knowledge and Beliefs About the Intervention

CHW knowledge and beliefs about the intervention also influenced its implementation. In CFIR, this construct focuses on users' ideas about the intervention and the value that they place on the intervention itself [18]. Study staff reported that CHWs were enthusiastic about the technology to enhance TB contact investigation during training and initial implementation. However, after experiencing technology failures during the pilot phase, their trust in the app declined, and they reported developing an expectation that the technology, especially the fingerprinting features, would fail, even after the errors had been corrected. They began to avoid using what they perceived as problematic aspects of the technology whenever possible.

Challenge: Goals and Feedback

Finally, the complex structure of the data coming out of the app slowed the development and implementation of a data management plan, making it difficult to provide timely feedback to study staff on progress towards key performance indicators for delivery of contact investigation, preventing the use of data for project management and quality assurance. *Goals and feedback*, a subconstruct in the inner setting domain, focuses on how well objectives are articulated and communicated to key stakeholders during implementation [18]. Software systems may not format data for easy management and timely dissemination to stakeholders, as others have observed [25]. In our project, consolidating, cleaning, analyzing, and disseminating data was time-consuming and required staff with advanced training in epidemiology and data management, as well as external data management and analysis apps, including Excel (Microsoft Corporation) and STATA (Stata Corporation). Extracting a master dataset required integration of 17 different worksheets using multiple unique identifiers in a complex hexadecimal format. Once created, reports could initially be

run as needed, but when intermittent but automated software upgrades to CommCareHQ introduced new data export formats, the files had to be reprogrammed several times. These barriers limited the ability of local staff to interact with the data directly.

In addition to the challenges with data structure, missing data was another challenge that local staff identified as a barrier during implementation. The survey software did not differentiate between a failure to complete a service and a failure to simply record it, rendering many indicators inaccurate and requiring local study staff to spend substantial time working with CHWs to obtain missing data, reducing time to engage with them about overall implementation goals. For example, data were often missing when a participant had declined HIV testing, often because the app did not prompt CHWs to submit a form stating whether HIV testing was not provided or was not accepted by the participant. HIV results were found to be missing at other times when mobile tablets failed to synchronize with the cloud-based server because of a lack of connectivity with the 3G network, or depletion of a tablet's pre-paid mobile service data plan. Missing data also interfered with dataset merges, preventing data updates until the missing data could be obtained, and preventing accurate projections or feedback on this key performance indicator.

Discussion

Overall, mHealth interventions offer great potential to facilitate innovative solutions to complex public health problems. Nevertheless, the very technological advances that are designed to improve health also introduce challenges. Here, using a qualitative approach and drawing on the findings of prior process-oriented analyses [26,27], we identified three main challenges related to technology that other mHealth innovators are likely to encounter when creating and implementing technological solutions in resource-constrained settings: maintaining adaptability and reducing complexity; maintaining positive beliefs about the intervention among those who deliver it; and setting goals and providing feedback in a timely and comprehensible manner to key stakeholders.

Our first significant finding was that hardware and software requirements could severely limit the adaptability of mHealth technology in a resource-limited setting. Complex digital interventions, defined as those that perform multiple interacting functions, may require specialized hardware and software elements to function effectively. Hardware inevitably requires replacement, especially with intensive use, and software may become outdated as vulnerabilities emerge or operating systems are upgraded. Thus, sourcing the components and expertise locally to maintain mHealth systems is critical for these technologies to deliver on and sustain their promise. This is in tension with the fact that necessary technologies may be either less reliable or non-existent in local settings, even though these technologies have great potential to create meaningful change in local health systems [32]. A study in Malawi that evaluated an mHealth intervention for community case management of children with acute conditions identified similar challenges [33]. Although that program appeared efficacious, key stakeholders across the Malawian health sector identified a lack of integration with local programs and the limited capacity of local ICT and management as a significant barrier to the reliability and sustainability of the program.

Similarly, a recent systematic review of mHealth technologies in developing countries identified a lack of infrastructure and local equipment as a key barrier to scale-up of mHealth programs across multiple geographic locations [34]. As a result of our experiences, we would argue that if relevant local or regional design, manufacturing, and ICT support is routinely unavailable, this capacity should ideally be developed and nurtured to achieve the successful implementation of mHealth interventions. Utilizing experts from around the globe to develop ICT, computer engineering, and informatics programs is paramount to expand local human resources to support high-quality mHealth implementation on the ground in resource-constrained settings. This may help avert many of the challenges identified here before implementation even begins. However, it will likely take several years to build sufficient infrastructure in resource-constrained settings and utilizing external mHealth experts is not always feasible or acceptable. Therefore, we would recommend that architects of new mHealth programs simplify their designs, choose interchangeable hardware components and local suppliers, and train as many local ICT staff as possible as part of their program implementation. Furthermore, we recommend that mHealth designers apply the principles of CFIR throughout the design process to anticipate better and prevent challenges in the complexity and adaptability domains [21]. In practice, this could lead to a stepwise approach to implementation to ensure that each component is successfully introduced to ensure better sustainability.

Our second significant finding was that while end users and implementers may have an overall positive view of mobile health, trust in and use of specific platforms is likely to fade over time if end users experience more than occasional failures and frustrations. In our study, CHWs reported exploiting loopholes in the mobile app to avoid fingerprinting procedures because they found them to be highly error-prone [27]. Such failures introduced embarrassing delays for participants, further undermining the motivation of CHWs to use the technology. A previous study from South Africa similarly found that doctors and nurses were resistant to change from paper-based to electronic medical records because of a fear of disrupting patient care [35]. Our findings suggest that first impressions are critical and that rigorous quality assurance checks before "going live" with new technology may be more effective than pilot studies, as these can disappoint end users, sometimes permanently. This idea is captured in another theoretical model, the Technology Acceptance Model (TAM) [36]. The TAM states that an individual's behavioral intention regarding the use of a technology is determined by the perceived usefulness and perceived ease of use, both of which may be impacted by end users' first impressions of a technology and trust in that technology over time. After software updates or app changes, it is essential to rapidly conduct quality assurance testing and quickly resolve errors for the end user to ensure that the technology is useful and easy to use. Implementers must maintain the trust of end users by evaluating their beliefs and continually incorporating their input about the intervention at

every step from design to implementation. In Sierra Leone, continuous feedback from end users to designers on how to eliminate errors and improve usability within the health care system was successfully used to refine a country-wide disease surveillance system in the wake of the Ebola crisis [37]. This is a promising approach for accelerating innovation and maintaining the satisfaction and trust of end users over time.

Our last significant finding was that the complexity of our data management system delayed feedback and prevented engaging stakeholders from achieving timelier, data-driven improvements to services. Future researchers and public health programs must consider ease of data management and accessibility to local managers when selecting software systems or considering customization. Once a software system has been selected, we recommend having a specific feedback plan in place in order to communicate goals to every level of stakeholder before implementation initiation to ensure that any mHealth technology selected will allow for appropriate communication between implementers and end users about implementation progress as well as ensure mHealth acceptability throughout. In addition, vendors that design and maintain survey software programs should prioritize consistency and usability in data management portals to facilitate continuous quality improvement activities by implementers.

This study had a few limitations. Although we collaborated with health administrators in the national TB program and local public health delivery network, we did not consult members of the Ministry of Health in charge of Uganda's digital health strategy. This gap may explain why we did not identify elements from CFIR's outer setting domain that may be highly relevant to implementing and scaling mHealth technologies. Although our focus in this project was on determining the effectiveness of our mHealth strategy before going to scale, researchers and public health officials within disease-specific areas should consider earlier consultation with policymakers to accelerate learning and adoption of best practices. Second, we used one particular software system and a limited range of hardware, which might limit the generalizability of this case study to other platforms. However, other projects in Sub-Saharan Africa have reported similar challenges with other mHealth platforms, leading us to believe that our findings transcend any specific software or hardware [24,25]. Third, new technologies are continually being developed and rapidly improved while academic evaluations such as this one move more slowly, potentially decreasing the relevance of our findings as newer technologies may not face the same challenges our study team experienced.

Our study also had important strengths. First, our study outlines operational and resource-related challenges to implementation that are specific to ICT in resource-constrained settings, complementing existing literature [38]. This work may also help address a critical gap in the TB literature, in that both the World Health Organization and a recent systematic review have called for structured evaluations of the implementation of digital health and TB [39,40]. Second, we used the CFIR to classify our challenges, thus increasing the generalizability of our results. The CFIR offers a standard set of principles for not only understanding but also anticipating implementation challenges independent of the setting. Third, we utilized information from a wide range of stakeholders, from implementing staff to end users over 4 years, resulting in a robust analysis of implementation challenges.

Overall, mHealth technologies have the potential to enhance public health programs regardless of disease or setting. However, the unique implementation challenges posed by these technologies should be examined, especially in low-resource settings, where ICT services have not yet proliferated. Implementation frameworks such as CFIR should be rigorously applied to evaluate mHealth interventions to measure not only their effectiveness but also key implementation outcomes such as fidelity and sustainability [15].

Conflicts of Interest

JEH reports consulting fees from Merck. All other authors have no conflicts of interest to declare.

References

1. 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]
2. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. PLoS One 2013;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
3. Betjeman TJ, Soghoian SE, Foran MP. mHealth in Sub-Saharan Africa. Int J Telemed Appl 2013;2013:482324 [FREE Full text] [doi: 10.1155/2013/482324] [Medline: 24369460]
4. Ggita JM, Ojok C, Meyer AJ, Farr K, Shete PB, 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 [FREE Full text] [doi: 10.5588/ijtld.17.0521] [Medline: 29663958]
5. Zurovac D, Otieno G, Kigen S, Mbithi AM, Muturi A, Snow RW, et al. Ownership and use of mobile phones among health workers, caregivers of sick children and adult patients in Kenya: cross-sectional national survey. Global Health 2013 May 14;9:20 [FREE Full text] [doi: 10.1186/1744-8603-9-20] [Medline: 23672301]

6. Hermans SM, Elbireer S, Tibakabikoba H, Hoefman BJ, Manabe YC. Text messaging to decrease tuberculosis treatment attrition in TB-HIV coinfection in Uganda. Patient Prefer Adherence 2017;11:1479-1487 [FREE Full text] [doi: 10.2147/PPA.S135540] [Medline: 28919720]
7. Hirsch-Moverman Y, Daftary A, Yuengling KA, Saito S, Ntoane M, Frederix K, et al. Using mHealth for HIV/TB Treatment Support in Lesotho: Enhancing Patient-Provider Communication in the START Study. J Acquir Immune Defic Syndr 2017 Jan 01;74 Suppl 1:S37-S43 [FREE Full text] [doi: 10.1097/QAI.0000000000001202] [Medline: 27930610]
8. Yah CS, Tambo E, Khayeka-Wandabwa C, Ngogang JY. Impact of telemonitoring approaches on integrated HIV and TB diagnosis and treatment interventions in sub-Saharan Africa: a scoping review. Health Promot Perspect 2017;7(2):60-65 [FREE Full text] [doi: 10.15171/hpp.2017.12] [Medline: 28326285]
9. Ridgeway K, Dulli LS, Murray KR, Silverstein H, Dal SL, Olsen P, et al. Interventions to improve antiretroviral therapy adherence among adolescents in low- and middle-income countries: A systematic review of the literature. PLoS One 2018;13(1):e0189770 [FREE Full text] [doi: 10.1371/journal.pone.0189770] [Medline: 29293523]
10. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
11. Ngabo F, Nguimfack J, Nwaigwe F, Mugeni C, Muhoza D, Wilson DR, et al. Designing and Implementing an Innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. Pan Afr Med J 2012;13:31 [FREE Full text] [Medline: 23330022]
12. Drake AL, Unger JA, Ronen K, Matemo D, Perrier T, DeRenzi B, et al. Evaluation of mHealth strategies to optimize adherence and efficacy of Option B+ prevention of mother-to-child HIV transmission: Rationale, design and methods of a 3-armed randomized controlled trial. Contemp Clin Trials 2017 Dec;57:44-50. [doi: 10.1016/j.cct.2017.03.007] [Medline: 28315480]
13. Cowan J, Michel C, Manhiça I, Mutaquiha C, Monivo C, Saize D, et al. Remote monitoring of Xpert® MTB/RIF testing in Mozambique: results of programmatic implementation of GxAlert. Int J Tuberc Lung Dis 2016 Mar;20(3):335-341. [doi: 10.5588/ijtld.15.0535] [Medline: 27046714]
14. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [Medline: 25361730]
15. Labrique A, Vasudevan L, Chang LW, Mehl G. H_pe for mHealth: more. Int J Med Inform 2013 May;82(5):467-469 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.016] [Medline: 23279850]
16. 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]
17. Pérez D, Van DSP, 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]
18. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
19. Rogers EM. Diffusion of Innovations. In: New York: Free Press. New York, United States: Free Press; 2003.
20. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
21. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the Consolidated Framework for Implementation Research. Implement Sci 2016 May 17;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
22. Birken SA, Powell BJ, Presseau J, Kirk MA, Lorencatto F, Gould NJ, et al. Combined use of the Consolidated Framework for Implementation Research (CFIR) and the Theoretical Domains Framework (TDF): a systematic review. Implement Sci 2017 Dec 05;12(1):2 [FREE Full text] [doi: 10.1186/s13012-016-0534-z] [Medline: 28057049]
23. Means AR, Kemp CG, Gwayi-Chore M, Gimbel S, Soi C, Sherr K, et al. Evaluating and optimizing the consolidated framework for implementation research (CFIR) for use in low- and middle-income countries: a systematic review. Implement Sci 2020 Mar 12;15(1):17 [FREE Full text] [doi: 10.1186/s13012-020-0977-0] [Medline: 32164692]
24. Bardosh KL, Murray M, Khaemba AM, Smillie K, Lester R. Operationalizing mHealth to improve patient care: a qualitative implementation science evaluation of the WelTel texting intervention in Canada and Kenya. Global Health 2017 Dec 06;13(1):87 [FREE Full text] [doi: 10.1186/s12992-017-0311-z] [Medline: 29208026]
25. Brown W, Giguere R, Sheinfil A, Ibitoye M, Balan I, Ho T, et al. Challenges and solutions implementing an SMS text message-based survey CASI and adherence reminders in an international biomedical HIV PrEP study (MTN 017). J Biomed Inform 2018 Apr;80:78-86 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.018] [Medline: 29501908]
26. Meyer AJ, Babirye D, Armstrong-Hough M, Mark D, Ayakaka I, Katamba A, et al. Text Messages Sent to Household Tuberculosis Contacts in Kampala, Uganda: Process Evaluation. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e10239 [FREE Full text] [doi: 10.2196/10239] [Medline: 30459147]

27. White EB, Meyer AJ, Ggita JM, Babirye D, Mark D, Ayakaka I, et al. Feasibility, Acceptability, and Adoption of Digital Fingerprinting During Contact Investigation for Tuberculosis in Kampala, Uganda: A Parallel-Convergent Mixed-Methods Analysis. J Med Internet Res 2018 Nov 15;20(11):e11541 [FREE Full text] [doi: 10.2196/11541] [Medline: 30442637]
28. Commission U. Post, Broadcasting and Telecommunications Market & Industry Q2 Report. 2018. URL: https://www.ucc.co.ug/wp-content/uploads/2017/09/Communication-Sector-Performance-for-the-Quarter-ending-June-2018.pdf [accessed 2020-06-25]
29. Kiberu VM, Mars M, Scott RE. Barriers and opportunities to implementation of sustainable e-Health programmes in Uganda: A literature review. Afr J Prim Health Care Fam Med 2017 May 29;9(1):e1-e10 [FREE Full text] [doi: 10.4102/phcfm.v9i1.1277] [Medline: 28582996]
30. Davis JL, Turimumahoro P, Meyer AJ, Ayakaka I, Ochom E, Ggita J, et al. Home-based tuberculosis contact investigation in Uganda: a household randomised trial. ERJ Open Res 2019 Jul;5(3) [FREE Full text] [doi: 10.1183/23120541.00112-2019] [Medline: 31367636]
31. Ggita JM, Katahoire A, Meyer AJ, Nansubuga E, Nalugwa T, Turimumahoro P, et al. Experiences and intentions of Ugandan household tuberculosis contacts receiving test results via text message: an exploratory study. BMC Public Health 2020 Mar 12;20(1):310 [FREE Full text] [doi: 10.1186/s12889-020-8427-0] [Medline: 32164612]
32. World Health Organization. Digital health for the END TB Strategy - an agenda for action. Geneva, Switzerland: World Health Organization, The END TB Strategy, European Respiratory Society. 2015. URL: https://www.who.int/tb/areas-of-work/digital-health/Digital_health_EndTBstrategy.pdf?ua=1 [accessed 2020-06-25]
33. Dharmayat KI, Tran T, Hardy V, Chirambo BG, Thompson MJ, Ide N, et al. Sustainability of 'mHealth' interventions in sub- Saharan Africa: a stakeholder analysis of an electronic community case management project in Malawi. Malawi Med J 2019 Sep;31(3):177-183 [FREE Full text] [doi: 10.4314/mmj.v31i3.3] [Medline: 31839886]
34. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the Use of Mobile Health in Improving Health Outcomes in Developing Countries: Systematic Review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
35. Cline GB, Luiz JM. Information technology systems in public sector health facilities in developing countries: the case of South Africa. BMC Med Inform Decis Mak 2013;13:13 [FREE Full text] [doi: 10.1186/1472-6947-13-13] [Medline: 23347433]
36. Tao D. Understanding intention to use electronic information resources: A theoretical extension of the technology acceptance model (TAM). AMIA Annu Symp Proc 2008 Nov 06:717-721 [FREE Full text] [Medline: 18999300]
37. Martin DW, Sloan ML, Gleason BL, de Wit L, Vandi MA, Kargbo DK, et al. Implementing Nationwide Facility-based Electronic Disease Surveillance in Sierra Leone: Lessons Learned. Health Secur 2020 Jan;18(S1):S72-S80. [doi: 10.1089/hs.2019.0081] [Medline: 32004124]
38. van Olmen J, Erwin E, García-Ulloa AC, Meessen B, Miranda JJ, Bobrow K, et al. Implementation barriers for mHealth for non-communicable diseases management in low and middle income countries: a scoping review and field-based views from implementers. Wellcome Open Res 2020;5:7 [FREE Full text] [doi: 10.12688/wellcomeopenres.15581.2] [Medline: 32399497]
39. Falzon D, Timimi H, Kurosinski P, Migliori GB, Van GW, Denkinger C, et al. Digital health for the End TB Strategy: developing priority products and making them work. Eur Respir J 2016 Jul;48(1):29-45 [FREE Full text] [doi: 10.1183/13993003.00424-2016] [Medline: 27230443]
40. Ngwatu BK, Nsengiyumva NP, Oxlade O, Mappin-Kasirer B, Nguyen NL, Jaramillo E, Collaborative group on the impact of digital technologies on TB. The impact of digital health technologies on tuberculosis treatment: a systematic review. Eur Respir J 2018 Jan;51(1) [FREE Full text] [doi: 10.1183/13993003.01596-2017] [Medline: 29326332]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
CHW: community health workers
eHealth: electronic health
ICT: information and communication technologies
IT: information technology
mHealth: mobile health
QA: quality assurance
TB: tuberculosis

Edited by G Eysenbach; submitted 22.04.20; peer-reviewed by R Scott, M Mars, C Kruse, R Lester; comments to author 13.05.20; revised version received 01.06.20; accepted 21.06.20; published 13.07.20.

Please cite as:
Meyer AJ, Armstrong-Hough M, Babirye D, Mark D, Turimumahoro P, Ayakaka I, Haberer JE, Katamba A, Davis JL
Implementing mHealth Interventions in a Resource-Constrained Setting: Case Study From Uganda
JMIR Mhealth Uhealth 2020;8(7):e19552
URL: http://mhealth.jmir.org/2020/7/e19552/
doi:10.2196/19552
PMID:32673262

Original Paper

Automated and Interviewer-Administered Mobile Phone Surveys in Burkina Faso: Sociodemographic Differences Among Female Mobile Phone Survey Respondents and Nonrespondents

Abigail R Greenleaf[1], PhD; Aliou Gadiaga[2], MSc; Yoonjoung Choi[3], DrPH; Georges Guiella[2], PhD; Shani Turke[4], MSPH; Noelle Battle[4], MPH; Saifuddin Ahmed[4], PhD; Caroline Moreau[4,5], MD

[1]ICAP at Columbia University, New York, NY, United States

[2]Institut Supérieur des Sciences de la Population, University of Ouagadougou, Ouagadougou, Burkina Faso

[3]iSquared, Information x Insight, Severna Park, MD, United States

[4]Department of Population, Family and Reproductive Health, Johns Hopkins University, Baltimore, MD, United States

[5]Soins et santé, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France

Corresponding Author:
Abigail R Greenleaf, PhD
ICAP at Columbia University
722 W 168th St
New York, NY, 10032
United States
Phone: 1 4439553694
Email: arg2177@cumc.columbia.edu

Abstract

Background: The remarkable growth of cell phone ownership in low- and middle-income countries has generated significant interest in using cell phones for conducting surveys through computer-assisted telephone interviews, live interviewer-administered surveys, or automated surveys (ie, interactive voice response).

Objective: This study aimed to compare, by mode, the sociodemographic characteristics of cell phone owners who completed a follow-up phone survey with those who did not complete the survey.

Methods: The study was based on a nationally representative sample of women aged 15 to 49 years who reported cell phone ownership during a household survey in Burkina Faso in 2016. Female cell phone owners were randomized to participate in a computer-assisted telephone interview or hybrid interactive voice response follow-up phone survey 11 months after baseline interviews. Completion of the phone survey was defined as participants responding to more than 50% of questions in the phone survey. We investigated sociodemographic characteristics associated with cell phone survey completion using multivariable logistic regression models, stratifying the analysis by survey mode and by directly comparing computer-assisted telephone interview and hybrid interactive voice response respondents.

Results: A total of 1766 women were called for the phone survey between November 5 and 17, 2017. In both the computer-assisted telephone interview and hybrid interactive voice response samples, women in urban communities and women with secondary education or higher were more likely to complete the survey than their rural and less-educated counterparts. Compared directly, women who completed the hybrid interactive voice response survey had higher odds of having a secondary education than those who completed computer-assisted telephone interviews (odds ratio 1.7, 95% CI 1.1-2.6).

Conclusions: In Burkina Faso, computer-assisted telephone interviews are the preferred method of conducting cell phone surveys owing to less sample distortion and a higher response rate compared with a hybrid interactive voice response survey.

(JMIR Mhealth Uhealth 2020;8(7):e17891) doi:10.2196/17891

KEYWORDS

cell phone; mHealth; Africa South of the Sahara; Burkina Faso; methodology, survey, nonrespondents, survey methods, interviews, telephone

Introduction

Background

Cell phone ownership in sub-Saharan Africa (SSA) has quadrupled over the past 10 years, rising to 444 million unique subscribers in 2017 [1]. Concomitant to increased cell phone ownership in SSA is a rise in the use of cell phones for remote data collection [2]. One of the first attempts to create a nationally representative estimate from an interactive voice response (IVR) survey in SSA took place in Ethiopia, Mozambique, and Zimbabwe in 2015 [3]. Since this study, phone survey research expanded and includes the use of computer-assisted telephone interviews (CATIs) and SMS surveys, but the optimal mode for nationally representative cell phone surveys is yet to be identified.

A CATI involves a live interviewer who administers the full survey, usually in a call center. In contrast, an IVR survey is conducted without an interviewer and requires that the respondents use their keypad to answer a prerecorded question or prompt (eg, "If yes, press 1. If no, press 2"). A hybrid IVR survey is a modified version of an IVR survey: a live interviewer opens the call, confirms respondent eligibility, administers consent, and explains how to respond to an IVR question before transferring the respondent to the IVR survey. Finally, an SMS approach asks the respondent to answer questions via a text message [4].

Although these remote modes hold the possibility of more rapid, cost-effective data collection, a transition from face-to-face (FTF) survey to cell phone data collection raises issues of survey quality. To operationalize survey quality, survey researchers devised the concept of total survey error, comprising 5 components: frame, nonresponse, specification, measurement, and data processing errors [5,6].

This analysis addresses nonresponse error, which occurs when people who are sampled but not interviewed differ in a non-negligible way from those who are interviewed [5]. In telephone surveys, nonresponse generally stems from 3 causes: (1) failure to contact sampled respondents, (2) refusal to participate, and (3) ability or language constraints [7]. Nonresponse bias occurs when the outcome of interest is systematically different between people who are sampled but who do not complete the survey and those who complete the survey [5,8].

Although there is a well-established body of literature on causes of error in phone surveys in high-income countries [9,10], nonresponse error and bias in cell phone surveys have been minimally explored in SSA because of the recency of the approach in the continent. Among the few phone survey studies conducted in low- and middle-income countries (LMICs), the representativeness of a sample is rarely considered. Rather, studies have mainly addressed the feasibility and measurement error [11]. Studies that attempt to profile nonrespondents often assess sample distortion by comparing phone survey respondents with a reference population based on census or Demographic and Health Survey (DHS) data or comparing early and late responders [6,12,13]. Moreover, 2 recent studies in SSA (Ghana: IVR and Cote d'Ivoire: CATI) found that completers of cell phone surveys were more likely to be young, urban, educated, and male compared with distributions in a representative sample survey [12,13]. Although comparing a phone study sample and a DHS sample can provide some insight on nonrespondents, the comparison cannot distinguish a frame error (cell phone ownership) from a nonresponse error (not answering the survey phone call among people who own phones) because the sample frame is unknown. The paucity of remote data collection studies in SSA and, in particular, studies using a known representative sampling frame to assess nonresponse bias is a notable knowledge gap.

To address this gap, we conducted a study in Burkina Faso using the Performance Monitoring and Accountability 2020 (PMA2020) platform. Since 2014, the PMA2020 platform has annually tracked family planning indicators and evaluated the impact of specific family planning programs using FTF surveys among a nationally representative probabilistic sample. The PMA2020 survey has completed 5 rounds of data collection in Burkina Faso and is exploring new approaches, including cell phone surveys, to collect quality data at a lower cost [6,11,12] The 2016 PMA2020 survey in Burkina Faso found that 86% of households owned a cell phone and 47% of women reported personal cell phone ownership [14].

Objectives

The objective of this research was to identify sociodemographic characteristics related to phone survey completion among a representative sample of female phone owners in Burkina Faso. We examined these questions for 2 phone survey modes: CATI and hybrid IVR.

Methods

Study Design

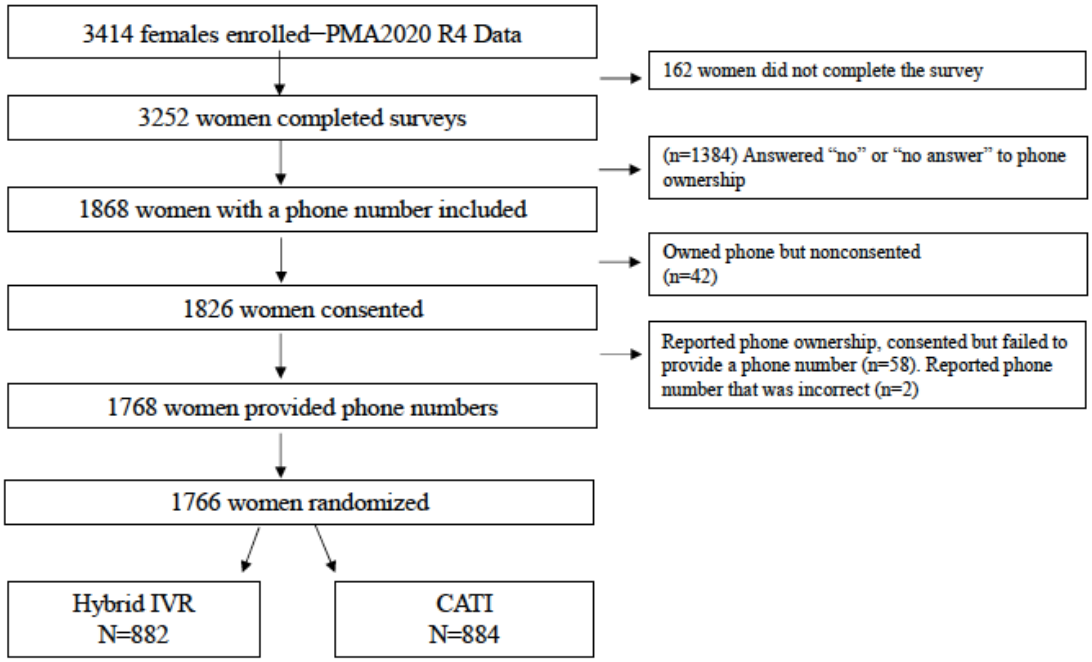
The study used a nationally representative sample of women of reproductive age, who owned a cell phone, identified in the Burkina Faso PMA2020 Round 4 (R4) survey. These female cell phone owners were randomized to receive a phone follow-up survey using either a CATI or a hybrid IVR. Hybrid IVR was preferred over IVR because it was hypothesized that having a live interviewer introduce the survey would increase response rates and improve data quality, given that only 30% of women in Burkina Faso are literate [15]. Phone follow-up occurred 11 months after the PMA2020 R4 FTF survey. The phone survey was a shortened version of PMA2020's standard FTF questionnaire.

Participants

The PMA2020 R4 survey in Burkina Faso used a two-stage stratified cluster design, starting with a selection of 83 enumeration areas using probability proportional to size sampling. The enumeration areas were stratified by rural or urban geographies; then, 35 of the approximately 200 households within each enumeration area were randomly selected. Within each sampled household, all women aged 15 to 49 years were eligible to participate, and those who provided written consent were included. Women who met the FTF inclusion criteria, had a cell phone, consented to be contacted

for a follow-up survey, and provided a phone number were eligible for follow-up. Of the 3215 female respondents interviewed during PMA2020 R4 (95.4% response rate), 1839 (57.20%) owned a cell phone; among them, 1766 (96.03%) consented to be contacted again for this study (Figure 1). The aforementioned 1766 women were assigned a random number in Stata (using the *generate random* command; StataCorp 2017) and then sorted by random numbers (from smallest to largest). Women were divided into 6 language groups (5 languages and one *other language* category) and then divided again into groups of approximately 120 women, assigned to an arm (hybrid IVR: n=882 or CATI: n=884), and then finally assigned to an interviewer. During the phone follow-up, women heard an abbreviated standardized verbal disclosure (consent) of key study information that emphasized the voluntary nature of the survey, then consented or refused.

Figure 1. Study enrollment flow chart.



Study Procedures

A total of 15 trained interviewers conducted the phone follow-up survey, operating from a call center in Ouagadougou 7 days a week between 8 AM and 8 PM. Interviewers called all eligible women (consenting female cell phone owners from PMA2020 R4).

Respondents were called up to 6 times at varying times of the day and at least once during the weekend. If a respondent answered and wanted to be called back within 15 min, the operator could accommodate the request, otherwise the respondent was called back the next day. A respondent was not called back if she refused to consent to the study or if someone answered the call who did not know the respondent. All respondents saw the same phone number appearing on their cell phone, which was identifiable as a landline phone number. The call center phone number could not be called back. Women who completed the survey were sent the equivalent of US $1 phone credit the day after completing the interview.

Questionnaire

The phone surveys (CATI and hybrid IVR) used the same question wording as the PMA2020 FTF questionnaire with only minor modifications. They included 17 questions: 5 introductory questions to identify the respondent, 4 demographic questions (age, area of residence, marital status, and parity), 5 questions about the respondent's awareness of modern contraceptive methods (intrauterine device, implant, condom, pills, and injectables), and 3 questions on contraceptive use (current use, previous method used, or pregnancy intention). After consent was administered, women receiving the CATI were asked if the interviewer could call the participant back should the call drop. For women receiving the hybrid IVR survey, after consent, they heard an explanation of what to expect during an IVR survey and then were asked to press 1 on their keypad. If the respondent was unable to press 1, she was considered incapable of participating in the IVR survey, and the interview ended. If the woman pressed 1, she was transferred to a prerecorded voice providing instructions about how to repeat or skip a question. Then, she was asked a practice question about the country that she currently lives in. From this point on, the questionnaires for both modes were identical until the last question. The last question in the CATI asked the respondent which region she lived in, and the last IVR question asked the respondent to enter her age.

Of the 17 survey questions, 1 question was numeric (age), 3 were multiple choice, and 12 were binary with a yes or no response option. The key question "What are you or your partner currently using to delay or avoid a pregnancy?" was field coded by interviewers in the CATI, and was presented as a multiple-choice question with 6 response options in the IVR survey. For the IVR survey, the school question was divided into 2 questions: first, asking if the respondent had ever been to school, and second, if the respondent answered affirmatively,

asking her up to what level she had studied. Both the CATI and hybrid IVR surveys were offered in 5 languages. Each respondent received a call in the language used in her PMA2020 FTF interview. The questionnaire was written in French in the software, and CATI interviewers translated in real time, after having developed and refined consistent translations in the local language during the training.

Ethical Approval

Both the FTF PMA2020 survey and phone survey received institutional review board approval from the Johns Hopkins Bloomberg School of Public Health and from the ethical committee in Burkina Faso, *Comité d'éthique pour la recherche en santé*.

American Association for Public Opinion Research Measures

Using the American Association for Public Opinion Research (AAPOR) call disposition codes [16] and study-specific considerations, we defined 8 respondent categories. All participants were eligible and were categorized as respondents or nonrespondents. Our outcome of interest was *completing the phone survey*, defined as respondents answering 50% or more of the relevant survey questions (eg, those assigned partial and complete call disposition codes). The follow-up data were only used to classify respondents into 8 groups (disposition codes) and thus calculate survey outcome rates.

Respondent Characteristics Measures

Sociodemographic information was collected during PMA2020 R4 and therefore available for all eligible women, regardless of whether the woman responded to the phone follow-up survey. Sociodemographic variables included age, categorized into 5-year age groups for descriptive analysis and into 4 age groups for logistic regressions (15-19, 20-29, 30-39, and 40-49 years). Additional covariates were current union status (in union—ie, currently married or living with a partner vs not in union), residential area (urban vs rural), highest grade level attended (none, primary, and secondary or higher), household wealth (lowest, middle, and highest tertiles), having electricity (yes vs no), and parity (ever given birth vs never given birth). Household wealth in PMA2020 surveys is a summary measure of the relative wealth of a household in the country based on household assets and housing conditions. The indicator is constructed using a principal component analysis, which is widely used in population-based surveys such as DHS surveys [17].

Statistical Analysis

To examine whether randomized groups (women assigned to the CATI vs hybrid IVR survey) were similar according to their background characteristics at baseline (PMA2020 R4 survey), we explored the distribution of selected demographic and socioeconomic variables by mode. We used American Association for Public Opinion Research call disposition classifications [16] to classify respondents and calculated 4 AAPOR's phone survey outcome rates (Table 1).

Table 1. Final disposition code by study arm (computer-assisted telephone interview vs hybrid interactive voice response) among female cell phone owners in Burkina Faso.

Participants, American Association for Public Opinion Research categorization[a], and explanation		Hybrid interactive voice response (n=882), n (%)	Computer-assisted telephone interview (n=884), n (%)
Nonrespondents			
NC[b] (2.20)	Noncontact (did not pick up)	225 (25.5)	228 (25.8)
O[c] (2.36)	Noncontact (someone picked up the phone call, but interviewer never spoke with the woman)	173 (19.6)	134 (15.2)
R[d] (2.12)	Refusal preconsent	90 (10.2)	54 (6.1)
R (2.111)	Refusal	14 (1.6)	6 (0.7)
R (2.121)	Break-off interactive voice response	31 (3.5)	N/A[e]
R (2.12)	Break-off (consented but <50% of relevant questions answered)	174 (19.7)	17 (1.9)
Respondents			
P[f] (1.2)	Partial (50%-80% of relevant questions answered)	18 (2.0)	5 (0.6)
I[g] (1.1)	Complete (more than 80% of relevant questions answered)	157 (17.8)	440 (49.8)

[a]Final disposition codes for random digit dial telephone surveys.

[b]NC: noncontact.

[c]O: other.

[d]R: refusal and break-off.

[e]N/A: not applicable.

[f]P: partial interview.

[g]I: complete interview.

Analyses were stratified by mode of data collection (CATI and hybrid IVR). All eligible women were included in the analysis. We compared the distribution of the sociodemographic characteristics between completers and noncompleters using chi-square tests. We then conducted multivariable logistic regression models to identify the independent factors associated with survey completion. We fitted the models separately for each survey mode (hybrid IVR and CATI). We did not include marital status or parity in the multivariable models because these characteristics were not significantly related to completion in the bivariate analyses. Electricity was also not included because of its high correlation with wealth. Finally, using multivariable logistic regression, we pooled hybrid IVR and CATI survey completers to compare the distribution of survey mode by background characteristics to evaluate whether background characteristics were different by mode. All analyses were conducted using Stata version 15.

Results

Respondent Characteristics

The average age of the 1766 women enrolled in this study—all of whom were contacted by phone at least once between November 5 and 17, 2017—was 28.5 years (SD 8.9) (Table 2). Overall, 42.02% (742/1766) of women had never attended school, 19.85% (351/1766) had attended primary school, and 38.12% (674/1766) had secondary or higher education. Of these, 65.27% (1154/1766) were married and 71.91% (1270/1766) were parous. Moreover, 66.57% (1177/1766) of the women were living in rural areas and 61.99% (1096/1766) had electricity in the household. Finally, 66.57% (1177/1766) were in the highest wealth tertile, with 18.10% (320/1766) in the middle tertile and 15.33% (271/1766) in the lowest tertile. After randomization, women in the hybrid IVR and CATI groups were similar across all sociodemographic characteristics examined in this study (Table 2).

Table 2. Sample characteristics of female cell phone owners in Burkina Faso, overall and by study arm.

Characteristics	Total study population (N=1766)	Hybrid interactive voice response arm (n=882)	Computer-assisted telephone interview arm (n=884)
Age (years)			
Mean (SD)	28.5 (8.9)	28.6 (9.1)	28.4 (8.7)
Range, n (%)			
15-19	338 (19.12)	163 (18.5)	175 (19.8)
20-24	351 (19.85)	175 (19.8)	175 (19.8)
25-29	342 (19.34)	176 (19.9)	166 (18.8)
30-34	265 (14.99)	126 (14.3)	139 (15.7)
35-39	207 (11.71)	192 (11.6)	105 (11.9)
40-44	158 (8.94)	82 (9.3)	76 (8.6)
45-49	107 (6.05)	58 (6.6)	48 (5.4)
Residential area, n (%)			
Urban	591 (33.43)	294 (33.3)	295 (33.4)
Rural	1177 (66.57)	588 (66.7)	589 (66.6)
Marital status, n (%)			
Currently not in union	614 (34.73)	312 (35.4)	301 (34.1)
Currently in union	1154 (65.27)	568 (64.4)	583 (65.9)
Highest school attended, n (%)			
Never	743 (42.02)	370 (42.0)	370 (41.9)
Primary	351 (19.85)	181 (20.5)	170 (19.2)
Secondary or higher	674 (38.12)	330 (37.4)	344 (38.9)
HH[a] wealth (tertile), n (%)			
Lowest	271 (15.33)	137 (15.5)	133 (15.1)
Middle	320 (18.10)	155 (17.6)	165 (18.7)
High	1177 (66.57)	590 (66.9)	586 (66.3)
Parity, n (%)			
Yes	1270 (71.91)	639 (72.5)	630 (71.3)
No	496 (28.08)	243 (27.5)	254 (28.7)
HH electricity, n (%)			
Yes	1096 (61.99)	548 (62.1)	548 (62.0)
No	672 (38.01)	334 (37.9)	336 (38.0)

[a]HH: household.

American Association for Public Opinion Research Outcomes

The percentage of eligible women who did not answer any of the 6 calls was the same among the 2 modes (hybrid IVR: 25.5% (225/882); and CATI: 25.8% (228/884); NC in Table 1). Overall, 47.11% (832/1766) women consented to participate in the phone interview when called: 43.1% (380/882) of women in the hybrid IVR arm and 52.2% (462/884) of women in the CATI arm (P<.001; AAPOR categories: R [2.121], R [2.12], P [1.2], and I [1.1] in Table 1). Break-off, defined as consenting but answering less than 50% of the questions, was substantially higher for the women in the hybrid IVR arm 23.2% (205/882) than those in the CATI arm 1.9% (17/884). Altogether, among all women randomized to the hybrid IVR arm, 19.8% (175/882) completed the survey (partial interviews: 18/882, 2.0%; complete interviews: 157/882, 17.8%), whereas 50.4% (445/884) of those randomized to the CATI arm completed the survey (partial interviews: 5/884, 0.6%; and complete interviews: 440/884, 49.8%).

The 4 essential AAPOR survey outcome rates (response, cooperation, refusal, and contact) based on disposition codes are presented in Table 3. The contact rate, the proportion of all cases in which a responsible member of the unit was reached by the survey, was similar between the 2 modes (54.8% for

hybrid IVR and 59.1% for CATI). The refusal rate for hybrid IVR surveys (11.9%) was almost double the refusal rate for CATI (6.8%), although the difference was not statistically significant (*P*=.23). The cooperation rate, defined as completed interviews over contacted respondents, for CATIs (85.2%) dwarfs the cooperation rate for hybrid IVR surveys (35.9%; cooperation rate 2). Finally, the response rate, a measure of the number of partial or completed interviews over all eligible respondents, was twice as high in the CATI arm compared with the hybrid IVR arm. Specifically, the response rate among women in the CATI arm was 50.3%, compared with a response rate of 19.8% among women in the hybrid IVR arm (response rate 6).

Table 3. Follow-up phone call outcome rates by study arm (computer-assisted telephone interview vs hybrid interactive voice response) among female cell phone owners in Burkina Faso.

Outcome rates	Hybrid interactive voice response arm (n=882)	Computer-assisted telephone interview arm (n=884)
Response rates		
Response rate 5:	156 (17.7)	439 (49.7)
Response rate 6:	175 (19.8)	445 (50.3)
Cooperation rates		
Cooperation rate 1:	284 (32.2)	743 (84.0)
Cooperation rate 2:	317 (35.9)	753 (85.2)
Refusal rate		
Refusal rate 3:	105 (11.9)	60 (6.8)
Contact rate		
Contact rate 3:	483 (54.8)	522 (59.1)

Bivariate and Multivariable Logistic Regressions

In the bivariate analyses (Table 4), we found that among hybrid IVR respondents, survey completion was not dependent on age (*P*=.48) but was lower among women living in rural areas than those living in urban areas (27/294, 9.2% vs 148/588, 25%; *P*=.002). The proportion of women who completed the survey was lower among women with no education compared with those with secondary education or higher (46/371, 12.4% vs 94/330, 28.5%; *P*<.001). Twice as many women who had household electricity completed the survey compared with women who had no electricity in the household (133/548, 24.3% vs 42/334, 12.6%; *P*<.001), and 3 times as many women in the highest tertile completed the survey compared with those in the lowest tertile (12/137, 8.8% vs 146/590, 25%; *P*<.001).

The patterns of completion were similar for CATIs, with residential area, wealth tertile, and household electricity being associated with completion. However, education was unrelated to CATI completion 45.9%, (170/370) of those with no education vs 53.8% (185/344) of those with secondary education or higher; *P*=.84. On the other hand, survey completion was dependent on age, with only 33% of women aged 15 to 19 years completing the survey, compared with 45.1% to 63.3% of women aged 20 to 49 years (*P*<.001). The proportion of urban women who completed the survey was greater than the proportion of rural women who completed the survey (333/589, 56.5% vs 112/295, 38.0%; *P*<.001).

The results from the multivariable logistic regression models comparing women who completed and those who did not complete the hybrid IVR survey and CATI are presented in Table 5. Women who were aged 30 years and older were more likely to complete the hybrid IVR survey than women aged 15 to 19 years, as well as women with secondary education, who had 2.5 times the odds of completing the survey relative to women with no education (95% CI 1.6-3.9). Urban women had twice the odds of completing the survey compared with rural women (95% CI 1.1. 3.6).

The odds of CATI completion were elevated for women aged 20 years and older compared with teenagers. Women in urban communities were more likely to complete the CATI survey compared with their rural counterparts (odds ratio [OR] 1.7, 95% CI 1.1-2.5) as were women with a secondary education or more who had 40% higher odds of completing the survey than women with no education (OR 1.4, 95% CI 1.1-2.0).

In multivariable logistic regression among survey completers (n=620), which assessed the survey mode according to women's background characteristics, the only significant difference was education: women who attended secondary or higher education were 70% more likely to be hybrid IVR completers than those who had no education (OR 1.7, 95% CI 1.1-2.6; Table 6).

Table 4. Percent completing the hybrid interactive voice response or computer-assisted telephone interview survey by women's sociodemographic characteristics among female cell phone owners randomized to each arm.

Characteristics	Hybrid interactive voice response			Computer-assisted telephone interview		
	Participants who did not complete the survey	Participants who completed the survey	*P* value	Participants who did not complete the survey	Participants who completed the survey	*P* value
Total, n (%)	707 (80)	175 (20)	N/A[c]	439 (49)	445 (51)	N/A
Age (years), n (%)			.48			<.001
15-19	137 (84.1)	26 (15.9)		116 (66.3)	59 (33.7)	
20-24	137 (78.3)	38 (21.7)		96 (54.9)	79 (45.1)	
25-29	145 (82.4)	31 (17.6)		66 (39.8)	100 (60.2)	
30-34	94 (74.6)	32 (25.4)		55 (39.6)	84 (60.4)	
35-39	84 (82.4)	18 (17.6)		56 (53.3)	49 (46.7)	
40-44	65 (79.3)	17 (20.7)		28 (36.8)	48 (63.2)	
45-49	45 (77.6)	13 (22.4)		22 (45.8)	26 (54.2)	
Residential area, n (%)			<.001			<.001
Rural	267 (90.8)	27 (9.2)		183 (62.0)	112 (38.0)	
Urban	440 (74.8)	148 (25.2)		256 (43.5)	333 (56.5)	
Marital status, n (%)			.28			.94
Currently not in union	244 (78.2)	68 (21.8)		150 (49.8)	151 (50.2)	
Currently in union	463 (81.2)	107 (18.8)		289 (49.6)	294 (50.4)	
Highest school attended, n (%)			<.001			.08
Never	325 (87.6)	46 (12.4)		200 (54.1)	170 (45.9)	
Primary	146 (80.7)	35 (19.3)		80 (47.1)	90 (52.9)	
Secondary or higher	236 (71.5)	94 (28.5)		159 (46.2)	185 (53.8)	
HH[a] wealth (tertile), n (%)			<.001			<.001
Lowest	125 (91.2)	12 (8.8)		86 (64.7)	47 (35.3)	
Middle	138 (89.0)	17 (11.0)		92 (55.8)	73 (44.2)	
Highest	444 (75.3)	146 (24.7)		261 (44.5)	325 (55.5)	
Parity, n (%)			.50			.13
No	200 (78.7)	54 (21.3)		143 (53.6)	124 (46.4)	
Yes	507 (80.7)	121 (19.3)		296 (48.0)	321 (52.0)	
HH electricity, n (%)			<.001			.005
No	292 (87.4)	42 (12.6)		187 (55.7)	149 (44.3)	
Yes	415 (75.7)	133 (24.3)		252 (46.0)	296 (54.0)	

[a]HH: household.

Table 5. Odds of completing the phone follow-up survey by background characteristics: multivariable logistic regression analyses among the 2 study arms. Italics indicates *P*<.05.

Characteristics	Hybrid interactive voice response completers versus noncompleters (n=882), adjusted odds ratio (95% CI)	Computer-assisted telephone interview completers versus noncompleters (n=884), adjusted odds ratio (95% CI)
Age group (years)		
15-19	Reference	Reference
20-29	1.4 (0.8-2.3)	*2.3 (1.5-3.4)*
30-39	*2.1 (1.2-3.8)*	*2.7 (1.8-4.2)*
40-49	*2.1 (1.1-4.0)*	*3.8 (2.3-6.4)*
Residential area		
Rural	Reference	Reference
Urban	*2.0 (1.1-3.6)*	*1.7 (1.1-2.5)*
Highest school attended		
No education	Reference	Reference
Primary	1.5 (0.9-2.4)	1.3 (0.9-1.9)
Secondary or more	*2.5 (1.6-3.9)*	*1.4 (1.1-2.0)*
Household wealth (tertile)		
Lowest (reference all other groups)	0.8 (0.4-1.8)	0.7 (0.4-1.2)
Highest (reference all other groups)	1.3 (0.7-2.5)	1.1 (0.7-1.7)

Table 6. Odds of completing the hybrid interactive voice response survey compared with completing the computer-assisted telephone interview survey by background characteristics: multivariable logistic regression analyses. Italics indicates *P* value below .05.

Characteristics	Hybrid interactive voice response versus computer-assisted telephone interview among survey completers (N=620; reference group: CATI), adjusted odds ratio (95% CI)
Age group (years)	
15-19	Reference
20-29	0.9 (0.5-1.5)
30-39	0.9 (0.5-1.7)
40-49	1.1 (0.5-2.1)
Residential area	
Rural	Reference
Urban	1.4 (0.8-2.4)
Highest school attended	
No education	Reference
Primary	1.3 (0.8-2.2)
Secondary or more	*1.7 (1.1-2.6)*
Household wealth (tertile)	
Lowest (reference: all other groups)	1.1 (0.5-2.5)
Highest (reference: all other groups)	1.4 (0.8-2.7)

Discussion

This analysis offers 3 main findings. First, we found that the CATI response and cooperation rates were more than double the rates for hybrid IVR surveys because of high break-off postconsent among women assigned to the hybrid IVR arm. Second, the low contact rates resulted in sample distortion for both modes. Third, hybrid IVR survey completers had higher odds of having secondary education than the CATI completers, indicating additional sample distortion for hybrid IVR surveys.

Therefore, we identified CATIs as a better approach for phone surveys in settings similar to Burkina Faso because of the less profound distortion of completers. Greater sample distortion

among those randomized to the hybrid IVR survey is expected because the mode requires participants to answer the questionnaire without assistance, creating a higher cognitive burden on respondents than the CATI. Thus, responding is more difficult for those with lower education or who do not speak a majority language [18]. In this study, the response rate among women with no education randomized to the hybrid IVR survey was 12.3% compared with the response rate of 28.5% among women with a secondary or higher level of education. In the CATI survey, the response rates were comparable in all education groups.

This study illustrates that CATI phone follow-up surveys among women are feasible but suffer from noteworthy nonresponse. Failure to contact the sampled participants was the main cause of nonresponse, with 46% of hybrid IVR survey and 41% of CATI participants classified as noncontacts. Noncontact is also the main cause of nonresponse in high-income countries [8]. Refusal to participate was minimal in both arms. Finally, the hybrid IVR survey had increased nonresponse because of difficulties in responding to a phone keypad, with 3.5% of all eligible participants willing to participate but unable to navigate the phone keypad to begin the survey. Recent random digit dial (RDD) studies in the United States lost 9% [19] and 7% [18] of the sample when transferring respondents from the CATI to IVR survey. Among the hybrid IVR survey participants who were able to press 1 (and be transferred to IVR), there was a substantial break-off, with 20% answering less than half of the survey questions.

Studies using CATIs for remote data collection in Lebanon [20], Honduras, and Peru [21] have examined response rates and related sample distortion. These studies enrolled men and women during a baseline FTF survey who were recontacted using CATIs. The study in Lebanon had an 82% response rate but did not compare respondent and nonrespondent profiles [20]. The profile of completers in our study had a similar pattern as completers of phone surveys in Peru and Honduras, with the wealthy, educated, and urban also more likely to complete the phone follow-up [21]. The response rate at first contact was 33% in Peru and 59% in Honduras. However, these surveys, administered as part of a study conducted by the World Bank, found that young participants were more likely to complete the survey. Our results found an opposite association most likely because we enrolled only women [21]. We believe this is because gender norms result in different ownership and access in a context such as Burkina Faso. Young men may have the financial and social freedom to own a cell phone, whereas young women are less likely to be economically independent and to be given a phone by a parent, as their actions are more closely monitored. Female cell phone owners may still have less opportunity to use their phones compared with male cell phone owners because of the lack of resources to pay for communications and due to the possibility that women are more likely to have a gatekeeper to receive a phone call.

The study has notable strengths. The 2 modes used in our research, IVR and CATI, are rarely compared in low- or high-income countries [22]. Not only is the direct comparison of two remote modes rare but also it is uncommon for IVR studies in SSA to have a known sampling frame. Our sampling frame was a population-based FTF sample that allowed us to assess the characteristics related to completion, regardless of the woman's participation in the phone follow-up. The FTF survey provided identical measures for phone survey respondents and nonrespondents, whereas many studies evaluating nonresponse error rely on surveys that do not have directly comparable indicators for nonrespondents. Knowledge of the characteristics of phone nonrespondents is also valuable for the survey design to oversample women who are less likely to respond and adjust remote data collection estimates through weighting. The study design also allows us to compare respondents with the full FTF sample (including women who do not own cell phones), where we see a dramatic difference in demographic profile. For example, among hybrid IVR survey completers, 54% of the population had a secondary education (data not shown), whereas only 19% of the full FTF female sample had a secondary education [14]. Finally, the randomization of respondents to CATIs or hybrid IVR surveys allows a more robust comparison of nonresponse by phone survey mode.

The follow-up design of the survey meant that women had already participated in a PMA2020 survey and consented to follow-up, yielding higher response rates than an RDD survey. Thus, our four survey outcomes rates must be interpreted with caution. Phone follow-up surveys conducted after an FTF survey in SSA have only investigated response rates among populations, including both male and female respondents, but have generally found higher response rates than those in our CATI survey [6]. Our CATI response rate was likely much lower because of the 11-month span between enrollment and follow-up and because women are harder to reach than men via cell phones [23]. Research from LMICs shows that rapid follow-up (defined as less than a month since the first contact) after enrolling in an FTF survey is key to reduce noncontact, the main cause of nonresponse in our survey [21,24]. Thus, our response rate highlights the difference in response rates when following up after a greater amount of time to create a new cross-sectional study originating from an FTF compared with a cohort quickly contacted for phone follow-up after an FTF interview. The response rate for the hybrid IVR survey was higher than that of other IVR surveys in SSA, most likely because of the selection of cell phone owners in our survey and because live interviewers introduced the survey. The only response rates available for comparison with our hybrid IVR survey rates are from surveys that use RDD sampling, not follow-up surveys, and use IVR without human introduction. Surveys in Mozambique and Zimbabwe had response rates of 9% and 8% [3,23], respectively, and a more recent RDD in Ghana had a response rate of 21% [12].

The context of Burkina Faso is unique in some areas pertinent to cell phone–based data collection, including high language fractionalization, making remote data collection more difficult because of language discordance between respondents and interviewers. Low female literacy reduces cell phone survey modes to those relying on interviewers. However, these contextual factors are similar across many West African countries.

The compound annual growth rate (CAGR) for cell phone ownership in SSA was in double digits at the beginning of the decade [1]. However, the CAGR is now half the previous level. Future cell phone ownership growth will be among rural and young people, but the time of rapid increase is waning: the percentage of cell phone owners in SSA is projected to only increase by 2% between 2023 and 2025 to 52% of the population [1]. Thus, the differences in phone owners and nonowners will not be quickly erased, and caution must be exercised when trying to create nationally representative estimates from phone surveys.

Conclusions

We identified the characteristics related to CATI and hybrid IVR survey completion and concluded that among PMA2020's target population, that is, women of reproductive age, CATI results in a more representative sample yet still greatly distorted sample. CATI and hybrid IVR survey completers were not reflective of the target population's characteristics, in particular, education, geographic location, and age. This study is among the first to analyze phone survey nonresponse errors and to compare CATI and hybrid IVR surveys in SSA. Our results underscore that enrolling a representative sample of women in a phone survey in West Africa is challenging, regardless of cell phone mode, and should inform remote data collection efforts in West Africa.

Acknowledgments

This work was supported by the Bill & Melinda Gates Foundation (Grant Number: OPP1079004). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Authors' Contributions

ARG conducted the literature search. ARG, YC, GG, ST, NB, SA, and CM designed the study. Data collection was led by AG, with support from ARG, YC, GG, ST, NB, and CM. Data analysis was conducted by ARG and CM, with support from YC and SA. The manuscript was drafted by ARG, CM, and SA. All coauthors reviewed the manuscript and provided input.

Conflicts of Interest

None declared.

References

1. The Mobile Economy Sub-Saharan Africa. GSMA. 2018. URL: https://www.gsma.com/mobileeconomy/sub-saharan-africa/ [accessed 2018-07-17]
2. Poushter J, Oates R. Cell Phones in Africa: Communication Lifeline. Pew Research Center. 2015 Apr 15. URL: https://www.pewresearch.org/global/2015/04/15/cell-phones-in-africa-communication-lifeline/ [accessed 2018-05-01]
3. Leo B, Morello R, Mellon J, Peixoto T, Davenport S. Do Mobile Phone Surveys Work in Poor Countries? The Center for Global Development. 2015. URL: https://www.cgdev.org/publication/do-mobile-phone-surveys-work-poor-countries-working-paper-398 [accessed 2016-10-18]
4. Du X, Wang W, van Velthoven MH, Chen L, Scherpbier RW, Zhang Y, et al. mHealth series: text messaging data collection of infant and young child feeding practice in rural China - a feasibility study. J Glob Health 2013 Dec;3(2):020403 [FREE Full text] [doi: 10.7189/jogh.03.020403] [Medline: 24363921]
5. Biemer P. Overview of design issues: total survey error. In: Marsden WJ, editor. Handbook of Survey Research. Bingley, UK: Emerald Group Publishing; 2010.
6. Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017 May 5;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428] [Medline: 28476725]
7. Bethlehem J, Bakker B. The impact of nonresponse on survey quality. Stat J IAOS 2014;30:5. [doi: 10.3233/SJI-140831]
8. Groves R. Nonresponse rates and nonresponse bias in household surveys. Public Opin Q 2006;70(5):20. [doi: 10.1093/poq/nfl033]
9. Lavrakas P. New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the US With Respondents Reached via Cell Phone Numbers. American Association for Public Opinion Research. 2010. URL: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/2010AAPORCellPhoneTFReport.pdf [accessed 2020-05-28]
10. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. Annu Rev Public Health 2007;28:113-126. [doi: 10.1146/annurev.publhealth.28.021406.144059] [Medline: 17094769]
11. Greenleaf AR, Gibson DG, Khattar C, Labrique AB, Pariyo GW. Building the evidence base for remote data collection in low- and middle-income countries: comparing reliability and accuracy across survey modalities. J Med Internet Res 2017 May 5;19(5):e140 [FREE Full text] [doi: 10.2196/jmir.7331] [Medline: 28476728]

12. L'Engle K, Sefa E, Adimazoya EA, Yartey E, Lenzi R, Tarpo C, et al. Survey research with a random digit dial national mobile phone sample in Ghana: methods and sample quality. PLoS One 2018;13(1):e0190902 [FREE Full text] [doi: 10.1371/journal.pone.0190902] [Medline: 29351349]
13. Larmarange J. Feasibility and representativeness of a random sample mobile phone survey in Côte d'Ivoire. Population 2016;71(1):10. [doi: 10.3917/pope.1601.0119]
14. Greenleaf AR, Ahmed S, Moreau C, Guiella G, Choi Y. Cell phone ownership and modern contraceptive use in Burkina Faso: implications for research and interventions using mobile technology. Contraception 2019 Mar;99(3):170-174 [FREE Full text] [doi: 10.1016/j.contraception.2018.11.006] [Medline: 30468721]
15. Enquête sur les Indicateurs du Paludisme au Burkina Faso (EIPBF). The DHS Program. 2014. URL: https://dhsprogram.com/pubs/pdf/MIS19/MIS19.pdf [accessed 2020-05-28]
16. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. American Association for Public Opinion Research. 2016. URL: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2017-04-29]
17. Filmer D, Pritchett L. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography 2001 Mar;38(1):115-132. [doi: 10.1353/dem.2001.0003] [Medline: 11227840]
18. Kepple NJ, Freisthler B, Johnson-Motoyama M. Bias in child maltreatment self-reports using interactive voice response (IVR). Child Abuse Negl 2014 Oct;38(10):1694-1705 [FREE Full text] [doi: 10.1016/j.chiabu.2014.04.001] [Medline: 24819534]
19. Brooks MJ, Bear T, Hacker K, Ricci EM, Foulds A, Anderson H, et al. An alternative technique for youth risk surveillance outside of the school system. J Adolesc Health 2017 Jan;60(1):72-78. [doi: 10.1016/j.jadohealth.2016.08.014] [Medline: 27793728]
20. Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. Cell phone and face-to-face interview responses in population-based surveys. Field Methods 2014 Jul 14;27(1):39-54. [doi: 10.1177/1525822X14540084]
21. Ballivian A, Azevedo JP, Durbin W, Rios J, Godoy J, Borisova C, et al. Listening to LAC: Using Mobile Phones for High Frequency Data Collection. World Bank. 2013. URL: https://catalog.ihsn.org/index.php/catalog/4775/related_materials [accessed 2017-12-02]
22. Midanik LG, Greenfield TK. Interactive voice response versus computer-assisted telephone interviewing (CATI) surveys and sensitive questions: the 2005 national alcohol survey. J Stud Alcohol Drugs 2008 Jul;69(4):580-588. [doi: 10.15288/jsad.2008.69.580] [Medline: 18612574]
23. Bridging the Gender Gap: Mobile Access and Usage in Low- and Middle-income Countries. GSMA. 2015. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/02/GSMA-Bridging-the-Gender-Gap-2015.pdf [accessed 2020-05-28]
24. Cecatti JG, Camargo RP, Pacagnella RC, Giavarotti T, Souza JP, Parpinelli MA, et al. Computer-assisted telephone interviewing (CATI): using the telephone for obtaining information on reproductive health. Cad Saude Publica 2011 Sep;27(9):1801-1808 [FREE Full text] [doi: 10.1590/s0102-311x2011000900013] [Medline: 21986607]

Abbreviations

AAPOR: American Association for Public Opinion Research
CAGR: compound annual growth rate
CATI: computer-assisted telephone interview
DHS: Demographic and Health Survey
FTF: face-to-face
IVR: interactive voice response
LMIC: low- and middle-income country
OR: odds ratio
PMA2020: Performance Monitoring and Accountability 2020
R4: round 4
RDD: random digit dial
SSA: sub-Saharan Africa

Edited by G Eysenbach; submitted 19.01.20; peer-reviewed by K L'engle, R Lester; comments to author 11.03.20; revised version received 29.03.20; accepted 30.03.20; published 14.07.20.

Please cite as:
Greenleaf AR, Gadiaga A, Choi Y, Guiella G, Turke S, Battle N, Ahmed S, Moreau C
Automated and Interviewer-Administered Mobile Phone Surveys in Burkina Faso: Sociodemographic Differences Among Female Mobile Phone Survey Respondents and Nonrespondents
JMIR Mhealth Uhealth 2020;8(7):e17891
URL: https://mhealth.jmir.org/2020/7/e17891
doi:10.2196/17891
PMID:32673250

Original Paper

Feasibility and Acceptability of an Adapted Mobile Phone Message Program and Changes in Maternal and Newborn Health Knowledge in Four Provinces of Afghanistan: Single-Group Pre-Post Assessment Study

Victoria Lebrun[1], MSPH; Lisa Dulli[1], PhD; Sayed Omar Alami[2], BCS; Arzoo Sidiqi[2], BCS; Ahmad Shah Sultani[2], MCS; Sayed Haroon Rastagar[2], MD, MPH; Iftikhar Halimzai[3], MD; Sharif Ahmadzai[2], SA, MD; Catherine S Todd[1], MD, MPH

[1]Global Health, Population, and Nutrition, FHI 360, Durham, NC, United States

[2]Helping Mothers and Children Thrive in Afghanistan (HEMAYAT) Project, FHI 360, Kabul, Afghanistan

[3]Health Promotions Department, Ministry of Public Health, Kabul, Afghanistan

Corresponding Author:
Victoria Lebrun, MSPH
Global Health, Population, and Nutrition
FHI 360
359 Blackwell St #200
Durham, NC, 27701
United States
Phone: 1 9195447040 ext 11640
Email: vlebrun@fhi360.org

Abstract

Background: Mobile phone apps for health promotion have expanded in many low- and middle-income countries. Afghanistan, with high maternal and newborn morbidity and mortality rates, a fragile health infrastructure, and high levels of mobile phone ownership, is an ideal setting to examine the utility of such programs. We adapted messages of the Mobile Alliance for Maternal Action (MAMA) program, which was designed to promote healthy behaviors during pregnancy and a newborn's first year of life, to the Afghan context. We then piloted and assessed the program in the provinces of Kabul, Herat, Kandahar, and Balkh.

Objective: The aim of this study was to assess the feasibility and acceptability of the MAMA pilot program, and to examine changes in reported maternal, newborn, and child health (MNCH) knowledge and attitudes among participants from baseline to follow up.

Methods: We conducted a single-group study with data collected within 10 weeks of enrollment, and data collection was repeated approximately 6 months later. Data were collected through face-to-face interviews using structured questionnaires. Eligible participants included pregnant women who had registered to receive fully automated mobile health messages and their husbands. Assessment questionnaires queried sociodemographic details; knowledge, attitudes, and health care-seeking practices; and intervention experience and acceptability at follow up. The number of messages received by a given phone number was extracted from the program database. We descriptively analyzed the feasibility and acceptability data and compared the change in MNCH knowledge between baseline and follow-up measures using the McNemar Chi square test.

Results: Overall, 895 women were enrolled in the MAMA program. Data from 453/625 women (72.5% of the pretest sample) who received voice (n=302) or text (n=151) messages, and 276/427 men (64.6% of the pretest sample) who received voice (n=185) or text (n=91) messages contributing data at both time points were analyzed. At follow up, 699/729 (95.9%) participants were still enrolled in the MAMA program; voice message and SMS text messaging subscribers received 43 and 69 messages on average over the 6-month period, respectively. Participants who were voice message subscribers and female participants more commonly reported missing messages compared with the text message subscribers and men; predominant reasons for missed messages were the subscribers being busy with chores or not having their shared phone with them. Over 90% of men and women reported experiencing benefits from the program, mainly increased knowledge, and 226/453 (49.9%) of the female participants reported referring someone else to the program. Most of the participants (377/453, 83.2% women and 258/276, 93.5% men) believed it was beneficial to include husbands in the program. Joint decision making regarding maternal and child health care increased

overall. The proportions of participants with correct knowledge significantly increased for all but one MNCH measure at follow up.

Conclusions: This assessment indicates that the pilot MAMA program is feasible and acceptable in the Afghan context. Further research should be conducted to determine whether program participation leads to improved MNCH knowledge, health practices, and health service utilization in this fragile setting prior to larger scale up.

(JMIR Mhealth Uhealth 2020;8(7):e17535) doi:10.2196/17535

KEYWORDS

Afghanistan; mobile apps; pregnant women; maternal health; newborn health; social and behavior change; mHealth; voice message; SMS

Introduction

Afghanistan has one of the highest maternal, newborn, and child mortality rates among Asian countries [1] and faces important challenges to improving these and other health indicators [2]. Despite substantial investment in training and deploying large numbers of health workers, there is uneven geographic and socioeconomic coverage of essential interventions in the country, including antenatal care, skilled birth attendance, immunization, and family planning [3]. Uptake of these services, where available, is also suboptimal. Only 18% of women had at least 4 antenatal care visits and 51% reported skilled birth attendance at their last delivery in the 2015 Demographic and Health Survey [4]. A 2013 survey found that 51% of children aged 12-23 months were fully immunized based on Afghanistan's national immunization schedule [5]. Additionally, the contraceptive prevalence rate has stagnated at approximately 20% since 2010 [3,4].

Among the factors associated with low uptake of essential maternal, newborn, and child health (MNCH) interventions are illiteracy and lack of education [6]. Maternal and infant morbidity and mortality rates are disproportionately higher among rural populations, which have lower literacy levels and are less likely to access health services or to be reached by health messaging through mass media or health providers [4,7,8]. Women's access to health services is also hampered by traditional social norms, which position men as the primary health decision makers and limit women's freedom of movement [9,10]. Health decision making occurs within the family structure with men, and sometimes older women such as mothers-in-law, serving as gatekeepers who determine when and where health services for pregnant women and infants are sought [5,11-13].

Within Afghanistan, mobile phone ownership and use have expanded markedly in the last decade, creating a potential channel to reach women and their families with health information. In a recent nationally representative survey, 91% of households reported owning a mobile phone [14]. Mobile phone ownership is higher among men compared to women; however, a 2012 cross-sectional survey found that 80% of women reported routinely using a mobile phone [15], and evidence suggests that their phone use is increasing. A 2015 cross-sectional survey in Nangarhar province found that 92% of women routinely used mobile phones [16]. Mobile phone–based programming for various health education purposes has been documented globally with mixed success; however, to date, there is no literature on the use of such programs in Afghanistan. Short message service (SMS) text messaging recommendations and appointment reminders during pregnancy have been positively received and were shown to improve antenatal and postnatal care attendance in other low- and middle- income countries [17-20]; recorded voice messages in addition to, or in place of, SMS text messaging enable low literacy populations to access health information [21]. The 2019 World Health Organization guideline "Recommendations for digital interventions for health system strengthening" recommends targeted client communications through digital channels, including mobile phone messages, to promote reproductive, maternal, and newborn health [22].

The Mobile Alliance for Maternal Action (MAMA) program was launched in 2012 as a public-private partnership to scale up an evidence-based mobile health (mHealth) program that is already being used in two countries [23]. With support from Baby Center, Johnson & Johnson, the United States Agency for International Development (USAID), the mHealth Alliance, and the United Nations Foundation [23-25], MAMA implemented its mHealth approach focused on age- and stage-based messaging directed toward pregnant women and their family members in Bangladesh [24,25], India, South Africa [26], and Nigeria. The MAMA program relays essential MNCH messages promoting healthy practices and care-seeking for pregnant women and their infants up to 12 months of age [23]. These messages are available for adaptation and use through an application process to interested groups in different countries. Countries employing the program have developed context-appropriate business models and message delivery channels, and have also identified and targeted subpopulations for program promotion [23].

In this study, we assessed a pilot implementation of the MAMA program in four provinces of Afghanistan to determine whether it was feasible and acceptable to Afghan users, and whether users report changes in MNCH attitudes, health decision making, and knowledge of key MNCH concepts. The data from this assessment will guide program refinement, and are intended to improve program use and potential effectiveness at scale up in Afghanistan and adaptation in similar contexts.

Methods

Study Design

We conducted a single-group, baseline/follow-up study to assess the feasibility, acceptability, and potential effect of the MAMA

program on MNCH knowledge. The MAMA pilot program was implemented in 80 health post catchment areas in the Balkh, Herat, Kabul, Kandahar, and Nangarhar provinces of Afghanistan. Nangarhar province was not included in the program assessment as baseline recruitment was not possible within the study time frame due to insecurity. The selected pilot districts were semiurban or rural and predominantly agrarian. Eligible study participants included pregnant women and their husbands who enrolled in the MAMA program with the help of a community health worker between May and July 2018.

The study was reviewed and approved by the FHI 360 Protection of Human Subjects Committee (#1240522-2) and was approved by the Afghanistan Ministry of Public Health (MOPH) Institutional Review Board (protocol #444670) prior to implementation.

Program Design

The MAMA message program relays essential MNCH information to guide actions for pregnant women and families with children under 12 months of age through mobile phones. Educational messages are sent twice weekly and are timed to the stage of pregnancy or age of the newborn. In a report exploring MAMA program implementation experiences in four countries, several factors were noted that may impact feasibility and acceptability [23]. Specifically, the report highlights the need to adapt the content to specific contexts with local stakeholder inputs; customer enrollment requires partners with local presence; and the need for client-centered message delivery, such as selecting the best time for calls [23]. In 2018, MAMA program messages were adapted to the Afghan context by technical specialists in social and behavior change, MNCH, and community health from the USAID-funded Helping Mothers and Children Thrive in Afghanistan (HEMAYAT) project, the Health Promotions and Reproductive Health Departments of the Afghanistan MOPH, and members of the Health Promotions Technical Working Group. The resulting adapted and translated messages were recorded by female native Dari and Pashto speakers as voice messages, pretested, and revised. Following pretesting, final messages were reviewed and approved by the Health Promotions Technical Working Group. An automated messaging platform was developed by Parsa Technology in Kabul, Afghanistan, which allowed user selection of preferred time of day to receive twice-weekly messages, message type (voice message or SMS text message), and language (Dari or Pashto) that then sent messages to subscriber phone numbers across a variety of mobile networks.

For pilot program implementation, female community health workers were selected from communities with active health posts. Each selected community health worker was trained to recruit and enroll 10 pregnant women and offer enrollment to their husbands, in collaboration with the Provincial Health Directorate and the Basic Package of Health Services implementer. Recruitment was primarily achieved through home visits to prospective clients, although some participants were recruited at locations such as health posts and adult literacy classes. For this pilot program, eligibility for registration was confined to pregnant women and their husbands who lived within the catchment area of the health post of the recruiting community health worker and had access to a functional mobile phone. Potential clients were informed that program staff might call them to confirm message receipt and program functionality. The community health workers ensured that women and men who registered were eligible and provided clients with a choice of written text (SMS) messages or automated recorded voice calls, as well as a choice for time of day to receive the calls or text messages. If the couple shared a phone, the woman was offered these choices, as she was considered the primary subscriber. The first message was timed to the woman's current week of pregnancy and continued sequentially from that point. Messages encouraged clients to discuss MNCH care with their spouses and families, and promoted healthy practices such as exclusive breastfeeding and seeking health care services, including antenatal care, skilled attendance at delivery, and infant immunization.

Monitoring was conducted by HEMAYAT program staff and community health supervisors through three approaches: direct field supervision and mentoring during the community health worker-led registration process; monitoring of the data dashboard incorporated with the MAMA platform reflecting the number of messages sent, received, and nonresponsive numbers; and project staff calls to registered clients to confirm number validity, message receipt, and whether the registered client met the criteria for the pilot registration process. As the first and third approaches were employed early in pilot program implementation, system errors such as some mobile networks not receiving messages were identified and repaired.

Sampling and Recruitment

For baseline data collection, all households that had enrolled in the MAMA program in the four included provinces were approached and program subscribers were recruited to participate in the study no more than 10 weeks after enrolling in the MAMA program. The community health workers who had promoted the program and registered subscribers introduced the data collection team to female subscribers and their husbands or male heads of household. Eligible participants were pregnant women and their husbands subscribing to the pilot program at least 4 weeks earlier and verbally consenting to study participation. Following explanation of the assessment, participation was offered, and verbal informed consent was obtained from subscribers interested in study participation by data collectors of the same sex. For follow-up data collection, households were revisited approximately 6 months after the baseline survey.

Measures and Data Collection

Structured interviews using paper questionnaires were conducted with study participants by trained, sex-matched data collectors fluent in the language(s) predominantly spoken in that province. Translated questionnaires for women and men were field-tested with volunteers from different ethnic groups across implementation sites. The baseline questionnaire included sections detailing household characteristics, the program registration process, and MNCH-related knowledge and attitudes. Interviewers asked program participants about their attitudes toward discussing MNCH and specific practices, household decision making about MNCH care, and their

knowledge of several maternal and newborn health issues featured in MAMA messages. We developed the knowledge questions based on content used to measure MNCH knowledge among men and women in assessments and evaluations performed in Afghanistan both within our group and in larger household-level surveys [4,27]. Knowledge questions were closed-ended and were in the form of single-response (for questions that have one correct answer such as the number of months for which infants should be exclusively breastfed) or multiple-response (for questions with more than one correct answer, such as serious health problems that can occur during pregnancy) questions. Answer choices were not read to the participant. Knowledge scores for multiple-response questions were converted to a binary score with 1 assigned to any correct answer and 0 assigned to no correct answers.

Feasibility measures included the proportion of female respondents who had to ask permission to register and from whom they asked permission; whether they were asked what time of day they prefer to receive messages and whether they received messages at their preferred time; continued to subscribe and receive messages after 6 months and the number of messages that respondents reported receiving per week on average; and reports of missed messages with stated reasons for missing the MAMA messages. Acceptability measures included the proportion of respondents who reported any and specific benefits to their participation in MAMA; reported that a member of their household listened to or read the messages and who it was (eg, husband, mother-in-law); stated that including their husband or their mother-in-law was beneficial; recommended MAMA use to someone else; and discussed MAMA with other pregnant women or new mothers.

Trained male and female data collectors administered the study instruments to consenting participants of the same sex in a private room in the household. The same data collectors conducted interviews at baseline and follow-up sessions; community health workers introduced the data collection team at households during study recruitment and at baseline data collection, consistent with cultural norms. Baseline data were collected in July and August 2018. Follow-up interviews were conducted with the same participants in January and February 2019, using a similar structured questionnaire including the same knowledge and attitude questions and adding questions on exposure to and acceptability of the MAMA program. Additionally, the MAMA mobile platform and database automatically recorded the dates and time of messages that were sent and received. We extracted the overall number of messages received for each participant at follow-up data collection.

Data Analysis

We used STATA Version 15 (StataCorp LP, College Station, TX, USA) to descriptively analyze sociodemographic characteristics, message exposure, and acceptability measures, disaggregated by sex and message type (SMS text or voice message). Results were summarized at baseline and at follow up for attitudes, decision making, and knowledge of MNCH topics. Additionally, we conducted exploratory analyses to examine the change in selected MNCH knowledge measures between baseline and follow up using the McNemar Chi square test for paired data, with a two-sided alpha of .05 and 80% power.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Results

Participant Characteristics

Across the four provinces sampled, 895 pregnant women were registered in the MAMA pilot program from May to August 2018. Over two-thirds (625/895, 69.8%) of female program participants and 427 husbands who were confirmed subscribers to the program agreed to participate in the study. A total of 499/625 (79.8%) women and 306/427 (71.7%) men completed both baseline and follow-up questionnaires. In some households, more than one woman or man enrolled in the study. However, because telephone numbers were used as the unique identifier to link baseline and follow-up data for participants, if more than one person of the same sex had the same phone number, it was impossible to link the data from the two time points to the correct participant. Thus, due to this uncertainty, we removed records for both participants from the final dataset. Therefore, we present results from 453/625 women (72.5% of the pretest sample) and 276/427 men (64.6% of the pretest sample) with complete baseline and follow-up data, disaggregated by sex.

Among both women and men, about two-thirds of participants (487/729, 66.8%) opted to receive voice messages and one-third (242/729, 33.2%) chose to receive SMS text messages. Sociodemographic and household characteristics of participants indicate that many women did not own a phone and used someone else's phone to access the program, and most of the participants had primary-level or no education (Table 1). Over three-quarters of women who chose voice messages reported having no formal education, compared to one-third of women who selected SMS text messages. High numbers of SMS text message recipients reported having electricity, televisions, and mobile phones with internet access in their households compared to participants who chose to receive voice messages.

Table 1. Sociodemographic characteristics of participants, by sex and message modality, across four provinces in Afghanistan (N=729).

Characteristic	Women		Men	
	Voice (n=302)	Text messaging (n=151)	Voice (n=185)	Text messaging (N=91)
Province, n (%)				
Balkh	77 (25.5)	56 (37.1)	58 (31.4)	42 (46.2)
Herat	73 (24.2)	48 (31.8)	36 (19.5)	29 (31.9)
Kabul	66 (21.9)	38 (25.2)	28 (15.1)	12 (13.2)
Kandahar	86 (28.5)	9 (6.0)	63 (34.1)	8 (8.8)
Age (years), mean (SD)	28.6 (12.7)	26.3 (10.9)	32.5 (9.6)	30.5 (7.5)
Phone ownership, n (%)	176 (58.3)	104 (68.9)	185 (100.0)	88 (96.7)
Phone used to access MAMA[a], n (%)				
Own	123 (40.7)	70 (46.4)	134 (72.4)	57 (62.6)
Spouse's	146 (48.3)	67 (44.4)	42 (22.7)	29 (31.9)
Mother/father's	1 (0.3)	6 (4.0)	3 (1.6)	3 (3.3)
Mother-in-law/father-in-law's	20 (6.6)	1 (0.7)	3 (1.6)	0 (0)
Brother/sister's	2 (0.7)	5 (3.3)	1 (0.5)	2 (2.2)
Brother-in-law/sister-in-law's	7 (2.3)	2 (1.3)	0 (0)	0 (0)
Education level, n (%)				
No formal education	234 (77.5)	46 (30.5)	84 (45.4)	29 (31.9)
Primary	31 (10.3)	25 (16.5)	38 (20.5)	12 (13.2)
Secondary	25 (8.3)	25 (16.6)	19 (10.3)	14 (15.4)
High school	10 (3.3)	40 (26.5)	25 (13.5)	21 (23.1)
Higher	1 (0.3)	13 (8.6)	12 (6.5)	12 (13.2)
Vocational or Madrassa	1 (0.3)	2 (1.3)	7 (3.8)	3 (3.3)
Number of living children at pretest, n (%)				
0	53 (17.6)	37 (24.5)	39 (21.1)	26 (28.6)
1-2	82 (27.2)	56 (37.1)	40 (21.6)	35 (38.5)
3-4	84 (27.8)	40 (26.5)	54 (29.2)	19 (20.9)
5 or more	83 (27.5)	18 (11.9)	52 (28.1)	11 (12.1)
Household wealth status indicators[b], n (%)				
Electricity	228 (75.5)	136 (90.1)	153 (82.7)	85 (93.4)
Radio	72 (23.8)	36 (23.8)	69 (37.3)	24 (26.4)
Television	168 (55.6)	116 (76.8)	102 (55.1)	76 (83.5)
Mobile phone with internet	53 (17.6)	61 (40.4)	57 (30.8)	32 (35.2)
Mobile phone without internet	268 (88.7)	121 (80.1)	173 (93.5)	86 (94.5)
Birth outcome at posttest, n (%)				
Born alive and healthy	272 (89.2)	138 (95.6)	165 (89.2)	87 (95.6)
Stillbirth	13 (4.9)	4 (1.1)	9 (4.9)	1 (1.1)
Other[c]	17 (6.0)	9 (3.3)	11 (6.0)	3 (3.3)

[a]MAMA: Mobile Alliance for Maternal Action.

[b]Multiple responses allowed.

[c]Other responses included abortion, miscarriage, and women who had not yet given birth.

MAMA Program Feasibility

The majority (312/453, 68.9%) of female participants overall reported that they had to seek the permission of a gatekeeper to enroll in the MAMA program; this gatekeeper was usually the woman's husband, although in some cases it was the mother-in-law (Table 2). Some subscribers reported not being offered a preferred time to receive messages by the community health worker during MAMA registration, but most of the women who chose a time preferred to receive messages in the morning or at night (Table 2).

After 6 months of program participation, the automated database confirmed that across groups, 95%-97% of respondents had not cancelled their subscription and were continuing to receive messages (Table 2). Of the 22 women and 8 men who were no longer subscribers at follow up, the majority (21/30, 70%) did not have the phone or SIM card used at registration or it was broken; a few women (3/30, 10%) had problems related to phone sharing. The program was designed to send 2 messages each week, with a few exceptions; however, some respondents reported receiving more than 2 messages per week. A software error that was not detected during beta testing caused the automated system to erroneously send multiple messages per day to the same phone number. Affected subscribers notified the project through a helpline or by telling the community health workers who then called HEMAYAT project staff to report the error. This problem was identified and corrected within 48 hours of reporting. Subscribers of voice messages received fewer total messages and were more likely to report occasionally missing messages compared to those who received text messages; women also stated more often than men that they missed messages. Common reasons for missing messages included being busy with chores or that someone else had the phone, and these reasons were equally common for both sexes.

MAMA Program Acceptability

Most women and men cited multiple benefits when asked about perceived program benefits (Table 3); gaining information about health care for themselves (women), their wives (men), or their children were the most frequently stated benefits. Women often reported that the ability to obtain health information at home was also a benefit, and about one-quarter of the respondents of both sexes spontaneously mentioned that they learned new health information by subscribing to MAMA. Nearly all men and over 80% of women (n=377) agreed that it was beneficial to include husbands in the MAMA program; the majority of women in the voice message group and approximately half of the women in the SMS text messaging group said their husbands had listened to or read the messages (Table 3). More than half of the participants agreed that including their mothers-in-law was beneficial, but 119 women (about 25%) stated that it was not beneficial. Most women reported having discussed the program with a peer, with about half of women and over one-third of men reporting having recommended the program to a friend or relative.

Table 2. Feasibility of Mobile Alliance for Maternal Action (MAMA) program registration and use among participants, by sex and message modality, in four provinces of Afghanistan (N=729).

Feasibility measure	Women		Men	
	Voice (n=302)	Text messaging (n=151)	Voice (n=185)	Text messaging (n=91)
Woman had to ask permission to register[a]	213 (70.5)	99 (65.6)	131 (70.8)	47 (51.7)
Who was asked for permission[b]				
Husband	174 (78.4)	83 (79.8)	96 (69.6)	34 (66.7)
Mother-in-law	32 (14.4)	18 (17.3)	34 (24.6)	14 (27.5)
Father-in-law	13 (5.9)	3 (2.9)	8 (5.8)	3 (5.9)
Brother-in-law	2 (0.9)	0 (0)	0 (0)	0 (0)
Other (specify):	1 (0.45)	0 (0)	0 (0)	0 (0)
Received message at preferred time at baseline, n (%)				
Not offered the choice of a preferred time	101 (33.4)	39 (25.8)	40 (21.6)	18 (19.8)
Yes	136 (45.0)	89 (58.9)	76 (41.1)	38 (41.8)
No	49 (16.2)	18 (11.9)	11 (5.9)	9 (9.9)
Refused or don’t know	16 (5.3)	10 (6.6)	58 (31.4)	26 (28.6)
Time of day preferred (baseline), n (%)				
Morning	88 (29.1)	43 (28.5)	40 (21.6)	23 (25.3)
Afternoon	27 (8.9)	18 (11.9)	18 (9.7)	8 (8.8)
Evening	31 (10.3)	9 (6)	17 (9.2)	2 (2.2)
Night	55 (18.2)	39 (25.8)	12 (6.5)	7 (7.7)
Anytime	0 (0)	3 (2)	9 (4.9)	9 (9.9)
No choice given/don’t know	101 (33.4)	39 (25.8)	89 (48.1)	42 (46.2)
Current MAMA subscription at posttest[c], n (%)	286 (94.7)	145 (96)	180 (97.3)	88 (96.7)
Total messages received, mean (SD)	43 (17.6)	69.4 (13)	43.7 (17.7)	68.5 (12.5)
Typical number messages received per week[d], n (%)				
0	2 (0.7)	2 (1.3)	5 (2.7)	1 (1.1)
1-2	240 (79.5)	137 (90.7)	104 (56.2)	62 (68.1)
3-5	47 (15.6)	12 (8)	63 (34.1)	25 (27.5)
6+	13 (4.3)	0 (0)	13 (7)	3 (3.3)
Ever missed messages, n (%)	142 (47)	36 (23.8)	49 (26.5)	14 (15.4)
Reasons for missing messages, among those who reported missing[b], n (%)	N=142	N=36	N=49	N=14
No balance or charge	14 (7.7)	1 (2.3)	13 (17.8)	4 (19.1)
Busy with chores	73 (40.3)	18 (41.9)	37 (50.7)	8 (38.1)
Someone else had phone	67 (37)	18 (41.9)	19 (26)	7 (33.3)
Wrong time	7 (3.9)	0 (0)	0 (0)	1 (4.8)
Someone else took the call	5 (2.8)	3 (7)	0 (0)	0 (0)
Other	15 (8.3)	3 (7)	4 (5.5)	1 (4.8)

[a]Male participants were asked whether their wives were required to obtain permission to register.

[b]Multiple responses allowed.

[c]Subscription and messages received data were extracted from the MAMA system database. All other data presented were based on self-reporting.

[d]Men were asked how many messages their wives received per week.

Table 3. Mobile Alliance for Maternal Action (MAMA) program acceptability among participants, by sex and message type, in four provinces of Afghanistan (N=729).

Acceptability measure	Women		Men	
	Voice (n=302)	Text messaging (n=151)	Voice (n=185)	Text messaging (n=91)
Reported benefits of MAMA[a], n (%)				
Information for own (wife's) health	218 (72.2)	108 (71.5)	109 (58.9)	60 (65.9)
Information for child's health	201 (66.6)	107 (70.9)	114 (61.6)	58 (63.7)
Able to get health information at home	121 (40.1)	47 (31.1)	56 (30.3)	23 (25.3)
Learned new health information	85 (28.2)	39 (25.8)	47 (25.4)	14 (15.4)
Other	11 (3.6)	2 (1.3)	10 (5.4)	8 (8.8)
None	10 (3.3)	4 (2.7)	8 (6)	0 (0)
Others in household have listened to/read messages, n (%)				
No one	79 (26.2)	53 (35.1)	116 (62.7)	64 (70.3)
Husband	182 (60.3)	73 (48.3)	N/A[b]	N/A
Mother-in-law/Mother (for husband)	76 (25.2)	16 (10.6)	47 (25.4)	17 (18.7)
Sister(s)-in-law	28 (9.3)	14 (9.3)	9 (4.9)	6 (6.6)
Sister(s)	26 (8.6)	9 (6)	12 (6.5)	14 (15.4)
Father-in-law/father (for husband)	19 (6.3)	2 (1.3)	14 (7.6)	3 (3.3)
Other	12 (4)	2 (1.3)	17 (9.2)	5 (5.5)
Agree that including husband is beneficial, n (%)				
Yes	256 (84.8)	121 (80.1)	173 (93.5)	85 (93.1)
No	31 (10.3)	21 (13.9)	4 (2.2)	2 (2.2)
Don't know	15 (5)	9 (6)	8 (4.3)	4 (4.4)
Agree that including mother-in-law (husband's mother) in program is beneficial, n (%)				
Yes	151 (50)	77 (51)	114 (61.6)	62 (68.1)
No	76 (25.2)	43 (28.5)	38 (20.5)	15 (16.5)
Don't know	29 (9.6)	6 (4)	33 (17.8)	14 (15.4)
Mother-in-law didn't listen to messages	46 (15.2)	25 (16.7)	N/A	N/A
Benefits of including mother-in-law[a], n (%)				
Mother-in-law helps participant understand messages	89 (29.5)	52 (34.4)	65 (35.1)	36 (39.6)
Helps participant follow instructions in messages	48 (15.9)	25 (16.6)	35 (18.9)	27 (29.7)
Increased mother-in-law's awareness of participant's health needs	64 (21.2)	26 (17.2)	63 (34.1)	21 (23.1)
Needs information on health care during pregnancy	68 (22.5)	17 (11.3)	25 (13.5)	8 (8.8)
Other	1 (0.3)	2 (1.3)	1 (0.5)	3 (3.3)
Discussed MAMA with other pregnant women or new mothers, n (%)	175 (58)	88 (58.3)	N/A	N/A
Recommended the MAMA program, n (%)	145 (48)	81 (53.6)	67 (36.2)	39 (42.9)
Who MAMA was recommended to[a], n (%)				
Family member	66 (45.5)	41 (50.6)	24 (35.8)	17 (43.6)
Friend	55 (37.9)	32 (39.5)	42 (62.7)	17 (43.6)
Neighbor	68 (46.9)	31 (38.3)	39 (58.2)	12 (30.8)
Others	3 (2.1)	0	2 (3.0)	3 (7.7)

[a]Multiple responses possible.

[b]N/A: not applicable; option/item not asked of men.

Attitudes, Decision Making, and Knowledge

We assessed attitudes toward MNCH, including comfort discussing maternal and newborn care with family members, correct knowledge and acceptance of exclusive breastfeeding and birth spacing, and attitudes about utilizing facility-based care during pregnancy and delivery (Table 4). The proportion of participants who agreed that they feel knowledgeable discussing health care for pregnancy and newborns with their spouse was high at baseline and increased slightly at follow up; however, there was little or no change in women feeling knowledgeable in discussing pregnancy or infant care with their mothers-in-law. Participant attitudes about breastfeeding and birth spacing improved between measures for both women and men; however, 118 (42.8%) of men at follow up still agreed that exclusive breastfeeding will not result in appropriate infant growth. The proportion of respondents who agreed that their mother-in-law or mother (for men) supported delivering their babies in a health facility decreased at posttest.

With respect to decision making about MNCH care, and when to seek services, reported joint decision making by respondents and their spouses together increased between time points (Table 5). In general, there was a wide range of responses to questions about who makes health care decisions with “respondent,” “spouse,” “respondent and spouse jointly,” and “mother-in-law/mother” all being common choices.

We examined differences between baseline and follow up in participants’ MNCH knowledge that overlapped with the MAMA message content. Generally, knowledge was higher for all indicators among women compared to men at both time points (Table 6). Knowledge for all measured topics increased between time points for both sexes with one exception, knowledge of the maximum time the lactational amenorrhea method has a reliable contraceptive effect; knowledge differences between women and men diminished on several items at follow up. Knowledge of applying chlorhexidine to a newborn’s umbilical cord stump improved substantially and knowing to start breastfeeding within 1 hour after delivery improved substantially among men.

Table 4. Attitudes toward maternal, newborn, and child health care at baseline and follow up, by participant sex, in four provinces of Afghanistan (N=729).

Item	Women (n=453)		Men (n=276)	
	Baseline, n (%)	Follow up, n (%)	Baseline, n (%)	Follow up, n (%)
Feel knowledgeable discussing health care for pregnancy with spouse				
Agree	395 (87.2)	405 (89.4)	228 (82.6)	247 (89.5)
Neutral	33 (7.3)	25 (5.5)	21 (7.6)	18 (6.5)
Disagree	19 (4.2)	20 (4.4)	10 (3.6)	5 (1.8)
Refused	6 (1.3)	3 (0.7)	17 (6.2)	6 (2.2)
Feel knowledgeable discussing health care for baby with spouse				
Agree	395 (87.2)	413 (91.2)	224 (81.2)	248 (89.9)
Neutral	33 (7.3)	24 (5.3)	28 (10.1)	13 (4.7)
Disagree	19 (4.2)	14 (3.1)	12 (4.4)	11 (4)
Refused	6 (1.3)	2 (0.4)	12 (4.4)	4 (1.5)
Feel knowledgeable discussing pregnancy health care with mother-in-law (mother)				
Agree	333 (73.5)	334 (73.7)	206 (74.6)	222 (80.4)
Neutral	42 (9.3)	62 (13.7)	29 (10.5)	11 (4)
Disagree	35 (7.7)	32 (7.1)	21 (7.6)	29 (10.5)
Refused	43 (9.5)	25 (5.5)	20 (7.3)	14 (5.1)
Feel knowledgeable discussing baby health care with mother-in-law (mother)				
Agree	333 (73.5)	327 (72.2)	215 (77.9)	219 (79.4)
Neutral	42 (9.3)	56 (12.4)	27 (9.8)	11 (4)
Disagree	36 (8)	43 (9.5)	18 (6.5)	33 (12)
Refused	42 (9.3)	27 (6)	16 (5.8)	13 (4.7)
Believe it is against Islam to use birth-spacing methods				
Agree	94 (20.8)	67 (14.8)	73 (26.5)	35 (12.7)
Neutral	31 (6.8)	12 (2.7)	26 (9.4)	17 (6.2)
Disagree	310 (68.4)	359 (79.3)	158 (57.3)	205 (74.3)
Refused	18 (4)	15 (3.3)	19 (6.9)	19 (6.9)
Believe exclusive breastfeeding is inadequate; babies need supplemental foods/liquids				
Agree	171 (37.8)	76 (16.8)	138 (50)	118 (42.8)
Neutral	44 (9.7)	31 (6.8)	29 (10.5)	39 (14.1)
Disagree	220 (48.6)	341 (75.3)	95 (34.4)	113 (40.9)
Refused	18 (4)	5 (1.1)	14 (5.1)	6 (2.2)
Believe mother-in-law (mother) supports health facility delivery				
Agree	360 (79.5)	340 (75.1)	236 (85.5)	226 (81.9)
Neutral	26 (5.7)	47 (10.4)	11 (4)	13 (4.7)
Disagree	32 (7.1)	40 (8.8)	19 (6.9)	16 (5.8)
Refused	35 (7.7)	26 (5.7)	10 (3.6)	21 (7.6)

Table 5. Reported health decision makers within the household among participants, by sex, across four provinces in Afghanistan (N=729).

Health decision maker	Women (n=453)		Men (n=276)	
	Baseline, n (%)	Follow up, n (%)	Baseline, n (%)	Follow up, n (%)
Regarding mother's health				
Respondent	100 (22.1)	83 (18.3)	65 (23.6)	79 (28.6)
Spouse	151 (33.3)	139 (30.7)	41 (14.9)	29 (10.5)
Respondent and spouse jointly	73 (16.1)	114 (25.2)	89 (32.3)	104 (37.7)
Mother-in-law/mother	80 (17.7)	76 (16.8)	57 (20.7)	42 (15.2)
Husband/wife and mother-in-law/mother jointly	18 (4)	8 (1.8)	4 (1.5)	0 (0)
Other relative	27 (6)	33 (7.3)	17 (6.2)	21 (7.6)
No response/refused	4 (0.9)	0 (0)	3 (1.1)	1 (0.4)
Regarding infant/child health				
Respondent	106 (23.4)	110 (24.3)	54 (19.6)	49 (17.8)
Spouse	125 (27.6)	109 (24.1)	60 (21.7)	49 (17.8)
Respondent and spouse jointly	93 (20.5)	127 (28)	89 (32.3)	116 (42)
Mother-in-law/mother	86 (19)	73 (16.1)	55 (19.9)	41 (14.9)
Husband/wife and mother-in-law/mother jointly	21 (4.6)	3 (0.7)	2 (0.7)	4 (1.5)
Other relative	16 (3.5)	31 (6.8)	15 (5.4)	17 (6.2)
No response/refused	6 (1.3)	0 (0)	1 (0.4)	0 (0)
Regarding care-seeking				
Respondent	82 (18.1)	81 (17.9)	90 (32.6)	74 (26.8)
Spouse	214 (47.2)	152 (33.6)	28 (10.1)	27 (9.8)
Respondent and spouse jointly	64 (14.1)	115 (25.4)	93 (33.7)	118 (42.8)
Mother-in-law/mother	48 (10.6)	64 (14.1)	34 (12.3)	28 (10.1)
Husband/wife and mother-in-law/mother jointly	11 (2.4)	12 (2.7)	2 (0.7)	4 (1.5)
Other relative	31 (6.8)	29 (6.4)	27 (9.8)	24 (8.7)
No response/refused	3 (0.7)	0 (0)	2 (0.7)	1 (0.4)

Table 6. Comparison of reported maternal, newborn, and child health care awareness and knowledge differences by time point and sex across four provinces in Afghanistan with the paired McNemar Chi square test (N=729).

Item	Women (n=453)			Men (n=276)		
	Baseline, n (%)	Follow up, n (%)	*P* value	Baseline, n (%)	Follow up, n (%)	*P* value
Knew any reason to take iron supplements	265 (58.5)	377 (83.2)	<.001	129 (46.7)	161 (58.3)	.004
Knew ≥4 antenatal care visits recommended	297 (65.6)	345 (76.2)	<.001	164 (59.4)	188 (68.1)	.03
Knew ≥1 pregnancy warning signs	382 (84.3)	436 (96.3)	<.001	197 (71.4)	246 (89.1)	<.001
Knew ≥1 childbirth warning signs	395 (87.2)	433 (95.6)	<.001	211 (76.5)	252 (91.3)	<.001
Knew ≥1 way to keep baby warm	351 (77.5)	427 (94.3)	<.001	185 (67.0)	242 (87.7)	<.001
Knew to apply chlorhexidine to cord	201 (44.4)	303 (66.9)	<.001	80 (29.0)	183 (66.3)	<.001
Knew when to start breastfeeding	400 (88.3)	422 (93.2)	.01	151 (54.7)	222 (80.4)	<.001
Knew recommended duration to practice exclusive breastfeeding	392 (86.5)	427 (94.3)	<.001	130 (47.1)	207 (75.0)	<.001
Knew maximum time LAM[a] can be practiced	63 (13.9)	57 (12.6)	.50	22 (8.0)	39 (14.1)	.02

[a]LAM: lactational amenorrhea method.

Discussion

Principal Results

Our results indicate that the adapted MAMA program was feasible and acceptable to implement in Afghanistan. Consistent with other evidence from Afghanistan, most participants preferred to receive voice calls rather than SMS text messages [16]. Nearly all participants continued to receive messages 6 to 8 months after enrollment and reported that they benefitted from the program. Women who received voice messages more often reported that their husband or mother-in-law engaged with the messages compared to women who received text messages. Text message recipients received more messages on average than those who received voice messages, but this is explained by the fact that the character limit for text messages required some MAMA messages to be divided into two or three separate texts. In open-ended responses, participants who said that they had missed messages or were no longer subscribers usually cited reasons related to not having the phone or SIM card that was used to enroll in the program. Phone sharing within households also appears to have contributed to missed messages. However, at follow up, less than 3% of women and men stated they were not receiving any messages on a weekly basis, suggesting technical viability of the program in this and possibly other fragile contexts. Men reported that they felt more knowledgeable discussing MNCH topics at follow up. Reported joint decision making between spouses about MNCH care modestly increased. Knowledge of several lifesaving MNCH issues and interventions, including pregnancy warning signs, newborn umbilical cord care, and breastfeeding, improved over the 6-month assessment period among both women and men.

Comparison With Prior Work

The proliferation of mobile phones is an important development in the Afghan context [4,28]; however, phone sharing practices, low literacy, a weak communications infrastructure, and social norms that limit women's ability to communicate, make, and act on health decisions are all factors that may hinder the success of an mHealth program in such a context. This study is one of few conducted in fragile settings such as Afghanistan to examine the feasibility and acceptability of using mobile phones for health education and to promote behavior change. Moreover, despite evidence of the primacy of male heads-of-household in health decision making [13], this is the only study conducted in Afghanistan to date that describes the implementation of an mHealth program that includes both women and men.

A recent review found mixed results and insufficient evidence that mHealth interventions improve MNCH outcomes [29]. However, many studies have tried to assess changes in health outcomes such as rates of antenatal care attendance, as opposed to more proximal measures such as knowledge and attitudes [26]. Additionally, most interventions have focused on only one stage of MNCH care (pregnancy, birth, postpartum, or infant care) [29], whereas the present evaluation spans the pregnancy and postpartum periods.

Another unique aspect of this study is the comparison of voice to SMS text message formats regarding participant characteristics, feasibility, and acceptability measures. In several studies conducted in a variety of contexts, voice calls were preferred over SMS text messages [30]; however, only one study conducted in Malawi compared SMS text and voice messages for improving the knowledge and uptake of maternal and newborn health practices [21]. The authors found that SMS text messages had a higher rate of successful message delivery and SMS participants were significantly more likely to report a behavior change compared to voice message participants; however, some participants could only access voice messages due to low literacy. In our study, a surprisingly large number of MAMA subscribers opted to receive text messages, despite low literacy. Additional research is needed to better understand the factors influencing message format selection and the benefits and limitations of each of these methods of message delivery in this context.

Two papers have presented findings from different analyses of the MAMA program in Bangladesh, called Aponjon, which is operated by the nonprofit social enterprise Dnet [24,25]. The first reported results were from a cross-sectional survey conducted among 255 women who had subscribed to Aponjon during pregnancy for at least 3 months and 389 matched controls for each subscriber, as well as 345 women who subscribed to Aponjon during the postpartum period and 455 matched nonsubscribers [24]. The researchers compared results on >75 knowledge and behavioral indicators between exposed and nonexposed women, finding statistically significant differences for 24 of the indicators. However, there was no clear pattern in the results to indicate an association between a specific knowledge area or behavior and the intervention. The authors reported that when comparing degrees of exposure (categorized as nonexposure, 3-5 months exposure, and 6-12 months exposure), they observed statistically significant associations between 6-12 month exposure and scoring in the 50th percentile or greater on composite measures of newborn knowledge and newborn health practices compared to those with no program exposure [25]. The second paper reported the results of a subanalysis comparing perinatal outcomes between women who started using the Aponjon program during pregnancy with those of women who enrolled in Aponjon postpartum. The authors found that exposure to program messages during pregnancy was not associated with skilled birth attendance, initiating breastfeeding within 2 hours after delivery, timing of the first bath for the baby, or frequency of postnatal care visits [24]. Although our study design precludes attributing changes in knowledge or attitudes to the intervention, the changes observed among both sexes are promising and warrant further study with research designed to test the effectiveness of the intervention on knowledge and behavioral outcomes.

Limitations

There are several limitations of this study. We acknowledge that the single-group study design precludes attributing changes in knowledge, attitudes, and decision making to the MAMA intervention, and therefore we treat these results as exploratory. The ability to generalize our findings is also limited. Those who opted to subscribe to the MAMA program may differ substantially from those who opted not to subscribe; however, community health workers reported that very few of the women they approached did not subscribe. Additionally, our analyses

included only 55% of women who subscribed to the pilot program; therefore, these results may not reflect subscribers as a whole. Participants who could be located and consented to be interviewed at both time points may have had higher engagement and more positive views of the program. Our measurement of MNCH knowledge items reflects participant knowledge at different stages of partial exposure to the program. Because the baseline interview was administered in some cases up to 10 weeks after enrollment and participants were at various stages of pregnancy, some participants were already exposed to some of the knowledge items we assessed at baseline, while others were not. At follow up, participants had similarly varying exposure to newborn care messages. Similarly, due to phone sharing within households, we elected not to assess whether knowledge change was proportionate to the number of messages received. Finally, there may have been confounding effects through exposure to other MNCH information channels during the pilot program. Community health workers at the community level and health care providers are the primary sources of MNCH messages and education in rural Afghanistan, although some households may also receive health information from other community-level volunteers (eg, Family Health Action Group members), radio, or television [31,32]. We are unaware of any other new health education programs implemented in the same communities at the same time as the MAMA pilot program. The existing sources may have created some bias, which we cannot rule out due to lack of a comparator group.

Conclusions

We found that Afghan women and men welcomed an mHealth MNCH educational program and were largely able to enroll and access weekly voice or text messages with few technological complications, and continued program use from pregnancy through early infancy. Participants reported that they benefited by gaining health knowledge, and many stated that they discussed the messages with their family and peers. Joint decision making between spouses appears to have increased during the intervention period, and including husbands in the program was more often described as beneficial by participants when compared with the discussion of messages with mothers-in-law; both are potential areas for further research in this cultural context. There were improvements in MNCH knowledge measures among both male and female participants; however, given the important limitations of the research that has been conducted on this approach, future research of this and similar programs should be designed to rigorously evaluate the effect of the intervention on MNCH knowledge and health behaviors. Prior to further effectiveness evaluation, we hope that the data from this assessment will be used to secure ongoing and sustained support to allow the MAMA program to be refined and implemented at a greater scale in Afghanistan.

Acknowledgments

We thank the participants and their families for their time and trust, and for permitting study workers into their homes and communities for data collection. We thank the Basic Package of Health Services implementing partner staff and the provincial health department program staff in Balkh, Herat, Kabul, and Kandahar provinces for their time and efforts in guiding study workers within communities and facilitating study activities. We also acknowledge the efforts of our study workers for intensive efforts in engaging families, community leaders, and community health workers in a time-sensitive fashion and for their attention to detail. Last, this publication has been made possible by the generous support of the American people as part of the United States Agency for International Development (USAID) HEMAYAT project and the USAID Afghanistan FP/MNCH Project (AID-306-A-15-00002). The contents of this manuscript are the responsibility of the authors and the HEMAYAT Project/Jhpiego Corporation and do not necessarily reflect the views of USAID or the US Government.

Conflicts of Interest

None declared.

References

1. WHO, World Bank Group, United Nations Population Division. Trends in maternal mortality: 1990 to 2015. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/194254/9789241565141_eng.pdf?sequence=1 [accessed 2020-03-24]
2. Salama P, Alwan A. Building health systems in fragile states: the instructive example of Afghanistan. Lancet Glob Health 2016 Jun;4(6):e351-e352 [FREE Full text] [doi: 10.1016/S2214-109X(16)30067-5] [Medline: 27198829]
3. Akseer N, Salehi AS, Hossain SMM, Mashal MT, Rasooly MH, Bhatti Z, et al. Achieving maternal and child health gains in Afghanistan: a Countdown to 2015 country case study. Lancet Glob Health 2016 Jun;4(6):e395-e413 [FREE Full text] [doi: 10.1016/S2214-109X(16)30002-X] [Medline: 27198844]
4. Central Statistics Organization, Ministry of Public Health, Kabul, Afghanistan, The DHS Program, ICF. Afghanistan Demographic and Health Survey 2015. 2017 Jan. URL: https://dhsprogram.com/pubs/pdf/FR323/FR323.pdf [accessed 2020-03-24]
5. Mugali RR, Mansoor F, Parwiz S, Ahmad F, Safi N, Higgins-Steele A, et al. Improving immunization in Afghanistan: results from a cross-sectional community-based survey to assess routine immunization coverage. BMC Public Health 2017 Apr 04;17(1):290 [FREE Full text] [doi: 10.1186/s12889-017-4193-z] [Medline: 28376806]
6. Hadi A, Mujaddidi N, Rahman T, Ahmed J. The inaccessibility and utilization of antenatal health-care services in Balkh Province of Afghanistan. Asia-Pacific Pop J 2007 Apr;22(1):29-42 [FREE Full text] [doi: 10.18356/633fb489-en]

7. Higgins-Steele A, Yousufi K, Sultana S, Ali AS, Varkey S. Ending Preventable Child Deaths from Pneumonia and Diarrhoea in Afghanistan: An Analysis of Intervention Coverage Scenarios Using the Lives Saved Tool. J Trop Med 2017;2017:3120854. [doi: 10.1155/2017/3120854] [Medline: 28298932]
8. Mayhew M, Hansen PM, Peters DH, Edward A, Singh LP, Dwivedi V, et al. Determinants of skilled birth attendant utilization in Afghanistan: a cross-sectional study. Am J Public Health 2008 Oct;98(10):1849-1856. [doi: 10.2105/AJPH.2007.123471] [Medline: 18703445]
9. Najafizada SAM, Bourgeault IL, Labonté R. Social Determinants of Maternal Health in Afghanistan: A Review. Cent Asian J Glob Health 2017;6(1):240 [FREE Full text] [doi: 10.5195/cajgh.2017.240] [Medline: 29138735]
10. Haider S, Todd C, Ahmadzai M, Rahimi S, Azfar P, Morris JL, et al. Childbearing and contraceptive decision making amongst Afghan men and women: a qualitative analysis. Health Care Women Int 2009 Oct;30(10):935-953. [doi: 10.1080/07399330903052129] [Medline: 19742366]
11. Amowitz LL, Reis C, Iacopino V. Maternal mortality in Herat Province, Afghanistan, in 2002: an indicator of women's human rights. JAMA 2002 Sep 11;288(10):1284-1291. [doi: 10.1001/jama.288.10.1284] [Medline: 12215139]
12. Newbrander W, Natiq K, Shahim S, Hamid N, Skena NB. Barriers to appropriate care for mothers and infants during the perinatal period in rural Afghanistan: a qualitative assessment. Glob Public Health 2014;9(Suppl 1):S93-S109 [FREE Full text] [doi: 10.1080/17441692.2013.827735] [Medline: 24003851]
13. van Egmond K, Naeem AJ, Verstraelen H, Bosmans M, Claeys P, Temmerman M. Reproductive health in Afghanistan: results of a knowledge, attitudes and practices survey among Afghan women in Kabul. Disasters 2004 Sep;28(3):269-282. [doi: 10.1111/j.0361-3666.2004.00258.x] [Medline: 15344941]
14. Akseer T, Hayat K. Afghanistan in 2019. A Survey of the Afghan People. San Francisco: The Asia Foundation; 2019. URL: https://asiafoundation.org/wp-content/uploads/2019/12/2019_Afghan_Survey_Full-Report.pdf [accessed 2020-03-24]
15. US Agency for International Development. A survey of Afghan women's access to mobile technology. 2013. URL: https://www.usaid.gov/sites/default/files/documents/1871/survey_afghan_women_mobile.pdf [accessed 2020-03-24]
16. Yamin F, Kaewkungwal J, Singhasivanon P, Lawpoolsri S. Women's Perceptions of Using Mobile Phones for Maternal and Child Health Support in Afghanistan: Cross-Sectional Survey. JMIR Mhealth Uhealth 2018 Apr 10;6(4):e76 [FREE Full text] [doi: 10.2196/mhealth.9504] [Medline: 29636317]
17. Feroz A, Perveen S, Aftab W. Role of mHealth applications for improving antenatal and postnatal care in low and middle income countries: a systematic review. BMC Health Serv Res 2017 Nov 07;17(1):704 [FREE Full text] [doi: 10.1186/s12913-017-2664-7] [Medline: 29115992]
18. Shiferaw S, Spigt M, Tekie M, Abdullah M, Fantahun M, Dinant G. The Effects of a Locally Developed mHealth Intervention on Delivery and Postnatal Care Utilization; A Prospective Controlled Evaluation among Health Centres in Ethiopia. PLoS One 2016;11(7):e0158600 [FREE Full text] [doi: 10.1371/journal.pone.0158600] [Medline: 27383186]
19. Skinner D, Delobelle P, Pappin M, Pieterse D, Esterhuizen TM, Barron P, et al. User assessments and the use of information from MomConnect, a mobile phone text-based information service, by pregnant women and new mothers in South Africa. BMJ Glob Health 2018;3(Suppl 2):e000561 [FREE Full text] [doi: 10.1136/bmjgh-2017-000561] [Medline: 29713504]
20. Hazra A, Khan ME, Mondal SK. Mobile Phone Messaging to Husbands to Improve Maternal and Child Health Behavior in India. J Health Commun 2018;23(6):542-549. [doi: 10.1080/10810730.2018.1483444] [Medline: 29902122]
21. Crawford J, Larsen-Cooper E, Jezman Z, Cunningham SC, Bancroft E. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. Glob Health Sci Pract 2014 Feb;2(1):35-46 [FREE Full text] [doi: 10.9745/GHSP-D-13-00155] [Medline: 25276561]
22. World Health Organization. Recommendations on digital interventions for health system strengthening. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-03-24]
23. Maternal and Child Survival Program. Lessons from Country Programs Implementing the Mobile Alliance for Maternal Action Programs in Bangladesh, South Africa, India and Nigeria, 2010-2016. Baltimore: Jhpiego; 2017. URL: https://www.ictworks.org/wp-content/uploads/2018/09/MAMA-Full-Report.pdf [accessed 2020-03-24]
24. Alam M, D'Este C, Banwell C, Lokuge K. The impact of mobile phone based messages on maternal and child healthcare behaviour: a retrospective cross-sectional survey in Bangladesh. BMC Health Serv Res 2017 Jun 24;17(1):434 [FREE Full text] [doi: 10.1186/s12913-017-2361-6] [Medline: 28645278]
25. Chowdhury ME, Shiblee SI, Jones HE. Does mHealth voice messaging work for improving knowledge and practice of maternal and newborn healthcare? BMC Med Inform Decis Mak 2019 Sep 05;19(1):179 [FREE Full text] [doi: 10.1186/s12911-019-0903-z] [Medline: 31488159]
26. 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 Jul;29(7):890-897. [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]
27. Alami SO, Dulli L, Dal Santo L, Rastagar SH, Seddiqi S, Hemat S, et al. Getting health information to internally displaced youth in Afghanistan: can mobile phone technology bridge the gap? Gates Open Res 2019 Jun 4;3:1483. [doi: 10.12688/gatesopenres.13008.1]

28. Akseer T, Maxwell-Jones C. Afghanistan in 2018: A Survey of the Afghan People. San Francisco: The Asia Foundation; 2018. URL: https://asiafoundation.org/wp-content/uploads/2018/12/2018_Afghan-Survey_fullReport-12.4.18.pdf [accessed 2020-03-24]
29. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review. JMIR Mhealth Uhealth 2018 Jan 09;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]
30. Amankwaa I, Boateng D, Quansah DY, Akuoko CP, Evans C. Effectiveness of short message services and voice call interventions for antiretroviral therapy adherence and other outcomes: A systematic review and meta-analysis. PLoS One 2018;13(9):e0204091 [FREE Full text] [doi: 10.1371/journal.pone.0204091] [Medline: 30240417]
31. Najafizada SAM, Labonté R, Bourgeault IL. Community health workers of Afghanistan: a qualitative study of a national program. Confl Health 2014;8:26 [FREE Full text] [doi: 10.1186/1752-1505-8-26] [Medline: 25904976]
32. Edward A, Branchini C, Aitken I, Roach M, Osei-Bonsu K, Arwal SH. Toward universal coverage in Afghanistan: A multi-stakeholder assessment of capacity investments in the community health worker system. Soc Sci Med 2015 Nov;145:173-183 [FREE Full text] [doi: 10.1016/j.socscimed.2015.06.011] [Medline: 26141453]

Abbreviations

HEMAYAT: Helping Mothers and Children Thrive in Afghanistan
MAMA: Mobile Alliance for Maternal Action
mHealth: mobile health
MNCH: maternal, newborn, and child health
MOPH: Ministry of Public Health
SMS: short message service
USAID: United States Agency for International Development

Edited by G Eysenbach; submitted 18.12.19; peer-reviewed by M Pensak, C Reis, F Yamin; comments to author 05.02.20; revised version received 01.05.20; accepted 13.05.20; published 20.07.20.

Please cite as:
Lebrun V, Dulli L, Alami SO, Sidiqi A, Sultani AS, Rastagar SH, Halimzai I, Ahmadzai S, Todd CS
Feasibility and Acceptability of an Adapted Mobile Phone Message Program and Changes in Maternal and Newborn Health Knowledge in Four Provinces of Afghanistan: Single-Group Pre-Post Assessment Study
JMIR Mhealth Uhealth 2020;8(7):e17535
URL: https://mhealth.jmir.org/2020/7/e17535
doi:10.2196/17535
PMID:32706690

Original Paper

The Effect of Women's Differential Access to Messages on Their Adoption of Mobile Health Services and Pregnancy Behavior in Bangladesh: Retrospective Cross-Sectional Study

Mafruha Alam[1], MPH; Cathy Banwell[1], PhD; Kamalini Lokuge[1], PhD

Australian National University, ACT, Australia

Corresponding Author:
Mafruha Alam, MPH
Australian National University
62 Mills Road, National Centre for Epidemiology and Population Health
Research School of Population Health, Australian National University
ACT, 2601
Australia
Phone: 61 6125 5602
Email: mafruha.alam@gmail.com

Abstract

Background: Text or voice messages have been used as a popular method for improving women's knowledge on birth preparedness and newborn health care practices worldwide. The Aponjon service in Bangladesh provides twice-weekly messages to female subscribers about their pregnancy and newborn care on mobile phones that they own or share with family members. It is important to understand whether women's singular access to a phone affects their service satisfaction and the adoption of health messages before deploying such interventions in resource-limited settings.

Objective: This study aims to evaluate the effect of women's singular and shared access to mobile phone messages on their service utilization and perceived behavioral change around birth preparedness and pregnancy care.

Methods: In 2014, Aponjon conducted a retrospective cross-sectional survey of 459 female subscribers who received text or voice messages during their pregnancy by themselves (n=253) or with family members (n=206). We performed multivariable regression analyses to investigate the association between pregnant women's differential access to messages and other socioeconomic factors and outcomes of service satisfaction, ability to recall service short code, ability to identify danger signs of pregnancy, preference for skilled delivery, arrangement of a blood donor for delivery and pregnancy complications, maternal nutrition, use of potable drinking water, and washing hands with soap for hygiene.

Results: In the multivariable analysis, women who had singular access to messages had higher odds of reporting high satisfaction (odds ratio [OR] 1.72, 95% CI 1.12-2.63; *P*=.01), recalling the service short code (OR 2.88, 95% CI 1.90-4.36; *P*<.001), consuming nutritious food 5 times a day (OR 1.58, 95% CI 1.04-2.40; *P*=.03), and following the instructions of Aponjon on drinking potable water (OR 1.90, 95% CI 1.17-3.09; *P*=.01) than women who shared access with family members. Women's differential access to messages did not affect their knowledge of danger signs and preparedness around delivery. Adolescent women and women aged 20-24 years had lower odds of planning safe deliveries than older women (aged≥25 years). Secondary education was statistically significantly associated with women's ability to recall the short code and pregnancy danger signs, plan safe delivery, and select blood donors for emergencies. Higher family income was associated with women's satisfaction, recognition of danger signs, and arrangement of blood donors and nutritious diet. Women who received more than 4 antenatal care visits had higher odds of liking the service, preferring skilled delivery, recalling danger signs, and consuming nutritious food.

Conclusions: The capacity of women to independently access mobile phone messages can improve their adoption of mobile health services and some pregnancy health care practices. A holistic approach and equitable support are required to improve access to resources and knowledge of delivery preparedness among low-literate and younger women in low-income households.

(JMIR Mhealth Uhealth 2020;8(7):e17665) doi:10.2196/17665

KEYWORDS

mHealth; inequality; access; pregnancy

Introduction

Background

The World Health Organization (WHO) defined mobile health (mHealth) as a medical and public health practice that is supported by mobile devices and other wireless devices [1]. Many developing countries around the world have recognized that mHealth can provide crucial lifesaving information in remote settings where the health workforce is scarce and have adopted appropriate mHealth interventions [1]. Mobile phone reminders and text messages have helped improve facility utilization for maternal and neonatal care in developing countries [2,3]. However, mHealth implementers have had difficulty reaching out to women who had no access to phones [2]. Approximately 21% of women worldwide have no phone, mainly because of cost, husband's disapproval, and lack of technical knowledge on how to operate a phone [4]. Women's non–phone ownership is most prevalent in Africa, the Middle East, and South Asia [5], regions known to contribute heavily to the global burden of maternal and neonatal deaths [6-8].

Bangladesh, a South Asian country with a population of approximately 160 million people in an area of 147,570 km^2, has made significant progress in reducing the under-5 mortality rate and the maternal mortality ratio (MMR), which used to be one of the highest in the world [9,10]. However, MMR and neonatal mortality rate (NMR) remain at 170 per 100,000 live births and 23 per 1000 live births, respectively, which are far behind the targets set by the millennium development goals [9,10]. Similar to other member countries of the United Nations, Bangladesh needs continued investment and innovative approaches to reduce the MMR to 70 per 100,000 live births and NMR to 12 per 1000 live births, including improving health and well-being of all people of all ages by 2030 under the postmillennium sustainable development agenda [11,12]. mHealth initiatives could boost Bangladesh's progress as mobile phone subscription has increased exponentially since its introduction in the early 90s [13]. Low-income households in rural settings are increasingly subscribing to mobile phones because of competitive pricing among telecom operators and mobile phone manufacturers [13]. Although the gender gap in mobile phone ownership is closing, it still exists. A study in rural Bangladesh revealed that ownership of mobile phones was almost 1.8 times more among men than women, and men were more likely to own a phone at an earlier age [14]. Women in low-income households are expected to share phones with family members more often than men [4,14,15]. Mobile phone acquisition was higher in households where only the husband was literate compared with households where only the wife was literate [13].

Inequity in access to technology such as mobile phones because of social, cultural, and economic differences is broadly known as the *digital divide* [13]. mHealth implementers in resource-limited settings need to address the *digital divide*, as women with shared or no access to phones and limited education may never receive health messages, especially voice messages that cannot be stored and read later like text messages [14-16]. In resource-limited settings where women do not have personal mobile phones, providing health workers or midwives with mobile phones for improved counseling and referrals has been tested [17,18]. However, this approach does not address the severe scarcity of skilled health workforce in remote areas [19]. As *phone sharing* is a common scenario in low-income households, it is important to understand how *shared access* to mHealth messages impacts women's adoption of recommended maternal and newborn health and well-being practices [15]. Previous studies reported low to moderate changes in maternal facility utilization among women who had *any form of access* to mobile phones compared with women with *no access* [20-22]. There is a dearth of literature evaluating women's *independent* and *shared* access to targeted messages on well-being practices during maternity.

Objectives

To cover the gap in existing knowledge on the impact of the *digital divide* on pregnancy and birth preparedness, we conducted a study of female subscribers enrolled in a mobile phone–based health education messaging service during their pregnancy. The objective of our study was to investigate the association between how women accessed messages and their (1) satisfaction and intention to access the mHealth service and (2) their knowledge and practice around birth preparedness and pregnancy wellness.

Methods

The Intervention

The *Aponjon* (meaning someone very dear or close in Bangla) service was the first national mobile phone–based health education service for pregnant women and mothers of 0- to 11-month-old babies in Bangladesh that was deployed as the first project of the global initiative the Mobile Alliance for Maternal Action in 2012 and had received endorsement from the Bangladeshi government and financial support from the United States Agency for International Development [23,24]. The Aponjon service typically provides a range of maternal and newborn well-being information to registered Bangladeshi women according to their gestational stage [23]. Any pregnant woman can enroll in the service at any stage of her pregnancy to receive pregnancy messages and then in the postpartum service for another year after childbirth [23]. Usually during pregnancy or after childbirth, potential subscribers are contacted by community health workers of partner nongovernment organizations and included in the service with their consents [23,25]. Alternatively, potential subscribers can enroll in the service by directly calling the service short code, which has been advertised widely through television commercials, billboards, newspaper advertisements, and leaflets [23,25]. Ideally, it is expected that a pregnant woman will enroll in the service during pregnancy and, at the end of her pregnancy, will be upgraded to the maternal service for another year on successful birth of a child. Subscribers are reminded to contact the service call center by dialing the service short code for a service upgrade and complain or queries.

Pregnant women receive a range of information on birth preparedness, including recognition of danger symptoms of pregnancy, labor and delivery, decision making around delivery

places and skilled birth attendants, arrangement of blood donors for complications during delivery and pregnancy-related emergencies, nutrition during pregnancy, hand washing procedure before eating, food preparation, and toilet and safe drinking water. The service was piloted for a year before rolling out in 2012 [23]. Subscriber women may choose between interactive voice resonance (IVR) and text messages as a service mode [25]. Each IVR message in Bangla is 1 min long, whereas text messages are in transliterated Bangla to support all mobile phone and contain up to 161 characters [25].

Aponjon female subscribers are enrolled in the service on confirmation that they have at least one mobile phone (own or shared) in their house that they can access at a certain time of the day. During enrollment, subscribers are asked to indicate the preferred time and days of the week to receive messages. The idea is that subscribers would have the handset for accessing messages at the specified time [23]. A female subscriber typically receives 2 messages a week from the service and an additional message for her husband or family member on a separate or the same mobile number if she includes them in the service [23,25]. The messages have been designed in accordance with the national and international guidelines on maternal, neonatal, and infant health care by accredited content experts [23,24].

Subscribers are typically charged 2.3 Bangladeshi Taka (BDT; US $0.03) per message with a *no fees* option for marginalized households [23,24]. Since mid-2013, the service has offered consultations with medically trained doctors through a 24-hour call center for clients [26]. Subscribers can consult doctors on maternal and newborn health issues for as long as they want, while being charged the normal call rate per minute [26]. In 2014, the service had more than one million subscriber bases [25].

Study Area

We analyzed data obtained from a cross-sectional survey as part of routine operations research conducted by Dnet, the implementing agency of Aponjon. The survey was conducted between February and April 2014 in selected subdistricts of Bogura, Bagerhat, Patuakhali, Chittagong, and Laxmipur districts in Bangladesh, which were purposively selected to reflect the remoteness, cultural diversity, geographical dispersion, and maximum acquisition of subscribers [25]. Administratively, Bangladesh is divided into 8 divisions, 64 districts (*zillas*), and 545 subdistricts (*upazillas*) [9,27]. The subdistricts or *upazillas* are further divided into urban and rural areas; rural areas in an *upazilla* consist of union *parishads* (UP), whereas *mouzas* (cluster of villages) make up each UP [28,29]. Urban areas in an *upazilla* are divided into wards and *mahallas* (cluster of households) within wards [28,29]. Bogura is a northern district in Bangladesh, which is also known as the industrial powerhouse of the North Bengal [30]. Bagerhat and Patuakhali are 2 districts in southwestern Bangladesh and lie in the fringe of Bay of Bengal [31,32]. Chittagong (Chattagram) is a district in the southeastern part of Bangladesh, known for the sea port and hill tracts [33], whereas Laxmipur is a district in the southern part of Bangladesh [34]. The average size of households in these districts varied from 3.8 to 5.1 (Bogura: 3.8, Bagerhat: 4.13, Chittagong: 5.1, Laxmipur: 4.71, and Patuakhali: 4.41) in 2011 [30-34]. Among the selected districts, Bagerhat and Chittagong have the highest average literacy rate at 58.98% (male: 59.97% and female: 57.99%) and 58.90% (male: 61.1% and female: 56.7%), respectively [31,33].

Sampling

The survey included subscriber women at different stages of pregnancy. Pregnant women were eligible to participate if they received pregnancy messages for at least two months and did not undergo a planned or unplanned pregnancy termination. Pregnant women who had just given birth to newborn baby but had not upgraded the service for the postpartum period were also eligible for the survey. Adolescent women aged less than 18 years were excluded from the survey.

A sampling frame with details of approximately 2274 potential survey respondents (who matched the inclusion criteria) was prepared from the Aponjon service database. The survey database contained information such as subscriber's name, address, cell phone number, age, date of beginning of last menstruation period, enrollment date, and type of access and socioeconomic information. On the basis of formative research experience, Dnet estimated a priori that a sample size of approximately 400 respondents was required, with an anticipated ratio of shared to independent access of 1:2, to detect a difference in proportions with outcomes of interest between shared and independent access of 15%, with 80% power and a 5% significance level. Assuming availability of pregnant women at home and consent rate of 60%, it was expected that 660 subscribers would need to be contacted for the study.

Owing to the lack of availability of an adequate number of community health workers who could assist field researchers in identifying households of subscribers, only 839 subscribers were potentially available to be contacted from the list. Eligible subscribers were randomly sampled district by district from the existing list until the proposed number of respondents had been recruited. A group of 24 field researchers, 2 field supervisors, and 1 central coordinator conducted the survey. A pair of male and female researchers conducted each interview at the respondent woman's house after ensuring privacy [25]. Each interview lasted approximately 1 hour. Before visiting the respondents at their home, researchers contacted the respondents over the phone to make an appointment. Initially, researchers read aloud information on the survey, confidentiality issues, benefits, and possible risks associated with participating in the survey. Respondents had the right to withdraw from the survey at any time and could refuse to answer any question. Verbal and written consents were received before each interview. Identification of respondents such as name, location, and cell phone number was replaced with IDs to maintain anonymity. The interviews were conducted in Bangla. Eligible respondents who could not be reached over the phone for an appointment or were not at home when field researchers visited or who did not want to participate were excluded from the survey [25]. Mostly, pregnant women who had relocated to have their birth at their parent's house outside the study area could not be interviewed.

Respondents were administered semistructured questionnaires that contained questions regarding pregnancy behavioral outcomes as well as service-related questions such as how women accessed the service, satisfaction with the service, recall of service short code, and perceived knowledge and benefits from the service.

All research instruments and consent forms for this service were reviewed by an international institutional review board (IRB) and received an exemption from IRB review. The authors received approval to analyze the survey data of Aponjon service from Dnet, the implementing agency of Aponjon service, and the Science and Medical Delegated Ethics Review Committee of the Australian National University.

Measures

The explanatory variable *type of access* was derived from the question, "Who accessed messages?," with responses options being "me," "me and family member," or "family member." We recorded the responses to *women=1* and *women or family member*=0. We hypothesized that women who accessed messages alone were more empowered and would access messages more regularly than women who shared access with family members, which would impact women's adoption of recommended practices in these 2 groups.

Our outcome of interest, satisfaction of respondents, was measured by asking them to rate the service on a 5-point Likert scale, ranging from 1 (very bad) to 5 (excellent). We recoded *high satisfaction* by categorizing scores 4 and 5 as *high=1* and scores 1 to 3 as *low=0*.

The outcome variable *recalls short code* was determined from a service-related question, "Please tell me the service short code (the number you see flash on your mobile phone when Aponjon messages come)." Respondents who were able to recall the 5 digits correctly were coded as *knows short code=1* and who could not were codes as *does not know short code=0*. We considered this variable to demonstrate women's familiarity with the service.

We measured the respondents' knowledge of pregnancy danger signs. They were asked, "Can you tell me the danger signs?" These included high fever, severe headache, blurry vision, convulsions, vaginal bleeding, severe lower abdominal pain, swelling in hands and feet, hypertension, and less fetal movement. We generated a binary variable *able to recall danger signs* defined as *yes=1* if respondents could recall at least two signs and *no=0* for respondents who could not recall.

WHO's birth preparedness and complication readiness (BPCR) plan recommends that pregnant women and their families are aware of planning the following elements before delivery: the place of birth, the birth attendant who will provide assistance during delivery, the location of the closest facility for delivery and emergency, arrangement of funds for any expenses related to delivery and emergency, supplies and materials necessary to bring to the facility, an identified labor and birth companion, an identified person who will look after the home and other children while pregnant woman is away, arrangement of transport to a facility for delivery and emergency, and the identification of compatible blood donors for delivery and emergency [35]. The outcome variable regarding birth preparedness around delivery for our study was derived from 3 sequential questions asked to pregnant women. The first question was, "Have you decided where your baby will be born?," with 2 responses *yes* or *no*, and the second question for women who responded *yes* to first question, "(If decided) where do you want your baby to be born?" with 2 possible responses *hospital or health facility* and *home*. The third question was, "(For home based delivery) who will deliver your baby?," with 2 responses *untrained birth attendant/relatives* and *trained birth attendant*. Responses to these 3 questions were grouped into binary responses as *Planned skilled delivery=1* and *Planned untrained birth attendants at home or made no plans yet=0*.

Outcome variable "selected blood donor for delivery or pregnancy emergency" was derived from the question, "Did you select a blood donor for delivery or complications during pregnancy?" Responses such as "yes I did" was coded as *yes=1*, and other responses such as "no," "I did not know," "haven't thought about it," or "I don't find it's necessary" were coded as *no=0*.

Having a balanced diet with essential macro and micronutrients is immensely important for pregnant women. Malnourished mothers are likely to give birth to premature babies with low birth weights who are at risk of dying within the first week of birth [9,36,37]. A maternal nutritional behavior–related binary outcome variable *consumed nutritious food 5 times a day* was derived from the question, "How many times in a day you eat one of these food items-vegetables, fruits, protein (such as milk, fish, meat, chicken, egg)?," with count responses such as 1, 2, 3, 4, 5, and 6. Reponses 5 and above were recoded as *consumed nutritious food 5 times a day=1*, and responses less than 5 were recoded as *no=0*.

Bangladesh, similar to other South Asian countries, has a supply of poor-quality drinking water, which is contaminated with microbial pathogens and pollutants (such as arsenic) that are responsible for diarrheal and other infectious diseases [38]. Aponjon advises female subscribers to ensure a clean source of drinking water and treat the water before drinking. A binary variable "followed Aponjon's instruction on drinking water" was constructed from 2 questions. The first question was, "What is your source of drinking water?," with 2 responses "I drink water that has been properly treated with various methods (such as boiling, filter)" and "I drink surface water directly from source (such as tap, pond)." The second question was, "Where did you learn about treating drinking water?," with 2 response options *Aponjon* and *Other*. The responses of these 2 questions were grouped into *yes=1* and *no=0* for our outcome variable.

Aponjon provides information on washing hands properly before food handling, taking meals, and after cleaning body parts. Women were asked, "How do you clean your hands for hygiene purposes?," with possible answers such as "only water" and "with soap or disinfectants." Women who replied "with soap or disinfectants" were asked, "Where did you learn to wash your hands in this way?," with possible answers "Aponjon," "Aponjon and others," and "other sources." Women who answered "Aponjon" or "Aponjon and others" were classified as "followed Aponjon's hand-washing procedure=1" and who

answered "other sources" as "followed hand-washing procedure learnt from other sources=0."

Other covariates considered for this study were respondent women's age (<20, 20-24, and ≥25 years), education (none or primary, junior secondary, and secondary or above), family income (BDT ≤10,000; 10,0001-20,000; and >20,000), first-time pregnancy (yes or no), place of residence (urban and rural), and districts (A, B, C, D, and E). We labeled district names to maintain anonymity. We also considered the frequency of antenatal care (ANC) visits (>4 ANC visits and ≤4 ANC visits) for all models. All these variables were selected because of their importance in the study or because of their demonstrated association with maternal and newborn health-related outcomes in mHealth interventions in previous research [9,20].

Statistical Analysis

The respondent characteristics are described using frequencies and percentages overall and by differential access groups (*women* and *women or family members*). Chi-squared tests were performed to examine the distribution of sample background characteristics and outcome variables by explanatory variables (who accessed messages).

We undertook multiple multivariable logistic regression analysis to investigate the relationship between *who accessed messages* and the outcomes of interest. Multicollinearity tests were performed using the variance inflation factor (VIF) test to assess the correlation between covariates. A VIF score of greater than 2 was set as a threshold, and variables at this threshold were dropped from the final models. We performed logistic regression for the following outcomes of interest: *high satisfaction on the service*, *recalled short code*, *planned delivery by skilled attendant at home or a hospital*, *able to recall danger signs*, *selected a blood donor for delivery or pregnancy complications*, *consumed nutritious food 5 times a day*, *followed Aponjon instruction on potable drinking water*, and *followed Aponjon's instruction on hand-washing*. Multivariable models were adjusted for women's age, education, family income, first pregnancy, ANC visits, place of residence, and district (to account for the sampling strategy). The Hosmer-Lemeshow test was used to test the goodness of fit for the models, and a model was considered a good fit when the *P* value was nonsignificant. Statistical significance was determined at a *P* value of less than .05. We used IBM SPSS Statistics for Windows version 24.0 for the analysis. The results are expressed as adjusted odds ratio (OR) with 95% CIs.

Results

Background Information of Respondents

Of the 687 subscribers contacted, approximately 66.8% (459/687) of them who had a successful live birth recently (209/459, 45.5%) or were in different stages of pregnancy (250/459, 54.5%) were interviewed. Socioeconomic characteristics of respondents are described in Table 1. Women's differential access to messages showed statistically significant differences in women's educational levels and districts. Approximately 98.9% (454/459) of women received voice messages over text. The outcome variables are described in Table 2.

Table 1. Background characteristics of participants by access to messages (N=459).

Variables	Who accessed messages		*P* value
	Women (n=253), n (%)	Women or family member (n=206), n (%)	
Age of respondents (years)			.35
<20	41 (16.2)	44 (21.4)	
20-24	127 (50.2)	95 (46.1)	
≥25	85 (33.6)	67 (32.5)	
Educational qualification of women			.005
None or primary education	77 (30.4)	88 (42.7)	
Junior secondary education	90 (35.6)	73 (35.4)	
Secondary school or higher	86 (34.0)	45 (21.9)	
Family monthly income, BDT[a] (US $)			.10
≤10,000 (118.15)	105 (41.5)	100 (48.5)	
10,000-20,000 (118.15-236.30)	95 (37.5)	87 (42.2)	
>20,000 (236.30)	53 (21.0)	19 (9.3)	
First-time pregnancy			.97
Yes	131 (51.8)	107 (51.9)	
No	122 (48.2)	99 (48.1)	
Place of residence			.99
Urban	140 (55.3)	114 (55.3)	
Rural	113 (44.7)	92 (44.7)	
Districts			.03
District A	30 (11.9)	31 (15.0)	
District B	68 (26.9)	39 (18.9)	
District C	65 (25.7)	41 (20.0)	
District D	36 (14.2)	29 (14.1)	
District E	54 (21.3)	66 (32.0)	
Number of antenatal care visits			.42
>4	153 (60.5)	132 (64.1)	
≤4	100 (39.5)	74 (35.9)	

[a]BDT: Bangladeshi Taka.

Table 2. Unadjusted outcome variables by type of access.

Outcome variables	Who accessed messages		*P* value
	Women, n (%)	Women or family member, n (%)	
High satisfaction (N=459)			.004
Yes	181 (71.5)	121 (58.7)	
No	72 (28.5)	85 (41.3)	
Recalled short code (n=458)			<.001
Yes	175 (69.4)	88 (42.7)	
No	77 (30.6)	118 (57.3)	
Able to recall danger signs (N=459)			.19
Yes	149 (58.9)	109 (52.9)	
No	104 (41.1)	97 (47.1)	
Planned skilled delivery (n=248)			.77
Yes	73 (49.7)	52 (51.5)	
No	74 (50.3)	49 (48.5)	
Identified blood donor (n=249)			.03
Yes	34 (22.8)	12 (12.0)	
No	115 (77.2)	88 (88.0)	
Consumed nutritious food 5 times a day (n=454)			.04
Yes	135 (54.4)	92 (44.7)	
No	113 (45.6)	114 (55.3)	
Followed instruction on potable drinking water (n=457)			.005
Yes	79 (31.2)	40 (19.6)	
No	174 (68.8)	164 (80.4)	
Followed Aponjon messages on hand washing (n=453)			.31
Yes	65 (26.1)	45 (22.1)	
No	184 (73.9)	159 (77.9)	

Effect on the Adoption of Mobile Health Services

Whether female subscribers were the sole receivers of the messages had statistically significantly increased their satisfaction with the service (OR 1.72, 95% CI 1.12-2.63; *P*=.01) and ability to recall the service short code correctly (OR 2.88; 95% CI 1.90-4.36; *P*<.001; Table 3).

Multivariable analysis of service satisfaction suggests that besides the *type of access*, women's middle-income (OR 1.96, 95% CI 1.24-3.11; *P*=.004) and higher-income (OR 2.48, 95% CI 1.25-4.93; *P*=.009) background and completion of more than 4 ANC visits (OR 2.86, 95% CI 1.82-4.48; *P*<.001) significantly impacted their satisfaction with the service than their counterparts (Table 3). Similarly, covariates such as completion of secondary education (OR 2.28, 95% CI 1.30-4.02; *P*=.004) and age group (20-24 years) were statistically significantly associated with women's ability to recall the short code correctly (OR 1.81, 95% CI 1.08-3.04; *P*=.02). Women from districts A (OR 3.88, 95% CI 1.50-10.00; *P*=.005), B (OR 2.39, 95% CI 1.17-4.86; *P*=.01), and D (OR 3.19, 95% CI 1.25-8.15; *P*=.01) had higher odds of expressing higher satisfaction than women from district E.

Table 3. Predictors of women's adoption of mobile health service.

Variables	High satisfaction		Recalled short code	
	Adjusted OR[a] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
Who accessed messages				
Women	1.72 (1.12-2.63)	.01	2.88 (1.90-4.36)	<.001
Women or family member (reference)	1 (N/A[b])	N/A	1 (N/A)	N/A
Age of respondents (years)				
<20	0.88 (0.43-1.85)	.89	1.01 (0.50-2.03)	.97
20-24	0.84 (0.49-1.43)	.84	1.81 (1.08-3.04)	.02
≥25 (reference)	1 (N/A)	N/A	1 (N/A)	N/A
Educational qualification of women				
No/primary (reference)	1 (N/A)	N/A	1 (N/A)	N/A
Junior secondary	0.96 (0.58-1.59)	.88	1.17 (0.72-1.90)	.51
Secondary or higher	1.08 (0.60-1.92)	.80	2.28 (1.30-4.02)	.004
Family monthly income, BDT[c] (US $)				
≤10,000 (118.15; reference)	1 (N/A)	N/A	1 (N/A)	N/A
10,001-20,000 (118.15-236.30)	1.96 (1.24-3.11)	.004	1.44 (0.92-2.26)	.11
>20,000 (236.30)	2.48 (1.25-4.93)	.009	1.66 (0.86-3.20)	.13
First-time pregnancy				
Yes	1.12 (0.66-1.89)	.66	1.10 (0.65-1.80)	.71
No (reference)	1 (N/A)	N/A	1 (N/A)	N/A
Number of antenatal care visits				
>4	2.86 (1.82-4.48)	<.001	1.17 (0.75-1.82)	.48
≤4 (reference)	1 (N/A)	N/A	1 (N/A)	N/A
Place of residence				
Urban	0.56 (0.29-1.09)	.09	1.34 (0.74-2.45)	.33
Rural (reference)	1 (N/A)	N/A	1 (N/A)	N/A
Districts				
District A	3.88 (1.50-10.00)	.005	1.49 (0.65-3.44)	.35
District B	2.39 (1.17-4.86)	.01	0.99 (0.52-1.93)	.99
District C	1.70 (0.73-3.95)	.16	1.08 (0.48-2.41)	.85
District D	3.19 (1.25-8.15)	.01	1.44 (0.60-3.44)	.41
District E (reference)	1 (N/A)	N/A	1 (N/A)	N/A

[a]OR: odds ratio.

[b]Not applicable.

[c]BDT: Bangladeshi Taka.

Effect on Readiness for Pregnancy and Delivery Complications

The type of access to messages did not show a statistically significant effect on women's readiness for pregnancy and delivery complications (Table 4). However, completion of secondary education was associated with women's ability to recall danger signs (OR 2.44, 95% CI 1.40-4.26; *P*=.001) and arrange skilled delivery (OR 2.37, 95% CI 1.05-5.32; *P*=.02) and blood donors (OR 5.33, 95% CI 1.74-16.32; *P*=.003). Similarly, women who received more than 4 ANC visits had higher odds of showing the expected behavioral outcomes (danger signs: OR 1.86, 95% CI 1.21-2.86; *P*=.004; skilled delivery: OR 2.13, 95% CI 1.15-3.96; *P*=.02; and blood donor: OR 3.71, 95% CI 1.52-9.02; *P*=.004) than women who had lower number of ANC visits. In addition, women from higher-income households had higher odds of recalling danger signs (OR 3.15, 95% CI 1.63-6.06; *P*=.001) and selecting blood donors (OR 4.48, 95% CI 1.42-14.08; *P*=.01) than women from the lower-income households. Age was statistically significantly

associated with women's delivery decisions; pregnant women who were aged less than 20 years (OR 0.17, 95% CI 0.06-0.47; *P*=.001) or aged 20 to 24 years (OR 0.31, 95% CI 0.15-0.63; *P*=.001) had lower odds of planning delivery with skilled birth attendant at home or a hospital than women who were aged 25 years or older. District was a confounder for 2 behavioral outcomes; women residing in district B had higher odds of planning skilled delivery (OR 4.59, 95% CI 1.77-11.90; *P*=.002) and arranging blood donors (OR 10.87, 95% CI 2.97-39.78; *P*<.001) than women from district E.

Table 4. Predictors of subscribers' readiness for pregnancy and delivery complications.

Variables	Able to recall danger signs		Planned skilled delivery		Selected blood donor for delivery and emergency	
	Adjusted (OR[a] 95% CI)	*P* value	Adjusted (OR 95% CI)	*P* value	Adjusted (OR 95% CI)	*P* value
Who accessed messages						
Women	1.07 (0.71-1.61)	.71	0.76 (0.42-1.37)	.37	1.62 (0.70-3.75)	.25
Women or family member (reference)	1 (N/A[b])	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Age of respondents (years)						
<20	0.74 (0.37-1.47)	.40	0.17 (0.06-0.47)	.001	1.34 (0.35-5.06)	.66
20-24	0.81 (0.48-1.35)	.43	0.31 (0.15-0.63)	.001	0.87 (0.34-2.21)	.77
≥25 (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Educational qualification of women						
No/primary (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Junior secondary	0.99 (0.62-1.59)	.97	1.34 (0.66-2.71)	.34	1.54 (0.54-4.44)	.41
Secondary or higher	2.44 (1.40-4.26)	.001	2.37 (1.05-5.32)	.02	5.33 (1.74-16.32)	.003
Family monthly income, BDT[c] (US $)						
≤10,000 (118.15; reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
10,001-20,000 (118.15-236.30)	2.08 (1.35-3.21)	.001	0.89 (0.47-1.70)	.74	1.86 (0.72-4.77)	.19
>20,000 (236.30)	3.15 (1.63-6.06)	.001	1.52 (0.62-3.68)	.35	4.48 (1.42-14.08)	.01
First-time pregnancy						
Yes	0.86 (0.52-1.41)	.55	1.61 (0.80-3.20)	.17	1.18 (0.46-3.03)	.72
No (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Number of antenatal care visits						
>4	1.86 (1.21-2.86)	.004	2.13 (1.15-3.96)	.016	3.71 (1.52-9.02)	.004
≤4 (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Place of residence						
Urban	1.00 (0.55-1.80)	.99	0.70 (0.32-1.55)	.38	0.45 (0.16-1.20)	.11
Rural (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Districts						
District A	0.80 (0.36-1.81)	.60	0.82 (0.25-2.66)	.74	2.21 (0.43-11.27)	.33
District B	1.11 (0.58-2.13)	.74	4.59 (1.77-11.90)	.002	10.87 (2.97-39.78)	<.001
District C	1.14 (0.51-2.52)	.74	0.94 (0.30-2.87)	.91	0.45 (0.08-2.46)	.36
District D	1.39 (0.59-3.29)	.44	1.56 (0.43-5.64)	.49	2.02 (0.38-10.62)	.40
District E (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A

[a]OR: odds ratio.

[b]Not applicable.

[c]BDT: Bangladeshi Taka.

Effect on Maternal Wellness Behavior at the Household Level

Women's sole access to mobile phone–based educational messages showed a statistically significant effect on the availability of nutritional elements 5 times a day (OR 1.58, 95% CI 1.04-2.40; *P*=.03) and access to potable drinking water (OR 1.90, 95% CI 1.17-3.09; *P*=.01; Table 5). Whether women solely accessed phone messages did not show a statistically significant association with their hand washing practices.

Higher family income (OR 2.40, 95% CI 1.25-4.61; *P*=.008) and more than 4 ANC visits (OR 3.46, 95% CI 2.22-5.40; *P*<.001) were other determinants of women's improved food consumption behavior (Table 5). District was a confounder for pregnant women's self-reported wellness behavioral outcomes at home; women from districts C (hand washing: OR 3.24, 95% CI 1.15-9.12; *P*=.02) and D (nutrition: OR 0.31, 95% CI 0.13-0.74; *P*=.008; drinking water: OR 7.26, 95% CI 2.78-18.9; *P*<.001; and hand washing: OR 10.73, 95% CI 3.75-30.63; *P*<.001) had higher odds of showing improved behavior than women from district E.

Table 5. Predictors of following maternal wellness instructions at the household level.

Variables	Consumed nutritious food 5 or more times per day		Followed Aponjon messages on drinking water		Followed Aponjon messages on hand washing	
	Adjusted OR[a] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
Who accessed messages						
Women	1.58 (1.04-2.40)	.03	1.90 (1.17-3.09)	.01	1.31 (0.80-2.13)	.28
Women or family member (reference)	1 (N/A[b])	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Age of respondents (years)						
<20	1.72 (0.85-3.48)	.13	1.19 (0.54-2.60)	.66	1.03 (0.46-2.34)	.93
20-24	1.11 (0.66-1.86)	.68	0.91 (0.51-1.64)	.77	0.82 (0.44-1.50)	.52
≥25 (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Educational qualification of women						
No or primary (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Junior secondary	1.52 (0.93-2.49)	.09	0.87 (0.49-1.53)	.63	0.82 (0.46-1.48)	.52
Secondary or higher	1.73 (0.99-3.03)	.05	0.80 (0.42-1.55)	.51	0.66 (0.34-1.28)	.22
Family monthly income, BDT[c] (US $)						
≤10,000 (118.15; reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
10,001-20,000 (118.15-236.30)	1.53 (0.98-2.40)	.06	1.25 (0.74-2.09)	.39	1.43 (0.85-2.42)	.17
>20,000 (236.30)	2.40 (1.25-4.61)	.008	0.80 (0.38-1.66)	.55	0.57 (0.24-1.31)	.18
First-time pregnancy						
Yes	0.87 (0.52-1.44)	.60	1.20 (0.68-2.12)	.51	0.95 (0.53-1.71)	.87
No (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Number of antenatal care visits						
>4	3.46 (2.22-5.40)	<.001	0.94 (0.58-1.52)	.80	0.66 (0.40-1.09)	.11
≤4 (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Place of residence						
Urban	1.26 (0.69-2.31)	.44	1.42 (0.73-2.74)	.29	1.59 (0.78-3.21)	.19
Rural (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A
Districts						
District A	0.89 (0.39-2.06)	.79	0.44 (0.13-1.47)	.18	1.49 (0.49-4.52)	.47
District B	0.52 (0.26-1.02)	.06	2.23 (1.01-4.94)	.04	1.74 (0.69-4.38)	.23
District C	0.62 (0.28-1.40)	.26	1.92 (0.75-4.92)	.17	3.24 (1.15-9.12)	.02
District D	0.31 (0.13-0.74)	.008	7.26 (2.78-18.9)	<.001	10.73 (3.75-30.63)	<.001
District E (reference)	1 (N/A)	N/A	1 (N/A)	N/A	1 (N/A)	N/A

[a]OR: odds ratio.

[b]Not applicable.

[c]BDT: Bangladeshi Taka.

Discussion

Principal Findings

We evaluated the association between women's unequal access to mobile phones and their perception about using mobile phone–based health services for pregnancy and birth preparedness in their socioeconomic context. Understanding the socioeconomic factors that contribute to women's access to mHealth services has implementation benefits in low-income settings where sharing a single phone among family members is a common phenomenon [4,14,15].

Our findings suggest that women's differential access to targeted messages positively affects their satisfaction and familiarity with the service short code and the adoption of certain recommended practices around pregnancy wellness, such as inclusion of nutritious elements in regular diets and drinking potable water. These findings were consistent when the models were adjusted for all covariates (age, education, income, first pregnancy, ANC frequency, and place of residence) and district. These findings confirm 2 aspects of the adoption of mobile phone–based educational service for pregnancy: (1) *sole and uninterrupted access to messages* provide target women an understanding of the service before they decide to adopt or reject a new innovation (mHealth) [39,40] and (2) women who access messages by themselves are in a better position to evaluate the context, need, complexity, and relative advantage of the information provided by mobile phones than women who occasionally or never access messages [39]. Our findings also suggest that women's age, education, and family income were statistically significantly associated with their mHealth experience. Although national initiatives and investment to improve employment opportunities and secondary school education should continue, mHealth implementers could ensure equity in accessing messages among subscribers by facilitating equitable approaches such as lending mobile phones to women from underprivileged households for the entire service period and program-specific training of clients, especially older women (aged ≥25 years) and women who have not completed secondary education, on operating mobile phones and accessing messages at the established times. We recommend additional equitable approaches, such as mHealth-supported cash vouchers for pregnant women and infants [41], to address prevailing food shortages in underprivileged households [42]. Furthermore, mobile phone–based messaging services have the potential to increase awareness of public health problems such as water-borne diseases and arsenic contamination in groundwater [38].

In our study, women's differential access to messages did not affect their choice of skilled birth attendant at home or a hospital, knowledge about the danger signs of pregnancy, and arrangement of blood donors for delivery and complications. Instead, we found, similar to previous studies, higher frequency of ANC visits and women's privileged background to have significantly improved maternal knowledge about the danger signs and adoption of BPCR measures [9,43]. We need to evaluate these findings carefully, as these behavioral outcomes may have been influenced by existing national campaigns and overall improvements in the utilization of maternal and child health care facilities [19,28,44]. For example, the rate of facility-based delivery has almost doubled from 23% in 2010 to 47% in 2016 [45], which could be a result of the government-funded voucher schemes to support marginalized pregnant women to have facility-based deliveries, checkups, tests, and arrangement of transport to facilities at free of cost [44]. Our recent exploratory study suggested that women, in general, were well informed about delivery and postpartum care guidelines by their community health workers, hospital staff, and national media during their regular ANC visits and were rather interested in receiving information on newborn care and nutrition from Aponjon service [46]. The study also suggested that although women from higher-income households could afford services at private hospitals, home-based deliveries were usually preferred by low-income households for uncomplicated births because of convenience, cost, and fear of C-section and were organized by elderly female family members and local (untrained) birth attendants [46,47].

Our findings are concerning though as younger women, especially who were in their late teens, did not have a plan for delivery with skilled birth attendant or at a facility. The national survey estimates that pregnancy-related mortality ratio among 15- to 19-year-old women has almost doubled from 75 deaths in 2010 to 144 deaths per 100,000 live births in 2016 and that first-time mothers are at increased risk of maternal deaths (215 per 100,000 live births) compared with mothers with live children [45]. Educating a large population of teenaged pregnant women on safe motherhood is challenging for the Bangladeshi government, especially in a situation where one-third of 15- to 19-year-old girls are likely to experience pregnancy before reaching adulthood [9,48]. Hence, the importance of repeat visits and family counseling of community health workers during pregnancy in low-income households, especially for adolescent women who are likely to rely on husbands and mothers-in-laws for delivery decisions, is undeniable [25,43,48,49].

Our research is limited by a number of factors. First, the results would be strengthened by a randomized control trial or pre-post quasi-experimental design rather than a cross-sectional survey. Small differences in the behavioral outcome associated with women's type of access could not be captured because of the power of the study that was designed a priori based on a moderate to large association; the study required a larger sample size for greater generalizability of our findings. Second, the study relied on respondents' self-reported behavioral outcomes rather than their attendance report at facilities or actual habits, and there may be problems with recall bias and overreporting. Third, the study could have benefited from system-generated data on the actual number of messages accessed by these 2 different groups of respondents, as a study in the United States reported a positive association between access to high frequency of messages and abstinence from alcohol consumption during the postpartum period [50].

Way Forward

Nevertheless, our research will be helpful for mHealth implementers who are working to reduce maternal and neonatal deaths in developing countries. Our research confirms that *women's sole access to messages* can change their perception of the service and adoption of maternal wellness messages. We infer from our findings that irrespective of how women access messages, one-way (voice) messages alone may not improve delivery decisions, which are controlled by socioeconomic circumstances of individual families and their personal experiences [46]. Therefore, mHealth implementers working in resource-limited settings need to consider cultural and socioeconomic constraints that limit women's access to mHealth services and all other maternal health care services and should adopt a holistic approach to ensure equity in their service. ANC visits (more than 4) remain an independent factor for women's

knowledge about pregnancy danger signs, birth preparedness, and maternal dietary diversity. mHealth services such as Aponjon may consider introducing mobile phone apps for health workers and a linked referral system to increase their efficiency in identifying high-risk mothers during monthly door-to-door ANC visits, maintaining follow-up visits, minimizing workload, and referring pregnant women to proper facilities [18,51-54]. Future studies should include anthropometric measurement and hospital data to understand the effect of pregnancy messages on birth outcomes [55,56].

Conclusions

Socioeconomic and cultural barriers to mobile phone access by women can be problematic for implementing mHealth services in resource-limited settings. Although the national policy targeting poverty alleviation and women's empowerment requires revisions, stand-alone mHealth implementers need to integrate with local health infrastructure and invest in building women's capacity to access mobile phones. We suggest ongoing research to monitor temporal trends in women's differential access to mobile phone messages in South Asian countries.

Acknowledgments

The authors acknowledge Dnet's (the implementing agency of Aponjon in Bangladesh) support to use their survey data on the Aponjon service and permission to conduct this research. Dnet received catalytic support from the United States Agency for International Development for launching the service. The authors would like to thank all the respondents and data enumerators of the survey. This research was conducted as part of a PhD study and was supported by the Research School of Population Health, the Australian National University (ANU), and an Australian Government Research Training Program Scholarship. The views in this publication are those of the authors only and do not necessarily reflect the views of the institutions involved in this publication.

Authors' Contributions

MA played a key role in the design, conception, data collection, coding, analysis, and interpretation of data and was a major contributor in writing the manuscript. MA worked for Dnet before beginning her PhD at ANU and was involved with Aponjon's research. CB and KL were involved in designing the work and critically revising the manuscript for important intellectual content. All authors proofread, approved the manuscript, and are accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2020-06-05]
2. Tamrat T, Kachnowski S. Special delivery: an analysis of mhealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012 Jul;16(5):1092-1101. [doi: 10.1007/s10995-011-0836-3] [Medline: 21688111]
3. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
4. Striving and Surviving: Exploring the Lives of Women at the Base of the Pyramid. GSMA. 2012. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2013/01/GSMA_mWomen_Striving_and_Surviving-Exploring_the_Lives_of_BOP_Women.pdf [accessed 2020-06-05]
5. Women & Mobile: A Global Opportunity - A Study on the Mobile Phone Gender Gap in Low and Middle-Income Countries. GSMA. 2010. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2013/01/GSMA_Women_and_Mobile-A_Global_Opportunity.pdf [accessed 2020-06-05]
6. Matermal Mortality: Key Facts 2018. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality [accessed 2020-06-05]
7. Trends in Maternal Mortality: 1990 to 2015 - Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/194254/9789241565141_eng.pdf?sequence=1 [accessed 2020-06-05]
8. Neonatal Mortality: The Neonatal Period is the Most Vulnerable Time for a Child. UNICEF DATA - Child Statistics. 2018. URL: https://data.unicef.org/topic/child-survival/neonatal-mortality/ [accessed 2020-06-05]
9. National Institute of Population Research and Training, Mitra Associates, and ICF International. Bangladesh Demographic and Health Survey 2014. Dhaka, Bangladesh, and Rockville, MD: NIPORT, Mitra and Associates, ICF International; 2016. URL: https://dhsprogram.com/pubs/pdf/FR311/FR311.pdf [accessed 2020-06-22]
10. Millennium Development Goals Bangladesh Progress Report. 2015. URL: https://www.bd.undp.org/content/bangladesh/en/home/library/mdg/mdg-progress-report-2015.html [accessed 2020-06-05]

11. Sustainable Development Goal 3: Ensure Healthy Lives and Promote Well-Being for All at All Ages. Sustainable Development Knowledge Platform. URL: https://sustainabledevelopment.un.org/sdg3 [accessed 2020-06-05]
12. Anand A, Roy N. Transitioning toward sustainable development goals: the role of household environment in influencing child health in sub-saharan Africa and South Asia using recent demographic health surveys. Front Public Health 2016;4:87 [FREE Full text] [doi: 10.3389/fpubh.2016.00087] [Medline: 27200339]
13. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile 'digital divide': changing determinants of household phone ownership over time in rural Bangladesh. JMIR Mhealth Uhealth 2015;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
14. Khatun F, Heywood AE, Hanifi SM, Rahman MS, Ray PK, Liaw S, et al. Gender differentials in readiness and use of mhealth services in a rural area of Bangladesh. BMC Health Serv Res 2017 Aug 18;17(1):573 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]
15. Poorman E, Gazmararian J, Elon L, Parker R. Is health literacy related to health behaviors and cell phone usage patterns among the text4baby target population? Arch Public Health 2014;72(1):13 [FREE Full text] [doi: 10.1186/2049-3258-72-13] [Medline: 24872883]
16. Crawford J, Larsen-Cooper E, Jezman Z, Cunningham SC, Bancroft E. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. Glob Health Sci Pract 2014 Feb;2(1):35-46 [FREE Full text] [doi: 10.9745/GHSP-D-13-00155] [Medline: 25276561]
17. Huq NL, Azmi AJ, Quaiyum MA, Hossain S. Toll free mobile communication: overcoming barriers in maternal and neonatal emergencies in rural Bangladesh. Reprod Health 2014 Jul 12;11:52 [FREE Full text] [doi: 10.1186/1742-4755-11-52] [Medline: 25015126]
18. Amoakoh-Coleman M, Borgstein AB, Sondaal SF, Grobbee DE, Miltenburg AS, Verwijs M, et al. Effectiveness of mhealth interventions targeting health care workers to improve pregnancy outcomes in low- and middle-income countries: a systematic review. J Med Internet Res 2016 Aug 19;18(8):e226 [FREE Full text] [doi: 10.2196/jmir.5533] [Medline: 27543152]
19. Bangladesh Health System Review. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/208214 [accessed 2020-06-05]
20. Jennings L, Omoni A, Akerele A, Ibrahim Y, Ekanem E. Disparities in mobile phone access and maternal health service utilization in Nigeria: a population-based survey. Int J Med Inform 2015 May;84(5):341-348. [doi: 10.1016/j.ijmedinf.2015.01.016] [Medline: 25737460]
21. Nie J, Unger JA, Thompson S, Hofstee M, Gu J, Mercer MA. Does mobile phone ownership predict better utilization of maternal and newborn health services? a cross-sectional study in Timor-Leste. BMC Pregnancy Childbirth 2016 Jul 23;16(1):183 [FREE Full text] [doi: 10.1186/s12884-016-0981-1] [Medline: 27448798]
22. Bishwajit G, Hoque MR, Yaya S. Disparities in the use of mobile phone for seeking childbirth services among women in the urban areas: Bangladesh urban health survey. BMC Med Inform Decis Mak 2017 Dec 29;17(1):182 [FREE Full text] [doi: 10.1186/s12911-017-0578-2] [Medline: 29284477]
23. Rajan R, Raihan A, Alam M, Agarwal S, Ahsan A, Bashir R, et al. MAMA ''Aponjon' Formative Research Report. Maternal and Child Health Integrated Program (MCHIP). Baltimore, MD, USA: Johns Hopkins University Global mHealth Initiative; 2013. URL: https://www.mchip.net/technical-resource/mama-aponjon-formative-research-report/ [accessed 2020-06-22]
24. Lessons from Country Programs Implementing the Mobile Alliance for Maternal Action Programs in Bangladesh, South Africa, India and Nigeria, 2010–2016. ICT Works. 2016. URL: https://www.ictworks.org/wp-content/uploads/2018/09/MAMA-Full-Report.pdf [accessed 2020-06-22]
25. Alam M, D'Este C, Banwell C, Lokuge K. The impact of mobile phone based messages on maternal and child healthcare behaviour: a retrospective cross-sectional survey in Bangladesh. BMC Health Serv Res 2017 Jun 24;17(1):434 [FREE Full text] [doi: 10.1186/s12913-017-2361-6] [Medline: 28645278]
26. Alam M, Banwell C, Olsen A, Lokuge K. Patients' and doctors' perceptions of a mobile phone-based consultation service for maternal, neonatal, and infant health care in Bangladesh: a mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11842 [FREE Full text] [doi: 10.2196/11842] [Medline: 31008716]
27. District Statistics Dhaka: Bangladesh Bureau of Statistics. Bangladesh Bureau of Statisics. 2019. URL: http://www.bbs.gov.bd/site/page/2888a55d-d686-4736-bad0-54b70462afda/- [accessed 2020-06-22]
28. World Health Organization, Ministry of Health and Family Welfare Bangladesh. Success Factors for Women's and Children's Health. Geneva: World Health Organization; 2015. URL: https://apps.who.int/iris/handle/10665/178623 [accessed 2020-06-05]
29. National Institute of Population Research and Training, MEASURE Evaluation, International Centre for Diarrheal Disease Research, Bangladesh. Bangladesh Maternal Mortality and Health Care Survey 2010. Dhaka, Bangladesh: NIPORT, MEASURE Evaluation, and ICDDR,B; 2010. URL: https://www.measureevaluation.org/resources/publications/tr-12-87 [accessed 2020-06-22]
30. Population and Housing Census District Statistics 2011: Bogura. Bangladesh Bureau of Statistics. 2011. URL: http://203.112.218.65:8008/WebTestApplication/userfiles/Image/PopCen2011/Comm_Bogra.pdf [accessed 2020-06-22]
31. Population and Housing Census District Statistics 2011: Bagerhat. Bangladesh Bureau of Statistics. 2011. URL: http://203.112.218.65:8008/WebTestApplication/userfiles/Image/PopCenZilz2011/Zila_Bagerhat.pdf [accessed 2020-06-22]

32. Population and Housing Census District Statistics 2011: Patuakhali. Bangladesh Bureau of Statistics. 2011. URL: http://203.112.218.65:8008/WebTestApplication/userfiles/Image/PopCenZilz2011/Zila_Patuakhali.pdf [accessed 2020-06-22]
33. Population and housing census district statistics 2011: Chittagong. Bangladesh Bureau of Statistics. 2011. URL: http://203.112.218.65:8008/WebTestApplication/userfiles/Image/PopCen2011/Com_Chittagong.pdf [accessed 2020-06-22]
34. Population and Housing Census District Statistics 2011: Laxmipur. Bangladesh Bureau of Statistics. 2011. URL: http://203.112.218.65:8008/WebTestApplication/userfiles/Image/PopCenZilz2011/Zila_Lakshmipur.pdf [accessed 2020-06-22]
35. World Health Organization. WHO Recommendations on Health Promotion Interventions for Maternal and Newborn Health. Geneva, Switzerland: World Health Organization; 2015. URL: https://www.who.int/maternal_child_adolescent/documents/health-promotion-interventions/en/ [accessed 2020-06-22]
36. Nutrition Situation Analysis Bangladesh. The UN Network for SUN. 2014. URL: https://www.unnetworkforsun.org/sites/default/files/2018-06/Bangladesh%20Full%20MNO_0.pdf [accessed 2020-06-05]
37. Baqui AH, Arifeen SE, Amin S, Black RE. Levels and correlates of maternal nutritional status in urban Bangladesh. Eur J Clin Nutr 1994 May;48(5):349-357. [Medline: 8055851]
38. Luby S. Water quality in South Asia. J Health Popul Nutr 2008 Jun;26(2):123-124 [FREE Full text] [Medline: 18686546]
39. Rogers EM. Diffusion of Innovations. New York, USA: Free Press; 1983.
40. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
41. Huda TM, Alam A, Tahsina T, Hasan MM, Khan J, Rahman MM, et al. Mobile-based nutrition counseling and unconditional cash transfers for improving maternal and child nutrition in Bangladesh: pilot study. JMIR Mhealth Uhealth 2018 Jul 18;6(7):e156 [FREE Full text] [doi: 10.2196/mhealth.8832] [Medline: 30021707]
42. Harris-Fry H, Azad K, Kuddus A, Shaha S, Nahar B, Hossen M, et al. Socio-economic determinants of household food security and women's dietary diversity in rural Bangladesh: a cross-sectional study. J Health Popul Nutr 2015 Jul 10;33:2 [FREE Full text] [doi: 10.1186/s41043-015-0022-0] [Medline: 26825273]
43. Pervin J, Nu UT, Rahman AM, Rahman M, Uddin B, Razzaque A, et al. Level and determinants of birth preparedness and complication readiness among pregnant women: a cross sectional study in a rural area in Bangladesh. PLoS One 2018;13(12):e0209076 [FREE Full text] [doi: 10.1371/journal.pone.0209076] [Medline: 30557336]
44. Health Bulletin. Directorate General of Health Services. 2018. URL: http://dghs.gov.bd/images/docs/Publicaations/HB%202018%20v2.pdf [accessed 2020-06-05]
45. National Institute of Population Research and Training, Mitra and Associates, and ICF International. Bangladesh Maternal Mortality and Health Care Survey 2016: Preliminary report. Dhaka, Bangladesh, and Chapel Hill, NC, USA: NIPORT, ICDDRB and MEASURE Evaluation; 2017. URL: https://www.measureevaluation.org/resources/publications/tr-17-218 [accessed 2020-06-22]
46. Alam M, Banwell C, Olsen A, Lokuge K. Adoption of mobile phone messages for delivery and newborn care in Bangladesh. Glob J Health Sci 2020 Feb 9;12(3):20. [doi: 10.5539/gjhs.v12n3p20]
47. Choudhury N, Ahmed SM. Maternal care practices among the ultra poor households in rural Bangladesh: a qualitative exploratory study. BMC Pregnancy Childbirth 2011 Mar 1;11:15 [FREE Full text] [doi: 10.1186/1471-2393-11-15] [Medline: 21362164]
48. Islam MM, Islam MK, Hasan MS, Hossain MB. Adolescent motherhood in Bangladesh: trends and determinants. PLoS One 2017;12(11):e0188294 [FREE Full text] [doi: 10.1371/journal.pone.0188294] [Medline: 29176807]
49. Akter MK, Yimyam S, Chareonsanti J, Tiansawad S. The challenges of prenatal care for Bangladeshi women: a qualitative study. Int Nurs Rev 2018 Dec;65(4):534-541. [doi: 10.1111/inr.12466] [Medline: 30176049]
50. Evans W, Nielsen PE, Szekely DR, Bihm JW, Murray EA, Snider J, et al. Dose-response effects of the text4baby mobile health program: randomized controlled trial. JMIR Mhealth Uhealth 2015 Jan 28;3(1):e12 [FREE Full text] [doi: 10.2196/mhealth.3909] [Medline: 25630361]
51. Alam M, Khanam T, Khan R. Assessing the Scope for Use of Mobile Based Solution to Improve Maternal and Child Health in Bangladesh: a Case Study. In: Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development. 2010 Presented at: ICTD'10; December 13-16, 2010; London, United Kingdom. [doi: 10.1145/2369220.2370755]
52. Bonnell S, Griggs A, Avila G, Mack J, Bush RA, Vignato J, et al. Community health workers and use of mhealth: improving identification of pregnancy complications and access to care in the Dominican Republic. Health Promot Pract 2018 May;19(3):331-340. [doi: 10.1177/1524839917708795] [Medline: 28578606]
53. Prinja S, Nimesh R, Gupta A, Bahuguna P, Gupta M, Thakur JS. Impact of m-health application used by community health volunteers on improving utilisation of maternal, new-born and child health care services in a rural area of Uttar Pradesh, India. Trop Med Int Health 2017 Jul;22(7):895-907 [FREE Full text] [doi: 10.1111/tmi.12895] [Medline: 28510997]
54. Mushamiri I, Luo C, Iiams-Hauser C, Ben Amor Y. Evaluation of the impact of a mobile health system on adherence to antenatal and postnatal care and prevention of mother-to-child transmission of HIV programs in Kenya. BMC Public Health 2015 Feb 7;15:102 [FREE Full text] [doi: 10.1186/s12889-015-1358-5] [Medline: 25886279]
55. Murthy N, Chandrasekharan S, Prakash MP, Kaonga NN, Peter J, Ganju A, et al. The impact of an mhealth voice message service (mMitra) on infant care knowledge, and practices among low-income women in India: findings from a

pseudo-randomized controlled trial. Matern Child Health J 2019 Dec;23(12):1658-1669 [FREE Full text] [doi: 10.1007/s10995-019-02805-5] [Medline: 31584144]

56. 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 Jul;29(7):890-897. [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]

Abbreviations

ANC: antenatal care
ANU: Australian National University
BDT: Bangladeshi Taka
BPCR: birth preparedness and complication readiness
IRB: international institutional review board
IVR: interactive voice resonance
mHealth: mobile health
MMR: maternal mortality ratio
NMR: neonatal mortality rate
OR: odds ratio
UP: union parishad
VIF: variance inflation factor
WHO: World Health Organization

Edited by G Eysenbach; submitted 02.01.20; peer-reviewed by Y Amor, B Loring; comments to author 20.03.20; revised version received 03.04.20; accepted 14.05.20; published 20.07.20.

Please cite as:
Alam M, Banwell C, Lokuge K
The Effect of Women's Differential Access to Messages on Their Adoption of Mobile Health Services and Pregnancy Behavior in Bangladesh: Retrospective Cross-Sectional Study
JMIR Mhealth Uhealth 2020;8(7):e17665
URL: https://mhealth.jmir.org/2020/7/e17665
doi:10.2196/17665
PMID:32706694

Original Paper

Digital Health and Inequalities in Access to Health Services in Bangladesh: Mixed Methods Study

Tanvir Ahmed[1,2], MBBS, MPH; Syed Jafar Raza Rizvi[3], BSc, MSc; Sabrina Rasheed[4], PhD; Mohammad Iqbal[5], MBBS; Abbas Bhuiya[4], PhD; Hilary Standing[1], PhD; Gerald Bloom[1], BPhil, MDCM; Linda Waldman[1], PhD

[1]Institute of Development Studies, Brighton, United Kingdom

[2]Department of Oncology and Medicine, Faculty of Medicine, Dentistry and Health, University of Sheffield, Sheffield, United Kingdom

[3]Johns Hopkins University, Dhaka, Bangladesh

[4]International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh

[5]Independent Public Health Expert, Dhaka, Bangladesh

Corresponding Author:
Tanvir Ahmed, MBBS, MPH
Department of Oncology and Medicine
Faculty of Medicine, Dentistry and Health
University of Sheffield
Jessop Wing Maternity Unit
Tree Root Walk, Broomhall
Sheffield, S10 2SF
United Kingdom
Phone: 44 7931163982
Email: tanvir79@outlook.com

Abstract

Background: Globally, the rapid growth of technology and its use as a development solution has generated much interest in digital health. In line with global trends, Bangladesh is also integrating technology into its health system to address disparities. Strong political endorsement and uptake of digital platforms by the government has influenced the rapid proliferation of such initiatives in the country. This paper aims to examine the implications of digital health on access to health care in Bangladesh, considering who uses electronic devices to access health information and services and why.

Objective: This study aims to understand how access to health care and related information through electronic means (digital health) is affected by sociodemographic determinants (ie, age, gender, education, socioeconomic status, and personal and household ownership of mobile phones) in a semiurban community in Bangladesh.

Methods: A cross-sectional survey of 854 households (between October 2013 and February 2014) and 20 focus group discussions (between February 2017 and March 2017) were conducted to understand (1) who owns electronic devices; (2) who, among the owners, uses these to access health information and services and why; (3) the awareness of electronic sources of health information; and (4) the role of intermediaries (family members or peers who helped to look for health information using electronic devices).

Results: A total of 90.3% (771/854) of households (471/854, 55.2% of respondents) owned electronic devices, mostly mobile phones. Among these, 7.2% (34/471) used them to access health information or services. Middle-aged (35-54 years), female, less (or not) educated, and poorer people used these devices the least (α=.05, α is the level of significance). The lack of awareness, discomfort, differences with regular care-seeking habits, lack of understanding and skills, and proximity to a health facility were the main reasons for not using devices to access digital health.

Conclusions: Although influenced by sociodemographic traits, access to digital health is not merely related to device ownership and technical skill. Rather, it is a combination of general health literacy, phone ownership, material resources, and technical skill as well as social recognition of health needs and inequity. This study's findings should serve as a basis for better integrating technology within the health system and ensuring equitable access to health care.

(JMIR Mhealth Uhealth 2020;8(7):e16473) doi:10.2196/16473

KEYWORDS

health equity; eHealth; mHealth; digital health; health technology; Bangladesh; developing countries

Introduction

Background

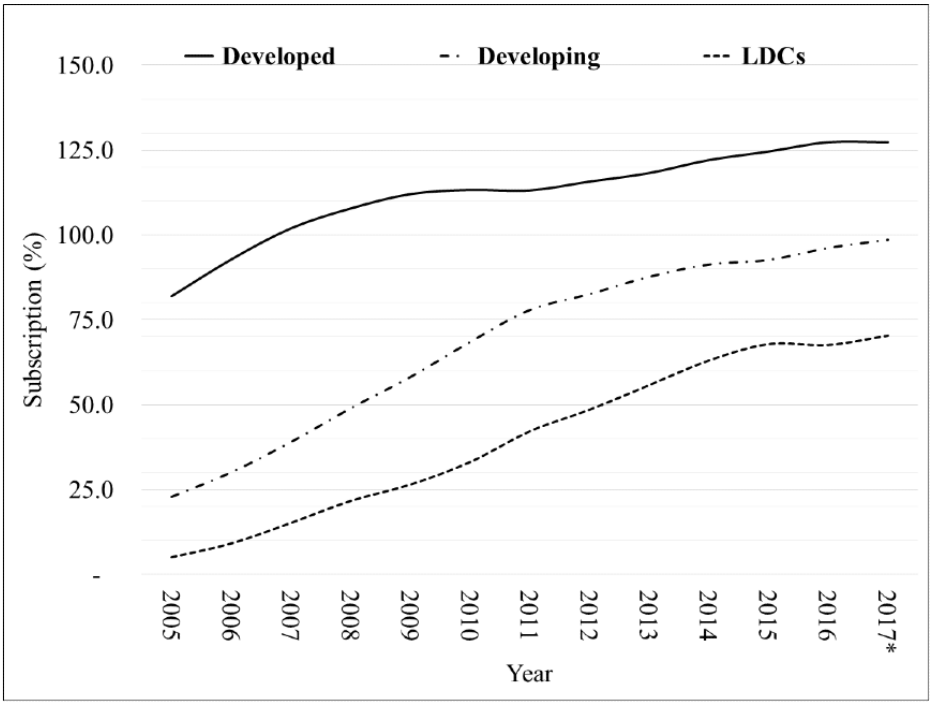
Globally, there has been an impressive growth in the number of cell phones and internet users, and the price of services and devices has decreased [1,2]. Mobile phones have become a thriving market, characterized by 7.7 billion (estimated) subscribers in 2017. Globally, the proportion of the population covered by at least a second generation (2G) network grew from 58% in 2001 to 95% in 2015. Internet penetration grew from 6.5% in 2000 to 43% in 2015, and the proportion of households with internet access at home increased from 18% in 2005 to 46% in 2015 [3,4]. The mobile-cellular subscription per 100 inhabitants has exceeded 100% in developed countries, and both developing and least developed countries (LDCs) are racing toward similar levels. Since 2005, most subscriptions have come from developing countries and LDCs, and the gap between these and developed countries is reducing (Figure 1) [3].

According to the International Telecommunication Union, a 10% increase in internet speed can increase economic growth by 1.3% in low- and middle-income countries (LMICs) [5,6]. Given such growth and potential, both national and personal spheres are now influenced by digital innovations, as governments strategize to expand coverage of developmental initiatives by reaching remote areas or individuals who engage in web-based shopping or personal banking. In 2014, the mobile-cellular industry generated US $3 trillion, which contributed to 3.8% of the global gross domestic product, and this number is expected to rise to 4.2% by 2020 [7]. This economic contribution came about largely by linking previously unconnected communities, financial inclusion through eCommerce (use of electronic means including mobile phones for financial purposes such as transferring money, everyday banking etc), and designing and delivering innovative solutions for improving quality of life (eg, sharing information, providing services, technologies for mass production). Examples include Tigo Kilimo by Tigo in Tanzania (launched in 2013) [7] and Airtel Green SIM in India (launched in 2007) for eAgriculture (use of electronic means including mobile phones to provide and access information regarding farming techniques and related products and market) [8,9] and TradeNet in Ghana [10] and bKash in Bangladesh [11,12] for eCommerce and mCommerce (using technology including mobile phones for financial transactions) as mobile wallets. M-Pesa by Vodafone is one of the largest mobile-based financial services in the world, used by millions across Africa, Europe, and Asia [13-15]. The information and communications technology (ICT) industry is now a popular area for employment worldwide. In 2014, the mobile industry directly employed approximately 13 million people and indirectly supported 12 million more [7,16-18]. Similar to other development domains, health is also exploring the potential of technology in improving the well-being of people, including health-related well-being, which can range from improvement in health ailments to acquiring the necessary information about health conditions or services.

The integration of technology within health began as electronic health (eHealth), the use of electronic platforms for the provision of health information and services, data collection and management, etc. When mobile phones are used to do the same, it is called mobile health (mHealth) [19]. Currently, because of the advent of artificial intelligence (AI) and big data, eHealth and mHealth are now called digital health. Although digital health indicates a much broader and smarter technology horizon, eHealth and mHealth remain the major forms of technology that are being endorsed for system integration to improve health system performance and increase peoples' access to health care [20]. Therefore, there is a growing number of conferences, workshops, websites, apps, and publications regarding how electronic platforms can be implemented and integrated within the health domain [1,2,21-23]. Bangladesh is also in the process of doing the same. Currently, the country has 4 mobile cellular operators and a significant subscription base (149 million). Evidence suggests that a large proportion of households own mobile phones (81%) [24-28]. The government's commitment to digital development, popularly known as Digital Bangladesh, has helped in fostering strategic and policy direction to adopt ICT in health care (and other development domains) [29-32]. Due to the large subscription base and strong political mandate, there are ≥42 internet-, SMS text messaging–, and call center–based eHealth and mHealth initiatives in the country providing awareness regarding maternal health, drug and alcohol abuse, smoking cessation, HIV/AIDS, and general health care [19,29,33-36]. Despite these initiatives, there is little systematic evidence of the impact of technology on equitable access to health services and information, especially in resource-poor settings such as Bangladesh.

Access to health includes availability, geographic location, affordability, and acceptability (by the community) of services. Considering these dimensions, the challenge is to ensure that everyone, irrespective of their social group, gender, etc, can access the required health care [37]. Bangladesh, despite its successes, has several health system challenges and prevailing health disparities, resulting in limited access to and utilization of quality health care [38,39]. Globally, one of the assumptions of the integration of technology for health care (and other development initiatives) is that digital health can contribute to equitable access to health care, especially in LMICs. However, there is evidence that access to electronic platforms can be hindered because of sociodemographic, gender, and geographic barriers—a state of disparity called the digital divide [40,41]. This divide bears great importance, especially when technology is being endorsed for system integration to improve access to health that can ensure universal health coverage (UHC) in Bangladesh [42]. Considering the growth of ICTs and the proliferation of digital health initiatives, a framework that defines the dimensions of access to health information and services through electronic means can contribute and is crucial to this integration. This is because the rapid growth of ICTs and related development solutions can often be misinterpreted as being equivalent to access and use.

Figure 1. Mobile-cellular subscription per 100 people by developed, developing, and least developed countries. LDC: least developed country.



Objective

This study aims to understand the degree to which ICTs enable Bangladeshi people to increase their access to information and improve their well-being by exploring the factors that influence whether people have access to, and use, digital health information and services (this refers to non-AI–, machine learning–, or big data–based initiatives, previously called eHealth and mHealth). It focuses on who owns electronic devices and uses their device to access health services and/or information and the implications of this for health equity. The findings can be useful in devising a framework to ensure equitable access to digital health information and services in Bangladesh and relevant contexts.

Methods

Study Setting, Population and Data Collection Tools and Techniques

The data presented in this paper were collected using a mixed methods design. The quantitative data, supported by research funding from the UK Economic and Social Research Council, comes from a household survey to understand the role of ICTs in health information seeking conducted between October 2013 and February 2014 in a subdistrict of Bangladesh called Mirzapur in the Tangail district. Mirzapur was chosen because it is semiurban with typical sociodemographic characteristics of other such periurban subdistricts in the country [43]. As there were no previous variance estimates of outcome variables, 0.5 was used to calculate the required sample size to obtain 95% CIs with plus or minus 10% precision and a design effect of 2. Thus, a sample size of 2401 was obtained, which finally became 2527, considering a 5% nonresponse rate. This was calculated for 3 settings: urban (Dhaka), rural (Chakaria), and semiurban (Mirzapur). Therefore, each setting ended up with a subsample of 842. In Mirzapur, the sample size reached 854 households. Sample households were selected from 28 villages (30 households per village) using systematic cluster sampling from a Health and Demographic Surveillance System (HDSS) [44]. Approximately 81% of the information was gathered from the household head or the spouse of the head. Where neither the head nor the spouse was present or able to take part, an adult child of the head or the child's spouse was interviewed. The survey questionnaire was developed and pretested both in Bangla and English (the Bangla version was used to collect actual data) on a browser-based, open-source platform using locally available 7-inch tablets with Android version 4.0. The use of digital means to seek health information or services was explained to the respondents as if the respondent (and/or any member of the household) had heard of or used phones or computers to seek information about any health issues or services via voice call, SMS text messaging, or internet browsing.

To further unpack the survey patterns and findings, 20 focus group discussions (FGDs) were conducted from February to March 2017 as part of the first author's PhD fieldwork. Respondents, who had never used electronic devices to seek health services or information, were selected from 6 FGD groups constructed from the HDSS using typical demographic and economic traits: 8 rich and poor young female and male (college students) groups, 8 rich and poor middle-aged female and male groups, and 4 rich and poor elderly female and male groups.

Data Analysis

The principal component analysis technique was used to elicit asset scores, and its distribution was arranged in quintiles to create the income groups. The lowest 2 quintiles were grouped as poor, and the highest 2 quintiles were grouped as rich. The distribution of the middle quintile was halved, and the lower half added to the poor group and the upper half to the rich group

to remove the complexity of constructing the FGD groups. As it was a qualitative exercise, further discussion was conducted with the group during the FGDs to subjectively triangulate the accuracy and related homogeneity of the participants' income status. Each group consisted of 4 to 5 participants, and the discussions took place at locations preferred by the participants, that is, households (for female respondents) and local gathering places (for men). Each session lasted approximately 30 min, facilitated by the first author (male and a PhD student). During discussions, although agreements in opinions were sought, disagreements were also noted along with the respondents' body language. A female notetaker was present during discussions of the female groups. Discussions were conducted and later transcribed in Bangla and then into English for internal consistency.

The sociodemographic analysis was conducted by identifying the frequencies of the respective characteristics/traits. Information on socioeconomic status (SES) was collected from the HDSS as asset scores, a popular method called the asset index. The range of scores was divided into 3 equal categories (lowest as poor, middle, and highest as rich), and frequencies of a sample belonging to the respective categories were identified. The ownership of electronic devices and the use of devices were analyzed for (1) seeking information; (2) health services or information; and (3) demographic factors such as age, gender, education, and income to stratify the results. A quantitative analysis, including tests of significant, was performed using STATA version 14 (made by StataCorp), and Microsoft Excel was used to construct graphs and some amount of data management. The patterns indicated by the quantitative analysis were further explored qualitatively using content analysis techniques [45,46]. Altogether, 3 broad themes emerged: (1) reasons for not using mobile phones or other electronic devices, (2) the extent of awareness of electronic sources of health information, and (3) the role of intermediaries (family members or peers who helped to look for health information using electronic devices). These were analyzed by reference to the gender, SES, and education of respondents. Both analyses were part of PhD studies supported by the International Development Research Centre.

Ethical Consideration

Ethical approval was obtained from the institutional review boards of the survey partners: International Centre for Diarrhoeal Disease Research, Bangladesh, and Institute of Development Studies, University of Sussex (for qualitative data collection). Before data collection, the study was explained to the participants, and written consent or thumbprints were obtained. During data collection (both qualitative and quantitative), informed consent was obtained from all participants.

Results

Sociemographic Profile of the Respondents and Ownership of Electronic Devices

Tables 1-3 show the sociodemographic characteristics and ownership of electronic devices of the survey households and respondents. The mean age of the respondents was 41.3 (SD 14.5) years, and 67.2% (574/854) of the respondents were between the ages of 25 and 54 years. A total of 38.3% (327/854) had less than primary education, and a small proportion (26/854, 3.0%) were graduates and above. More than half (494/854, 57.9%) of the households had 4 to 6 members, representing the usual Bangladeshi family size [47]. In all, 71.9% (614/854) of the respondents were female. There were more female respondents because the survey was conducted during the daytime when male members were away. A total of 73.1% (625/854) of the respondents were not employed (being a housewife is not considered a job in Bangladesh), which was not intentional. Conversely, 77.3% (660/854) of the household heads were employed. A total of 95.9% (819/854) of households had a regular income, and 92.6% (791/854) of households had no social security card (in Bangladesh, having a social security card indicates that the corresponding household belongs to the extremely poor socioeconomic group), a common feature of semiurban Bangladesh. In all, 55.2% (471/854) of respondents had personal electronic devices and 90.3% (771/854) of the households owned at least one, which was almost always a mobile phone. The personal ownership was considerably low, perhaps because the survey population had male:female ratio of 1:3 and in a context like Bangladesh, women have less access (in this case ownership) to devices compared with men. However, in the case of household device ownership, family members tend to share electronic devices, and there was always a chance of a lack or less use of devices due to the influence of other family members. Therefore, to understand the equity implications of access to electronic devices, although lower, personal ownership was considered in this study rather than shared ownership.

Table 1. Demographic statistics of the respondent (N=854).

Parameter	Values
Age (years), mean (SD)	41.3 (14.5)
16-24, n (%)	97 (11.4)
25-34, n (%)	215 (25.2)
35-44, n (%)	190 (22.3)
45-54, n (%)	169 (19.8)
55-64, n (%)	110 (12.9)
>65, n (%)	73 (8.6)
Gender, n (%)	
Male	240 (28.1)
Female	614 (71.9)
Education, n (%)	
Less than primary	327 (38.3)
Primary	206 (24.1)
Secondary	254 (29.7)
Higher secondary	41 (4.8)
Graduation and above	26 (3.0)
Respondent employment status, n (%)	
Yes	229 (26.9)
No	625 (73.1)

Table 2. Household-level information of the respondents (N=854).

Parameter	Values, n (%)
Members per household	
1-3	257 (30.1)
4-6	494 (57.9)
>7	103 (12.1)
Household head working status	
Yes	660 (77.3)
No	194 (22.7)
Socioeconomic status of the household	
Poor	295 (34.5)
Middle	276 (32.3)
Rich	283 (33.1)
Presence of menial labor[a]	
Yes	35 (4.1)
No	819 (95.9)
Household's social security card	
Yes	61 (7.1)
No	791 (92.6)
Don't know	2 (0.2)

[a]Refers to jobs such as housemaid and unskilled day laborer.

Table 3. Electronic device ownership of the respondents (N=854).

Parameter	Household, n (%)	Personal, n (%)
Ownership of devices		
Total	771 (90.28)	471 (55.2)
Mobile	751 (87.9)	454 (53.2)
Laptop	2 (0.2)	0 (0.0)
Both	18 (2.1)	17 (2.0)

Access to Health-Related Information or Services by Respondents Who Owned Electronic Devices and Influence of Sociodemographic Characteristics

Tables 4 and 5 show that everyone who owned mobile phones or laptops/computers had used those to communicate or seek information, which included day-to-day conversations to seeking specific services, such as agriculture or other government-related information. It also showed that the predominant means of seeking information was voice calls, followed by SMS text messaging. Analysis of health information seeking for any health concern, including information only, or services showed that of all people who sought health information, a number (22/34, 65%) reported seeking information/services for minor health issues. Although everyone had used their device to seek some form of information, only 7.2% (34/471) sought health services or health information.

The use of electronic devices by the people who owned personal devices was stratified by sociodemographic characteristics of age (P=.02), sex (P=.41), education (P=.004), and SES (P<.01; Table 6). Although overall use was low, the pattern suggests that among those who used their devices, people who were middle-aged (35 to 54 years), women, and poorer people who had less or no education used them less than others. The difference in use was found to be significant on the chi-square test.

Table 4. Percentage of use and awareness of the use of devices to communicate and seek information by personal device owners (cell phones, laptops, or both; n=471).

Parameter[a]	Used and aware, n (%)	Not used nor aware, n (%)
Any information		
Total	471 (100.0)	0 (0.0)
Voice call	471 (100.0)	0 (0.0)
SMS text messaging	226 (48.0)	245 (52.0)
Internet	26 (6.0)	445 (94.5)
Health-related information	34 (7.0)	437 (92.8)

[a]Multiple responses.

Table 5. Percentage of communication device use to seek health information (n=34).

Parameter	Values, n (%)
Minor health issues	22 (64.7)
Serious health issues	12 (35.3)

Table 6. Percentage of personal device owners who sought health information or services by age, sex, education, and socioeconomic status in Mirzapur (n=471).

Parameter	Used and aware, n (%)	Not used nor aware, n (%)	*P* value
Age in year[a]			.02
Young adult	5 (8.2)	56 (91.8)	
Adult	15 (9.9)	137 (90.1)	
Middle age	5 (2.8)	175 (97.2)	
Elderly and above	9 (11.5)	69 (88.5)	
Gender			.41
Male	13 (7.9)	151 (92.1)	
Female	21 (6.8)	286 (93.2)	
Education[a]			.004
No education	5 (4)	120 (96)	
Primary	4 (3.4)	113 (96.6)	
Secondary	14 (8.4)	153 (91.6)	
Higher secondary and above	11 (17.7)	51 (82.3)	
Socioeconomic status[a]			<.001
Poor	2 (1.4)	137 (98.6)	
Middle	8 (6)	126 (94)	
Rich	24 (12.1)	174 (87.9)	

[a]Statistically significant (*P* value is less than .05).

Barriers to Accessing Health Information and/or Services by the Owners of Electronic Devices

People who did not use electronic means to access health care or information were further interviewed over FGDs to explore their reasons. Almost everyone has accessed some form of electronic information at some point using mobile phones, mostly through voice calls to an office (ie, local agriculture office or bank) to ask for information. However, despite ample promotion and publicity, the provision of electronic health information or services and the words digital health were unfamiliar to many. One male student mentioned:

> *I have never heard the word digital health until today. No one told us that one can get health-related information in this way. But sometimes we make calls to some office to know about things. In this way we can get information easily.*

Female students had slightly different views. They preferred to call their friends and/or family for information, as one explained:

> *We use mobile phones to talk about many things. If we need to know about something, we call our friends or elders. But I can't remember if we have ever talked about digital health.*

Nonetheless, some of them also use mobile phones to seek health-related information despite a lack of familiarity with formal words such as digital health. Most of the participants had asked for advice from a family member or from someone who had relevant knowledge. As a result, most were reluctant to use electronic means, and only a few had used their device to seek health information, such as the internet and call center. This indicates that people use their phones to ask advice from friends, family, or social acquaintances but are not aware of or do not use formal digital health services. As one of the middle-aged female participants said:

> *We are ignorant people, we don't understand all these. Besides we don't need this [digital health], it's enough if you can receive and make a call using a phone.*

The reasons mentioned in the FGDs for not using electronic means to seek health information and/or services can be summarized into 4 major reasons.

Awareness of Digital Health Services

Many participants were not aware of eHealth services. This was usual in Bangladesh, given the ongoing promotion of the telecom industry in Bangladesh. A few younger respondents could mention GrameenPhone's health helpline (789), but most of them had a general lack of understanding of eHealth services. Respondents knew that 789 is a service to call physicians using GrameenPhone mobile phones but did not know how it worked. Other than lack of skill to navigate the device and the platform, lack of awareness regarding associated costs and how to talk about personal health ailments was a major concern for the participants.

Most of the participants were confused by the promotional activities undertaken by telecom companies. Mirzapur has an abundance of kiosks and shops that offer a range of mobile phone–related services and products with colorful banners and

posters displaying information about these services. Considering the overflow of information on display, FGD participants described it as "difficult to distil information related to digital health services." One of the male participants said:

> *The local shops are full of pictures and words about hundreds of offers. Among those, I don't remember any explaining the availability of such type of health service [digital health]. If we can't find one, how can we be aware that such services exist?*

The young and educated participants were more aware of digital health services compared with others. However, even among this group, most did not know much about these services. Some of them were aware of social media–based health initiatives and promotions. Almost everyone had a Facebook account and had seen adverts and information related to health. Although social media (including Facebook) mostly presented information on diet, healthy lifestyles, and beauty tips, some serious issues such as cancer, HIV, and AIDS have also been presented. As one of the female students explained:

> *We don't know what it is [eHealth] and how can we get health information through it. Or how does it work, how much money it takes etc. Most of us use Facebook or at least have seen it. I sometimes get posts related to beauty or diet-related tips. Sometimes I get information on cancer. But I don't know what eHealth is. I think if digital health services can be made as easy as Facebook, then everyone will come to know about it.*

Personal Comfort and Acceptance

During FGDs, respondents (mostly women) expressed specific concerns regarding not knowing the counselor/provider personally and were hesitant to talk about personal issues and illnesses. One female respondent said:

> *We are rural ignorant people. How can we talk about illnesses to someone, whom we don't know or see? We are shy and just can't do it.*

Middle-aged respondents expressed their lack of trust in the accuracy and quality of digital health information and preferred face-to-face interactions with the person providing the information and advice. Almost everyone preferred to discuss health care needs with their friends and family first, then with local drug sellers and village doctors (quacks) before taking them to formal medical providers. If someone advised them, only then did they consider seeking health information or service electronically. Young and educated respondents were more inclined to use eHealth as they perceived it as ensuring one's confidentiality and privacy. However, they had concerns about cost. A young male participant said:

> *It takes up money from mobile account balance. Both internet and talk time are expensive. But it's true that you can say many things over a phone which is rather difficult in a face to face consultation with a person whom you know.*

Literacy and Skills

Some participants mentioned that they lacked the skills needed to access digital health effectively. This ranged from proficiency in and with English (and Roman letters) and technical ability to navigate the device and its software (ie, specific app, browser, and internet settings). For example, calling a call center entails the ability to press specific numbers in response to questions or directions. Alternatively, browsing the internet requires English literacy and technical skills to set their devices for internet use. One of the middle-aged participants said:

> *It's easier for young people. They know how to do this using their mobile phones or computers. They also have the skills to do that. That's why I don't have internet in my phone.*

Some young participants also felt that lack of proficiency in English is a barrier to accessing information and services electronically. One of them mentioned:

> *We are Bengali people and Bangla is our language. We are not very good at English which, in my opinion, is our main problem*

Some of the young participants mentioned that family members had asked them (or their friends) to look for health information electronically, but they never looked for it themselves. One of the elderly respondents said:

> *We are old and that's why we don't know much about it. We can only receive and make calls. Sometimes if someone sends an SMS, we take the phone to the other members of the house to find out what it is. We do the same when we want to know something about the phone.*

Proximity to Health Facilities

One of the reasons why respondents did not engage with digital health in Mirzapur was the availability of and access to conventional, formal health services within their vicinity. Mirzapur has a public hospital, the Upazila (subdistrict) Health Complex, and a philanthropic hospital (Kumudini). For any medical emergency, anyone can visit these hospitals instead of using a call center or other digital health services. During discussions, participants agreed that it could be one of the reasons for their reluctance to use digital health services, including information. One of the middle-aged participants mentioned:

> *Kumudini hospital was established long time ago and is just beside our house. It is much easier and more comfortable for us to visit this hospital when needed. Besides we also have the upazila hospital.*

Discussion

Principal Findings

This study aims to examine the current linear understanding that access to technology results in access to health care by all. Considering digital health services as a proxy for the integration and implementation of technology for providing health care and information, it has used sociodemographic characteristics to

explore which population groups have access to technology and have accessed digital health services. In more elaborate terms, it focuses on the degree to which ICT and technology has enabled Bangladeshi people to increase their access to information and improve their well-being by exploring the factors that influence whether people have access to electronic devices (namely, mobile phones or personal computers/laptops) and use these devices to access health services and/or information digitally. The findings show that although there is high household ownership (771/854, 90.3%) of mobile phones, personal ownership is much lower (471/854, 55.2%). Everyone who owned a personal mobile phone used it to seek information and services electronically, but only a small proportion (34/471, 7.2%) used it to seek health-related information or services. Although the data suggested younger men and those with a higher education and SES chose to use their devices to access digital health, there was little statistical evidence that sociodemographic factors influenced the use of digital health information and services. According to the findings from the FGDs, nonuse of devices to seek health-related information or services was related to the perception of digital health as an unfamiliar health care–seeking model. Other factors were lack of technological skills and related literacy to seek electronic information or services, associated cost of accessing information, lack of awareness about digital health services, and proximity to functioning health centers.

In the context of rural Bangladesh, previous works have reported slightly lower household ownership of mobile phones, but with an upward trend [28,31,48]. The data reported in this paper were collected later, and in a semiurban context (Mirzapur), which is adjacent to Dhaka. Mirzapur is thus likely to have greater access to technology and resources than rural Bangladesh. The high ownership of electronic devices found in this paper is consistent with what others have reported. However, if ownership of devices is used as a proxy for access to digital technology, the data show that only about half of the respondents have personal devices. As there is a dearth of evidence regarding the personal ownership of devices in Bangladesh, it was not possible to compare the findings with the situation in other parts of the country. Nevertheless, the general idea that high household ownership and subscription to mobile-cellular networks means high access to technology needs to be reconsidered and explored further.

Socio-Demography of the Use of Digital Health

Although this paper has shown that only a small proportion (7.2%) of owners have used their phones to access digital health for health services or information, in the rural context, use of devices to access digital health has been reported to be even lower (2%) [28]. The difference in the spread of technology over time and context and related access (semiurban versus rural) can be the reason for this difference. However, such low use of devices generally to access digital health does not indicate that everyone is unable to access services or information digitally. In line with these findings, globally (including Bangladesh), male, young, educated, and wealthier groups are more likely to use their electronic devices to seek general information and health information or services [49-52]. A recently published paper, based on the findings from Mirzapur, reported that the use of mobile phones to access health information at least once in the last 12 months was 45% in college students (young and educated adults) compared with 18% in the general population. It also reported that internet users were predominantly (two-third) male phone owners [26]. Therefore, this paper strongly argues that any attempt to integrate technology in the health systems of Bangladesh (and similar contexts) and to endorse related digital health innovations must take into account sociodemographic attributes and the fact that services are more likely to be accessed by young, educated, and male populations. Although this represents a potential disparity in access to digital health, it also positions young and educated people to help the diffusion of technology in the community as change agents and therefore paves the way toward the much discussed and desired integration of technology into the health system to meet the challenges of UHC in Bangladesh and related contexts [42].

Other Factors Influencing the Use of Digital Health

The other reasons for the low uptake of digital health include lack of awareness about digital health in terms of its modus operandi and availability, lack of personal comfort and acceptance of this form of health service or information, lack of literacy and skill for using digital health technology, and proximity to other health facilities that provide emergency care. However, these reasons are not mutually exclusive, and the relationship within and between them must be examined further in terms of underlying equity challenges. Awareness regarding digital health initiatives is possibly the first and foremost of these reasons, yet communicating the potential of digital health is not, as discussed above, straightforward. Many middle-aged residents have access to household resources and the relevant educational achievements (Table 1) that would make accessing digital health possible, but they are not sufficiently informed and do not have the technological skills (Table 2). They are simultaneously disinclined to pursue health services or information provided in this manner because they are unfamiliar with and do not trust the mode of delivery. Moreover, should they or their family members have a health need, particularly an urgent or emergency one, they would be able to access Mirzapur's other health facilities. Young people, by contrast, are aware of the potential of digital health and have the relevant skills and literary sophistication required (Table 1). As indicated by others, they do not have the material resources and influence that would support and facilitate access to digital health (ie, financial resources and decision-making capacity in health care need) [26]. As a result, they tend to use this service when, as shown above, older people who have the necessary resources request that young people engage with digital health services or information. Young people, like the older generations, have access to Mirzapur health facilities when there is an emergency; however, their primary health concerns are, as others have reported and the earlier discussion indicates, private, nonurgent, and often deemed unnecessary, such that their concerns are dismissed and they are treated with disrespect [26].

The importance of sociodemographic and economic factors and related material factors was highlighted in other technology-based health interventions in Bangladesh. One of the popular mHealth interventions in Bangladesh called

Aponjon, supported by the United States Agency for International Development and implemented in 2012, was designed to provide voice or text message–based pregnancy and newborn care. After studying the impact of the intervention, it was concluded that exposure to Aponjon messages was not associated with improved maternal and child health outcomes, such as the presence of skilled birth attendance at birth, breast-feeding practices, or facility delivery. Rather, higher sociodemographic and income status of the mother was more significantly associated with reported improvements [53]. Another intervention study was piloted to understand the feasibility and acceptability of a mobile phone-based intervention that combined counseling and the payment of cash transfers to improve the perception of both maternal nutrition during pregnancy and nutrition of the infant in rural Bangladesh. Although the study could not conclude the intervention as effective, it highlighted other material issues such as charging the mobile phone or spending cash (obtained from the project) to feed one's family rather than focusing on the nutrition of the pregnant mother or infant as major barriers to the effectiveness of such mHealth interventions [54].

Conclusions

There is high awareness and use of electronic devices to access electronic information and/or services across various sociodemographic categories in Bangladesh. Global trends suggest that access to the use of mobile phones and computers is expanding, including contexts like Bangladesh [15,31]. However, Mirzapur's respondents continued to avoid digital health initiatives. Why? The only paper that has tried to explain the reason for such low use in Bangladesh concluded that, although the community has some technological readiness and will to use mHealth, lack of adequate human resources and technological abilities of the people may have restricted the use of electronic devices for health services or information [55].

Inequitable access to health care is one of the major health system concerns for LMICs, including Bangladesh. Expansion of coverage of health care provision and implementing efficient policy making and governance activities is critical for improving population access to health care. At the same time, adopting an equity focus can ensure the inclusion of various population groups. The central question in this paper is, therefore, does high access to technology mean that various sociodemographic groups can access digital health services, thereby having higher access to health care and information? The answer lies in how technology interfaces with other social determinants of health to produce equity and inequity. Shankardass et al [56] argued that health inequities are caused by complex social, economic, and political factors (ie, the influence of gatekeepers, affordability of services, provision of quality health care, and strategies to secure access of poor and vulnerable groups to health). These factors limit recognition of the need for and creation of proequity policies. Digital health promises to address access to health services and information [57], and, in demonstrating substantial growth in technological access, it gives the impression that challenges in access to health services and information are being addressed, leading to a decrease in the digital divide. However, as shown in this paper, this approach of ubiquitous access to technology obscures how inequity in access to health plays out. As Embrett and Randall [58] argue, addressing health inequity is dependent on generating public awareness to develop sufficient political incentives for change. However, the lack of access to health services and health information that, for instance, young people experience is not socially acknowledged. Digital health information and services offer some potential to address this challenge, with young people having appropriate awareness, sufficient skills, education, and literacy that would make this an attractive option, but they lack the English sophistication required to articulate health needs and use, and they do not have the necessary resources to turn this into a reality. Addressing inequity in digital health requires action to increase not only device ownership and the technical skills necessary to operate these devices but also the material resources to encourage their use and social recognition of health needs and inequity.

This paper presents evidence on the implications of the adoption of digital health services on equitable access to health. To capitalize on the growth of technology, it is important to recognize that without appropriate recognition, digital health services may result in underused services and can influence further disparities such as information-rich poor groups. This means that without the required approach, whereas some social groups will have more access to health care because of their better health, technology, and general literacy, others may become vulnerable and marginalized with restricted access. This should serve as a basis for any future attempt in devising (and adopting) operational frameworks that envision both accountability and equity for effective integration of ICT platforms (digital health) to address disparities and related health systems challenges for Bangladesh and other LMICs.

Limitations of the Study

This survey was undertaken with the expectation of high levels of digital health usage. As a result, insufficient attention was paid to the use of mobile phones for health information through casual conversations among peers and kin.

Acknowledgments

This research has been funded by the ESRC/DFID Poverty Alleviation Scheme, grant number ES/J018651/1. International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) acknowledges with gratitude the commitment of DFID and the ESRC (UK) to its research efforts. icddr,b is also grateful to the Governments of Bangladesh, Canada, Sweden and the UK for providing core/unrestricted support. The authors also gratefully acknowledge the support of Tamanna Sharmin, Nazib uz Zaman Khan, Mohammad Selim and other members from the icddr,b team for various phases of data collection, management and analysis of the data. We would like to thank Drs. Simon Batchelor and Nigel Scott for their contributions to the survey. Finally, we would

like to thank the International Development Research Centre (IDRC), Canada for partially funding the PhD which produced this paper.

Conflicts of Interest

None declared.

References

1. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the center for health market innovations. Bull World Health Organ 2012 May 1;90(5):332-340 [FREE Full text] [doi: 10.2471/BLT.11.099820] [Medline: 22589566]
2. mHealth New Horizons for Health Through Mobile Technologies. World Health Orgnaization. 2011. URL: http://www.who.int/about/ [accessed 2018-04-26]
3. Statistics. ITU: Committed to Connecting the World. 2018 Nov 13. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/default.aspx [accessed 2020-06-08]
4. ICT Facts & Figures. ITU: Committed to Connecting the World. 2019 Jan 13. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf [accessed 2020-06-08]
5. Minges M. Exploring the Relationship Between Broadband and Economic Growth. The World Bank. 2016. URL: http://documents.worldbank.org/curated/en/178701467988875888/pdf/102955-WP-Box394845B-PUBLIC-WDR16-BP-Exploring-the-Relationship-between-Broadband-and-Economic-Growth-Minges.pdf [accessed 2018-04-25]
6. Katz R. Impact of Broadband On the Economy: Research to Date and Policy Issues. ITU: Committed to Connecting the World. 2012. URL: https://www.itu.int/ITU-D/treg/broadband/ITU-BB-Reports_Impact-of-Broadband-on-the-Economy.pdf [accessed 2018-04-25]
7. The Mobile Economy. GSMA. 2015. URL: https://www.gsma.com/mobileeconomy/global/2015/GSMA_Global_Mobile_Economy_Report_2015.pdf [accessed 2018-04-25]
8. Comparing Agent-Led versus Self-Register Subscription Models for mAgri Services: Mobile for Development. GSMA. 2018. URL: https://www.gsma.com/mobilefordevelopment/programme/magri/comparing-agent-led-versus-self-register-subscription-models-magri-services/ [accessed 2018-04-25]
9. Case Study: Airtel Green SIM. GSMA. 2015. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2015/03/GSMA_Case_IKSL_web2.pdf [accessed 2018-04-25]
10. de Wulf L. Tradenet in Ghana Best Practice of the Use of Information Technology. The World Bank. 2004. URL: http://documents.worldbank.org/curated/en/137601468771018721/pdf/313620GH0trade11WDR20050bkgd0paper1.pdf [accessed 2018-04-25]
11. Company Profile. bKash. 2017. URL: https://www.bkash.com/about/company-profile [accessed 2018-04-26]
12. Quadir K. The Story of bKash. The Daily Star. 2015. URL: https://www.thedailystar.net/the-story-of-bkash-21235 [accessed 2018-04-26]
13. About Vodafone Group. Vodafone Group. 2019 Jan 13. URL: http://www.vodafone.com/content/index/about/about-us/money_transfer.html [accessed 2020-06-08]
14. Camner G, Pulver C, Sjöblom E. What Makes a Successful Mobile Money Implementation? Learnings from M-PESA in Kenya and Tanzania. GSMA. 2009. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2012/06/mpesa_case_study9983.pdf [accessed 2018-04-25]
15. Aker JC, Mbiti IM. Mobile phones and economic development in Africa. J Econ Perspect 2010 Aug 1;24(3):207-232. [doi: 10.1257/jep.24.3.207]
16. Khalil M, Dongier P, D'Costa V, Zhen-Wei QC, Smith P, Sudan R. Information and Communications for Development: Extending Reach and Increasing Impact. The World Bank. 2009. URL: http://documents.worldbank.org/curated/en/645821468337815208/pdf/487910PUB0EPI1101Official0Use0Only1.pdf [accessed 2018-04-26]
17. Connecting to Work: How ICTs Are Expanding Job Opportunities Worldwide. The World Bank. 2013. URL: http://www.worldbank.org/en/news/feature/2013/09/10/how-icts-are-expanding-job-opportunities [accessed 2018-04-26]
18. Raftree L. ICT Opportunities and Barriers for Youth in Developing Countries. UNICEF Office of Research - Innocenti. 2018. URL: https://www.unicef-irc.org/article/1009-ict-opportunities-and-barriers-for-youth-in-developing-countries.html [accessed 2018-04-26]
19. Ahmed T, Lucas H, Khan AS, Islam R, Bhuiya A, Iqbal M. eHealth and mhealth initiatives in Bangladesh: a scoping study. BMC Health Serv Res 2014 Jun 16;14:260 [FREE Full text] [doi: 10.1186/1472-6963-14-260] [Medline: 24934164]
20. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
21. Akter S, Ray P. mHealth - an ultimate platform to serve the unserved. Yearb Med Inform 2010;19(1):94-100. [Medline: 20938579]

22. Ryu S. Book review: mhealth: new horizons for health through mobile technologies: based on the findings of the second global survey on ehealth (global observatory for ehealth series, volume 3). Healthc Inform Res 2012;18(3):231 [FREE Full text] [doi: 10.4258/hir.2012.18.3.231]
23. Chetley A, Davies J, Trude B, Mcconnell H, Ramirez R, Shields T, et al, editors. Improving Health, Connecting People : the Role of ICTs in the Health Sector of Developing Countries - A Framework Paper. Washington, DC: World Bank; 2006.
24. Mobile Phone Subscribers in Bangladesh February, 2018. BTRC: Bangladesh Telecommunication Regulatory. 2018. URL: http://www.btrc.gov.bd/content/mobile-phone-subscribers-bangladesh-february-2018 [accessed 2018-04-26]
25. Ahmed T, Bloom G, Iqbal M, Lucas H, Rasheed S, Waldman L. E-health and M-Health in Bangladesh: Opportunities and Challenges Internet. Evidence Report No. 60; Horizon Scanning (Tomorrow Today). Brighton: Institute of Development Studies; 2014.
26. Waldman L, Ahmed T, Scott N, Akter S, Standing H, Rasheed S. 'We have the internet in our hands': Bangladeshi college students' use of ICTs for health information. Global Health 2018 Mar 20;14(1):31 [FREE Full text] [doi: 10.1186/s12992-018-0349-6] [Medline: 29554929]
27. Paina L, Wilkinson A, Tetui M, Ekirapa-Kiracho E, Barman D, Ahmed T, et al. Using theories of change to inform implementation of health systems research and innovation: experiences of future health systems consortium partners in Bangladesh, India and Uganda. Health Res Policy Syst 2017 Dec 28;15(Suppl 2):109 [FREE Full text] [doi: 10.1186/s12961-017-0272-y] [Medline: 29297374]
28. Khatun F, Hanifi S, Iqbal M, Rasheed S, Rahman MS, Ahmed T, et al. Prospects of mhealth services in Bangladesh: recent evidence from Chakaria. PLoS One 2014;9(11):e111413 [FREE Full text] [doi: 10.1371/journal.pone.0111413] [Medline: 25375255]
29. Our eHealth: The Roadmap of Impressive Development. DGHS. 2012. URL: http://www.dghs.gov.bd/licts_file/images/eHealth/our_eHealth_Aug_2012.pdf [accessed 2020-06-23]
30. The Millennium Development Goals: Bangladesh Progress Report 2011. Planipolis - UNESCO. 2012. URL: http://planipolis.iiep.unesco.org/upload/Bangladesh/Bangladesh_MDG_2011.pdf [accessed 2020-06-23]
31. Country Overview: Bangladesh. GSMA Intelligence. 2014. URL: https://www.gsmaintelligence.com/research/?file=140820-bangladesh.pdf&download [accessed 2018-04-26]
32. National ICT Policy - 2009. Ministry of Science and Information & Communication Technology, Government of the People's Republic of Bangladesh. 2009. URL: http://www.bcs.org.bd/img/upload/page/11.pdf [accessed 2014-12-01]
33. Afroz T. Delivery of Mobile Phone Aided Health Services in Rural Bangladesh: A Study on Two Upazilas Internet. Master in Public Policy and Governance (MPPG). 2012. URL: http://mppg-nsu.org/attachments/396_13 [accessed 2020-06-23]
34. Ansari N. Stories from rural Bangladesh: presenting impact of the mobile phone based 'health help line' service in rural Bangladesh. In: Svensson J, Wicander G, editors. The 2nd International Conference on M4D Mobile Communication Technology for Development. Kampala, Uganda: Karlstad University Studies; 2010.
35. Bangladesh eHealth Inventory Report. -: Bangladesh Knowledge Management Initiative (BKMI); 2014.
36. Uddin J, Biswas T, Adhikary G, Ali W, Alam N, Palit R, et al. Impact of mobile phone-based technology to improve health, population and nutrition services in rural Bangladesh: a study protocol. BMC Med Inform Decis Mak 2017 Jul 6;17(1):101 [FREE Full text] [doi: 10.1186/s12911-017-0502-9] [Medline: 28683742]
37. Peters DH, Garg A, Bloom G, Walker DG, Brieger WR, Rahman MH. Poverty and access to health care in developing countries. Ann N Y Acad Sci 2008;1136:161-171. [doi: 10.1196/annals.1425.011] [Medline: 17954679]
38. Ahmed S, Hossain M, Rajachowdhury AM, Bhuiya A. The health workforce crisis in Bangladesh: shortage, inappropriate skill-mix and inequitable distribution. Hum Resour Health 2011 Jan 22;9:3 [FREE Full text] [doi: 10.1186/1478-4491-9-3] [Medline: 21255446]
39. Chowdhury AM, Bhuiya A, Chowdhury ME, Rasheed S, Hussain Z, Chen LC. The Bangladesh paradox: exceptional health achievement despite economic poverty. Lancet 2013 Nov 23;382(9906):1734-1745. [doi: 10.1016/S0140-6736(13)62148-0] [Medline: 24268002]
40. Understanding the Digital Divide. Paris, France: Organisation for Economic Co-Operation and Development (OECD); 2001.
41. Wilhelm A. myth. Inf Soc 2002 Oct;18(5):415-416. [doi: 10.1080/01972240290108212]
42. Adams AM, Ahmed T, El Arifeen S, Evans TG, Huda T, Reichenbach L, Lancet Bangladesh Team. Innovation for universal health coverage in Bangladesh: a call to action. Lancet 2013 Dec 21;382(9910):2104-2111. [doi: 10.1016/S0140-6736(13)62150-9] [Medline: 24268605]
43. Mirzapur Upazila. Banglapedia. 2015. URL: http://en.banglapedia.org/index.php?title=Mirzapur_Upazila [accessed 2018-04-30]
44. Das S, Nasrin D, Ahmed S, Wu Y, Ferdous F, Farzana F, et al. Health care-seeking behavior for childhood diarrhea in Mirzapur, rural Bangladesh. Am J Trop Med Hyg 2013 Jul;89(1 Suppl):62-68 [FREE Full text] [doi: 10.4269/ajtmh.13-0107] [Medline: 23629937]
45. Krippendorff K. Content Analysis: an Introduction to Its Methodology. Thousand Oaks, CA: Sage Publications; 2004.
46. Mayring P. Qualitative Content Analysis. Forum Qualitative Sozialforschung. 2000. URL: http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385#gcit [accessed 2018-04-30]

47. Bangladesh Demographic and Health Survey 2014. The DHS Program. 2014. URL: https://dhsprogram.com/pubs/pdf/FR311/FR311.pdf [accessed 2020-06-23]
48. Tran MC, Labrique AB, Mehra S, Ali H, Shaikh S, Mitra M, et al. Analyzing the mobile 'digital divide': changing determinants of household phone ownership over time in rural Bangladesh. JMIR Mhealth Uhealth 2015 Feb 25;3(1):e24 [FREE Full text] [doi: 10.2196/mhealth.3663] [Medline: 25720457]
49. Shah R, Jaisinghani P. Toward a More Inclusive Digital Economy. MIT Press Journals. 2014. URL: https://www.mitpressjournals.org/userimages/ContentEditor/1421439266694/INNOVATIONS-DIGITAL-INCLUSION-FOREWORD.pdf [accessed 2018-05-08]
50. Ac'lar A. Exploring the Aspects of Digital Divide in a Developing Country. Issues in Informing Science and Information Technology. 2011. URL: http://iisit.org/Vol8/IISITv8p231-244Acilar248.pdf [accessed 2018-05-08]
51. Bridging the Gender Gap: Mobile Access and Usage in Lowand Middle-income Countries. GSMA. 2015. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2016/02/Connected-Women-Gender-Gap.pdf [accessed 2020-06-23]
52. Khan NU, Rasheed S, Sharmin T, Ahmed T, Mahmood SS, Khatun F, et al. Experience of using mhealth to link village doctors with physicians: lessons from Chakaria, Bangladesh. BMC Med Inform Decis Mak 2015 Aug 5;15:62 [FREE Full text] [doi: 10.1186/s12911-015-0188-9] [Medline: 26242574]
53. Alam M, D'Este C, Banwell C, Lokuge K. The impact of mobile phone based messages on maternal and child healthcare behaviour: a retrospective cross-sectional survey in Bangladesh. BMC Health Serv Res 2017 Jun 24;17(1):434 [FREE Full text] [doi: 10.1186/s12913-017-2361-6] [Medline: 28645278]
54. Huda TM, Alam A, Tahsina T, Hasan MM, Khan J, Rahman MM, et al. Mobile-based nutrition counseling and unconditional cash transfers for improving maternal and child nutrition in Bangladesh: pilot study. JMIR Mhealth Uhealth 2018 Jul 18;6(7):e156 [FREE Full text] [doi: 10.2196/mhealth.8832] [Medline: 30021707]
55. Khatun F, Heywood AE, Ray PK, Hanifi S, Bhuiya A, Liaw S. Determinants of readiness to adopt mhealth in a rural community of Bangladesh. Int J Med Inform 2015 Oct;84(10):847-856. [doi: 10.1016/j.ijmedinf.2015.06.008] [Medline: 26194141]
56. Shankardass K, Lofters A, Kirst M, Quiñonez C. Public awareness of income-related health inequalities in Ontario, Canada. Int J Equity Health 2012 May 21;11:26 [FREE Full text] [doi: 10.1186/1475-9276-11-26] [Medline: 22613058]
57. Phua KH, Sheikh K, Tang S, Lin V. Editorial - health systems of Asia: equity, governance and social impact. Soc Sci Med 2015 Nov;145:141-144. [doi: 10.1016/j.socscimed.2015.10.027] [Medline: 26522185]
58. Embrett MG, Randall G. Social determinants of health and health equity policy research: exploring the use, misuse, and nonuse of policy analysis theory. Soc Sci Med 2014 May;108:147-155. [doi: 10.1016/j.socscimed.2014.03.004] [Medline: 24641879]

Abbreviations

AI: artificial intelligence
FGD: focus group discussion
HDSS: Health and Demographic Surveillance System
ICT: information and communications technology
LDC: least developed country
LMIC: low- and middle-income country
SES: socioeconomic status
UHC: universal health coverage

Edited by G Eysenbach; submitted 02.10.19; peer-reviewed by H Lucas, MAB Siddique, B Ghose, T Tahsina; comments to author 26.11.19; revised version received 29.02.20; accepted 29.02.20; published 21.07.20.

Please cite as:
Ahmed T, Rizvi SJR, Rasheed S, Iqbal M, Bhuiya A, Standing H, Bloom G, Waldman L
Digital Health and Inequalities in Access to Health Services in Bangladesh: Mixed Methods Study
JMIR Mhealth Uhealth 2020;8(7):e16473
URL: http://mhealth.jmir.org/2020/7/e16473/
doi:10.2196/16473
PMID:32706736

Original Paper

Maternal Parenting Electronic Diary in the Context of a Home Visit Intervention for Adolescent Mothers in an Urban Deprived Area of São Paulo, Brazil: Randomized Controlled Trial

Daniel Fatori[1*], PhD; Adriana Argeu[1*]; Helena Brentani[1*], MD, PhD; Anna Chiesa[2*], PhD; Lislaine Fracolli[2*], PhD; Alicia Matijasevich[3*], MD, PhD; Euripedes C Miguel[1*], MD, PhD; Guilherme Polanczyk[1*], MD, PhD

[1]Department of Psychiatry, University of Sao Paulo Medical School, Sao Paulo, Brazil

[2]Collective Health Nursing Department, School of Nursing, University of Sao Paulo, Sao Paulo, Brazil

[3]Department of Preventive Medicine, University of Sao Paulo Medical School, Sao Paulo, Brazil

[*]all authors contributed equally

Corresponding Author:
Daniel Fatori, PhD
Department of Psychiatry
University of Sao Paulo Medical School
785 R Dr Ovídio Pires de Campos
Sao Paulo
Brazil
Phone: 55 11961626183
Email: daniel.fatori@gmail.com

Abstract

Background: Pregnancy during adolescence is prevalent in low- and middle-income countries (LMICs), which is associated with various adverse outcomes that can be prevented with home visiting programs. However, testing these interventions in LMICs can be challenging due to limited resources. The use of electronic data collection via smartphones can be an alternative and ideal low-cost method to measure outcomes in an environment with adverse conditions.

Objective: Our study had two objectives: to test the efficacy of a nurse home visiting intervention on maternal parenting and well-being measured by an electronic daily diary (eDiary), and to investigate the compliance rate of the eDiary measurement method.

Methods: We conducted a randomized controlled trial to test the efficacy of Primeiros Laços, a nurse home visiting program, for adolescent mothers living in an urban deprived area of São Paulo, Brazil. A total of 169 pregnant adolescents were assessed for eligibility criteria, 80 of whom were included and randomized to the intervention (n=40) and control group (care as usual, n=40). Primeiros Laços is a home visiting intervention delivered by trained nurses tailored to first-time pregnant adolescents and their children, starting during the first 16 weeks of pregnancy until the child reaches 24 months of age. Participants were assessed by blind interviewers at 8-16 weeks of pregnancy (baseline), 30 weeks of pregnancy, and when the child was 3, 6, and 12 months of age. At 18 months, participants were assessed regarding maternal parenting and parental well-being using a 7-consecutive-day eDiary. The smartphone app was programmed to notify participants every day at 9:00 PM over a period of seven days.

Results: We were able to contact 57/80 (71%) participants (29 from the intervention group and 28 from the control group) when the child was 18 months of age. Forty-eight of the 57 participants (84%) completed at least one day of the eDiary protocol. The daily compliance rate ranged from 49% to 70%. Our analyses showed a significant effect of the intervention on parental well-being (B=0.32, 95% CI [0.06, 0.58], *P*=.02) and the maternal parenting behavior of the mother telling a story or singing to the child (odds ratio=2.33, 95% CI [1.20, 4.50], *P*=.01).Our analyses showed a significant effect of the intervention on parental well-being (B=0.32, *P*=.02) and the maternal parenting behavior of the mother telling a story or singing to the child (odds ratio=2.33, *P*=.01).

Conclusions: The Primeiros Laços intervention improved maternal parenting and parental well-being, demonstrating its promise for low-income adolescent mothers. The compliance rate of the eDiary assessment showed that it was generally accepted by adolescent mothers with limited resources. Future studies can implement ambulatory assessment in LMICs via smartphones to measure mother and child behaviors.

Trial Registration: ClinicalTrials.gov NCT02807818; https://clinicaltrials.gov/ct2/show/NCT02807818

(JMIR Mhealth Uhealth 2020;8(7):e13686) doi:10.2196/13686

KEYWORDS

mHealth; early childhood development; maternal care; randomized clinical trial; daily diary; ambulatory assessment

Introduction

Every year, 16 million 15 to 19-year-old adolescents give birth, predominantly in low and middle-income countries (LMICs) [1]. Pregnancy during adolescence is associated with higher risks of eclampsia, endometritis, infections, low birth weight [2], increased maternal and neonatal mortality [3], and morbidity such as maternal depression, parental stress, and insecure attachment [4,5]. Children of adolescent mothers are also at increased risk for impaired early childhood development and behavioral functioning, as well as insecure and disorganized attachment. Interaction between adolescent mothers and their children is reported to be different from that of adult mothers. In particular, adolescent mothers have been shown to provide less verbal stimulation and are less sensitive to the child's needs [6]. Hence, developing and testing interventions focused on preventing adverse outcomes among this high-risk group is a timely need, especially in LMICs where an estimated 250 million children below the age of 5 do not meet their potential [7]. The basis of robust intervention studies is the adequate and accurate measurement of outcomes. However, testing early childhood development (ECD) interventions in LMICs can be challenging due to limited resources.

Studies focusing on maternal parenting usually rely on direct or recorded observations of dyads by trained and certified experts [8] or self-reported questionnaires based on retrospective behavior. However, these methods have important limitations. Observation of dyads is a lengthy and costly assessment, which is difficult to implement in contexts where resources (eg, trained experts, available space) are scarce and accessibility is limited, such as in LMICs. Retrospective self-reported questionnaires are subject to memory bias and lack ecological validity, resulting in potentially inaccurate data (information bias). The use of ambulatory assessment delivered via smartphones can be an alternative data collection method to circumvent these problems.

Ambulatory assessment encompasses methods designed to study people in their natural environment. Owing to the ubiquitous presence of smartphones, ambulatory assessment methods are now most commonly delivered via mobile technology. Subjects can be notified to answer questions during specific times of the day throughout days, weeks, or months. Given the numerous advantages of ambulatory assessment, its use has recently grown in many fields [9-11]. Nevertheless, the field of ECD has been slow in adopting ambulatory assessment, with few initiatives conducted in developed countries. One study used the ambulatory assessment method to assess the relationship between crying and fussing at 12 months of age along with the physical health of the child and emotional security of the mother [12]. Another study validated an ambulatory assessment protocol to evaluate parental discipline in children aged 18-36 months [13]. However, to our knowledge, no study conducted in an LMIC has tested the potential of ambulatory assessment in the context of ECD programs. Moreover, ambulatory assessment methods have only been adopted in observational studies to date, and there has been no randomized trial in the ECD field that used ambulatory assessment via mobile technology as an outcome. Most studies testing mobile technology in the field of ECD are interventions related to various outcomes such as maternal depression [14], infant feeding [15-17], physical activity [18], and maternal health [19-22].

Therefore, we implemented an electronic daily diary (eDiary), a specific type of ambulatory assessment through which participants report a set of behaviors that occurred during the day to measure maternal parenting and parental well-being, in the context of a randomized controlled trial. In particular, we assessed the efficacy of Primeiros Laços, a nurse home visiting program for adolescent mothers living in an urban deprived area of São Paulo, Brazil. The intervention aimed to foster the mother-child relationship and improve child development. The present study had two objectives: (1) to test the efficacy of a nurse home visiting intervention on child maternal parenting and parental well-being measured by an eDiary, and (2) to investigate the compliance rate of the eDiary measurement method. The first hypothesis was that mothers who receive the home visiting intervention will present a higher frequency of maternal parenting behaviors and higher scores of well-being. The second hypothesis was that the eDiary method will present a compliance rate similar to that reported in previous studies, despite the fact that our sample is composed of low-income adolescent mothers living in adverse conditions.

Methods

Study Design and Participants

The present study is part of a parallel group randomized controlled trial that was originally designed to test the efficacy of Primeiros Laços, a nurse home visiting program for adolescent mothers living in an urban deprived area of São Paulo, Brazil. Previous findings from this trial can be found elsewhere [23-26]. From June to September 2015, a total of 169 pregnant youth were assessed for eligibility criteria, 80 of whom were included and randomized to the intervention (n=40) and control (care as usual, n=40) groups. Inclusion criteria were: (a) aged 14 to 19 years old, (b) first pregnancy, (c) between 8-16 weeks of gestation, (d) low socioeconomic status (classes C, D, E according to the widely used Brazilian ABEP scale [27]), and (e) living in the western region of the city of São Paulo. Participants were recruited in the primary health care system. To avoid unbalanced groups, randomization was stratified according to the primary health care unit type and grandmother years of schooling. The allocation ratio was 1:1.

Intervention

Primeiros Laços is a home visiting intervention delivered by trained nurses that is tailored to first-time adolescent pregnant women and their children, which starts during the first 16 weeks of pregnancy until the child reaches 24 months of age. The

frequency of visits was (a) biweekly during gestation and from 2 to 20 months of the child's age; (b) weekly during the first and last month of pregnancy, and during puerperium; and (c) monthly from 21 to 24 months of the child's age. The Primeiros Laços program is based on three theoretical frameworks: attachment theory [28], self-efficacy theory [29], and the bioecological model [30]. This intervention adopts a learning-based approach and aims to have the child at the center of the mother's life, fostering her capacity to perceive and react to the child's needs, with the objective of improving maternal sensitivity to the child's behaviors and emotions. Primeiros Laços was developed by our team based on the Brazilian program Janelas de Oportunidades [31,32], Minding the Baby program [33], and Nurse-Family Partnership [34]. The development of the program was also informed with input from key national stakeholders involved in early childhood and maternal health research and advocacy.

The content is directed to five domains. The first domain is health and social care, in which the nurse provides education regarding maternal and child health such as nutrition, hygiene, common pathologies in childhood, domestic care, vaccination, prevention of accidents, and child development. The second domain is environmental health, in which the nurse provides support and help to identify resources to guarantee adequate living conditions, safe housing, daycare and school, and access to health services. The third domain is life course, which involves life course planning to help participants achieve life goals such as finishing high school, finding a part-time job, starting college, and postponing the birth of a second child. The needs and goals of the participants were discussed individually to respect their personal goals and wishes. Nurses also help mothers gain access to primary care services and government-sponsored social programs. The fourth domain includes parenting skills, which involves education on parenting skills and child behavior considering each development stage, aiming to develop a sensitive and responsive pattern of care. The last domain is family and social support, as nurses frequently highlighted the role of family members and friends to support parental needs. Nurses were supervised weekly by the developers of the program (senior nurses and a child psychologist) to discuss cases, plan visits, and to guarantee fidelity and quality of visits. More details about the intervention can be found elsewhere [35,36].

Care as Usual

Participants allocated to the control group received care from Unified Health System (Sistema Único de Saúde), Brazil's public health system [37], according to national guidelines of the Ministry of Health [38-40] in line with World Health Organization guidelines. Prenatal and postnatal care is delivered by health units of the primary care system free of charge, focusing on preventive interventions, early detection of gestational risk, and referral to specialized health services in cases of high-risk pregnancies. Participants from the intervention group also had access to public health care provided by the Unified Health System.

Assessment

Participants were assessed by interviewers blinded to group allocation at 8-16 weeks of pregnancy (baseline), 30 weeks of pregnancy, and when the child was 3, 6, 12, 18, and 24 months of age. Interviews were conducted by trained psychologists who underwent a 1-month training program provided by senior psychiatrists, psychologists, and pediatricians. Before study interviews commenced, interviewers trained by assessing volunteers living in the western region of São Paulo. Ambulatory assessment was administered when the child was 18 months old. At baseline (8-16 weeks of pregnancy), we measured depression symptoms using the Beck Depression Inventory [41,42] and anxiety symptoms were assessed using the Beck Anxiety Inventory [43,44], which have both been validated in Brazil. Family food insecurity was measured using the abridged version of the Brazilian Food Insecurity Scale (Escala de Insegurança Alimentar) [45], a widely used scale adopted in national epidemiological studies [46].

eDiary

We used the LifeData system [47], a mobile technology suite for smartphone secure data collection available for Android and iOS. Our eDiary protocol comprised a set of questions designed via a web-based dashboard. The smartphone app was programmed to notify participants every day at 9:00 PM over a period of 7 days. Participants had up to 120 minutes to complete the eDiary protocol and received up to three reminders. Trained psychologists visited participants to provide support on downloading, installing, and using the app. The eDiary protocol was shown and explained to all participants using a demonstration version installed on the psychologist's smartphone. Participants who did not own a smartphone were provided one for the duration of the eDiary protocol (n=18).

The present study focused on the following questions of the eDiary protocol: (1) How do you feel about the day today? (score 1-100, general maternal well-being) (2) Did you take care of your child or spend some time with him/her today? (3) What was it like taking care of your child today? (score 1-100, parental well-being) (4) Did you read or show a book to your child today? (5) Did you tell stories or sing to your child today? (6) Did you go out or go for a stroll with your child today? (7) Did you play with your child today? (8) Did you talk to your child today? (9) Did you eat or have a meal with your child today? (10) Did you kiss, hug, tickle (or have any other physical contact with) your child today? (11) How many hours did you spend with your child today? Maternal parenting behavior was operationalized as parenting practices related to mother-child interaction and positive activities known to stimulate child development. Questions 2, and 4 to 11 were related to maternal parenting. Screenshots of the eDiary protocol can be found elsewhere (imgur.com/a/eTMAp84).

Statistical Analysis

The sample size was calculated based on the difference in the electroencephalogram alpha wave frequency between the groups (30%) with a probability of type I error of 5% and statistical power of 80%. Frequencies and distribution of eDiary outcomes are provided for each of the 7 days of the intervention. The

Fisher exact test and *t* tests were used to examine differences between groups (intervention vs control) at each time point. To examine intervention effects on eDiary outcomes, we used generalized estimating equation models [48]. Continuous outcomes were normalized using z-scores. Generalized estimating equation models were used to examine mean differences in normalized continuous outcomes and differences in predictive probabilities for categorical outcomes, both over multiple time points. The quasi-likelihood under the independence model criterion was used to indicate adequate correlation structures for longitudinal data that show a better fit to each model [49]. Fitted models were used to estimate and plot marginal mean scores for continuous outcomes and predictive probabilities for categorical outcomes at each time point. Time was entered as a continuous covariate in all models. Time trends were verified, and all models presented a linear trend [50]. All randomized participants that were contacted when their child was 18 months of age were included in the analysis, except for participants with missing data at all time points (n=9). Statistical tests were all two-sided and *P* values <.05 were considered statistically significant. Parameter estimates are reported with the 95% CI. Analyses were conducted using STATA 15.1 software.

Ethical Considerations

Our study was approved by the Ethics Committee of the University of São Paulo Medical School, University Hospital of the University of São Paulo, and São Paulo Municipal Health Department. All participants and their primary caregiver signed written informed consent forms. The study was registered at clinicaltrials.gov (NCT02807818).

Results

We were able to contact 57/80 (71%) participants (29 in the intervention group and 28 in the control group) when the child was 18 months old. Figure 1 shows the Consolidated Standards of Reporting Trials (CONSORT) flow chart for participant selection and grouping throughout the trial. Twenty participants discontinued the intervention. Forty-eight of the 57 (84%) participants completed at least one day of the eDiary protocol. The daily compliance rate ranged from 49% to 70% (Figure 2). We tested for potential baseline differences between the analysis group (n=48) and the remaining participants (n=32) but did not find statistically significant differences (Table 1). In addition, there was no association of specific days of the week with response status. Participants who received a smartphone for the duration of the study (n=18) also did not show a different response rate from those who already owned a smartphone.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram.

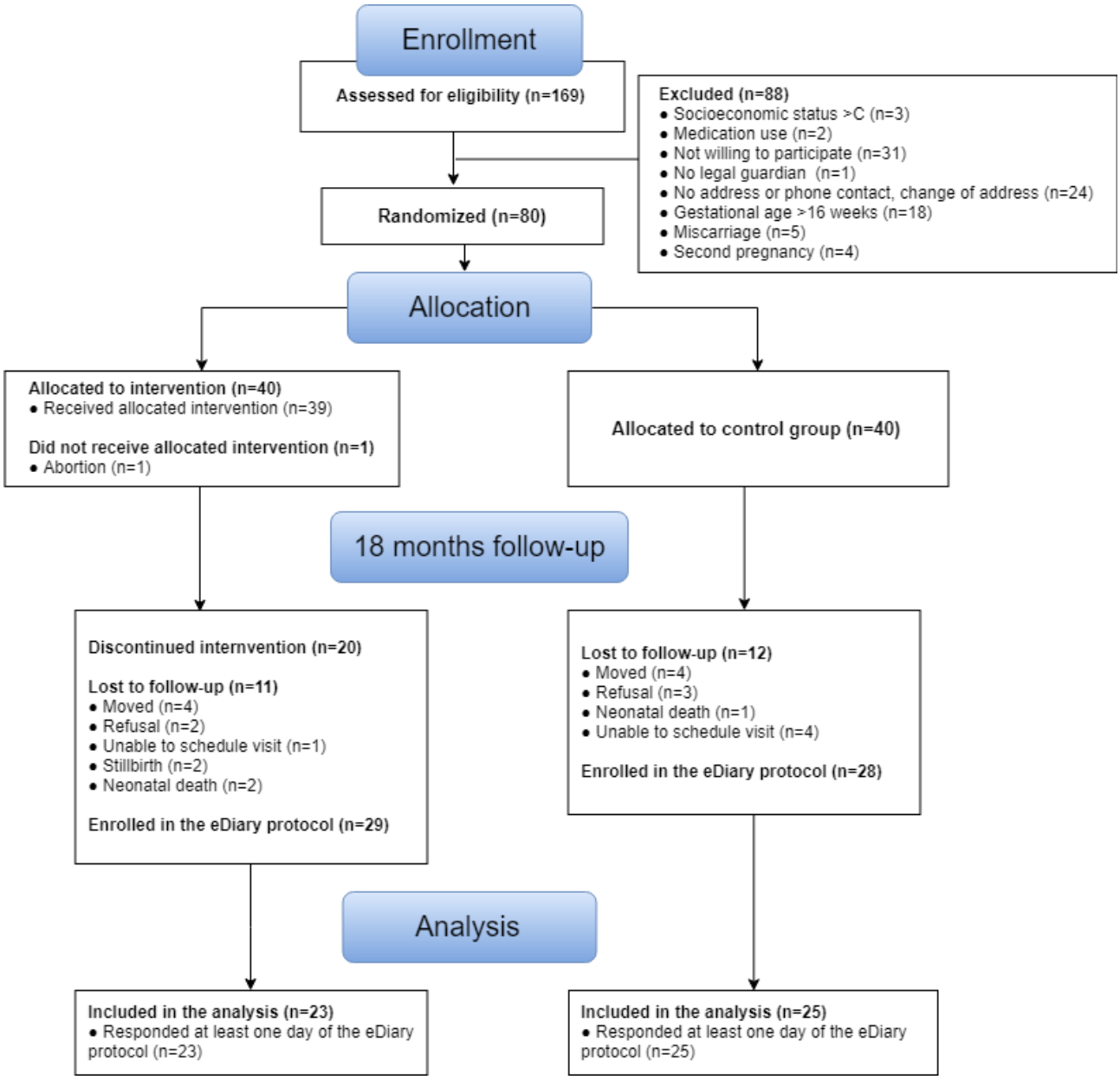


Figure 2. eDiary compliance rate per day (n=57).

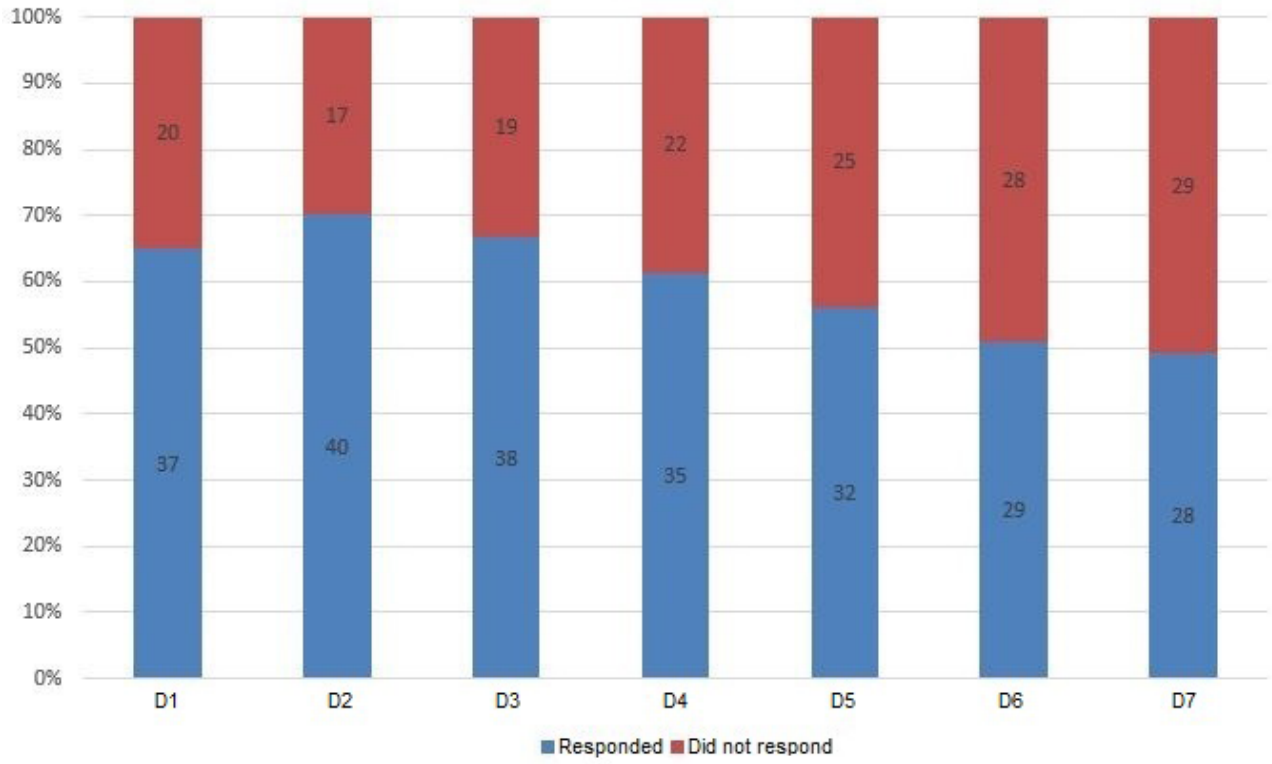


Table 1. Baseline sample characteristics (N=80).

Baseline characteristics	Included for analysis (n=48)	Lost to follow up (n=32)	*P* value
Intervention group, n (%)	23 (48)	17 (53)	.82
Maternal age, mean (SD)	17.0 (1.4)	17.2 (0.9)	.54
Family SES[a] (D and E status), n (%)	16 (33)	14 (44)	.36
Maternal educational level (illiterate), n (%)	5 (10)	7 (22)	.21
Grandmother educational level (illiterate), n (%)	24 (10)	19 (44)	.49
Family income (0-300), n (%)	2 (4)	1 (4)	.85
Substance use during gestation, n (%)	17 (35.)	10 (31)	.81
Food insecurity, n (%)	21 (44)	13 (41)	.82
Maternal depression, n (%)	12 (25)	5 (16)	.41
Maternal anxiety, n (%)	13 (27)	6 (19)	.43
Child sex (male), n (%)	25 (52)	15 (56)	.81

[a]SES: socioeconomic status.

Tables 2 and 3 depict the total sample distribution of eDiary data, as well as the distribution for each group. The total sample grand mean for continuous outcomes across the 1-week period was 80.3 for general well-being, 89.1 for parental well-being, and 843.8 for time spent in minutes with the child. Univariate analysis showed the following significant associations: time spent with the child at day 4 (*P*=.02), read to child at day 6 (*P*=.04), and telling a story or singing at day 2 (*P*=.04) and at day 6 (*P*=.047). Figure 3 and Figure 4 depict the results of the generalized estimating equations models with trajectories of outcomes. Overall, the results showed a significant effect of the intervention on parental well-being (B=0.32, 95% CI [0.06, 0.58], *P*=.01) and increased behavior of the mother telling a story or singing to the child (odds ratio=2.33, 95% CI [1.20, 4.50], *P*=.01) (Figure 4).

Table 2. Results of continuous outcomes per day.

Outcome	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
General well-being, mean (SD)							
TS[a]	80.1 (22.7)	84.0 (19.0)	72.1 (28.1)	74.1 (24.6)	86.1 (21.4)	84.7 (17.9)	81.3 (27.6)
C[b]	79.3 (20.9)	83.8 (16.7)	68.9 (27.7)	72.4 (24.8)	87.3 (15.3)	79.7 (17.2)	77.3 (24.5)
I[c]	80.8 (24.8)	84.0 (21.2)	74.9 (28.9)	76.3 (25.2)	84.9 (26.9)	89.1 (17.9)	84.3 (30.2)
Parental well-being, mean (SD)							
TS	90.0 (14.0)	93.9 (10.7)	85.1 (20.5)	83.0 (20.3)	90.7 (12.1)	93.4 (12.6)	87.4 (18.8)
C	88.7 (13.8)	91.0 (13.1)	82.8 (22.1)	81.3 (20.8)	88.7 (13.6)	86.9 (16.1)	84.6 (15.6)
I	91.1 (14.4)	96.6 (7.1)	87.3 (19.1)	85.0 (20.2)	92.9 (10.3)	99.3 (1.5)	89.5 (21.1)
Time spent with child (minutes), mean (SD)							
TS	868.7 (500.3)	819.2 (519.8)	676.2 (572.1)	978.5 (482.4)	921.0 (500.9)	867.4 (533.0)	775.5 (588.5)
C	851.1 (458.4)	817.5 (496.3)	691.0 (554.3)	804.4 (493.5)	955.7 (450.1)	712.4 (489.0)	821.8 (525.2)
I	884.4 (547.2)	820.7 (554.7)	663.0 (602.3)	1189.9 (386.6)	883.9 (563.6)	1010.5 (550.2)	742.5 (647.3)

[a]TS: total sample.

[b]C: control group.

[c]I: intervention group.

Table 3. Frequency of categorical outcomes per day.a

Outcome	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Taking care of the child, n (%)							
TS[b]	36 (97)	37 (95)	37 (97)	32 (97)	31 (100)	25 (89)	24 (92)
C[c]	17 (47)	18 (49)	18 (49)	17 (53)	16 (52)	12 (48)	10 (42)
I[d]	19 (53)	19 (51)	19 (51)	15 (47)	15 (48)	13 (52)	14 (58)
Reading or showing a book to the child, n (%)							
TS	4 (11)	6 (16)	5 (14)	7 (22)	8 (26)	5 (20)	6 (25)
C	1 (25)	2 (33)	2 (40)	3 (43)	3 (38)	0	2 (33)
I	3 (75)	4 (67)	3 (60)	4 (57)	5 (63)	5 (100)	4 (67)
Telling stories or singing to the child, n (%)							
TS	19 (53)	23 (62)	18 (49)	12 (38)	14 (45)	12 (48)	11 (46)
C	7 (37)	8 (35)	8 (44)	6 (50)	6 (43)	3 (25)	4 (36)
I	12 (63)	15 (65)	10 (56)	6 (50)	8 (57)	9 (75)	7 (64)
Going out or for a stroll with the child, n (%)							
TS	29 (81)	18 (49)	23 (62)	20 (63)	23 (74)	12 (48)	12 (50)
C	16 (55)	8 (44)	14 (61)	12 (60)	12 (52)	4 (33)	3 (25)
I	13 (45)	10 (56)	9 (39)	8 (40)	11 (48)	8 (67)	9 (75)
Playing with the child, n (%)							
TS	36 (100)	33 (89)	35 (46)	31 (97)	29 (94)	20 (80)	20 (83)
C	17 (47)	15 (45)	17 (49)	16 (52)	15 (52)	9 (45)	8 (40)
I	19 (53)	18 (53)	18 (51)	15 (48)	14 (48)	11 (55)	12 (60)
Talking to the child, n (%)							
TS	36 (100)	37 (100)	36 (97)	31 (97)	30 (97)	23 (92)	19 (79)
C	17 (47)	18 (49)	17 (47)	16 (52)	15 (50)	10 (44)	7 (37)
I	19 (53)	19 (51)	19 (53)	15 (48)	15 (50)	13 (57)	12 (63)
Eating/ having meals with the child, n (%)							
TS	34 (94)	34 (92)	32 (89)	29 (91)	30 (97)	22 (88)	20 (83)
C	17 (50)	17 (50)	15 (47)	16 (55)	16 (53)	10 (46)	8 (40)
I	17 (50)	17 (50)	17 (53)	13 (45)	14 (47)	12 (55)	12 (60)
Physical contact with the child, n (%)							
TS	36 (100)	37 (100)	36 (100)	32 (100)	31 (100)	24 (96)	23 (96)
C	17 (47)	18 (49)	17 (47)	17 (53)	16 (52)	11 (46)	9 (39)
I	19 (53)	19 (51)	19 (53)	15 (47)	15 (48)	13 (54)	14 (61)

[a]Sample size varied for different categories and days due to nonresponses, ranging from 10 to 18 and from 13 to 21 in the control and intervention group, respectively.

[b]TS: total sample.

[c]C: control group.

[d]I: intervention group.

Figure 3. Effect of intervention on continuous outcomes: mother well-being (B=0.18, 95% CI [-0.15, 0.51],*P*=.286), parental well-being (B=0.32, 95% CI [0.06, 0.58], *P*=.015), total time spent with the child (B=0.05, 95% CI [-0.43, 0.52], *P*=.845).

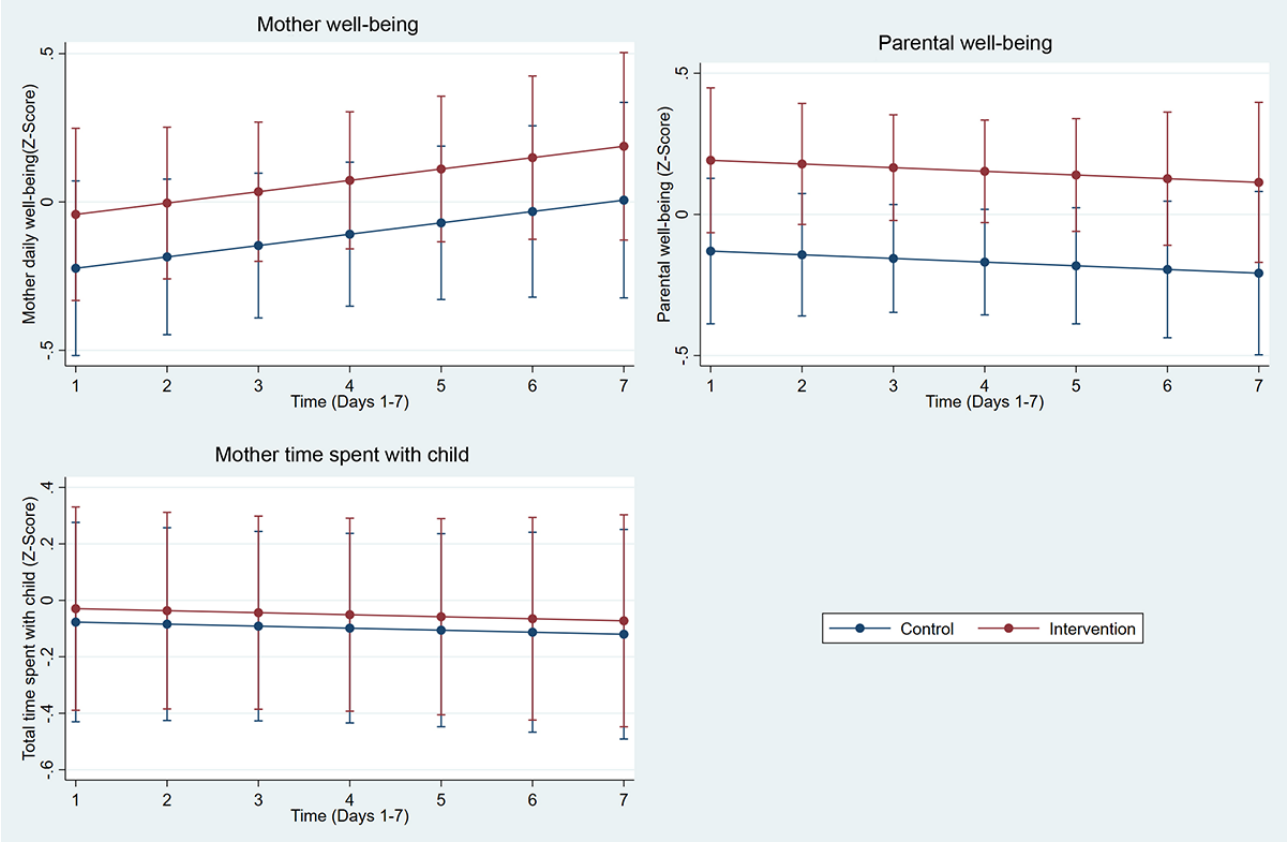
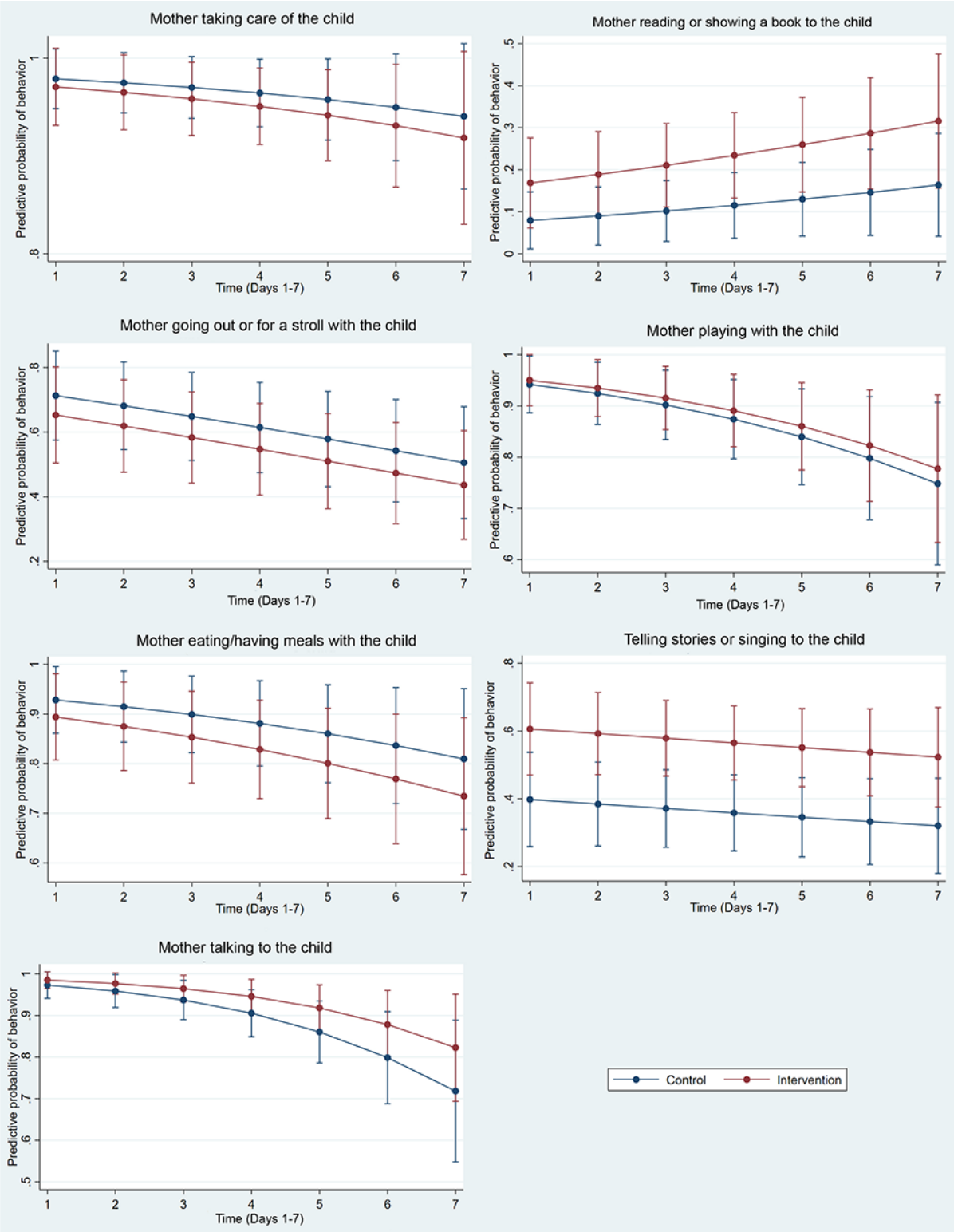


Figure 4. Effect of intervention on categorical outcomes: mother taking care of the child (OR=0.71, 95% CI [0.19, 2.60], *P*=.61), mother reading or showing a book to the child (OR=2.35, 95% CI [0.90, 6.11], *P*=.08), mother telling stories or singing to the child (OR=2.33, 95% CI [1.20, 4.50]. *P*=.01), mother going out or for a stroll with the child (OR=0.76, 95% CI [0.33, 1.72], *P*=.51), mother playing with the child (OR=1.17, 95% CI [0.44, 3.17], *P*=.75), mother talking to the child (OR=1.82, 95% CI [0.69. 4.77], *P*=.22), mother eating/having meals with the child (OR=0.65, 95% CI [0.22, 1.90], *P*=.43).



Discussion

Principal Findings

This is the first ECD randomized controlled trial testing an intervention using a smartphone-based ambulatory assessment to measure outcomes. The results showed that our nurse home visiting program had a positive effect on parental well-being and maternal parenting as measured by a smartphone eDiary. Previous studies have shown that maternal parenting behaviors in adolescent mothers [6] and low-income mothers [51,52] are less frequent. Consequently, it was surprising to notice a ceiling

effect on some maternal parenting behaviors such as taking care of the child, talking to the child, and playing with the child, regardless of being part of the intervention or control group. Early childhood research specifically focused on maternal parenting is surprisingly lacking in Brazil, making it difficult to directly compare findings. Brazilian studies related to early life outcomes conducted in the past decade have mainly focused on maternal health [53] and older children [54]. However, we observed that the frequencies of telling stories or singing to the child and especially reading a book to the child were low overall (38%-62% and 11%-26%, respectively).

It is worth mentioning that Brazil is a country currently facing many educational challenges. The last report from the Programme for International Student Assessment (PISA) showed that reading performance among students has not improved in the last 18 years [55]. In the last PISA, Brazil and other Latin American countries ranked at the bottom of the international ranking on reading skills, below the top 50. Overall, 29% of the Brazilian population is functionally illiterate, presenting only basic reading and writing skills [56], and 44% of the Brazilian population have not fully or partially read a book in the last 3 months [57]. Telling stories and reading books are behaviors deeply connected to past behaviors and previous experiences. In a social context in which people have difficulties reading and understanding books, the ability and practice of telling stories may be impaired. Moreover, low-income families typically have less access to libraries [52] and usually cannot afford to buy books. Even though our intervention emphasized the importance of reading and telling stories, these barriers may have prevented participants from engaging in such activities.

We also found an intervention effect on daily reports of parental well-being. Parenting a child is known to have positive effects on adult parents such as higher levels of life satisfaction, happiness, general feeling of positive emotion, and more meaning in life [58]. However, adolescent mothers are at higher risk for maternal depression and impairment in parenting [6,59], and also experience more parental stress [60]. The effect of Primeiros Laços on parental well-being may be related to the parenting skills component of the intervention. After the child was born, the content of the program emphasized child development and early cognitive and socioemotional stimulation. During home visits, nurses showed the participants activities that were appropriate for the child's age to enhance the mother-child bond and to provide adequate stimulation. We hypothesized that these specific activities would buffer the effects of parental stress and have a positive influence on parental well-being.

The compliance rate of the eDiary assessment was 84% (participants who responded at least one day of assessment), demonstrating that it was generally accepted by adolescent mothers living in an urban deprived area. This compliance rate was similar to that reported in other studies conducted with participants in the same age range. For example, a study conducted in Canada using a compliance criterion of participants responding to questions less than 7 times reported a rate of 90.4% [12]. A recent review on ecological momentary assessment studies with adolescents showed that among clinical studies with less than 3 assessments per day, the average compliance rate was 73.5%. This same review also showed that studies with more prompts per day (6 or more) had a higher compliance rate [61], suggesting that future studies should implement similar methods. Furthermore, a recent analysis of a pooled dataset from 10 ambulatory assessment studies comprising more than 1700 participants showed that compliance declined across days [62], which is similar to the pattern in the compliance rate observed in the present study. Even though a higher compliance rate across days would be ideal, we believe that our findings show that assessments conducted via smartphones can be adequately conducted in low-income areas. This method can be a low-cost alternative for clinical trials in LMICs, as well as a potential monitoring method linked to the public health system. In this way, health professionals could be notified when the frequency of parenting behaviors falls below the expectation to adequately stimulate child development, potentially leading to customized interventions.

Limitations

Our findings should be viewed in light of some limitations. The sample size calculation of our clinical trial was based on different outcomes than those reported herein. The eDiary outcomes that were analyzed are secondary outcomes of the clinical trial, which were conceptualized and implemented after the clinical trial started. Therefore, our study may not have had sufficient power to find differences in these secondary outcomes, resulting in potential false negatives when compared to primary outcomes analyses. Additionally, our sample size limited the scope of our analyses. For instance, it would be interesting to use a structural equation modeling approach such as growth curves to validate the constructs assessed; however, the small sample sizes did not allow for convergence of these models. Moreover, the lack of variability and a ceiling effect in some variables may have also influenced these potential analyses. Therefore, even though we presented findings suggesting that we adequately measured the constructs of interest, we were not able to validate these measures quantitatively. However, we found that maternal parenting behaviors were negatively associated with maternal depression measured at the same time point [25], which is an expected finding since it is known that depressive symptoms can influence parenting. In addition, our assessment of maternal parenting was based exclusively on the frequency of behaviors, and we did not collect data on the quality of the interaction between the participants and their children. The quality of maternal parenting is known to be crucial for child development and the mother-child relationship. In addition, the use of eDiary self-report assessment may have influenced the findings due to effects of social desirability bias [63] or the participants' awareness of being part of an ECD study [64]. However, this is unlikely since we detected low rates of some maternal parenting behaviors that would be considered socially approved. Moreover, we included a subgroup of participants who did not own a smartphone at the time of trial commencement. This was surprising since 86% of participants reported owning a smartphone at baseline. We found that it is a common practice among this population to frequently exchange, borrow, or buy smartphones. Furthermore, some participants owned old smartphones with limited hardware in terms of performance as well as with old operating systems,

preventing them from properly using the smartphone app. We were able to include these participants by lending them a smartphone, which could have influenced our results. However, we did not find differences in outcomes between participants with their own smartphone compared with those who borrowed smartphones. Future implementations of smartphone-based assessments in low-income areas should consider that some people may not own a smartphone. This approach also may not be suited for rural areas, especially in LMICs, where poverty is more predominant.

Conclusion

Our findings demonstrate the efficacy of Primeiros Laços, a nurse home visiting program for pregnant youth living in an urban deprived area, on improving maternal parenting and well-being assessed by an eDiary. Assessments with the eDiary were successfully conducted in this specific population. The implications are two-fold. First, Primeiros Laços is a promising intervention to promote maternal parenting and well-being among low-income adolescent mothers. As a structured and manualized intervention, this could be implemented in the primary care system, potentially benefiting millions of Brazilian mothers and children. Other LMICs similar to Brazil may also benefit from the program given appropriate adaptations for language, culture, and context. Second, our findings demonstrate the potential for future ECD intervention studies to implement ambulatory assessment in LMICs via smartphones for measuring mother and child behaviors. More frequent assessments of maternal parenting behaviors and well-being should be implemented to further enhance temporal and ecological validity, and to also expand the scope of measured behaviors.

Acknowledgments

This work was supported by Grand Challenges Canada (GCC), Fundação Maria Cecília Souto Vidigal (0722-03), the Bill & Melinda Gates Foundation (OPP1142172), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2016/13451-9), Conselho Nacional de Desenvolvimento Científico e Tecnológico (420267/2016-6), and Companhia Brasileira de Metalurgia e Mineração. We thank all of the study participants and our research team members: Flávia Afonso, Prof. Alexandre Ferraro, Ana Simon, Patrícia Pereira, Andrea Nascimento, Angela Bagattini, Letícia Silva, Cecília Silva, Maria Aparecida, Fernanda Alarcão, and Kátia Brito.

Conflicts of Interest

GP has served as a consultant to Shire and Johnson & Johnson. He has served on the speakers' bureau of Shire. He has received royalties from Editora Manole. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 16720 KB - mhealth_v8i7e13686_app1.pdf]

References

1. Darroch J, Woog V, Bankole A, Ashford L. Guttmacher Institute. New York: Guttmacher Institute; 2016. Costs and Benefits of Meeting the Contraceptive Needs of Adolescents URL: https://www.guttmacher.org/sites/default/files/report_pdf/adding-it-up-adolescents-report.pdf [accessed 2020-06-01]
2. Ganchimeg T, Ota E, Morisaki N, Laopaiboon M, Lumbiganon P, Zhang J, WHO Multicountry Survey on Maternal Newborn Health Research Network. Pregnancy and childbirth outcomes among adolescent mothers: a World Health Organization multicountry study. BJOG 2014 Mar;121(Suppl 1):40-48. [doi: 10.1111/1471-0528.12630] [Medline: 24641534]
3. Conde-Agudelo A, Belizán JM, Lammers C. Maternal-perinatal morbidity and mortality associated with adolescent pregnancy in Latin America: Cross-sectional study. Am J Obstet Gynecol 2005 Feb;192(2):342-349. [doi: 10.1016/j.ajog.2004.10.593] [Medline: 15695970]
4. Sommer K, Whitman TL, Borkowski JG, Schellenbach C, Maxwell S, Keogh D. Cognitive readiness and adolescent parenting. Dev Psychol 1993;29(2):389-398. [doi: 10.1037/0012-1649.29.2.389]
5. Azevedo WFD, Diniz MB, Fonseca ESVB, Azevedo LMRD, Evangelista CB. Complications in adolescent pregnancy: systematic review of the literature. Einstein (Sao Paulo) 2015;13(4):618-626 [FREE Full text] [doi: 10.1590/S1679-45082015RW3127] [Medline: 26061075]
6. Grau J, Wilson K, Smith E, Castellanos P, Duran P. Adolescent Parenting: Risk and Protective Factors in the Context of Poverty. In: King R, Maholmes V, editors. The Oxford Handbook of Poverty and Child Development. Oxford: Oxford University Press; 2012.
7. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, Lancet Early Childhood Development Series Steering Committee. Early childhood development coming of age: science through the life course. Lancet 2017 Jan 07;389(10064):77-90 [FREE Full text] [doi: 10.1016/S0140-6736(16)31389-7] [Medline: 27717614]

8. Carter A, Garrity-Rokous F, Chazan-Cohen R, Little C, Briggs-Gowan M. Maternal depression and comorbidity: predicting early parenting, attachment security, and toddler social-emotional problems and competencies. J Am Acad Child Adolesc Psychiatry 2001 Jan;40(1):18-26. [doi: 10.1097/00004583-200101000-00012] [Medline: 11195555]
9. van Os J, Lataster T, Delespaul P, Wichers M, Myin-Germeys I. Evidence that a psychopathology interactome has diagnostic value, predicting clinical needs: an experience sampling study. PLoS One 2014;9(1):e86652 [FREE Full text] [doi: 10.1371/journal.pone.0086652] [Medline: 24466189]
10. Kramer I, Simons CJP, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014 Feb;13(1):68-77. [doi: 10.1002/wps.20090] [Medline: 24497255]
11. Killingsworth MA, Gilbert DT. A wandering mind is an unhappy mind. Science 2010 Nov 12;330(6006):932. [doi: 10.1126/science.1192439] [Medline: 21071660]
12. Sawada N, Gagné FM, Séguin L, Kramer MS, McNamara H, Platt RW, et al. Maternal prenatal felt security and infant health at birth interact to predict infant fussing and crying at 12 months postpartum. Health Psychol 2015 Aug;34(8):811-819. [doi: 10.1037/hea0000152] [Medline: 25528180]
13. Passini CM, Pihet S, Favez N, Schoebi D. Assessment of parental discipline in daily life. J Fam Psychol 2013 Apr;27(2):324-329. [doi: 10.1037/a0031504] [Medline: 23458698]
14. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using Smartphone-Based Psychoeducation to Reduce Postnatal Depression Among First-Time Mothers: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]
15. de Neigreiros da Silva NV, Pontes CM, Cabral de Sousa NF, de Vasconcelos MGL. [Health Technologies and their contributions to the promotion of breastfeeding: an integrative review of the literature]. Cien Saude Colet 2019 Feb;24(2):589-602 [FREE Full text] [doi: 10.1590/1413-81232018242.03022017] [Medline: 30726390]
16. Litterbach E, Russell CG, Taki S, Denney-Wilson E, Campbell KJ, Laws RA. Factors Influencing Engagement and Behavioral Determinants of Infant Feeding in an mHealth Program: Qualitative Evaluation of the Growing Healthy Program. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e196 [FREE Full text] [doi: 10.2196/mhealth.8515] [Medline: 29254908]
17. Russell CG, Denney-Wilson E, Laws RA, Abbott G, Zheng M, Lymer SJ, et al. Impact of the Growing Healthy mHealth Program on Maternal Feeding Practices, Infant Food Preferences, and Satiety Responsiveness: Quasi-Experimental Study. JMIR Mhealth Uhealth 2018 Apr 25;6(4):e77 [FREE Full text] [doi: 10.2196/mhealth.9303] [Medline: 29695373]
18. 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 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
19. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review. JMIR Mhealth Uhealth 2018 Jan 09;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]
20. Daly LM, Horey D, Middleton PF, Boyle FM, Flenady V. The Effect of Mobile App Interventions on Influencing Healthy Maternal Behavior and Improving Perinatal Health Outcomes: Systematic Review. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10012 [FREE Full text] [doi: 10.2196/10012] [Medline: 30093368]
21. Chan KL, Chen M. Effects of Social Media and Mobile Health Apps on Pregnancy Care: Meta-Analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
22. Tarqui-Mamani C, Sanabria-Rojas H, Portugal-Benavides WJ, García JC, Castro-Garay W, Escalante-Lazo R, et al. [Effectiveness of mobile technology and weight gain in pregnant women in Callao, Peru]. Rev Salud Publica (Bogota) 2018;20(1):67-72 [FREE Full text] [doi: 10.15446/rsap.V20n1.63488] [Medline: 30183887]
23. Fatori D, Matijasevitch A, Brentani H, Constantino Miguel E, Polanczyk G. 3.30 The effects on child development of a nurse home visitation program for pregnant youth living in adverse conditions: a randomized control trial. J Am Acad Child Adolescent Psych 2019 Oct;58(10):S204. [doi: 10.1016/j.jaac.2019.08.193]
24. Shephard E, Fatori D, Mauro LR, de Medeiros Filho MV, Hoexter MQ, Chiesa AM, et al. Effects of Maternal Psychopathology and Education Level on Neurocognitive Development in Infants of Adolescent Mothers Living in Poverty in Brazil. Biol Psychiatry Cogn Neurosci Neuroimaging 2019 Oct;4(10):925-934 [FREE Full text] [doi: 10.1016/j.bpsc.2019.05.009] [Medline: 31345780]
25. Fatori D, Matijasevitch A, Brentani H, Constantino Miguel E, Polanczyk G. 3.29 Maternal care and its association with depression and stress: a smartphone daily diary study in the context of a home-visiting intervention for adolescent mothers living in adverse conditions in Sao Paulo, Brazil. J Am Acad Child Adolescent Psych 2019 Oct;58(10):S204. [doi: 10.1016/j.jaac.2019.08.192]
26. Fatori D, Polanczyk GV, Miguel E, Matijasevich A. 5.57 Maternal Depression Trajectories in Adolescent Mothers Living in a Poor Urban Area and Their Association With Parental Stress, Infant Behavioral Problems, and Psychological Violence. J Am Acad Child Adolescent Psych 2017 Oct;56(10):S274. [doi: 10.1016/j.jaac.2017.09.340]
27. Associação Brasileira das Empresas de Pesquisas. São Paulo: Associação Brasileira das Empresas de Pesquisas; 2015. Critério Brasil 2015 e atualização da distribuição de classes para 2016 URL: http://www.abep.org/criterio-brasil [accessed 2020-06-01]

28. Bowlby J. Attachment and loss: retrospect and prospect. Am J Orthopsychiatry 1982 Oct;52(4):664-678. [doi: 10.1111/j.1939-0025.1982.tb01456.x] [Medline: 7148988]
29. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
30. Bronfenbrenner U. Ecology of the family as a context for human development: Research perspectives. Dev Psychol 1986 Nov;22(6):723-742. [doi: 10.1037/0012-1649.22.6.723]
31. Chiesa AM, Fracolli LA, De la O Ramallo Veríssimo M, Campos Pavone Zoboli L, Keismanas de Ávila L, Pina de Oliveira AA. Building health care technologies based on health promotion. Rev Esc Enferm USP 2009 Dec;43(spe2):1352-1357. [doi: 10.1590/S0080-62342009000600036]
32. Martins J, Verissimo M, Oliveira MA. Health agents' evaluation of the project instruments in: "our children - windows of opportunities". Texto Contexto - Enferm 2008;17(1):106-114 [FREE Full text] [doi: 10.1590/S0104-07072008000100012]
33. Sadler LS, Slade A, Close N, Webb DL, Simpson T, Fennie K, et al. Minding the Baby: Enhancing reflectiveness to improve early health and relationship outcomes in an interdisciplinary home visiting program. Infant Ment Health J 2013 Sep 01;34(5):391-405 [FREE Full text] [doi: 10.1002/imhj.21406] [Medline: 24049219]
34. Olds DL. The nurse-family partnership: An evidence-based preventive intervention. Infant Ment Health J 2006 Jan;27(1):5-25. [doi: 10.1002/imhj.20077] [Medline: 28640426]
35. Pinheiro DGM, Coelho TPB, Silva CFAD, Silva LAD, Chiesa AM, Fracolli LA. Validação de conteúdo do program de visita domiciliar para a díade mãe-criança. Cogitare Enferm 2018 Jul 05;23(2). [doi: 10.5380/ce.v23i2.54055]
36. Fracolli LA, Reticena KDO, Abreu FCPD, Chiesa AM. A implementação de um programa de visitas domiciliárias com foco na parentalidade: um relato de experiência. Rev Esc Enferm USP 2018 Aug 23;52:1-8 [FREE Full text] [doi: 10.1590/s1980-220x2017044003361]
37. Paim J, Travassos C, Almeida C, Bahia L, Macinko J. The Brazilian health system: history, advances, and challenges. Lancet 2011 May 21;377(9779):1778-1797. [doi: 10.1016/S0140-6736(11)60054-8] [Medline: 21561655]
38. Ministério da Saúde. Biblioteca Virtual em Saúde. Brasília; 2012. Atenção ao pré-natal de baixo risco URL: http://bvsms.saude.gov.br/bvs/publicacoes/cadernos_atencao_basica_32_prenatal.pdf [accessed 2020-06-01]
39. Ministério da Saúde. Ministério da Saúde. Brasília; 2018. Caderneta da Gestante URL: https://portalarquivos2.saude.gov.br/images/pdf/2018/agosto/31/Caderneta-da-Gestante-2018.pdf [accessed 2020-06-01]
40. Ministério da Saúde. Secretaria de Atenção à Saúde, Departamento de Ações Programáticas Estratégicas. Brasília; 2012. Gestação de Alto Risco Manual Técnico URL: http://bvsms.saude.gov.br/bvs/publicacoes/manual_tecnico_gestacao_alto_risco.pdf [accessed 2020-06-01]
41. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clin Psychol Rev 1988 Jan;8(1):77-100. [doi: 10.1016/0272-7358(88)90050-5]
42. Richter P, Werner J, Heerlein A, Kraus A, Sauer H. On the validity of the Beck Depression Inventory. A review. Psychopathology 1998;31(3):160-168. [doi: 10.1159/000066239] [Medline: 9636945]
43. Gorenstein C, Andrade L. Validation of a Portuguese version of the Beck Depression Inventory and the State-Trait Anxiety Inventory in Brazilian subjects. Braz J Med Biol Res 1996 Apr;29(4):453-457. [Medline: 8736107]
44. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
45. Santos LPD, Lindemann IL, Motta JVDS, Mintem G, Bender E, Gigante DP. Proposal of a short-form version of the Brazilian food insecurity scale. Rev Saude Publica 2014 Oct;48(5):783-789 [FREE Full text] [doi: 10.1590/s0034-8910.2014048005195] [Medline: 25372169]
46. Corrêa AMS. Insegurança alimentar medida a partir da percepção das pessoas. Estud Av 2007 Aug;21(60):143-154. [doi: 10.1590/S0103-40142007000200012]
47. LifeData. Real-time insights for research and clinical trials URL: https://www.lifedatacorp.com/ [accessed 2018-09-04]
48. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
49. Cui J. QIC Program and Model Selection in GEE Analyses. Stata J 2018 Nov 19;7(2):209-220. [doi: 10.1177/1536867X0700700205]
50. Twisk J. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. Cambridge: Cambridge University Press; 2013.
51. Shah R, Sobotka SA, Chen Y, Msall ME. Positive Parenting Practices, Health Disparities, and Developmental Progress. Pediatrics 2015 Aug;136(2):318-326 [FREE Full text] [doi: 10.1542/peds.2014-3390] [Medline: 26216325]
52. Milteer RM, Ginsburg KR, Council On Communications And Media Committee On Psychosocial Aspects Of Child And Family Health, Mulligan DA. The importance of play in promoting healthy child development and maintaining strong parent-child bond: focus on children in poverty. Pediatrics 2012 Jan;129(1):e204-e213. [doi: 10.1542/peds.2011-2953] [Medline: 22201149]
53. Victora CG, Aquino EML, do Carmo Leal M, Monteiro CA, Barros FC, Szwarcwald CL. Maternal and child health in Brazil: progress and challenges. Lancet 2011 May 28;377(9780):1863-1876. [doi: 10.1016/S0140-6736(11)60138-4] [Medline: 21561656]

54. de Oliveira EA, Barros FC, da Silva Anselmi LD, Piccinini CA. The Quality of Home Environment in Brazil: An Ecological Model. J Child Fam Stud 2006 Aug 22;15(5):631-642. [doi: 10.1007/s10826-006-9063-8]
55. Programme for International Student Assessment (PISA).: Organisation for Economic Co-operation and Development; 2015. Brazil URL: http://www.compareyourcountry.org/pisa/country/BRA?lg=en [accessed 2020-06-01]
56. Estadão Conteúdo. Negócios. Porto Alegre: Editora Globo; 2018. Três em cada 10 são analfabetos funcionais no Brasil, aponta estudo URL: https://epocanegocios.globo.com/Brasil/noticia/2018/08/epoca-negocios-tres-em-cada-10-sao-analfabetos-funcionais-no-pais-aponta-estudo.html [accessed 2018-11-13]
57. Rodrigues MF. O Estado de São Paulo. São Paulo: Grupo Estado; 2016. 44% da população brasileira não lê e 30% nunca comprou um livro, aponta pesquisa Retratos da Leitura URL: https://cultura.estadao.com.br/blogs/babel/44-da-populacao-brasileira-nao-le-e-30-nunca-comprou-um-livro-aponta-pesquisa-retratos-da-leitura/ [accessed 2018-11-13]
58. Nelson SK, Kushlev K, English T, Dunn EW, Lyubomirsky S. In defense of parenthood: children are associated with more joy than misery. Psychol Sci 2013 Jan 01;24(1):3-10. [doi: 10.1177/0956797612447798] [Medline: 23201970]
59. Hodgkinson S, Beers L, Southammakosane C, Lewin A. Addressing the mental health needs of pregnant and parenting adolescents. Pediatrics 2014 Jan;133(1):114-122 [FREE Full text] [doi: 10.1542/peds.2013-0927] [Medline: 24298010]
60. Larson NC. Parenting Stress Among Adolescent Mothers in the Transition to Adulthood. Child Adolescent Soc Work J 2004 Oct;21(5):457-476. [doi: 10.1023/B:CASW.0000043359.38426.b4]
61. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance With Mobile Ecological Momentary Assessment Protocols in Children and Adolescents: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
62. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Response compliance and predictors thereof in studies using the experience sampling method. Psychol Assess 2019 Feb;31(2):226-235. [doi: 10.1037/pas0000662] [Medline: 30394762]
63. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217. [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
64. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ECD: early childhood development
eDiary: electronic daily diary
LMIC: low and middle-income country
PISA: Programme for International Student Assessment

Edited by G Eysenbach; submitted 12.02.19; peer-reviewed by W Curioso, K Stasiak, M McKinley, S Maingot, A Fairman; comments to author 02.10.19; revised version received 27.11.19; accepted 07.02.20; published 28.07.20.

Please cite as:
Fatori D, Argeu A, Brentani H, Chiesa A, Fracolli L, Matijasevich A, Miguel EC, Polanczyk G
Maternal Parenting Electronic Diary in the Context of a Home Visit Intervention for Adolescent Mothers in an Urban Deprived Area of São Paulo, Brazil: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e13686
URL: https://mhealth.jmir.org/2020/7/e13686
doi:10.2196/13686
PMID:32720906

Original Paper

Evaluating Network Readiness for mHealth Interventions Using the Beacon Mobile Phone App: Application Development and Validation Study

Thomas Foster Scherr[1*], PhD; Carson Paige Moore[1*], BSc; Philip Thuma[2], MD; David Wilson Wright[1], PhD

[1]Department of Chemistry, Vanderbilt University, Nashville, TN, United States

[2]Macha Research Trust, Choma, Zambia

[*]these authors contributed equally

Corresponding Author:
Thomas Foster Scherr, PhD
Department of Chemistry
Vanderbilt University
7300 Stevenson Center
1234 Stevenson Center Lane
Nashville, TN, 37235
United States
Phone: 1 615 322 5516
Email: Thomas.f.scherr@vanderbilt.edu

Abstract

Background: Mobile health (mHealth) interventions have the potential to transform the global health care landscape. The processing power of mobile devices continues to increase, and growth of mobile phone use has been observed worldwide. Uncertainty remains among key stakeholders and decision makers as to whether global health interventions can successfully tap into this trend. However, when correctly implemented, mHealth can reduce geographic, financial, and social barriers to quality health care.

Objective: The aim of this study was to design and test Beacon, a mobile phone–based tool for evaluating mHealth readiness in global health interventions. Here, we present the results of an application validation study designed to understand the mobile network landscape in and around Macha, Zambia, in 2019.

Methods: Beacon was developed as an automated mobile phone app that continually collects spatiotemporal data and measures indicators of network performance. Beacon was used in and around Macha, Zambia, in 2019. Results were collected, even in the absence of network connectivity, and asynchronously uploaded to a database for further analysis.

Results: Beacon was used to evaluate three mobile phone networks around Macha. Carriers A and B completed 6820/7034 (97.0%) and 6701/7034 (95.3%) downloads and 1349/1608 (83.9%) and 1431/1608 (89.0%) uploads, respectively, while Carrier C completed only 62/1373 (4.5%) file downloads and 0/1373 (0.0%) file uploads. File downloads generally occurred within 4 to 12 seconds, and their maximum download speeds occurred between 2 AM and 5 AM. A decrease in network performance, demonstrated by increases in upload and download durations, was observed beginning at 5 PM and continued throughout the evening.

Conclusions: Beacon was able to compare the performance of different cellular networks, show times of day when cellular networks experience heavy loads and slow down, and identify geographic “dead zones” with limited or no cellular service. Beacon is a ready-to-use tool that could be used by organizations that are considering implementing mHealth interventions in low- and middle-income countries but are questioning the feasibility of the interventions, including infrastructure and cost. It could also be used by organizations that are looking to optimize the delivery of an existing mHealth intervention with improved logistics management.

(JMIR Mhealth Uhealth 2020;8(7):e18413) doi:10.2196/18413

KEYWORDS

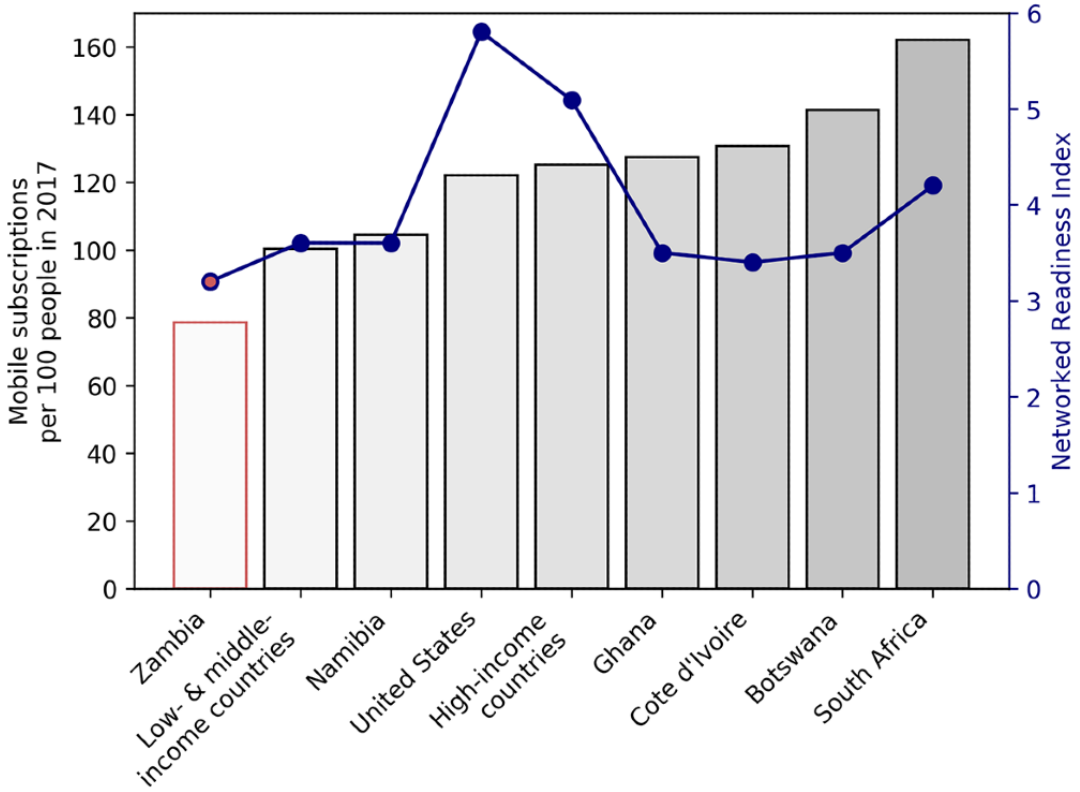
mHealth; network readiness; network assessment; mobile network

Introduction

A technological revolution is rapidly changing the way health care is approached worldwide. Previously insurmountable logistical challenges such as supply chain management [1], cold reagent storage [2,3], and disease surveillance [4] are being solved by a continuous stream of innovations [5,6]. These solutions, which can range from simple SMS text message appointment reminders [7,8] to more sophisticated approaches such as mobile phone–based clinical-grade electrocardiograms [9], have broad uses and can be tailored to address many scenarios and user groups [5,6,10]. This new class of health care interventions, called mobile health (mHealth), takes advantage of the prevalence of network-connected mobile devices and mobile apps and is being adapted to meet challenges present in low- and middle-income countries to improve health outcomes [11]. With 5.3 billion mobile subscribers worldwide and 90% of the global population covered by wireless signals, a niche has developed for the use of mobile phones in public health campaigns [12-14]. mHealth approaches are growing in popularity in high-income countries, where they provide patients with improved convenience and enable providers to engage their patients in near real-time [15-17]. However, mHealth approaches are also being used in global health settings, where they are implemented to decrease costs, minimize barriers to facilitating care, and provide more useful surveillance data [18-20].

The growth of global mHealth has naturally followed the surge in mobile phone uptake. The ubiquity of mobile devices in low- and middle-income countries increased significantly between 2000 and 2015, and it is expected to expand further [21]. In Africa specifically, the average number of mobile phone subscriptions per 100 people increased from 3 to 85 over that same period [21]. In some low- and middle-income countries, the number of mobile subscriptions per person is higher than in many developed countries, including the United States (Figure 1).

Figure 1. The number of mobile subscriptions per 100 people in 2017 and the World Economic Forum Networked Readiness Index (NRI) for several African nations, the United States, World Bank low- to middle-income countries, and World Bank high-income countries. Data for Zambia, the country where this study was performed, are emphasized in red. The NRI data for low- to middle-income countries and high-income countries are averages of the available NRIs for World Bank countries in each classification (several countries in each group do not have a listed NRI).



In Zambia, where we conducted our study, the World Bank reported a prevalence of 78.6 mobile phone subscriptions per 100 people. Despite the availability of mobile phones in Zambia, the country has a Networked Readiness Index (NRI) value of 3.2/6.0, ranking 116th worldwide. The NRI is a measurement that quantifies a nation's use of technology to increase international competitiveness. Zambia has approximately one-third fewer mobile phones per capita than the United States and a conspicuously lower NRI score; therefore, it is easy to formulate a skeptical opinion of the feasibility of mHealth in countries such as Zambia. While it is important to note that mobile phone penetration or mobile network performance alone does not infer broad mHealth readiness, network performance remains a critical factor in preparing for mHealth interventions. Nontechnical factors such as uptake by policy makers and health care practitioners as well as social and cultural acceptance remain integral to the success of mHealth interventions; however, we have chosen to focus on a specific technical indicator [22].

Although mobile phone use has penetrated developing nations to an extent that was unimaginable two decades ago, the implementation of new mHealth interventions in these nations has been met with apprehension from key stakeholders and decision makers [23-26]. Another major source of skepticism is the lack of growth of current projects in the field. Concerns regarding poor connectivity and limitations of accessibility have hampered the success of previous mHealth interventions, and these barriers prevent pilot projects from developing into sustainable, large-scale programs. In Uganda alone, between 2008 and 2009, 23 mHealth projects did not progress beyond pilot testing [12,27]. The failure of these programs to develop beyond preliminary investigations into widespread, integrated mHealth systems increases hesitation, and mobile networks remain an underused health care resource in low- and middle-income countries.

Pilot studies of mHealth interventions often fail to move forward because they cannot overcome the limitations of working with mobile devices in resource-limited settings. Each setting has a unique mobile landscape, and successful strategies are sometimes not generalizable not only from country to country but even within a country or region of interest. Therefore, although the network infrastructure may be in place, even with its clear room for improvement, mHealth pilot studies have not yet succeeded in harnessing that network to produce improved health outcomes. While several factors contribute to overall mHealth readiness, a need remains for tools that can be employed to provide a more complete understanding of the mobile landscape to inform mHealth interventions, particularly in the more remote areas of low- and middle-income countries, where this information is scarce. In this study, we present a mobile phone–based tool, Beacon, to evaluate a setting's capacity for basic mHealth interventions.

Methods

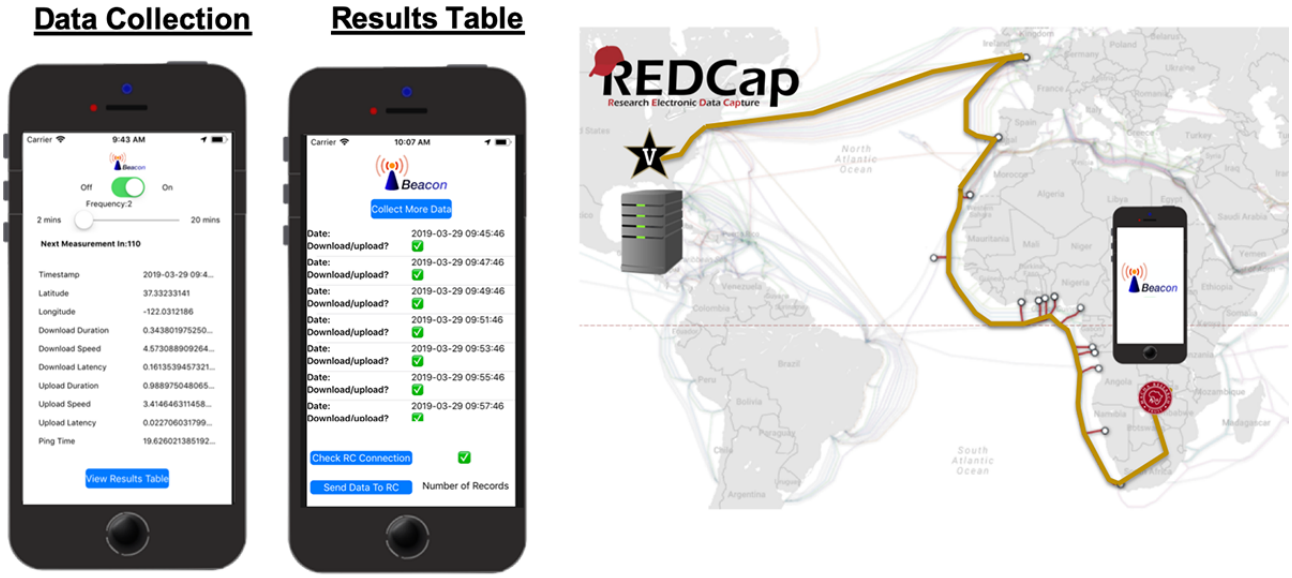
Purpose

We envisioned the Beacon app as a platform to evaluate the infrastructure of a global health setting for readiness to implement mobile health interventions. While there are several existing methods to measure cellular signals, these technologies are often limited by a lack of generalizability, inability to differentiate errors, failure to permanently record any data collected, and requirement for repeated manual intervention to initiate data collection. To rigorously document and assess a region's mHealth readiness, it is critical for the ideal tool to be broadly accessible, reliable and to have automated data recording capabilities. As such, the software was designed to collect spatiotemporal data on cellular network capabilities and mimic certain necessities, such as data transmission to and from research databases or electronic medical records. In this case, the files that were repeatedly uploaded and downloaded in the software were small camera phone images (3.37 MB and 1.57 MB in size) of malaria rapid diagnostic tests (RDTs); upload and download of such images is a frequent occurrence in our diagnostic work in malaria elimination campaigns. The images were then used repeatedly to mimic RDT image transfer as an example of a potential mHealth app. However, these files could be readily modified to fit an alternative mHealth intervention, including a mock patient record or a video file.

Software

A custom app was developed in the Swift programming language using the XCode integrated development environment and deployed through the Apple App Store's beta testing program, TestFlight. The app has two screens: a main screen for data collection and a results screen for data transmission. Collected data are reported as time, latitude, longitude, ping duration, download latency, total download duration, download speed, upload latency, total upload duration, and upload speed (Multimedia Appendix 1). The main view of the app has a simple interface with four components (Figure 2): 1) an on/off switch that allows the user to start and stop data collection, 2) a slider that allows the user to select the frequency of data collection, 3) a table containing the most recent data, and 4) a button to send the user to the Results table, where they can transmit their results to the Research Electronic Data Capture (REDCap) electronic research database [28]. The user can toggle data collection on using the on/off switch, at which point a timer is started with the duration specified by the frequency selected by the user. When the timer expires, the software asynchronously performs the following events using third party libraries (Alamofire v4.7 and PlainPing v0.4): a timestamp is recorded, GPS coordinates are recorded, a ping is sent to a physical server located in the Wright Laboratory at Vanderbilt University, a preloaded image of a rapid diagnostic test is downloaded from the server to the mobile phone, and a different preloaded rapid diagnostic test image is uploaded from the mobile phone to the server. For work in remote regions with unknown connectivity, the geolocation and timestamp are still recorded in the event that an upload, download, or ping fails. All the data describing each interaction between the mobile phone and physical server are entered into a data structure that can be visualized on the results screen. On the results screen, a table displays an entry for each set of collected values. The table row entry shows the timestamp of the data and a check/x symbol depending on if there were any errors in the data entry in that row. At the bottom of the screen, a button ("Check RC Connection") enables the user to check whether their mobile device was able to connect to REDCap for long-term data storage. If this connection failed, the user is notified, and the user's data remain in storage for later upload. If a connection can be established, another button ("Send Data to RC") is enabled; when this button is clicked, the entire data structure describing the interactions of the mobile phone with the VU server is transferred from the app to the web-based REDCap platform. Upon successful data upload to REDCap, the number of data entries uploaded is shown and the data table is cleared.

Figure 2. (Left) The two main screens in the Beacon app: the data collection screen and the results table and data transmission screen. (Right) A representative signal trace from the Macha Research Trust to Vanderbilt University, adapted from [13]. Using Beacon, files were downloaded and uploaded to a server at Vanderbilt University. After sufficient data aggregation, data were transferred to REDCap for storage. REDCap: Research Electronic Data Capture.



Hardware

The software was tested for compatibility on multiple iOS devices, including an iPhone 8 and an iPhone X. The data collection in Macha was performed on multiple iPhone SE (2017) phones, all running iOS 12.1.3. While Apple products are less common globally than they are in the United States, this model has comparable hardware to that in commonly used mobile devices, including the processor, memory, and camera. Therefore, this iPhone is a reasonable device for this testing, with the understanding that the software can readily be ported to Android devices, which are more common in low- and middle-income countries. Data collection was performed on the three primary cellular networks operating in and around Macha, Zambia: Airtel, MTN, and Zamtel. While evaluation of the performance of a particular carrier is an expected use of Beacon in practice, in this work, we randomly assigned each network an alphabetical code to avoid explicitly identifying network performance.

Data Collection Using the Beacon App

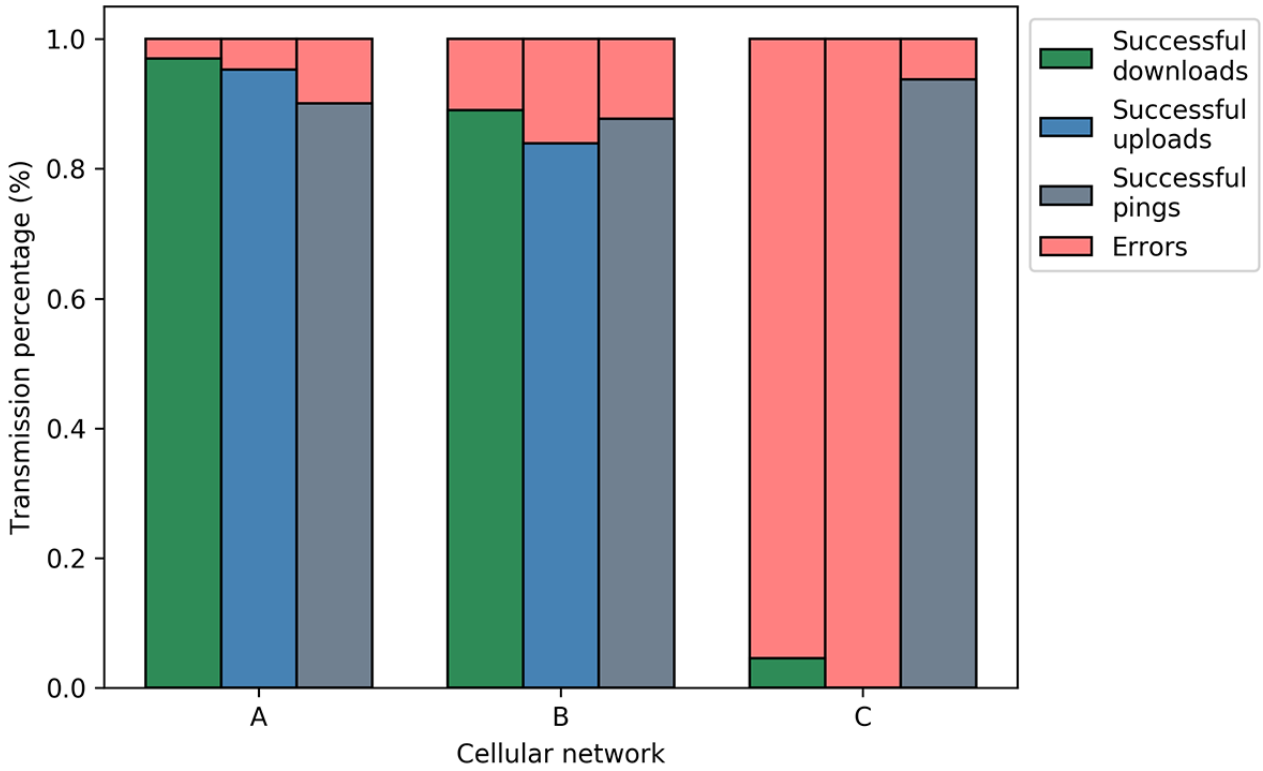
After the timer was initiated in the app, the mobile device was placed in Guided Access mode. This mode allows the user to turn off interactions with the screen or device buttons, allowing the app to collect data uninterrupted by inadvertent screen taps. The app was iteratively and collaboratively refined over the course of 2 weeks before data collection began. Data were collected by three separate users during the course of routine field work in and around the Macha Research Trust between the months of February and August 2019. Users were provided instructions and documentation on how to use the app but were not given any other instructions related to data collection frequencies, locations to visit or to avoid, or how often to upload their results to REDCap. Data were downloaded from REDCap for offline statistical analysis.

Results

Screening Mobile Carrier Efficacy Using Beacon

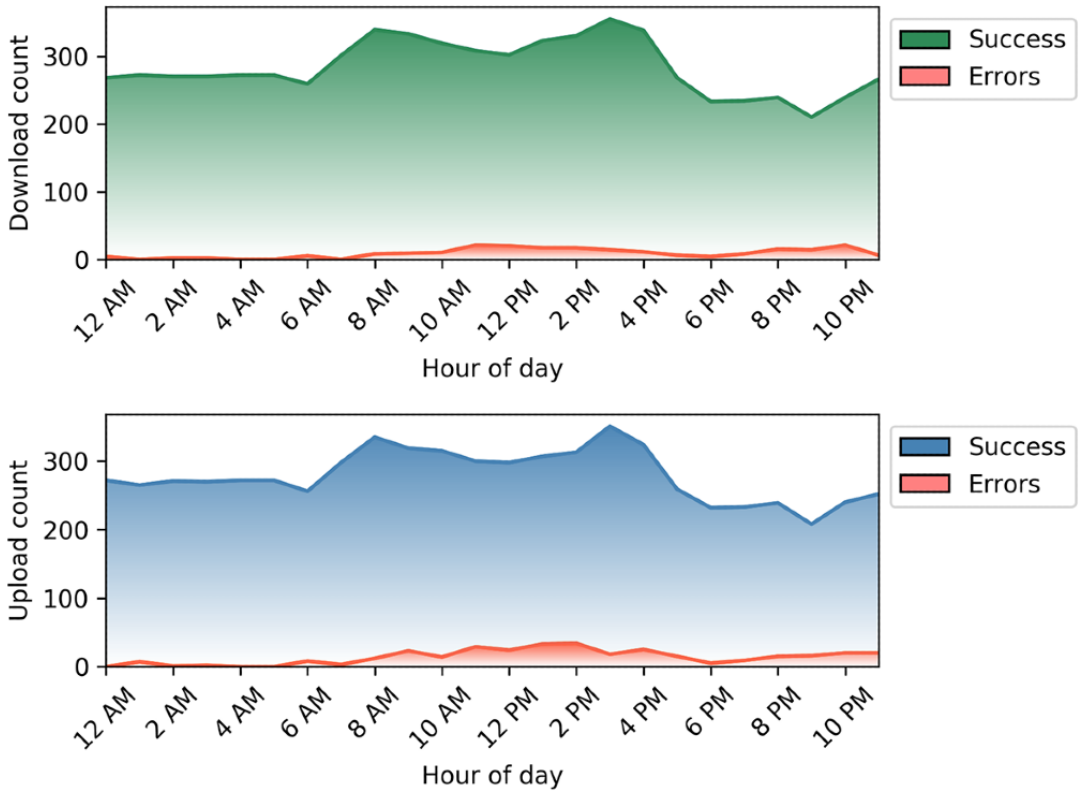
The Beacon app was used to compare the three primary mobile phone carriers operating in or near Macha, Zambia. Figure 3 shows the results of this comparison, delineated by the transmission success rate of each individual modality: downloads, uploads, and pings. Carriers A and B performed with high success; Carrier A completed 6820/7034 (97.0%) successful downloads and 1431/1608 (89.0%) successful uploads, while Carrier B completed 6701/7034 (95.3%) successful downloads and 1349/1608 (83.9%) successful uploads. However, Carrier C was observed to successfully complete only 62/1373 (4.5%) file downloads and 0/1373 (0.0%) file uploads despite a high rate of successful pings. These results agree with anecdotal evidence reported by Macha residents. The remaining data shown in this report focus on Carrier Network A and on upload duration and download duration as two of the most tangible metrics of cellular network performance.

Figure 3. Success and error rates of different modalities of data transfer (download, upload, ping) across three different cellular provider networks.



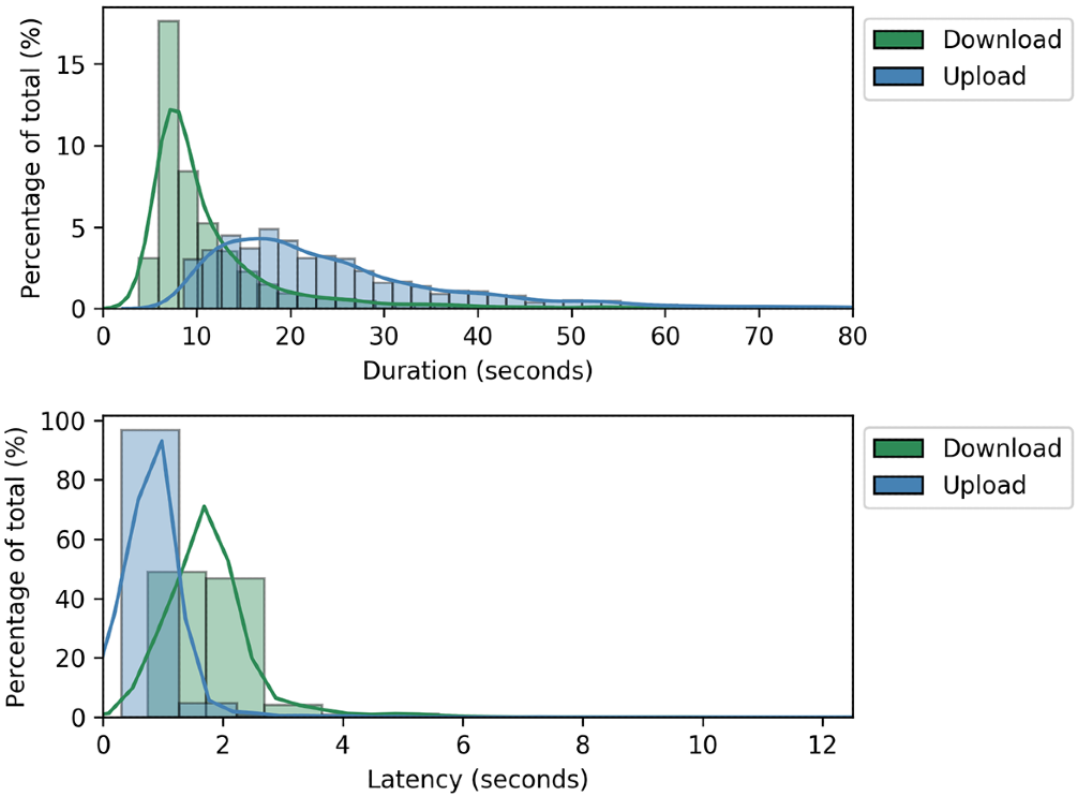
Data collection was not fixed at a single defined frequency throughout the entire study, resulting in a nonuniform distribution of record counts based on the day and time of collection (Figure 4). The largest amounts of data were collected at midday and on weekends. As the system was designed, the number of download attempts matched the number of upload attempts (Multimedia Appendix 2). Errors were mostly observed to increase proportionally as the total data count increased; thus, we did not see sampling bias in our results. However, there was slight disagreement between the number of download errors and the number of upload errors in our results, indicating that success of one mode of transmission did not guarantee success of another.

Figure 4. Download (top) and upload (bottom) success and error counts for cellular network A grouped by hour of the day.



The distributions of the upload and download durations recorded for successful data transmissions on Carrier A are shown in Figure 5. On average, downloads occurred faster and in a narrower time range than uploads. To this point, 50% of downloads were completed in under 8.5 seconds and 75% were completed in under 13 seconds, whereas 50% and 75% of uploads were completed in 21 and 30.5 seconds, respectively; the fastest and slowest download times were 3.7 seconds and 21.2 seconds, respectively, while the fastest and slowest upload times were 8.4 and 53.8 seconds, respectively. The average download latency was larger than the average upload latency (1.8 seconds and 0.8 seconds, respectively).

Figure 5. Distributions of upload and download duration (top) and latency (bottom) recorded for Carrier A over the course of Beacon testing. Data outside 1.5 times the interquartile range of the lower and upper quartiles were considered outliers and were removed from the analysis.



Spatial Resolution of Cellular Data Transfer Capabilities in Rural Zambia

All data collected were tagged with both a GPS location and a timestamp. A single representative day of data collection is shown in Figure 6. As can be seen, the signal strength can be overlaid on a map as a point map with respect to the coordinates from which the data were collected. In and around our study site, the signal strength was found to be high, with most downloads lasting between 4 and 12 seconds. For the data shown, an abrupt increase in upload duration begins just after 5 PM local time. These download durations were more than 1.5 times the upper quartile and are statistical outliers. More variation, including several other outliers, can be observed in the temporal upload duration results; however, the increased variation shown in the representative data matches the overall observed trends shown in Figure 5.

Our primary study site was located in Macha (a rural village in the Southern Province of Zambia), and the overwhelming majority of our data were collected in and around Macha. However, users were not geographically constrained, and data were collected at other sites, including a larger town (Choma, 50 kilometers away) and a larger city (Livingstone, 190 km away). The durations of successful downloads and uploads are shown as a function of the haversine distance from the Macha Research Trust in Figure 7. While the distance measured from Macha does not have explicit directionality and thus cannot explicitly confirm location by itself (eg, if a user traveled 50 km in the opposite direction, this plot would still show the data points next to Choma), confirmation of the GPS coordinates shows that the overwhelming majority of data points near 50 km and 190 km from the Macha Research Trust were collected in Choma and Livingstone, respectively. Near the Macha Research Trust, the full ranges of download and upload durations were observed; meanwhile, in Choma and Livingstone, the cellular signal was generally more consistent, but outliers were still occasionally observed. In the regions between these cities, higher numbers of errors were reported, which is probably directly related to the distance from the cell towers along the routes traveled and thus more indirectly related to the lack of densely populated areas between cities.

Figure 6. A representative 24-hour period of data collection: (top) download speed mapped by GPS coordinates, where download speed is denoted on a scale from green (high) to yellow (low); (middle) download duration over time; and (bottom) upload duration over time for the same period. Errors are denoted with red x symbols. DL: download; UL: upload

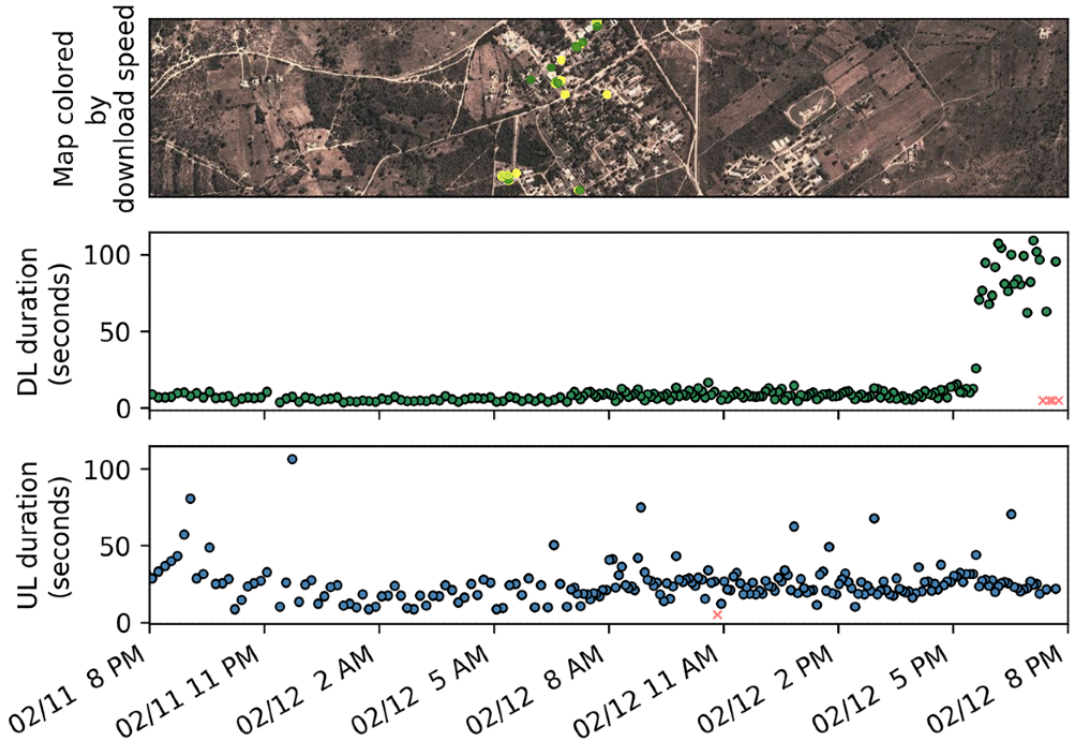
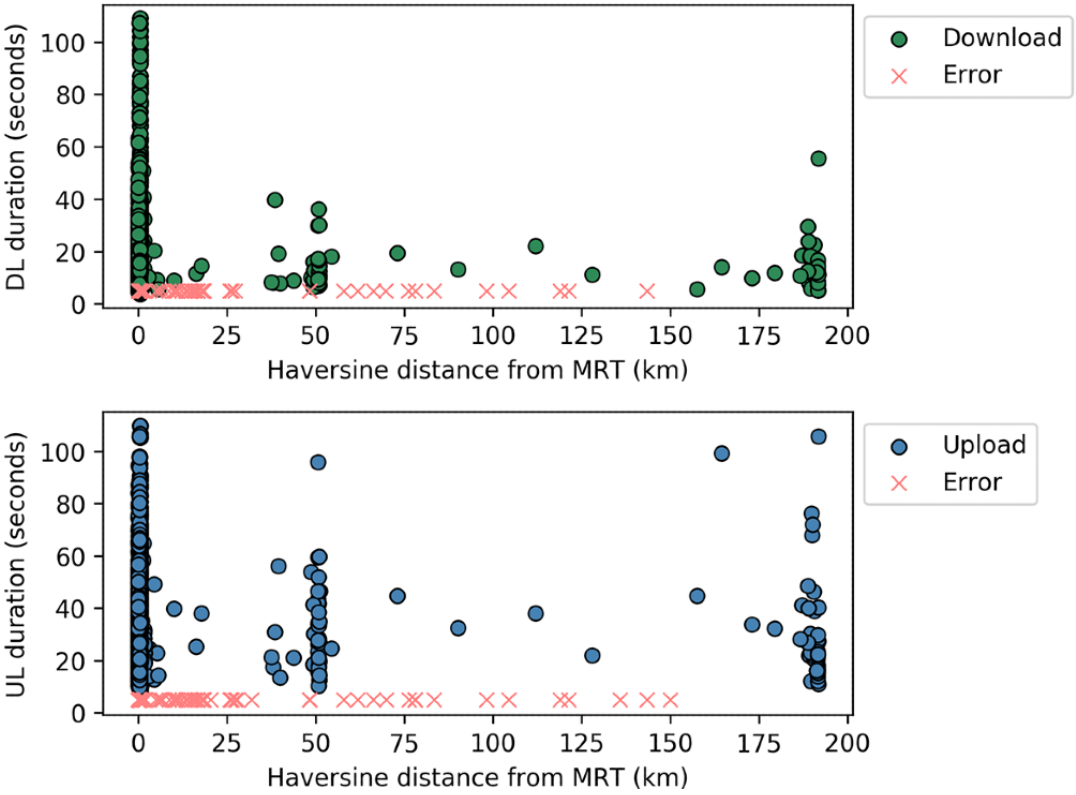


Figure 7. Download (top) and upload (bottom) durations and the errors in each mode as a function of the haversine distance from the Macha Research Trust. DL: download; km: kilometers; MRT: Macha Research Trust; UL: upload.

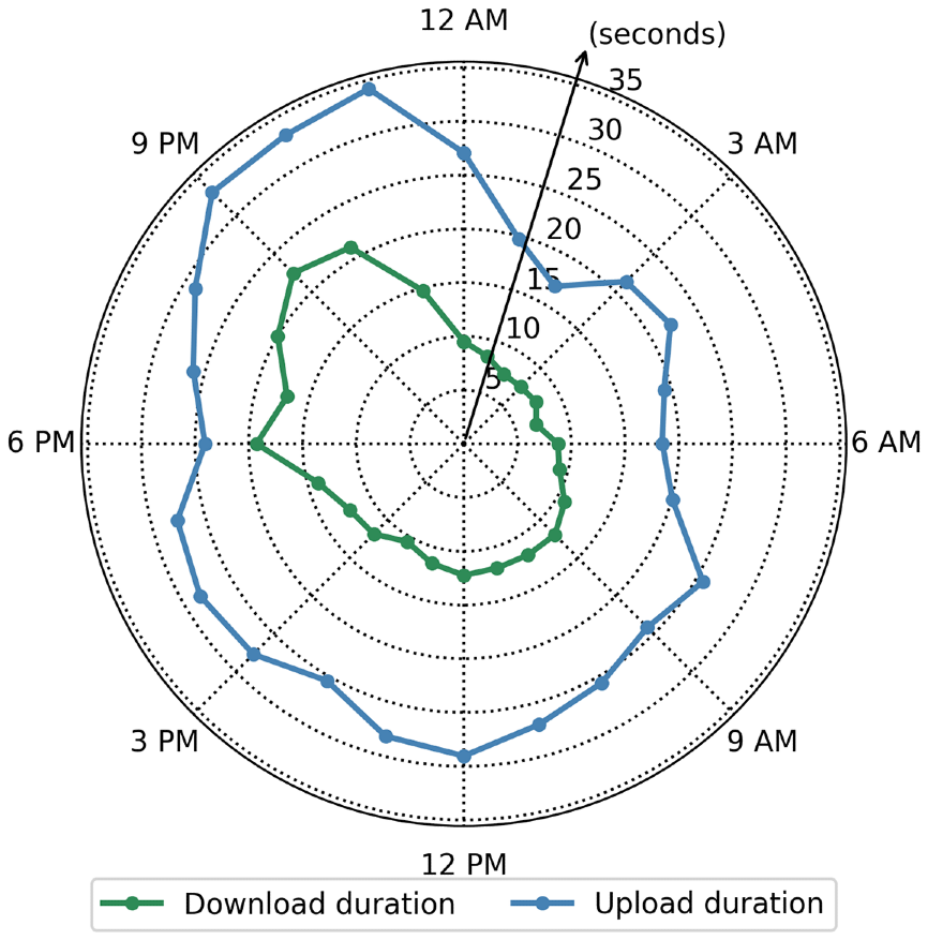


Temporal Distribution of Cellular Data Transfer Capabilities in Rural Zambia

When the data are grouped by hour of the day (Figure 8), the download durations are short early in the morning (from 2 AM to 5 AM), followed by an increase from 5 AM to 8 AM. The durations remain roughly constant from 8 AM to 4 PM, with a slight decrease at 2 PM. Download durations experience a significant increase in the evening hours beginning at 5 PM and lasting through 11 PM, until they begin to decrease again (12

AM to 2 AM). A similar but slightly less pronounced and more noisy trend was observed for upload duration. When the data are grouped by day of the week, on average, the download durations are fastest on Monday (10.8 seconds) and slowest on Tuesday (15.5 seconds); on average, the upload durations are fastest on Sunday (21.8 seconds) and slowest on Friday (31.6 seconds; Multimedia Appendix 3). Upload and download latency reflected a similar, although more noisy, trend (Multimedia Appendix 4).

Figure 8. Average download and upload durations for each hour of the day.



Discussion

Principal Findings

In this study, we developed and validated a tool designed to assist readiness evaluation of prospective sites for the implementation of mHealth intervention. In this study, we utilized the Beacon app to examine mobile network performance in and around Macha, Zambia. During this study, we observed several key characteristics of the Macha mobile network, including hourly variations in the network signal and spatial distribution of the signal strength. However, we note that each mHealth site is unique and will have its own variations in network performance. Our results in Macha, Zambia, may be similar to those in other low- and middle-income countries; however, we recommend that each new site be investigated before mHealth interventions are pursued.

The Beacon app enables spatiotemporal mapping of two of the most practical measures of network performance: upload speed and download speed. We acknowledge that network performance alone does not confer mHealth readiness; however, understanding the digital landscape of a potential mHealth intervention site is a critical step in assessing the feasibility of future mHealth studies. Data collection in Beacon is fully automated, and data are stored remotely in an easily accessible research database. The Beacon app can detect errors in either upload or download attempts and log them with the same geographic and temporal precision provided for successful attempts. The app can be readily tailored based on the needs of researchers to customize data collection and aggregation to fit the intervention of interest.

Beacon was tested in Macha, Zambia, and was critical in developing a more well-rounded understanding of the mobile landscape around this rural village. In this study, we were able to identify the best-performing cellular network for use in potential mHealth interventions and, in this case, validate the anecdotal experiences of Macha residents. Although we observed significant differences between the available networks, variations in network performance are to be expected in any environment. Network performance is a function of many elements, including but not limited to the number and locations of available cell towers, obstructions between towers (eg, buildings), cell tower hardware, mobile unit hardware, and current network load at a given time [29]. Many of these factors

can be determined before or during testing and, combined with Beacon verification, can further influence decisions reached by mHealth researchers.

In addition to a comparative evaluation of multiple networks, Beacon was used to quantitatively measure upload and download metrics around the test site. From anecdotal experience, download speed in Macha was observed to be faster than upload speed, and this was confirmed during data collection. In most scenarios, this is to be expected due to the asymmetric design and operation of cellular networks [30]. However, despite the increased speed of downloads compared to uploads, download latency was observed to be higher than upload latency. This result is counterintuitive in light of the observed trends in total download/upload duration but may be related to the asymmetry of network configurations. However, overall latency in both directions was low and did not exceed the acceptable latency for single-item transfer, which is a positive finding given the physical distance between the mobile phone (Macha, Zambia) and the home server (Vanderbilt University, Nashville, TN) [31].

Even the simple raw data from Beacon can be transformed into more information-dense results; for instance, data can be grouped temporally (by time of day and day of week). We saw no discernible trend for upload or download speed based on the day of the week; however, the hourly data did appear to correlate with times of high human activity in the catchment area. Between approximately midnight and 6 AM, upload and download durations were at their lowest level, and they began to increase between 6 AM and 8 AM. This is presumably related to many people waking and beginning to use their mobile devices. During working hours, this moderate activity remained and continued until approximately 5 PM, when mobile phone users leave work and more freely engage with their mobile devices. These data are useful for mHealth interventions, as they provide both an optimal window for large-volume data transfer or other network activities and a proxy for when users are on their mobile devices. Based on this information, an mHealth implementation team may decide to schedule large data backups in the early hours of the morning, when the burden on the local network is low and it would thus be more capable of high-capacity use; they may also decide that lightweight SMS text message interventions will be more successful at times of higher network activity.

The spatial resolution of Beacon is also a strength for mHealth researchers who are planning interventions. Plots of signal strength mapped over topographical or satellite images can help visualize the digital landscape and can be useful to denote strength based on distance from landmarks such as cell towers, hospitals, and airports. This, in turn, can help inform researchers interested in implementing both localized and broadly spaced interventions. Further collaboration between mHealth researchers, local municipalities, and internet service providers could enable the use of these data to inform strategic infrastructure improvements.

As mentioned previously, users were not provided any instructions regarding the frequency of data collection. Among the strengths of the Beacon app are that it can be easily powered on and off and that the frequency of data collection can be adjusted over a broad range. While high-frequency data collection provides more granular temporal and spatial data, it does come at a higher overall cost. With a small file size and at the minimum collection rate (collection every 20 minutes), Beacon costs less than 30 kwacha per week (approximately US $2). These operating costs would be expected to increase proportionally with higher frequency or larger download/upload file size; however, high spatiotemporal data can still be collected for <US $100 per month. This potential increase in cost experienced by varying these parameters may more closely approximate the actual intervention at a site; therefore, it may prove to be worthwhile to mHealth investigators when using Beacon-based surveillance. Although we did not observe sampling bias during this initial test, we do recognize the potential for variable sampling frequency to introduce bias into a preliminary study. For instance, if rolling outages were to occur during a certain time of day in a study location and collection frequency was increased during this time, further analysis of sampling practices might be required to account for the higher volume of signal errors. This requires diligent monitoring and analysis or, as a more straightforward solution, the implementation of a strict data collection frequency that reflects the needs of the planned mHealth intervention.

Conclusions

Health care remains a considerable challenge in many parts of the world, particularly resource-limited countries. Even without continuous connectivity, mHealth interventions have genuine potential to improve global health outcomes. Although many countries have high mobile phone penetration, not all locations are capable of supporting mHealth. The lack of infrastructure and high costs may be prohibitive; however, proper planning and tailoring of these interventions based on the infrastructure available greatly increases the likelihood of success. The small cost of piloting a short signal mapping study, such as this one, would be negligible compared to the cost of blind implementation of an intervention that is destined to fail. Beacon represents a valuable mHealth tool that can help determine the readiness of a site for mHealth interventions. By combining temporal and spatial tags with metrics of network performance, researchers can better understand the digital landscape around their potential intervention site.

Acknowledgments

The authors would like to acknowledge preliminary data collection by Saidon Mbambara. Additionally, the authors would like to thank MF Richards for critical review of the manuscript. The authors would also like to acknowledge support in part from Fogarty International Center at the National Institutes for Health (1R21TW010635) and the Burroughs Wellcome Fund Collaborative

Research Travel Grant (Scherr, 2016). This work used REDCap, which is supported by the National Center for Advancing Translational Sciences at the National Institutes for Health (UL1TR000445).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Beacon application user interface and workflow.
[DOCX File , 587 KB - mhealth_v8i7e18413_app1.docx]

Multimedia Appendix 2
Download (top) and upload (bottom) duration grouped by hour of the day (left) and day of the week (right).
[PNG File , 199 KB - mhealth_v8i7e18413_app2.png]

Multimedia Appendix 3
Number of successful and unsuccessful download (top) and upload (bottom) data points collected, grouped by day of the week.
[PNG File , 79 KB - mhealth_v8i7e18413_app3.png]

Multimedia Appendix 4
Download (top) and upload (bottom) latency grouped by hour of the day (left) and day of the week (right).
[PNG File , 187 KB - mhealth_v8i7e18413_app4.png]

References

1. Githinji S, Kigen S, Memusi D, Nyandigisi A, Mbithi AM, Wamari A, et al. Reducing stock-outs of life saving malaria commodities using mobile phone text-messaging: SMS for life study in Kenya. PLoS One 2013 Jan;8(1):e54066 [FREE Full text] [doi: 10.1371/journal.pone.0054066] [Medline: 23349786]
2. Gilbert SS, Thakare N, Ramanujapuram A, Akkihal A. Assessing stability and performance of a digitally enabled supply chain: Retrospective of a pilot in Uttar Pradesh, India. Vaccine 2017 Apr;35(17):2203-2208. [doi: 10.1016/j.vaccine.2016.11.101] [Medline: 28364932]
3. Teng JE, Thomson DR, Lascher JS, Raymond M, Ivers LC. Using Mobile Health (mHealth) and Geospatial Mapping Technology in a Mass Campaign for Reactive Oral Cholera Vaccination in Rural Haiti. PLoS Negl Trop Dis 2014 Jul 31;8(7):e3050. [doi: 10.1371/journal.pntd.0003050]
4. Freifeld CC, Chunara R, Mekaru SR, Chan EH, Kass-Hout T, Ayala IA, et al. Participatory epidemiology: use of mobile phones for community-based health reporting. PLoS Med 2010 Dec;7(12):e1000376 [FREE Full text] [doi: 10.1371/journal.pmed.1000376] [Medline: 21151888]
5. Classification of Digital Health Intervention v1.0. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf;jsessionid=656ABE6BBE9CA00813F2ACD598792B8F?sequence=1 [accessed 2019-07-23]
6. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
7. Fairhurst K, Sheikh A. Texting appointment reminders to repeated non-attenders in primary care: randomised controlled study. Qual Saf Health Care 2008 Oct;17(5):373-376. [doi: 10.1136/qshc.2006.020768] [Medline: 18842978]
8. Déglise C, Suggs LS, Odermatt P. Short Message Service (SMS) Applications for Disease Prevention in Developing Countries. J Med Internet Res 2012 Jan 12;14(1):e3. [doi: 10.2196/jmir.1823] [Medline: 22262730]
9. Saxon L. Ubiquitous wireless ECG recording: a powerful tool physicians should embrace. J Cardiovasc Electrophysiol 2013 Apr;24(4):480-483. [doi: 10.1111/jce.12097] [Medline: 23421574]
10. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
11. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014 Oct 27;7(1):25606 [FREE Full text] [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
12. Lemaire J. Scaling Up Mobile Health: Elements Necessary for the Successful Scale Up of mHealth in Developing Countries. Advanced Development for Africa. 2011 Dec. URL: http://www.adaorganization.net/uploads/2/3/7/1/23713723/scaling_up_mobile_health_elements_necessary_for_the_successful_scale_up_of_mhealth_in_developing_countries.pdf [accessed 2019-07-20]

13. Submarine Cable Map. Submarine Cable Map.: TeleGeography URL: https://www.submarinecablemap.com/ [accessed 2020-03-17]
14. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Global Health Action 2017 Aug 25;10(sup3):1327686. [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
15. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and Use of mHealth Apps: A Study from England. Pharmacy (Basel) 2017 Jun 14;5(2) [FREE Full text] [doi: 10.3390/pharmacy5020033] [Medline: 28970445]
16. Shapiro-Mathews E, Barton AJ. Using the patient engagement framework to develop an institutional mobile health strategy. Clin Nurse Spec 2013;27(5):221-223. [doi: 10.1097/NUR.0b013e3182a0b9e2] [Medline: 23942098]
17. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
18. 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 Jun 13;10:49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]
19. Long L, Pariyo G, Kallander K. Digital Technologies for Health Workforce Development in Low- and Middle-Income Countries: A Scoping Review. Glob Health Sci Pract 2018 Oct 10;6(Supplement 1):S41-S48. [doi: 10.9745/ghsp-d-18-00167]
20. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014. [doi: 10.1111/tmi.12525] [Medline: 25881735]
21. Mobile cellular subscriptions (per 100 people) Internet. The World Bank Data.: The World Bank URL: https://data.worldbank.org/indicator/IT.CEL.SETS.P2 [accessed 2019-12-09]
22. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015 Mar;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
23. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. Bull World Health Organ 2012 May 1;90(5):332-340 [FREE Full text] [doi: 10.2471/BLT.11.099820] [Medline: 22589566]
24. Istepanian R, Woodward B. Mobile health (m-health): Evidence-based progress or scientific retrogression. In: Feng D, editor. Biomedical Information Technology. London: Academic Press, Elsevier; 2016:717-733.
25. Feng D, editor. Biomedical Information Technology. London: Academic Press, Elsevier; 2019:978.
26. Labrique A, Vasudevan L, Mehl G, Rosskam E, Hyder AA. Digital Health and Health Systems of the Future. Glob Health Sci Pract 2018 Oct 10;6(Supplement 1):S1-S4. [doi: 10.9745/ghsp-d-18-00342]
27. Siedner MJ, Lankowski A, Musinga D, Jackson J, Muzoora C, Hunt PW, et al. Optimizing network connectivity for mobile health technologies in sub-Saharan Africa. PLoS One 2012 Sep 28;7(9):e45643 [FREE Full text] [doi: 10.1371/journal.pone.0045643] [Medline: 23029155]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
29. Chen Y, Duffield N, Haffner P, Hsu WL, Jacobson G, Jin Y, et al. Understanding the complexity of 3G UMTS network performance. In: 2013 IFIP Networking Conference.: Institute of Electrical and Electronics Engineers; 2013 May Presented at: 2013 IFIP Networking Conference; 2013; Brooklyn, NY.
30. Agarwal S, Laifenfeld M, Trachtenberg A, Alanyali M. Using bandwidth sharing to fairly overcome channel asymmetry. In: Information Theory and Applications Workshop. 2006 Presented at: Information Theory and Applications Workshop; February 2006; San Diego, CA.
31. Charzinski J. Web Performance in Practice – Why We are Waiting. AEU-Int J Electron C 2001 Jan;55(1):37-45. [doi: 10.1078/1434-8411-00006]

Abbreviations

mHealth: mobile health
NRI: Network Readiness Index
RDT: rapid diagnostic test
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 25.02.20; peer-reviewed by B Eapen, D Mendes; comments to author 18.03.20; revised version received 24.04.20; accepted 13.05.20; published 28.07.20.

Please cite as:
Scherr TF, Moore CP, Thuma P, Wright DW
Evaluating Network Readiness for mHealth Interventions Using the Beacon Mobile Phone App: Application Development and Validation Study
JMIR Mhealth Uhealth 2020;8(7):e18413
URL: http://mhealth.jmir.org/2020/7/e18413/
doi:10.2196/18413
PMID:32720909

Original Paper

Mobile Health Apps in Pediatric Obesity Treatment: Process Outcomes From a Feasibility Study of a Multicomponent Intervention

Sarah Browne[1*], BSc, PhD; M-Tahar Kechadi[2], PhD; Shane O'Donnell[3*], PhD; Mckenzie Dow[1], BA; Louise Tully[4], BA, MA; Gerardine Doyle[5], PhD; Grace O'Malley[4,6], BSc, PhD

[1]School Public Health, Physiotherapy & Sports Science, University College Dublin, Dublin, Ireland

[2]Insight Centre for Data Analytics, School of Computer Science, University College Dublin, Dublin, Ireland

[3]School of Sociology, University College Dublin, Dublin, Ireland

[4]Division of Population Health Sciences, School of Physiotherapy, Royal College of Surgeons in Ireland, Dublin, Ireland

[5]UCD Michael Smurfit Graduate Business School, University College Dublin, Dublin, Ireland

[6]W82GO Child and Adolescent Weight Management Service, Children's Health Ireland at Temple Street, Dublin, Ireland

[*]these authors contributed equally

Corresponding Author:
Grace O'Malley, BSc, PhD
Division of Population Health Sciences
School of Physiotherapy
Royal College of Surgeons in Ireland
Beaux Lane House
Dublin, D02 YN77
Ireland
Phone: 353 014028591
Email: graceomalley@rcsi.ie

Abstract

Background: Multicomponent family interventions underline current best practice in childhood obesity treatment. Mobile health (mHealth) adjuncts that address eating and physical activity behaviors have shown promise in clinical studies.

Objective: This study aimed to describe process methods for applying an mHealth intervention to reduce the rate of eating and monitor physical activity among children with obesity.

Methods: The study protocol was designed to incorporate 2 mHealth apps as an adjunct to usual care treatment for obesity. Children and adolescents (aged 9-16 years) with obesity (BMI ≥98th centile) were recruited in person from a weight management service at a tertiary health care center in the Republic of Ireland. Eligible participants and their parents received information leaflets, and informed consent and assent were signed. Participants completed 2 weeks of baseline testing, including behavioral and quality of life questionnaires, anthropometry, rate of eating by Mandolean, and physical activity level using a smart watch and the myBigO smartphone app. Thereafter, participants were randomized to the (1) intervention (usual clinical care+Mandolean training to reduce the rate of eating) or (2) control (usual clinical care) groups. Gender and age group (9.0-12.9 years and 13.0-16.9 years) stratifications were applied. At the end of a 4-week treatment period, participants repeated the 2-week testing period. Process evaluation measures included recruitment, study retention, fidelity parameters, acceptability, and user satisfaction.

Results: A total of 20 participants were enrolled in the study. A web-based randomization system assigned 8 participants to the intervention group and 12 participants to the control group. Attrition rates were higher among the participants in the intervention group (5/8, 63%) than those in the control group (3/12, 25%). Intervention participants undertook a median of 1.0 training meal using Mandolean (25th centile 0, 75th centile 9.3), which represented 19.2% of planned intervention exposure. Only 50% (9/18) of participants with smart watches logged physical activity data. Significant differences in psychosocial profile were observed at baseline between the groups. The Child Behavior Checklist (CBCL) mean total score was 71.7 (SD 3.1) in the intervention group vs 57.6 (SD 6.6) in the control group, *t*-test $P<.001$, and also different among those who completed the planned protocol compared with those who withdrew early (CBCL mean total score 59.0, SD 9.3, vs 67.9, SD 5.6, respectively; *t*-test $P=.04$).

Conclusions: A high early attrition rate was a key barrier to full study implementation. Perceived task burden in combination with behavioral issues may have contributed to attrition. Low exposure to the experimental intervention was explained by poor acceptability of Mandolean as a home-based tool for treatment. Self-monitoring using myBigO and the smartwatch was acceptable among this cohort. Further technical and usability studies are needed to improve adherence in our patient group in the tertiary setting.

(JMIR Mhealth Uhealth 2020;8(7):e16925) doi:10.2196/16925

KEYWORDS

childhood obesity; diet therapy; mHealth; mobile phones; smartphones; appetite; satiety; rate of eating; accelerometer; physical activity

Introduction

Background

Global prevalence rates of childhood obesity were estimated at 7.8% for boys and 5.6% for girls in 2016, and prevalence is increasing in low-income countries and communities [1]. Interventions in childhood are critical, as children with obesity experience a range of physical and psychosocial health issues and are at a high risk of developing chronic disease in adulthood [2]. Diet, physical activity, and other behavioral interventions can be effective in terms of change in adiposity, and significant clinically relevant metabolic benefits have been demonstrated with a 0.25 reduction in BMI z-score [3], although meaningful reductions in cardiometabolic markers are observed with reductions of 0.15 [4,5]. Recent Cochrane meta-analyses of behavior change interventions reported 12-month reductions in BMI z-score of −0.06 units among children aged 6 to 11 years [6] and −0.13 units (95% CI −0.21 to −0.05) in adolescents aged 12 to 17 years [7].

There is growing evidence that eating behaviors, not simply driven by food choice, influence energy consumption, appetite, and satiety. Fast eating is associated with high body weight [8], and interventions to reduce the eating rate seem to enhance weight loss [9]. Eating food quickly may contribute to blunted responses to normal satiety signals, whereby an individual does not respond to gastric distension and gut peptide release, so that normal appetite suppression pathways do not function as expected during fast eating occasions [10,11]. A reduction in the eating rate aimed at reducing portion size and normalizing satiety signaling has been recently studied [10-16]. A study of adolescents aged 9 to 17 years found that an eating rate intervention enhanced weight loss at 12 months compared with usual care (change in BMI z-score of −0.27) [16]. Slowing eating rate can also reduce self-selected portion size with no reduction in postmeal satiety levels among children and teenagers [10,13,16]. A recent review appraised a number of commercial apps targeting appetite regulation [17]. Research-driven interventions include real-time technology-assisted tools for meal times, including utensils with vibrotactile feedback [12,14] and Mandolean, a plate scale measuring eating rate with real-time computer or smartphone feedback [13,15,16]. Mandolean has shown promise for the treatment of childhood obesity [16].

Physical activity in combination with dietary behavior change, rather than either in isolation, is the recommended component of interventions for childhood obesity [6,7,18]. The use of wearable accelerometers to measure physical activity is the accepted objective means of measurement in free-living individuals [19], which can be used to determine energy expenditure and requirements [20], and the time spent in high-intensity physical activity determines variation in childhood cardiometabolic risk factors [21]. One of the advantages of mobile health (mHealth) interventions compared with traditional approaches is that data from monitoring tools and participant engagement are provided objectively.

Despite improved technologies and access to mHealth tools for the purpose of monitoring health status and implementing interventions for health behavior change [22,23], challenges with adherence and exposure remain [24,25]. Planned exposure, impact, and potential outcomes are altered by participants' interaction with study tools and technology [13,15,24]. The importance of content, design, and testing periods with the target group has been emphasized as a means of enhancing engagement with mHealth apps [24,25]. Reporting process measures is increasingly important, as they contribute to moving the field forward and providing translational accuracy in research and practice [26].

Interventions that improve and expand treatment options for children with obesity are important because of challenges within traditional clinical care, including available time and resources that impact access for service users. mHealth tools provide adjunctive options to standard treatment approaches and can be beneficial for patients at home and their clinical team. However, engagement with devices and apps can act as a barrier to treatment [13,15,22].

Objective

The aim of this study was to determine the feasibility and acceptability of an intervention using 2 mHealth apps among children in the treatment for obesity in a tertiary outpatient setting. As diet and physical activity interventions are typically undertaken together, it was of interest to assess the acceptability of Mandolean in addition to a physical activity monitoring tool.

Methods

Study Design

This study was conducted to determine the feasibility and acceptability of a proposed mHealth intervention. The study was not registered as a trial; however, a randomized design was implemented to ascertain protocol feasibility for a proposed randomized controlled trial. We evaluated the process of using

2 mHealth smartphone apps among children and adolescents receiving treatment for obesity. Feasibility measures included recruitment rates and procedures, and retention rates. Fidelity encompassed intervention delivery and adherence to randomization and study procedures. Retrospective acceptability included objective measures (engagement with smartphone apps) and self-reported measures (system usability score surveys and verbal feedback).

Participants

Children and adolescents (aged 9-16 years), with a diagnosis of clinical obesity (BMI >98th percentile for age and sex), referred to the W82GO Child and Adolescent Obesity Service at Children's Health Ireland at Temple Street, Dublin, Ireland, were eligible to participate. Socioeconomic status was indicated by the Pobal HP Deprivation Index for Small Areas [27]. Children and adolescents were required to have access to a smartphone (phones with Android operating system version 6.0 or above were compatible with the smartwatch and myBigO at version 0 when used in the feasibility study and both Android and iOS were compatible with Mandolean, which did not undergo further development during the study). Smartphone literacy was assumed, and training was provided on study apps by a researcher at baseline who established competency in study tasks. Exclusion criteria included moderate or severe learning difficulties that would prevent the use of smartphone apps or giving informed assent, the child having a concurrent serious medical issue, if the parent or child was not proficient in understanding English, refusal by the child to give assent or parents/legal guardians to give informed consent to participate in the project, or if the child lived in Direct Provision (the system of asylum seeker accommodation used in the Republic of Ireland). Pregnancy and the use of medications known to affect weight also precluded participation.

Study Procedures

Health professionals working within the obesity service informed eligible participants about the study and provided a patient information leaflet to parent(s)/legal guardian(s) and their child. After 3 to 7 days, a researcher contacted the parent/guardian by phone to answer questions and check whether they wished to participate. Once written informed consent and child assent were received, a study appointment was offered for baseline assessment. A single dietetic researcher (SB) implemented the study protocol, coordinating, and providing communication. All participants met a researcher for scheduled face-to-face and phone communication and were also invited to call or email the researcher outside scheduled reviews. Study contacts at each time point (T), modality, and actions/measures for each contact are detailed later. At baseline, the researcher guided participants through a practice meal of their choice (brought to the hospital by participants) using Mandolean (Multimedia Appendix 1 shows a more detailed protocol). When participants completed a meal using Mandolean, data were available to the research team via a dedicated web-based clinical portal (Mandobase).

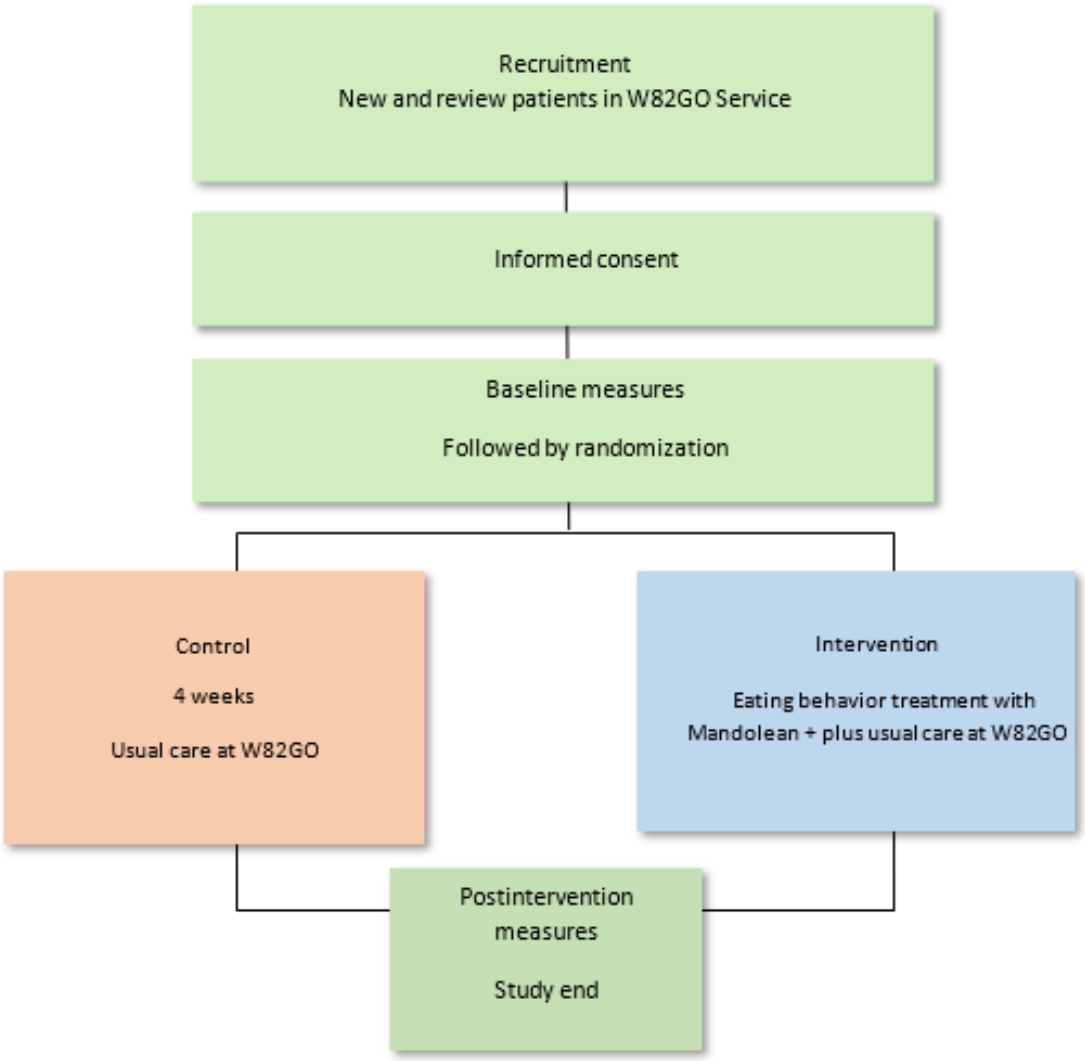
At present, the Big Data Against Childhood Obesity (BigO) project is testing the myBigO app and clinical portal [28]. The app aims to gather behavioral data alongside measures of environmental conditions (eg, urban built environment, infrastructure for physical activity, food marketing) among young people in general and an age-matched clinical cohort with obesity. Aggregation mechanisms are being developed to correlate population behaviors with environmental characteristics for the purpose of highlighting priority public health interventions [29].

All participants registered with the myBigO app and were set up with their smartwatch at baseline and postintervention. The default method of data synchronizing between the smartwatch and myBigO on the smartphone was a wireless internet connection (to avoid potential expense by using participants' personal mobile data). To ensure that the BigO system received regular accelerometer data, participants and their parent(s) were shown how to check the Bluetooth and internet connection between the smartwatch and phone, and they were asked to repeat this process every evening. For consistency, standard verbal, practical, and written instructions to take home were provided to participants and their caregivers (Multimedia Appendix 2). When participants completed a meal using Mandolean and wore the smartwatch, the data were available to the research team via web-based clinical portals (Mandobase and BigO clinical portal; see Multimedia Appendix 3 for detail of baseline and postintervention measures).

Usual Clinical Care

The W82GO Child and Adolescent Weight Management Service is a multidisciplinary obesity service that delivers efficacious obesity interventions [30]. Children and adolescents aged ≤16 years with a BMI >98th percentile are referred to the service by hospital physicians based at Children's Health Ireland at Temple Street. On referral, children and their caregivers are invited to a multidisciplinary clinic and undergo assessment by a pediatric dietitian, a pediatric physiotherapist, and a pediatric psychologist. On the basis of the needs of the child and family, a treatment plan is developed, and patients are offered either group-based treatment or treatment delivered in a 1:1, more traditional outpatient setting. Treatment is family based and was developed using contemporary scientific evidence [30]. Participants allocated to the usual care arm completed baseline testing, followed by 4 weeks of usual clinical care and subsequent retesting (Figure 1).

Figure 1. Flowchart of study protocol for a mobile health randomized feasibility study for an eating behavior intervention with children (aged 9.0-16.9 years) in the treatment of obesity in a tertiary health care setting. Baseline and postintervention measures include anthropometry, questionnaires (Child Behavior Checklist, Pediatric Quality of Life, Piers-Harris, Dutch Eating Behavior Questionnaire, and, at study end, evaluation questionnaires and System Usability Scale [31]), rate of eating using Mandolean, and physical activity levels with the smartwatch and myBigO app.



Intervention

Mandolean was developed by the Section of Applied Neuroendocrinology, Karolinska Institute, and Mando Group AB, Stockholm, Sweden. It consists of a plate scale that is wirelessly connected to a smartphone app with 2 main functions: (1) measures the rate of eating and (2) provides the user with visual feedback on slowing the rate of eating. The intervention arm involved usual care with additional training to reduce the rate of eating using Mandolean. Following randomization, the participants assigned to the intervention group received additional instruction on using the training functions of Mandolean for at least one meal per day (lunch, dinner, or both) over 4 weeks (minimum planned dose exposure: 28 meals). Using Mandolean training functions, the patient learns to adopt a typical pattern of eating and satiety by following the displayed *ideal* rate of eating, which they aim to match. The clinician used baseline data (usual portions sizes and rate of eating) to guide a *training meal* program for the user. The training aims to teach the patient to eat 280 to 350 g in 13 to 15 min and to perceive a level of satiety of *moderately full* by the end of the meal. A full description of the Mandolean training procedures is included in Multimedia Appendix 1. The use of Mandolean in this study integrated a number of behavior change components, categorized according to the behavior change technique taxonomy by Michie et al [32], including goals and planning, feedback and monitoring, social support, shaping knowledge, comparison of behavior, repetition and substitutions, and antecedents (Multimedia Appendix 4 gives the subtechniques used).

Sample Size

A sample size of 20 participants was the target, which was considered sufficient to evaluate the process measures in this randomized feasibility study and is in line with similar studies [10,13].

Randomization

Using a web-based randomization service (Sealed Envelope), participants were randomized to Mandolean eating behavior training intervention or control group (Figure 1) by 1 researcher (SB). Age (9-11.9 years and 12-16.0 years) and gender stratifications were applied, and participants' parents were

informed of their treatment group by phone after baseline measurements. Participants, therefore, were aware of their study allocation, as the intervention required exposure to new eating behavior training. At this point, further study review appointments were planned which are outlined in the results section.

Outcomes

The Consolidated Standards of Reporting Trials (CONSORT) extension for randomized feasibility studies was used to guide transparency and quality in reporting study measures (Multimedia Appendix 5) [33,34]. Trial process–related outcomes measured in this study addressed (1) feasibility (recruitment process, rates of recruitment to the study, and rates of retention and attrition of the study arms), (2) fidelity (adherence to randomization protocol, appointment attendance [number, modality, and duration of study appointments], dose delivered [study tasks planned and completed], dose received [training exposure logged in Mandolean clinical portal], and adherence to intervention procedures), and (3) acceptability (participant engagement with the Mandolean app [number of training meals completed], participants' engagement with the BigO physical activity monitoring app [volume of data collected], and scores from the system usability scale [SUS] questionnaires [31]).

Data Analysis

Statistical methods for quantitative measures included descriptive frequencies and standard *t* test comparisons between intervention and control groups and completer and noncompleter groups for baseline age, anthropometry, and questionnaire scores. Qualitative feedback from questionnaires and discussions with users was analyzed for content and categories created for the purpose of presenting key acceptability issues, challenges, and facilitators for users and health care professionals.

Ethics

The research protocol was reviewed by the research ethics committee at Temple St. Children's University Hospital and approved on August 08, 2018 (number 18.013). A pseudonymized patient identification coding system was incorporated and stored in an encrypted file at the clinical site, so that no personal patient information was shared or processed via mHealth apps. Data collected on the apps were locally transformed on the participants' mobile phones, and the transformed data, not containing identifiable information, were transmitted and stored to the respective clinical portals (for Mandolean and BigO).

Results

Participants

Participants were recruited between May 2018 and February 2019. Table 1 describes participants' characteristics and baseline assessment measures for the intervention and control groups. A total of 63% (5/8) of participants in the intervention group and 42% (5/12) of participants in the control group were categorized as being below average socioeconomic status (Multimedia Appendix 6 gives further detail). No significant differences between the intervention and control groups were noted for mean age, BMI, or BMI SD score (SDS). Differences in mean total score, externalizing behavior total score, and internalizing behavior total score for the parent-reported Child Behavior Checklist (CBCL) were observed (Table 1). A significantly higher score in mean baseline total T score for CBCL was observed between those who completed the study and those who did not, indicating more behavioral issues among those who withdrew. Owing to the high attrition rate, completer and noncompleter groups are also presented in Table 1.

Table 1. Characteristics of participants in a randomized feasibility study for a mobile health eating behavior training intervention at baseline and follow-up.

Participants' characteristics and baseline measures	Intervention (n=8)	Control (n=12)	Completed the study (n=12)	Did not complete the study (n=8)
Sex				
Male, n (%)	3 (37)	6 (50)	4 (33)	5 (63)
Female, n (%)	5 (63)	6 (50)	8 (67)	3 (37)
Age (years), mean (SD)	13.1 (2.3)	13.5 (2.3)	13.3 (2.7)	13.5 (1.5)
Baseline BMI (kg/m^2), mean (SD)	31.6 (3.9)	33.2 (5.9)	32.16 (5.7)	33.1 (4.6)
Baseline BMI SD score, mean (SD)	3.02 (0.27)	3.04 (0.60)	3.00 (0.56)	3.09 (0.37)
Stage of usual care (weeks), mean (SD)	40.1 (46.2)	17.7 (16.8)	26.9 (33.1)	26.3 (34.6)
Baseline physical and psychosocial health self-report				
Child or adolescent self-report questionnaire score, mean (SD)				
Physical health PedsQL[a] [35]	74.6 (17.1)	69.1 (15.1)	70.5 (17.6)	72.3 (13.3)
Psychosocial health PedsQL	49.0 (24.87)	64.7 (20.5)	63.3 (19.5)	51.2 (27.5)
DEBQ[b] external eating [36]	1.58 (0.72)	2.00 (0.62)	2.00 (0.56)	1.58 (0.79)
DEBQ emotional eating score	1.29 (0.76)	1.50 (0.53)	1.63 (0.61)	1.09 (0.52)
DEBQ restrained eating score	1.91 (0.81)	2.03 (0.30)	2.08 (0.30)	1.84 (0.79)
Parent self-report questionnaire score, mean (SD)				
CBCL[c,d,e] total T score [37]	71.7 (3.1)	57.6 (6.6)	59.0 (9.3)	67.9 (5.6)
CBCL externalizing behavior T score[f]	67.8 (4.7)	57.2 (7.8)	58.2 (7.5)	65.0 (8.7)
CBCL internalizing behavior T score[g]	64.3 (6.2)	53.8 (8.5)	56.1 (9.5)	60.3 (9.2)

[a]PedsQL: pediatric quality of life.

[b]DEBQ: Dutch eating behavior questionnaire.

[c]CBCL: Child Behavior Checklist.

[d]CBCL total T score. Intervention group versus control group: *t* test for equality of means, equal variances assumed $P<.001$.

[e]CBCL total T score. Completed the study group versus did not complete group: *t* test for equality of means, equal variances assumed $P<.04$.

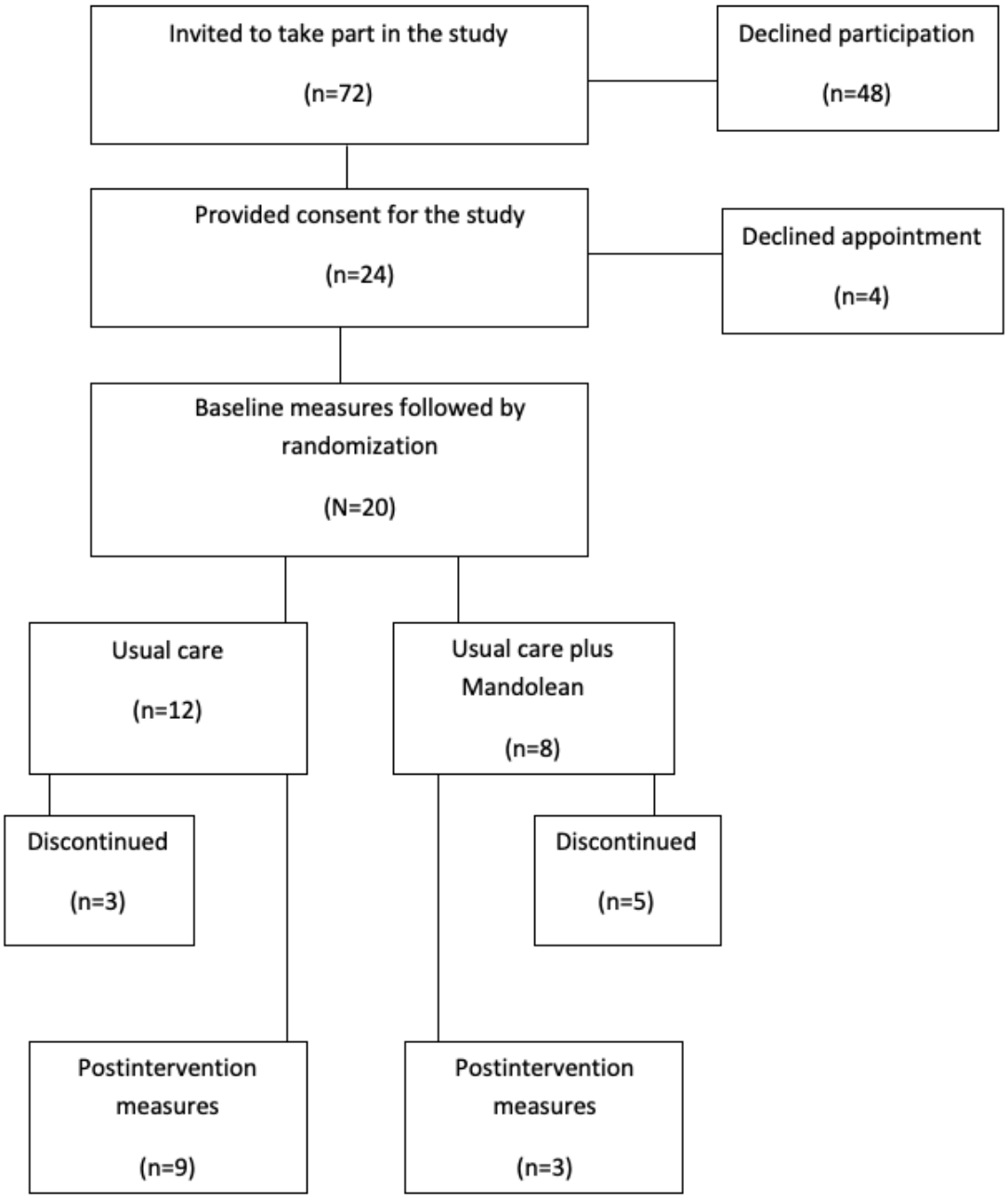
[f]CBCL externalizing T score. Intervention group versus control group: *t* test for equality of means, equal variances assumed $P=.02$.

[g]CBCL internalizing behaviour T score. Intervention group versus control group: *t* test for equality of means, equal variances assumed $P=.01$.

Feasibility: Rate of Recruitment

Children and adolescents were recruited between June 2018 and January 2019. One strategy of recruitment, which involved offering patients and their families study recruitment packs at their first multidisciplinary assessment appointment for the obesity service, was discontinued during the feasibility study. Families reported mixing up the recruitment pack with usual care information received on the same day and had not read the study information by the time the researcher contacted them some days later by phone. Instead, a researcher or clinician provided a 5-min information session and study information pack to parents and their children when they were established within the service and followed up with a phone call 3 to 7 days later. In total, 72 eligible parent-child dyads took recruitment packs for the study, and 33% (24/72) signed informed consent to participate. Following this 28% (20/72) attended for the first study appointment (Figure 2 shows participant flow through the study).

Figure 2. Consolidated standards of reporting trials diagram for pilot randomized trials.



Feasibility: Retention and Attrition

A total of 25% (3/8) of participants in the control arm and 63% (5/12) of participants in the intervention arm withdrew early. Characteristics, participation levels, and feedback from children and adolescents who dropped out of the feasibility study before completion are presented in more detail in Multimedia Appendix 7.

Fidelity: Adherence to the Randomization Protocol

The web-based randomization process resulted in males being underrepresented in the intervention group.

Fidelity: Appointment Attendance

Intervention participants attended study appointments in addition to usual care (Table 2). There was good adherence to planned face-to-face appointments at time point 1 (T1); however, there was mixed adherence thereafter, with the exception of phone reviews provided by the researcher. The time allocated to the study appointments was appropriate. Illness, school and family commitments, competing appointments at the hospital, and living a long distance from hospital (as perceived by families) were barriers to attending appointments. Final reviews were completed by phone for 3 families who lived at a distance to the hospital to minimize absenteeism from school/work.

Table 2. Fidelity with planned actions and measures at each time point.

Actions and measure	Time point					
	T1 (week 0)	T1a (T1+1 week)	T2 (T1+2 weeks)	T2a (T1+4 weeks)	T3 (T1+6 weeks)	T4 (T1+8 weeks)
Planned mode	In person	Phone	Phone or in person	Phone	In person	In person
Mode adherence[a], n (%)	20 (100)	16 (80)	17 (85)	12 (60)	8 (40)	9 (45)
Time allocated, min	60	15	30	15	30	30
Time allocated adherence[a], n (% of those who attended)	17 (85)	16 (100)	16 (94)	12 (100)	7 (88)	9 (100)
Smartwatch and myBigO setup[a], n (%)	12 (60)	5 (25)	1 (5)	N/A[b]	6 (30)	N/A
Mandolean app installation and baseline meal demonstration[a], n (%)	16 (80)	N/A	N/A	N/A	N/A	N/A
Mandolean intervention training meal demonstration (n=8), n (%)	N/A	N/A	4 (50)	N/A	N/A	N/A
Intervention verbal instructions (n=8), n (%)	N/A	N/A	5 (63)	N/A	N/A	N/A
Intervention standard instructions (n=8), n (%)	N/A	N/A	5 (63)	N/A	N/A	N/A
Anthropometry[a], n (%)	15 (75)	N/A	N/A	N/A	N/A	9 (45)
Sociodemographic data[a], n (%)	20 (100)	N/A	N/A	N/A	N/A	N/A
Questionnaires, n (%)						
Dutch eating behavior	19 (95)	N/A	N/A	N/A	N/A	N/A
Piers-Harris	19 (95)	N/A	N/A	N/A	N/A	N/A
Child behavior checklist parental questionnaire	19 (95)	N/A	N/A	N/A	N/A	N/A
Pediatric quality of life	19 (95)	N/A	N/A	N/A	N/A	N/A
Mobile health app usability questionnaires, n (%)	N/A	N/A	N/A	N/A	N/A	11 (55)

[a]Adherence to planned protocol expressed as % of 20 participants recruited to study at baseline unless otherwise indicated.

[b]N/A: not applicable at given time-point.

Fidelity: Dose Delivered and Adherence to Intervention Procedures

Table 2 shows adherence to intervention protocols at each time point. The reasons for an incomplete smartwatch setup at T1 included an incompatible smartphone (n=5), the parent could not remember personal account password to complete syncing with the app (n=2), and insufficient time (n=1). In total, 5 patients took written instructions for smartphone installation at T1 and completed the process at home on a compatible smartphone belonging to another parent or sibling. Two patients did not complete the Mandolean setup at T1 because of patient time constraints, and this was planned for time point 2 (T2), which was completed with 1 participant, and the other did not attend T2 or subsequent appointments. Two patients did not complete questionnaires at baseline, because of lack of time, and were asked to return by post or at the next appointment, of which 1 set was returned.

In terms of intervention implementation, of 8 patients randomized to Mandolean treatment, 4 received demonstration and instructions by the researcher in person; 1 patient was provided with instructions over the phone and written instructions by post, and 3 patients did not attend T2 to commence training (Table 2).

Fidelity: Dose Received

Smartwatch and myBigO Apps

An early version of the myBigO app used here (which accessed accelerometer data via smartwatches) was compatible with Android operating systems 6.0 and above. Of the 18 children set up with smartwatches at baseline, 11 were connected to parents' phones, 1 was connected with a sister's phone, and 6 children and adolescents used their own phones. Available data from the BigO system indicated that of the 18 smartwatch setups with myBigO, 50% (9/18) of participants contributed some data. Of those who contributed data, the range was highly variable from 0.3 to 9.2 days (mean 1.6, SD 2.9 days; median 0.2; IQR 2.0). Two participants did not wear the smartwatch at all after the baseline setup (one because of illness and the other was self-conscious about wearing at school) and subsequently dropped out. Two parents deleted the myBigO app at some point during the study (one because of lack of space on the phone and the other because the father thought the child was no longer using the watch/app). One child did not live with the parent who had myBigO and the smartwatch setup with a compatible smartphone and, as a result, did not synchronize regularly. Attrition (n=8), poor attendance at time point 3 (n=2), technical challenges resynchronizing watch to phone (n=3), and strap

breaking (n=1) were reasons for low usage postintervention. In addition, self-reported days wearing the watch among users at the postintervention stage was highly variable. Two patients reported sensory issues and disliked wearing the watch. A short battery life and forgetting to charge or wear were commonly reported by patients for not wearing as advised.

Mandolean

All participants completed at least one Mandolean baseline meal that measured their rate of eating. Three participants completed 1 baseline meal with the researcher at the hospital canteen and did not complete any at home. The remaining participants successfully completed some baseline measurement meals at home, with a range of usage (2-19 meals). Participants' engagement with Mandolean and exposure to the planned intervention are presented in Table 3.

Table 3. Summary statistics for the use of Mandolean device at baseline (all participants), intervention phase (intervention only), and postintervention (all participants).

Variable	Baseline meal frequency	Training meals frequency	Postintervention meals frequency
Number children/teenagers, n (%)	20 (100)	8 (40)	12 (60)
Mean (SD)	5.7 (5.2)	5.4 (9.2)	4.6 (5.2)
Median	5.0	1.0	3.5
25th centiles	1.3	0.0	0.5
75th centiles	9.0	9.3	7.8
Mean meals as percentage of planned meals	56	19	46
Unsuccessful meal attempts[a], mean (SD)	1.2 (2.1)	0.5 (0.9)	1.0 (0.8)

[a]Unsuccessful meal attempts were meals initiated and therefore registered on the clinical portal system, but ultimately did not record the rate of eating successfully.

Of the 8 participants allocated to the intervention training, 5 received training instructions and 3 did not engage with the training component. One participant subsequently dropped out of the study before training commenced; therefore, the final number of participants exposed to intervention training meals was 50% (4/8 of those randomized to training). Exposure at an individual level represented 7% (2/28 planned meals; male aged 12.3 years), 14% (4/28 planned meals; female aged 9.5 years), 39% (11/28 planned meals; female aged 15.2 years), and 93% (26/28 planned meals; female aged 11.6 years) of planned intervention.

Acceptability

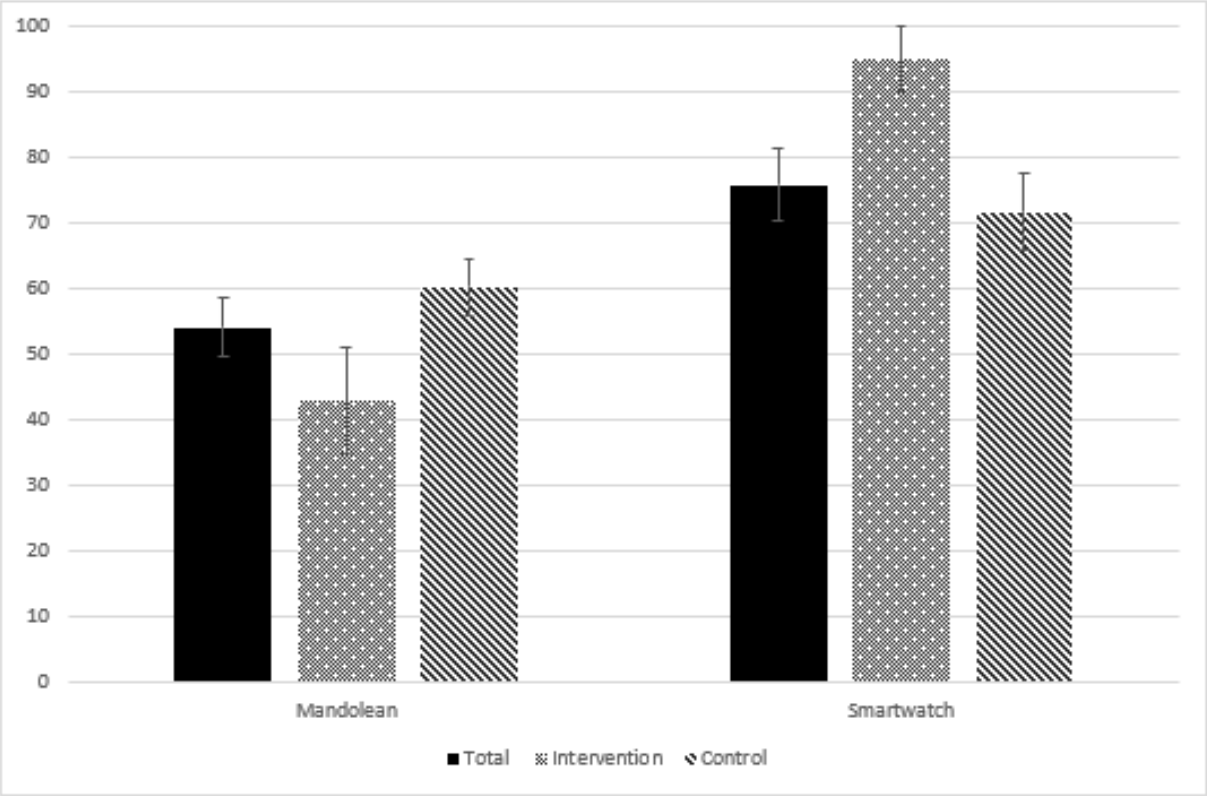
Participants' Engagement With Mandolean and the BigO Physical Activity Monitoring App

The findings relating to the dose received presented earlier indicated poor acceptability as measured through active engagement among users.

Self-Report Acceptability Measure: System Usability Scale

The mean SUS score results are illustrated in Figure 3. A score of 68 or greater is considered acceptable when assessing the user experience of technology [31]. Mandolean did not achieve a mean greater than 68 for the total group or within any subgroup, and the smartwatch achieved a mean of 68 or greater for all groups. A more detailed table of results is presented in Multimedia Appendix 8.

Figure 3. Bar chart showing participants' postintervention System Usability Scale scores for Mandolean and smartwatch for the total (n=11), intervention (n=4), and control (n=7) groups.



Acceptability: Qualitative Feedback From Participants and Their Parents

Qualitative feedback from participants and their parents was categorized for each piece of technology. The main barriers to using Mandolean were (1) connectivity issues; (2) difficult, awkward, or time consuming to set up, which interfered with family meal times; (3) incompatible with family routine (ie, no regular family meal times, summer holidays, or parental shift work); and (4) forgetting to use. A small number of children and parents reported becoming more aware of their speed of eating as a result of using Mandolean. Most participants enjoyed wearing the watch, liked the timekeeping function, and self-monitored their daily activity levels. The main drawbacks noted were (1) a short battery life, (2) sensory issues and finding the watch uncomfortable, and (3) feeling self-conscious at school. Participants' quotes and more details are presented in Multimedia Appendix 8.

Discussion

Principal Findings

We conducted a study to determine the feasibility and acceptability of a proposed intervention using 2 mHealth apps among children and adolescents being treated for obesity. The study process was documented thoroughly, and based on the observed results, we concluded the need for further technical usability testing in this population. A slow recruitment rate, high attrition rate, and low fidelity with planned interventions were the key outcomes informing feasibility. Greater psychosocial issues among the intervention and noncompleter groups, observed in the baseline behavioral questionnaire (CBCL), were also noteworthy in this cohort. Although we cannot imply causality about the effect on study engagement or attrition, this finding provides important contextual background about individual and group characteristics, which may have contributed to suboptimal usage of mHealth tools at home.

In terms of delivering study components, we found realistic time points and modes of delivery. Poor fidelity with participants' exposure to intervention components, in particular, the low number of participants randomized to the intervention who attended the necessary training, and the low level of engagement with training meals on Mandolean were the primary barriers to intervention implementation. Adherence with the smartwatch set up at baseline was for the majority achieved, whereas fidelity postintervention was problematic because of attrition and nonattendance for reviews.

We considered the high attrition rate as a signal of poor acceptability of the intervention, particularly when a greater number of intervention arm participants opted to leave the study before completion. The poor rating of Mandolean on the SUS scores provided further evidence. In contrast, positive feedback about using and understanding the smartwatch and the myBigO app was received. Despite the acceptability of the smartwatch, the wide range of exposure levels is suggestive of underlying barriers that need to be understood if we are to maximize

adherence to mHealth adjuncts to clinical care. Sekhon et al [38] proposed acceptability as multifaceted to include the following 7 component constructs: affective attitude, burden, perceived effectiveness, ethicality, intervention coherence, opportunity costs, and self-efficacy. Applying these constructs to our SUS and qualitative findings, we suggest that technical difficulties, perceived awkwardness, and time cost associated with using Mandolean contributed to a negative attitude among participants and their parent(s). These, in turn, possibly contributed to feelings of high burden and low confidence in completing the required study tasks. There is also the possibility that the perceived burden associated with Mandolean affected overall study task adherence, including the smartwatch and myBigO.

Comparison With Prior Work

Our study with Mandolean differs from previously published work in 2 technical ways: (1) we used a new mobile tool, with a smartphone app interface for Mandolean, whereas others plugged it into a computer [15,16], and (2) we had access to objective engagement data from a clinical portal facility that was not available for an earlier clinical study [16]. Compared with a community-based feasibility study that reported process measures, we had an older cohort (9-16.0 years vs 6-11.0 years), we recruited in the hospital setting, and we also incorporated an additional mHealth tool to measure physical activity levels [15]. Intervention exposure in a more recent study using smartphones mirrored our findings that engagement with Mandolean meals at home among adolescents varied considerably within the intervention group from 5 to 80 meals (median 28 meals), out of a planned 1 meal per day for 6 months [13]. A study of younger children found that just 19% achieved the minimum expected usage of 5 meals per week with Mandolean [15]. These studies and others [39] reported engagement issues among children and adolescents when mHealth apps are considered burdensome.

Individual factors contributing to poor adherence with wearing the smartwatch that we found, including early attrition, sensory issues, forgetting to charge, forgetting to wear, and feeling self-conscious, are similar to other studies using wearable devices with young people. Although rigorous research with smartwatches as activity trackers is in its infancy in comparison with traditional accelerometers [40], some of the same adherence issues may apply. For example, Jago et al [41] reported considerable variability in adhering to a 7-day accelerometer among children, finding that parents would forget to put on the accelerometer, some children found it uncomfortable, and some were self-conscious about wearing it at school; however, others were excited or interested in wearing it. We found similar barriers among our older group and facilitators among younger participants. Research with a commercial fitness tracker among adolescents with obesity reported a discontinuation rate of 68% before the end of the study that was linked to barriers to physical activity not being addressed by a tracker, seasonality, feelings of activity incompetence, and gradual withdrawal of parental and clinical support [42]. Our finding that high acceptability of a smartwatch does not translate to adherence is also reported by Phan et al [42]. The physical activity self-monitoring tool available on the smartwatch screen in our study was cited by a number of children and parents as a benefit because they were trying to increase their activity levels. Formal integration of recognized and important behavioral change techniques for *self-monitoring* and *goal setting* [32] could be maximized for future studies with research-led apps such as myBigO and smartwatches. Retention in clinical research, however, is challenging, and we note that a two-fold study initiation rate may be required to achieve meaningful physical activity data for clinicians treating children in treatment for obesity.

Although withdrawal of 63.5% of participants in the intervention arm here is considerably higher than that in previous studies using Mandolean [13,15,16], the psychosocial profile of participants has not been described by others. Therefore, we cannot assume that participants were similar across published trials. The most comparable health care setting and age group to our own is the study by Ford et al [16]. They demonstrated high compliance with the intervention, particularly among the Mandolean arm, with 83% of participants attending all study appointments. High-intensity contact (14 appointments in 12 months) was reported as a facilitator of retention [16]. Meta-analyses in the wider electronic health (eHealth) literature show attrition rates of 12% to 29% [43]. Attrition can depend on a range of factors known to be influential in clinical care (eg, school absenteeism, dissatisfaction with care components, demographic factors) [44-46], in addition to factors influencing dropout from eHealth-specific interventions (eg, registration requirements and male gender) [47-49]. One unique aspect of this study was the combination of 2 mHealth tools for measuring both dietary behavior and physical activity. Our attrition rate, therefore, may not be informative for all mHealth interventions with children and adolescents but certainly contributes insight where high participant engagement is expected among groups with complex health and psychological needs.

The predominant technical issues identified were difficulty connecting scale to phone, Mandolean failing to recognize the plate when weighing out food, and loss of connection midmeal, which others also report as a barrier to compliance at home [13,15]. User experiences such as *annoying*, *cumbersome*, and *time consuming* reported by Hamilton-Shield et al [15] reflected the feedback from our participants, despite using an updated mobile app version of Mandolean. We also found a number of other practical challenges similar to Hinton et al [13], including child impatience during setup, interruption of the flow of family mealtimes, difficulty with the portion size limit for training meals, and forgetting to use. A minority of parents in this study reported more positively that children adapted quickly to the routine of using Mandolean at mealtimes, particularly younger participants. In contrast, parents whose teenagers withdrew from the study early would have preferred them to complete the study. In a community study, by 4 to 10 weeks, some children lost interest, and a few parents were tired of using the Mandolean, particularly setting up, dishing food onto a plate, and weighing food [15]. Although their trial was a longer duration, we found that these reactions were evident earlier (from 2 weeks onward) and, in some cases, contributed to early withdrawal from the study.

Higher behavioral problems among the intervention group at baseline appeared to be a random outcome of allocation. The

subsequent high attrition within the group suggests that psychosocial issues in combination with the intervention burden discussed earlier may have played a role in early attrition. The reasons for attrition among children with obesity vary in the intervention literature. Some interventions report no differences between completers and noncompleters for behavior measures using CBCL [50,51]. Others have shown that the social competence of children aged 8 to 14 years who are overweight or obese, defined using the CBCL, was one of a number of predictors (including lower baseline weight and Caucasian parents) of BMI SDS reduction following a 12-month intervention [45]. Behavioral problems in children from disadvantaged areas have been linked to adherence in other conditions such as type 1 diabetes [51]. Behavioral problems are associated with a high risk of overweight and obesity among children, independent of other risk factors such as parental obesity, education, poverty, and race [52,53]; therefore, we expected to find behavioral issues among a substantial number of our participants. Participants here may not be representative of the general population of children and adolescents with obesity for a number of potential reasons, including obesity severity that prompted a referral to the specialist pediatric service, varying motivation depending on the stage of treatment, and interest in joining research studies based on health status. Although patients with known behavioral disorders were ineligible for participation in this study, a self-report tool at baseline assessing psychosocial and behavioral issues detected underlying behavioral issues that may have affected children's ability to partake fully in research tasks (and hence treatment tasks). On the basis of this experience, we recommend multidisciplinary baseline assessments to include behavioral measures for similar adjunct interventions with children receiving treatment for obesity.

This is the first published study in which myBigO with a smartwatch was incorporated into an intervention in a clinical setting. The process outcomes provide some lessons for future research and practice. To fully describe the feasibility phase process, we included all participants who attended a baseline appointment. The slow uptake and early attrition observed indicates that children in treatment for obesity may require greater choice or flexibility in how they contribute data and benefit personally from participation. Different approaches to recruitment and deployment need to be explored with different subgroups, including those with psychosocial or behavioral issues and different age groups (children vs early adolescents vs midadolescents), as they have varied motivations. Greater self-monitoring functionalities and reminders to charge and wear smartwatches may also improve adherence based on barriers reported. Given our challenges with a postintervention accelerometer period, one unbroken smartwatch exposure may be more realistic and preferable among children in treatment for obesity, reducing the need for extra face-to-face review appointments. However, clinicians wishing to observe behavior pre- and postintervention may wish to examine the feasibility of a longer intervention period in which participants would have time to implement and monitor goals based on baseline measures.

Limitations

This is a feasibility study with a small sample size and short intervention period. Recruitment was limited to children and adolescents attending 1 specialist pediatric obesity program, which may have limited recruitment rate and introduced population bias in terms of obesity severity and associated complications and participant motivation to participate in the research. The recruitment rate was also delayed at the study outset by the implementation of the General Data Protection Regulation in the European Union and Irish interpretation of the regulation for the purposes of health research. Poor adherence to treatment was evident at early stages; therefore, a longer intervention with Mandolean was unlikely to add additional benefit or knowledge in this group. The intervention group received additional training on Mandolean that the control group was not exposed to which heightened awareness about behaviors of interest and combined with task burden could well introduce unknown biases. A more structured and validated technical usability study may be of benefit to further understand the challenges children and their families face in using the mobile Mandolean system. Although the SUS has been used previously to evaluate adult user experience with wearable devices [54], the survey has not, to our knowledge, been validated in a pediatric population. However, it has advantages for use with children, as it is short and uncomplicated. It is suitable for use with small samples, provides space for comments, and a final score can be interpreted with an established reference standard. The smartwatch with myBigO was in the first stage of feasibility testing in this cohort, and, as such, the technology will continue to be developed based on users' needs. The attrition and engagement measures should be interpreted in the context of a feasibility study with 2 mHealth apps aimed at children with obesity and, as such, not indicative of the performance of either tool used in isolation.

Conclusions

Our study explored the process outcomes of using mHealth tools in an obesity treatment study and highlighted challenges and opportunities related to feasibility, fidelity, and acceptability. By transparently reporting feasibility using the CONSORT extension guidance [33], we reported potential challenges for mHealth interventions among children with obesity. mHealth interventions are rapidly progressing; however, we need to be cautious in terms of efficacy, burden to participants, and our responsibility to identify vulnerable subgroups at baseline. Challenges include task burden and adherence to complexity in mHealth systems recommended for use at home, particularly among families experiencing behavioral issues. The opportunities noted here include high acceptability of a self-monitoring physical activity system where data are shared between patients and healthcare workers. Children with obesity attending treatment have complex needs and, given health service limitations, would benefit from adjuncts to traditional treatment that can be implemented outside of clinical settings. Additional technical and usability studies are recommended to improve our understanding of adherence to treatment.

Acknowledgments

The research team wishes to acknowledge the research collaborators Anastasios Delopoulos, Eirini Lekka, Ioannis Ioakeimidis, Christos Diou, Isabel Perez, Daniel Ferri, and all the BigO Project Consortium Members; Cecilia Bergh, Per Söderston, and the team at Mando Group AB, Stockholm, Sweden; the W82GO clinical and administrative team, and the physiotherapy department at Children's Health Ireland, Temple Street, Dublin; and participating patients and their parents. The work leading to these results has received funding from the European Community's Health, Demographic Change, and Well-being program under grant agreement number 727688. The project in receipt of this grant is 'The BigO Project: Big Data Against Childhood Obesity' [28].

Conflicts of Interest

The researchers collaborated with Mando Group AB, Stockholm, Sweden, who provided Mandolean devices for the purpose of the study.

Multimedia Appendix 1
Protocol and process for training participants to use Mandolean.
[DOCX File , 20 KB - mhealth_v8i7e16925_app1.docx]

Multimedia Appendix 2
Written standard instructions provided to the participants.
[DOCX File , 536 KB - mhealth_v8i7e16925_app2.docx]

Multimedia Appendix 3
Description of baseline measurements.
[DOCX File , 18 KB - mhealth_v8i7e16925_app3.docx]

Multimedia Appendix 4
Behaviour Change Techniques incorporated into the intervention.
[DOCX File , 19 KB - mhealth_v8i7e16925_app4.docx]

Multimedia Appendix 5
Consolidated standards of reporting trials of electronic and mobile health apps and online telehealth checklist report.
[PDF File (Adobe PDF File), 1657 KB - mhealth_v8i7e16925_app5.pdf]

Multimedia Appendix 6
Bar chart showing participants' socioeconomic position.
[DOCX File , 19 KB - mhealth_v8i7e16925_app6.docx]

Multimedia Appendix 7
Characteristics and feedback from children and teenagers who dropped out before completion.
[DOCX File , 22 KB - mhealth_v8i7e16925_app7.docx]

Multimedia Appendix 8
System usability scores for Mandolean and a smartwatch, including qualitative comments.
[DOCX File , 24 KB - mhealth_v8i7e16925_app8.docx]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
2. Umer A, Kelley GA, Cottrell LE, Giacobbi P, Innes KE, Lilly CL. Childhood obesity and adult cardiovascular disease risk factors: a systematic review with meta-analysis. BMC Public Health 2017 Aug 29;17(1):683 [FREE Full text] [doi: 10.1186/s12889-017-4691-z] [Medline: 28851330]
3. Reinehr T, Lass N, Toschke C, Rothermel J, Lanzinger S, Holl RW. Which amount of BMI-SDS reduction is necessary to improve cardiovascular risk factors in overweight children? J Clin Endocrinol Metab 2016 Aug;101(8):3171-3179. [doi: 10.1210/jc.2016-1885] [Medline: 27285295]

4. Kolsgaard ML, Joner G, Brunborg C, Anderssen SA, Tonstad S, Andersen LF. Reduction in BMI z-score and improvement in cardiometabolic risk factors in obese children and adolescents. The Oslo adiposity intervention study-a hospital/public health nurse combined treatment. BMC Pediatr 2011 May 27;11(36):47 [FREE Full text] [doi: 10.1186/1471-2431-11-47] [Medline: 21619652]
5. Savoye M, Nowicka P, Shaw M, Yu S, Dziura J, Chavent G, et al. Long-term results of an obesity program in an ethnically diverse pediatric population. Pediatrics 2011 Mar;127(3):402-410 [FREE Full text] [doi: 10.1542/peds.2010-0697] [Medline: 21300674]
6. Mead E, Brown T, Rees K, Azevedo LB, Whittaker V, Jones D, et al. Diet, physical activity and behavioural interventions for the treatment of overweight or obese children from the age of 6 to 11 years. Cochrane Database Syst Rev 2017 Jun 22;6:CD012651 [FREE Full text] [doi: 10.1002/14651858.CD012651] [Medline: 28639319]
7. Al-Khudairy L, Loveman E, Colquitt JL, Mead E, Johnson RE, Fraser H, et al. Diet, physical activity and behavioural interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. Cochrane Database Syst Rev 2017 Jun 22;6:CD012691 [FREE Full text] [doi: 10.1002/14651858.CD012691] [Medline: 28639320]
8. Ohkuma T, Hirakawa Y, Nakamura U, Kiyohara Y, Kitazono T, Ninomiya T. Association between eating rate and obesity: a systematic review and meta-analysis. Int J Obes (Lond) 2015 Nov;39(11):1589-1596. [doi: 10.1038/ijo.2015.96] [Medline: 26100137]
9. Robinson E, Almiron-Roig E, Rutters F, de Graaf C, Forde CG, Smith C, et al. A systematic review and meta-analysis examining the effect of eating rate on energy intake and hunger. Am J Clin Nutr 2014 Jul;100(1):123-151. [doi: 10.3945/ajcn.113.081745] [Medline: 24847856]
10. Hawton K, Ferriday D, Rogers P, Toner P, Brooks J, Holly J, et al. Slow down: behavioural and physiological effects of reducing eating rate. Nutrients 2018 Dec 27;11(1):191 [FREE Full text] [doi: 10.3390/nu11010050] [Medline: 30591684]
11. Galhardo J, Hunt LP, Lightman SL, Sabin MA, Bergh C, Sodersten P, et al. Normalizing eating behavior reduces body weight and improves gastrointestinal hormonal secretion in obese adolescents. J Clin Endocrinol Metab 2012 Feb;97(2):E193-E201. [doi: 10.1210/jc.2011-1999] [Medline: 22162463]
12. Faith MS, Diewald LK, Crabbe S, Burgess B, Berkowitz RI. Reduced eating pace (RePace) behavioral intervention for children prone to or with obesity: does the turtle win the race? Obesity (Silver Spring) 2019 Jan;27(1):121-129 [FREE Full text] [doi: 10.1002/oby.22329] [Medline: 30515992]
13. Hinton EC, Birch LA, Barton J, Holly JM, Biernacka KM, Leary SD, et al. Using neuroimaging to investigate the impact of Mandolean training in young people with obesity: a pilot randomised controlled trial. BMC Pediatr 2018 Nov 22;18(1):366 [FREE Full text] [doi: 10.1186/s12887-018-1342-1] [Medline: 30466438]
14. Hermsen S, Frost JH, Robinson E, Higgs S, Mars M, Hermans RC. Evaluation of a smart fork to decelerate eating rate. J Acad Nutr Diet 2016 Jul;116(7):1066-1068. [doi: 10.1016/j.jand.2015.11.004] [Medline: 26785908]
15. Hamilton-Shield J, Goodred J, Powell L, Thorn J, Banks J, Hollinghurst S, et al. Changing eating behaviours to treat childhood obesity in the community using Mandolean: the community Mandolean randomised controlled trial (ComMando)-a pilot study. Health Technol Assess 2014 Jul;18(47):i-xxiii, 1 [FREE Full text] [doi: 10.3310/hta18470] [Medline: 25043221]
16. Ford AL, Bergh C, Södersten P, Sabin MA, Hollinghurst S, Hunt LP, et al. Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. Br Med J 2009 Jan 5;340:b5388. [doi: 10.1136/bmj.b5388] [Medline: 20051465]
17. Lyzwinski LN, Edirippulige S, Caffery L, Bambling M. Mindful eating mobile health apps: review and appraisal. JMIR Ment Health 2019 Aug 22;6(8):e12820 [FREE Full text] [doi: 10.2196/12820] [Medline: 31441431]
18. National Institute for Healthcare and Excellence. Weight Management: Lifestyle Services For Overweight or Obese Children and Young People. London, UK: National Institute for Healthcare and Excellence; 2013.
19. Cain KL, Sallis JF, Conway TL, van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health 2013 Mar;10(3):437-450 [FREE Full text] [doi: 10.1123/jpah.10.3.437] [Medline: 23620392]
20. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. Med Sci Sports Exerc 2004 Sep;36(9):1625-1631. [Medline: 15354047]
21. Tarp J, Child A, White T, Westgate K, Bugge A, Grøntved A, International Children's Accelerometry Database (ICAD) Collaborators. Physical activity intensity, bout-duration, and cardiometabolic risk markers in children and adolescents. Int J Obes (Lond) 2018 Sep;42(9):1639-1650 [FREE Full text] [doi: 10.1038/s41366-018-0152-8] [Medline: 30006582]
22. Tu AW, Watts AW, Chanoine JP, Panagiotopoulos C, Geller J, Brant R, et al. Does parental and adolescent participation in an e-health lifestyle modification intervention improves weight outcomes? BMC Public Health 2017 Apr 24;17(1):352 [FREE Full text] [doi: 10.1186/s12889-017-4220-0] [Medline: 28438202]
23. 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]
24. de Cock N, van Lippevelde W, Vangeel J, Notebaert M, Beullens K, Eggermont S, et al. Feasibility and impact study of a reward-based mobile application to improve adolescents' snacking habits. Public Health Nutr 2018 Aug;21(12):2329-2344. [doi: 10.1017/S1368980018000678] [Medline: 29607800]

25. O'Malley G, Dowdall G, Burls A, Perry IJ, Curran N. Exploring the usability of a mobile app for adolescent obesity management. JMIR Mhealth Uhealth 2014 Jun 27;2(2):e29 [FREE Full text] [doi: 10.2196/mhealth.3262] [Medline: 25098237]
26. Blatch-Jones AJ, Pek W, Kirkpatrick E, Ashton-Key M. Role of feasibility and pilot studies in randomised controlled trials: a cross-sectional study. BMJ Open 2018 Sep 25;8(9):e022233 [FREE Full text] [doi: 10.1136/bmjopen-2018-022233] [Medline: 30257847]
27. Haase T, Pratschke J. Pobal: Pobal Programmes and Initiatives throughout Ireland. 2017. The 2016 Pobal HP Deprivation Index for Small Areas (SA) URL: https://www.pobal.ie/app/uploads/2018/06/The-2016-Pobal-HP-Deprivation-Index-Introduction-07.pdf [accessed 2018-06-12]
28. The BigO Study. 2019. Big Data Against Childhood Obesity URL: https://bigoprogram.eu/ [accessed 2020-02-15]
29. Diou C, Sarafis I, Papapanagiotou V, Ioakimidis I, Delopoulos A. A methodology for obtaining objective measurements of population obesogenic behaviors in relation to the environment. J Stat 2019 Dec 10;35(4):677-690 [FREE Full text] [doi: 10.3233/sji-190537]
30. O'Malley G, Hussey J, Roche E. A pilot study to profile the lower limb musculoskeletal health in children with obesity. Pediatr Phys Ther 2012;24(3):292-298. [doi: 10.1097/PEP.0b013e31825c14f8] [Medline: 22735484]
31. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of ehealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
32. 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]
33. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
34. Eysenbach G. Consort-eHealth: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]
35. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the pediatric wuality of life inventory version 4.0 generic core scales in healthy and patient populations. Med Care 2001 Aug;39(8):800-812. [doi: 10.1097/00005650-200108000-00006] [Medline: 11468499]
36. van Strien T, Frijters JE, Bergers GP, Defares PB. The Dutch eating behavior questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. Int J Eat Disord 1986 Feb;5(2):295-315 [FREE Full text] [doi: 10.1002/1098-108x(198602)5:2<295::aid-eat2260050209>3.0.co;2-t]
37. Achenbach TM, Dumenci L. Advances in empirically based assessment: revised cross-informant syndromes and new DSM-oriented scales for the CBCL, YSR, and TRF: comment on Lengua, Sadowksi, Friedrich, and Fischer (2001). J Consult Clin Psychol 2001 Aug;69(4):699-702. [Medline: 11550735]
38. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
39. Jensen CD, Duncombe KM, Lott MA, Hunsaker SL, Duraccio KM, Woolford SJ. An evaluation of a smartphone-assisted behavioral weight control intervention for adolescents: pilot study. JMIR Mhealth Uhealth 2016 Aug 23;4(3):e102 [FREE Full text] [doi: 10.2196/mhealth.6034] [Medline: 27554704]
40. 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]
41. Jago R, Sebire SJ, Turner KM, Bentley GF, Goodred JK, Fox KR, et al. Feasibility trial evaluation of a physical activity and screen-viewing course for parents of 6 to 8 year-old children: TeamPlay. Int J Behav Nutr Phys Act 2013 Mar 4;10:31 [FREE Full text] [doi: 10.1186/1479-5868-10-31] [Medline: 23510646]
42. Phan TL, Barnini N, Xie S, Martinez A, Falini L, Abatemarco A, et al. Feasibility of using a commercial fitness tracker as an adjunct to family-based weight management treatment: pilot randomized trial. JMIR Mhealth Uhealth 2018 Nov 27;6(11):e10523 [FREE Full text] [doi: 10.2196/10523] [Medline: 30482743]
43. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity ehealth interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
44. Skelton JA, Goff DC, Ip E, Beech BM. Attrition in a multidisciplinary pediatric weight management clinic. Child Obes 2011 Jun 20;7(3):185-193 [FREE Full text] [doi: 10.1089/chi.2011.0010] [Medline: 21966612]
45. de Niet J, Timman R, Jongejan M, Passchier J, van den Akker E. Predictors of participant dropout at various stages of a pediatric lifestyle program. Pediatrics 2011 Jan;127(1):e164-e170. [doi: 10.1542/peds.2010-0272] [Medline: 21149433]
46. Cote MP, Byczkowski T, Kotagal U, Kirk S, Zeller M, Daniels S. Service quality and attrition: an examination of a pediatric obesity program. Int J Qual Health Care 2004 Apr;16(2):165-173. [doi: 10.1093/intqhc/mzh015] [Medline: 15051711]

47. van der Mispel C, Poppe L, Crombez G, Verloigne M, de Bourdeaudhuij I. A self-regulation-based ehealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
48. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international internet health interventions: a step towards global ehealth. Health Promot Int 2013 Sep;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
49. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595. [doi: 10.1093/her/cyp063] [Medline: 19897515]
50. Vignolo M, Rossi F, Bardazza G, Pistorio A, Parodi A, Spigno S, et al. Five-year follow-up of a cognitive-behavioural lifestyle multidisciplinary programme for childhood obesity outpatient treatment. Eur J Clin Nutr 2008 Sep;62(9):1047-1057. [doi: 10.1038/sj.ejcn.1602819] [Medline: 17554247]
51. Cohen DM, Lumley MA, Naar-King S, Partridge T, Cakan N. Child behavior problems and family functioning as predictors of adherence and glycemic control in economically disadvantaged children with type 1 diabetes: a prospective study. J Pediatr Psychol 2004;29(3):171-184. [doi: 10.1093/jpepsy/jsh019] [Medline: 15131135]
52. Aparicio E, Canals J, Arija V, de Henauw S, Michels N. The role of emotion regulation in childhood obesity: implications for prevention and treatment. Nutr Res Rev 2016 Jun;29(1):17-29. [doi: 10.1017/S0954422415000153] [Medline: 27045966]
53. Lumeng JC, Gannon K, Cabral HJ, Frank DA, Zuckerman B. Association between clinically meaningful behavior problems and overweight in children. Pediatrics 2003 Nov;112(5):1138-1145. [doi: 10.1542/peds.112.5.1138] [Medline: 14595059]
54. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. JMIR Mhealth Uhealth 2018 Nov 8;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]

Abbreviations

BigO: big data against childhood obesity
CBCL: Child Behavior Checklist
CONSORT: consolidated standards of reporting trials
eHealth: electronic health
mHealth: mobile health
SDS: SD score
SUS: system usability scale

Edited by G Eysenbach, S Nelson; submitted 05.11.19; peer-reviewed by J Alvarez Pitti, S Smith; comments to author 22.01.20; revised version received 11.03.20; accepted 29.03.20; published 08.07.20.

Please cite as:
Browne S, Kechadi MT, O'Donnell S, Dow M, Tully L, Doyle G, O'Malley G
Mobile Health Apps in Pediatric Obesity Treatment: Process Outcomes From a Feasibility Study of a Multicomponent Intervention
JMIR Mhealth Uhealth 2020;8(7):e16925
URL: https://mhealth.jmir.org/2020/7/e16925
doi:10.2196/16925
PMID:32673267

Original Paper

Acceptance, Usage, and Barriers of Electronic Patient-Reported Outcomes Among German Rheumatologists: Survey Study

Martin Krusche[1,2], MD; Philipp Klemm[2,3], MD; Manuel Grahammer[4], MSc; Johanna Mucke[2,5], MD; Diana Vossen[2,6], MD; Arnd Kleyer[2,7], MD; Philipp Sewerin[2,5], MD, PhD; Johannes Knitza[2,7], MD

[1]Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin, Berlin, Germany

[2]Working Group Young Rheumatology, German Society for Rheumatology, Berlin, Germany

[3]Department of Rheumatology, Immunology, Osteology and Physical Medicine, Justus Liebig University Gießen, Campus Kerckhoff, Bad Nauheim, Germany

[4]Abaton GmbH, Berlin, Germany

[5]Department of Rheumatology and Hiller Research Unit Rheumatology, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

[6]Rheinisches Rheumazentrum Meerbusch, St Elisabeth Hospital, Meerbusch, Germany

[7]Department of Internal Medicine 3 – Rheumatology and Immunology, Friedrich-Alexander University Erlangen-Nürnberg, University Hospital Erlangen, Erlangen, Germany

Corresponding Author:
Martin Krusche, MD
Department of Rheumatology and Clinical Immunology
Charité – Universitätsmedizin
Chariteplatz 1
Berlin, 10117
Germany
Phone: 49 30 4505130
Email: martin.krusche@charite.de

Abstract

Background: The use of patient-reported outcomes (PROs) allows for patient-centered, measurable, and transparent care. Electronic PROs (ePROs) have many benefits and hold great potential to improve current usage of PROs, yet limited evidence exists regarding their acceptance, usage, and barriers among rheumatologists.

Objective: This study aims to evaluate the current level of acceptance, usage, and barriers among German rheumatologists regarding the use of ePROs. The importance of different ePRO features for rheumatologists was investigated. Additionally, the most frequently used PROs for patients with rheumatoid arthritis (RA) were identified.

Methods: Data were collected via an online survey consisting of 18 questions. The survey was completed by members of the Working Group Young Rheumatology of the German Society for Rheumatology (Arbeitsgemeinschaft Junge Rheumatologie der Deutschen Gesellschaft für Rheumatologie [DGRh]) at the 2019 annual DGRh conference. Only members currently working in clinical adult rheumatology were eligible to complete the survey.

Results: A total of 119 rheumatologists completed the survey, of which 107 (89.9%) reported collecting PROs in routine practice and 28 (25.5%) already used ePROs. Additionally, 44% (43/97) were planning to switch to ePROs in the near future. The most commonly cited reason for not switching was the unawareness of suitable software solutions. Respondents were asked to rate the features of ePROs on a scale of 0 to 100 (0=unimportant, 100=important). The most important features were automatic score calculation and display (mean 77.50) and simple data transfer to medical reports (mean 76.90). When asked about PROs in RA, the respondents listed pain, morning stiffness, and patient global assessment as the most frequently used PROs.

Conclusions: The potential of ePROs is widely seen and there is great interest in them. Despite this, only a minority of physicians use ePROs, and the main reason for not implementing them was cited as the unawareness of suitable software solutions. Developers, patients, and rheumatologists should work closely together to help realize the full potential of ePROs and ensure a seamless integration into clinical practice.

(JMIR Mhealth Uhealth 2020;8(7):e18117) doi:10.2196/18117

KEYWORDS

electronic patient-reported outcome measures; eHealth; rheumatology; rheumatoid arthritis; patient perspective; mobile phone

Introduction

Patient-Reported Outcomes in Rheumatology

In order to effectively monitor treatment outcomes, it is crucial to include the patient's perspective. As quality of life reported by patients and clinical assessment by physicians can have divergent results [1], unaltered, direct patient-reported outcomes (PROs) have become an integral part of today's clinical routine [2,3] and play a crucial role in clinical studies [4]. The use of PROs allows patient-centered, measurable, and transparent care [3,5]. Various PROs have been established to reflect the individual's perceived state of health [6]. There are disease-specific PROs, such as the rheumatoid arthritis disease activity index [7], or more general PROs, such as the Health Assessment Questionnaire Disability Index (HAQ-DI) [8]. Various factors, such as symptom severity, physical status, patient satisfaction, and disease-specific comorbidity affect PROs and can be reported. PROs also make it possible to monitor symptoms (eg, fatigue or depression) that are otherwise difficult to measure but play a decisive role in patients' quality of life [9].

Treat-to-Target Approach to Increase Therapy Effectiveness

New and innovative therapies have significantly improved treatment results of patients with rheumatic diseases such as rheumatoid arthritis (RA) [10]; however, the effectiveness of new drugs seems to stagnate [11]. To maximize treatment effectiveness, rheumatologists must optimize other therapy variables. For example, the window of opportunity [12], that is, the time frame in which the treatment is most effective, should be respected [13]. In this context, the treat-to-target approach was developed [14]. The aim of this strategy is to define a treatment target at therapy initiation and to closely monitor treatment response in order to identify insufficient treatment success and modify the therapeutic strategy as needed. This approach represents a challenge for rheumatologists, as resources are limited [15]. In reality, therapies are not assessed frequently enough and, therefore, are all too often not adjusted to the current state of the disease [16,17]. Two important reasons for the current poor disease/treatment management are (1) the poor access to rheumatology specialists and (2) an increasing deficit of follow-up appointments for already-diagnosed patients. This situation is likely to worsen due to the current shortage of rheumatologists in Germany [18], and the trend indicates that it will become more difficult in the future [19].

Electronic Patient-Reported Outcomes on the Rise

Digitalization promises new ways to improve patient outcomes and shape a more efficient and transparent health care environment. Electronic PROs (ePROs) could help to realize the treat-to-target principle on a far larger scale, as the therapeutic outcomes can be evaluated by the patient more frequently and in any setting.

Currently, 49% of German rheumatologists use medical apps in their clinical routine [20]. Various mobile apps offer ePRO services [21,22]. ePROs have the potential to save valuable time and money, as no manual digitization and calculation are required.

Furthermore, ePROs can be recorded anywhere and anytime. Patients can receive reminders to complete the ePROs, which would help reduce the risk of patients being lost to follow-up or failing to complete questionnaires.

Additionally, data would only be collected during disease flare-ups (ie, when patients need to consult their rheumatologists); however, ePROs could also be collected on a regular basis (eg, biweekly) in order to establish benchmarks.

These data would enable rheumatologists to differentiate more precisely between a disease flare-up and a general insufficient response to treatment, thereby allowing treat-to-target strategies to be achievable on a regular basis.

The greatest advantage to digital outcome reporting is the continuous documentation of the individual's perceived state of health in a standardized, transparent, and validated way. Furthermore, if patients change physicians, they can easily send their data to their new provider (eg, referral to in-house care).

Aim of This Study

ePROs hold great potential to improve current treatment of rheumatic diseases; however, limited evidence exists regarding usage in clinical practice or hurdles towards acceptance in daily care among rheumatologists. A better understanding of the individual interests, motivations, fears, and circumstances of the treating physician is necessary to promote the further use of ePROs. The aim of this study, therefore, was to identify the current state of ePRO acceptance, usage, and barriers among German rheumatologists.

Methods

Data were collected via a survey of members of the Working Group Young Rheumatology (Arbeitsgemeinschaft Junge Rheumatologie [AGJR]) of the German Society of Rheumatology at the 2019 annual German Society of Rheumatology conference in Dresden, Germany. The web-based survey was created using SurveyMonkey (SurveyMonkey Inc). The survey was conducted via iPad (Apple Inc) at the AGJR congress booth. Only clinically active adult rheumatologists were asked to complete the survey. Prior to the conference, a task force consisting of AGJR members designed the questionnaire in a web-based consensus meeting after initial individual research of the current literature. Each individual reported the results of their respective deliberations and presented points assessed in the questionnaire that included proposals for the wording of individual questions. In total, 18 questions were formulated assessing acceptance, usage, and barriers concerning ePROs among German rheumatologists.

Results

In total, 724 members of the German Society of Rheumatology joined the conference and 119 adult rheumatologists participated in the survey. Of the 119 participants, 68 (57.1%) were male and 51 (42.9%) were female. Of the age makeup, 16 of 119 (13.4%) participants were aged between 20 and 30 years; 41 (34.4%), between 30 and 40 years; 23 (19.3%), between 40 and 50 years; 25 (21.1%), between 50 and 60 years; and 14 (11.8%), older than 60 years.

A total of 43 of the 119 (36.1%) participants were residents and 76 (63.9%) were consultants. Of the 43 residents, 35 (81%) were working in a university hospital, 6 (14%) were working in a private or state hospital, and 2 (5%) were employed in a private medical office. Of the 76 consultants, 26 (34%) worked at a university hospital, 21 (28%) worked in a private or state hospital, and 29 (38%) worked in a medical office. These characteristics can be seen in Table 1.

Table 1. Characteristics of the respondents.

Characteristics	Respondents, n (%)
Sex	
Female	51 (42.9)
Male	68 (57.1)
Age (years)	
20-30	16 (13.4)
31-40	41 (34.4)
41-50	23 (19.3)
51-60	25 (21.1)
>61	14 (11.8)
Workplace	
Resident	43 (36.1)
University hospital	35 (81.4)
Private or state hospital	6 (13.9)
Medical office	2 (4.6)
Consultant	76 (63.9)
University hospital	26 (34.2)
Private or state hospital	21 (27.6)
Medical office	29 (38.2)

Of the total participants, 89.9% (107/119) used any form of PRO daily; however, 10.1% (12/119) of respondents indicated that they did not collect PROs at all. Of the rheumatologists using PROs, 77.3% (92/107) of survey participants reported that they used pen and paper–based PROs. Furthermore, 23.5% (28/119) of respondents collected PROs electronically at each patient appointment and 4.2% (5/119) collected PROs electronically before patient contact.

Among respondents not currently using electronic collection, 44% (43/97) reported that they planned to use ePROs; however, 33% (32/97) were still undecided, and 23% (22/97) were not planning to use ePROs (see Table 2).

The question "Why are ePROs not used?" was answered by 68 physicians: 34% (23/68) stated that they did not know a specific software, 12% (8/68) indicated that the introduction of a software was too complicated, and 12% (8/68) thought the software was too expensive. Furthermore, 18% (12/68) of nonimplementers reported that using ePROs was too time-consuming, 16% (11/68) reported that patients preferred paper-based questionnaires, and 32% (22/68) stated "other reasons" (Table 2). Regarding the question "What do rheumatologists use PROs for in routine care?" 66.4 % (79/119) of respondents indicated clinical decision making; 39.5% (47/119), research; 23.5% (28/119), reimbursement; 21.8% (26/119), internal quality monitoring; and 16.8% (20/119), improving patient satisfaction. In addition, 5.9% (7/119) stated that PROs were used for external quality monitoring, and 8.4% (10/119) stated that they were collected but offered no further value (Table 2).

Table 2. Participant responses to 4 questions.

Question and answers	Responses, n (%)
Are you planning to use ePROs[a]?	
Yes	43 (44.3)
No	22 (22.7)
Undecided	32 (33)
Why not?[b]	
No software	23 (33.8)
Implementation too complicated	8 (11.8)
Patients prefer paper	11 (16.2)
I prefer paper	2 (2.9)
Expansive software	8 (11.8)
No proven benefit	1 (1.5)
No need	4 (5.9)
Requires too much time	12 (17.6)
Other reasons	22 (32.4)
What do you usually use PROs[c] for?[b]	
No usage	10 (8.4)
Clinical decisions	79 (66.4)
Research	47 (39.5)
Internal quality management	26 (21.9)
External quality management	7 (5.9)
To increase patient satisfaction	20 (16.8)
Reimbursement reasons	28 (23.5)
Other reasons	7 (5.9)
Do you have access to the following information before patient contact?[b]	
Total score of the recorded PROs	36 (30.3)
Prevalues	67 (56.3)
Histograms	11 (9.2)
Neither	40 (33.6)

[a]ePROs: electronic patient-reported outcomes.

[b]Multiple answers were possible.

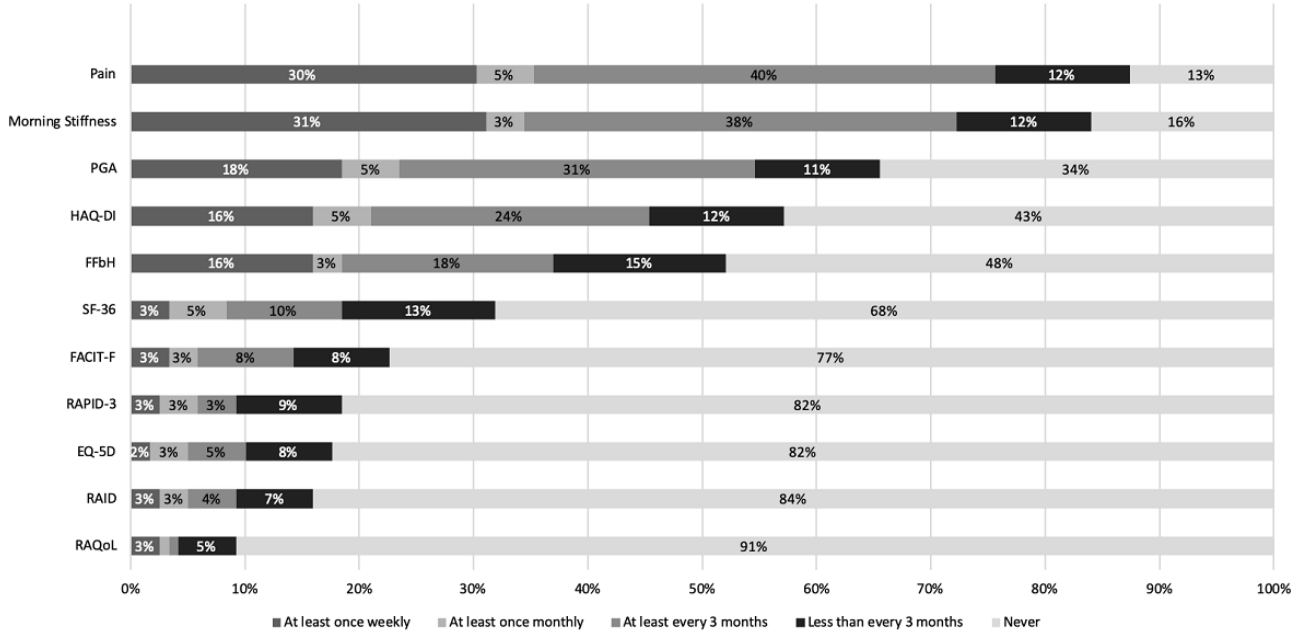
[c]PROs: patient-reported outcomes.

The most frequently used PROs for patients with RA were pain, morning stiffness, and patient global assessment, as seen in Figure 1.

In regards to reviewing the results of PROs, 16.8% (20/119) of respondents stated that they never review the results of the PROs before a patient consultation, 29.4% (35/119) stated that they review them sometimes, 31.1% (37/119) stated that they review PROs often, and 22.7% (27/119) claimed to review PROs before every patient visit.

Furthermore, rheumatologists were asked to specify why PROs were not always reviewed. This question was answered by 96 of the 119 (80.6%) participants, and multiple answers were possible. A total of 67% (62/93) of respondents said that this was due to lack of time, 27% (25/93) indicated that PRO results were often not available before the patient appointment, 15% (14/93) answered that the PROs were incompletely answered by the patient, and 12% (11/93) indicated that they trusted their own judgement more than the patient's. In addition, 12% (11/93) reported that the results of the PROs were often confusing.

Figure 1. Patient-reported outcomes being used in clinical practice and their respective frequency. EQ-5D: EuroQol Five-Dimensional Questionnaire; FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue; FFbH: Funktionsfragebogen Hannover; HAQ-DI: Health Assessment Questionnaire Disability Index; PGA: patient global assessment; RAID: Rheumatoid Arthritis Impact of Disease; RAPID-3: Routine Assessment of Patient Index Data 3; RAQoL: Rheumatoid Arthritis Quality of Life; SF-36: Short Form-36.



We further investigated whether doctors have access to the total score of all recorded PROs, previous PROs, or a histogram of prior values. In total, 30.3% (36/119) of respondents had access to the total score of the recorded PROs, 56.3% (67/119) had access to the previous values of the PROs, 9.2% (11/119) had access to histograms of the PROs, and 33.6% (40/119) did not have access to this information (Table 2). Furthermore, 86.6% (103/119) of rheumatologists reported that their patients did not have access to the results of their PROs. Additionally, we asked how often rheumatologists discussed the results of the PROs with their patients, on a scale of 0 (least often) to 100 (most often), and the mean score was 38.05 (SD 30.57).

Furthermore, the importance of different ePRO functions was analyzed on a scale of 0 to 100. Graphic display of the ePROs was rated with a mean score of 63.50 (SD 31.19). The respondents rated a simple data transfer (eg, from a digital source to the practice's computer system) a mean score of 76.90 (SD 30.07). Regarding an automatic notification to the physician if a critical threshold value of the ePROs was exceeded, the mean importance was 51.65 (SD 33.5). The notification of the patient if a critical threshold value was exceeded was rated a mean score of 34.55 (SD 30.61), as seen in Table 3.

Table 3. Rating of 6 subjects from 0 (lowest agreement) to 100 (highest agreement).

Question	n[a]	Mean (SD)
In how many cases do you discuss the PRO[b] results with your patients?	119	38.05 (30.57)
How important would the graphic display of ePROs[c] be to you?	119	63.50 (31.19)
How important would the automatic score calculation and display of ePROs be to you?	119	77.50 (27.64)
How important would the simple transfer of the ePROs to medical reports be to you?	119	76.90 (30.07)
How important would an automatic notification to yourself be for you if a critical threshold is exceeded by an ePRO?	119	51.65 (33.50)
How important would an automatic notification to the patient be for you if a critical threshold is exceeded by an ePRO?	119	34.55 (30.61)

[a]Number of participants who responded to the question.

[b]PRO: patient-reported outcome.

[c]ePROs: electronic patient-reported outcomes.

Discussion

Principal Findings

For the first time, this study showed that PROs are widely accepted and used by German rheumatologists; however, it also indicated that the use of ePROs is lagging. Only 10.1% (12/119) of respondents indicated that they do not use PROs at all. The most frequently used PROs were pain (104/119, 87.4%), morning stiffness (100/119, 84.0%), patient global assessment (78/119, 65.5%), HAQ-DI (68/119, 57.1%), and the Hannover functional questionnaire (Funktionalfragebogen Hannover)

(62/119, 52.1%). Other validated PRO instruments, such as the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) were rarely used (27/119, 22.7%). Compared to a previous European survey in 2008 [23], the percentage of rheumatologists in Germany who never use specific PROs was greater in 2019 than in 2008, especially among the use of the HAQ-DI (57% vs 80%) and Short Form-36 (SF-36) (51% vs 32%). These results could have multiple causes. We speculate that our survey consisted of more nonuniversity-based rheumatologists (58/119, 48.7%), who tend to more seldomly use PROs in their daily clinical practice. Interestingly, in 2019, some rheumatologists were using PROs on a weekly basis to allow for more continuous documentation.

The majority (92/119, 77.3%) of German rheumatologists use the traditional pen and paper–based score evaluation, but 23.5% (28/119) already use ePRO in clinical routine. In the future, ePROs will play a more significant role, as 77% (75/97) plan to implement or are undecided about implementing ePROs.

A further barrier stopping German rheumatologists from implementing ePROs is the perception of ePROs being time-consuming processes (12/68, 18%).

The importance of PROs for rheumatologists is reflected by the broad acceptance and use of PROs in clinical care. Only 16.8% (20/119) never review PROs, and the most commonly cited reason (62/93, 67%) for not reviewing the data was a lack of time. More than half (67/119, 56.3%) of the physicians only have access to the last evaluated scores, and only 30.3% (36/119) have access to the complete PRO data set.

A major finding of this study is that discussing the PRO result with the patient was not deemed important. On a rating scale from 0 (lowest importance) to 100 (highest importance) the subject of discussing the result with the patient was rated as a mean of 38.05 (SD 30.57). Additionally, 86.6% (103/119) of the rheumatologists indicated that their patients have no access to their PRO data at all. This indicated a lack of involvement of the patient in therapy and a lack of transparency regarding PROs. Patients' access to these results and active discussion with their rheumatologist could increase patient satisfaction, adherence, and empowerment, and it could provide a strong basis for better shared decision making (SDM).

The finding that only 9.2% (11/119) of the rheumatologists surveyed have access to PRO graph scores over time, despite the average mean score of 63.50 (SD 31.19) for the importance of having a graphic display of the ePRO scores, highlights the interest in this feature. Interestingly, rheumatologists in Germany prioritize an automatic score calculation with a graphic display over just graphically displaying the results (mean 77.50 vs mean 63.50).

Comparison With Prior Work

Our data shows a high acceptance of PROs and a lagging use of ePROs. The high usage of traditional PROs in the process of clinical decision making corresponds with the recommendations of the overarching principles of the treat-to-target concept, according to international and European guidelines [24,25]. PROs are being declared as a main treatment target in RA patients [26]. They are particularly important because objective scientific targets make little sense if the patient has subjectively ongoing symptoms. The concept defines quality of life as the primary therapeutic goal; therefore, the primary use of PROs for clinical decision making provides a solid justification for the use of PROs in a treat-to-target approach.

It was recently shown that the vast majority (96%) of patients are willing to share mobile app data for research purposes [27]. Integration of ePROs into existing registries should be actively pursued as a way to generate valuable primary research data and to motivate patients to continue using ePROs [21,28-30].

The main reason for not reviewing and thereby considering the PRO data for treatment is, according to our data, a lack of time (62/93, 67%). The German data align very well with the data from the United States, in which rheumatologists were asked why they do not qualitatively measure RA metrics routinely and 62.5% of respondents stated that a lack of time was the reason [31]. In Israel, 73% of rheumatologists mentioned time constraints as the reason for not implementing the treat-to-target concept [17]. These data are interestingly independent of PRO or clinical metric collection. One end point of future studies should be the time saved by the physician and the patient using ePROs compared with paper and pen–based scores. In addition to the potential time saved by using ePROs, the concept of SDM is an important aspect that could be achieved by continuous patient monitoring and transparent integration of the process into therapy management decisions. A recent study analyzing SDM in RA patients showed that there is still room for improvement [32]. Interestingly, PROs such as pain, morning stiffness, and patient global assessment were much more often compiled than others, such as SF-36, FACIT-F, or computerized adaptive testing. That is certainly due to the simpler data collection (fewer questions) in everyday clinical practice. Less frequently used PROs seem to be collected in clinical trials and have only found their way into clinical practice to a limited degree.

A notable finding of this work is that despite PROs being regularly compiled, a majority of respondents (103/119, 86.6%) stated that their patients have no access to their data. Furthermore, a need to notify patients when values exceed a critical threshold was also given a low score of importance (mean 34.55, SD 30.61). Whether this is also due to a lack of time or whether the physicians trust their own judgement more than the patient's is unclear, but potential approaches for improving patient care can be seen in this study. ePROs could be used as a basis for SDM. It is known that SDM increases adherence and leads to increased patient satisfaction [33]. As the World Health Organization states that 50% of chronically ill patients do not take their medication regularly, adherence and patient satisfaction play a key role in daily rheumatologic practice [34,35].

Particularly among patients, there is great interest in PROs. Navarro-Millán et al [36] showed a marked interest among patients in becoming more involved in the therapy and in collecting and sharing ePROs with their doctors when it facilitated communication. Other publications suggest that patients prefer entering ePROs from home [37]. Walker et al [38] showed high correlation of clinical parameters, like the

Simplified Disease Activity Index, Clinical Disease Activity Index, and Disease Activity Score-44, with the Routine Assessment of Patient Index Data 3, and a high willingness of the patients to monitor their targets via a smartphone. The fastest-growing group of smartphone users are older adults [39]. In Germany, 90% of patients with inflammatory rheumatic diseases regularly use a smartphone [27]. Wearable digital technologies enable objective, passive, and continuous ePROs. A growing body of evidence exists that these data can efficiently complement clinical routine procedures [40,41]. ePROs therefore promise to facilitate the realization of the treat-to-target approach.

Various barriers to implementing ePROs have been identified in the survey, among them being the availability and the price of corresponding software. Further difficulties could include the effort of adequate staff and patient training in managing and collecting ePROs.

Regarding the use of PROs for rheumatoid arthritis, our results are in line with previous European results [23].

Limitations

A total of 119 physicians participated in the survey, which represents approximately 15% of all German rheumatologists [18]. Survey participants tended to be both younger and still in training (rheumatology residents). A high proportion had a university employer, and participants were more often employed in a hospital than the German average. Therefore, the results may be positively biased.

In Germany, PROs are financially reimbursed by the health care system. Consequently, the proportion PROs collected is higher than in other countries [23], and a transfer of the results to other health systems should therefore be carried out cautiously.

Some 32% (22/68) of physicians who reported not using ePROs stated “other reasons” for nonuse. Unfortunately, data security issues were not specifically questioned, though it is also a possible hurdle for the adoption of ePROs. Therefore, the authors suggest that data security should be independently addressed in future research projects.

This question set dealt only with the use of PROs and ePROs, regardless of the method used to determine the outcomes (eg, via tablet in the doctor’s office, or independently via the patient’s smartphone). A distinct connection between the use of PROs and the use/implementation of the treat-to-target concept was not independently questioned; thus, only estimated conclusions can be made about the use of PROs for this concept.

Potential Vision for the Future

ePROs offer a great potential to overcome current health care obstacles. Automatic score calculations, reminders, and complete data sets save valuable time. Physicians could focus on interpreting the results and getting a preview of the patients, as the results are available before the consultation. In this respect, the treat-to-target approach is not only an option but can also easily be implemented for close monitoring and management (ie, tight control).

An automatic warning system could alert physicians or patients if the PROs exceed or fall below a critical threshold, thereby allowing treatment success to be monitored automatically as a means of supporting the patient and physician. If a patient exceeds the threshold value, a short-term check-up with the physician is recommended and more intensive treatment or a change in therapy may be required. If the patient stays below a threshold value, the next follow-up appointment may not be required as promptly, or the medication may be tapered to allow a more efficient use of limited resources. The recently passed Digital Supply Act in Germany allows physicians to prescribe digital health apps to patients, which are reimbursed by the country’s statutory health insurance. This reimbursement for using apps could foster the wide use of ePROs in clinical routine, as reimbursement appeared to increase usage by almost 10% in a previous study [23].

Conclusions

Our study results suggest a high acceptance and usage of PROs as a part of patient care in Germany. Nevertheless, some barriers to the usage of PROs were also found to exist, in particular the lack of time for the physician and the absence of data for prior values.

A majority of rheumatologists plan to implement ePROs for clinical decision making. The main barrier to implementing ePROs is a lack of knowledge about suitable software solutions, which matches expectations regarding ease of implementation and price. This work provides a rationale to investigate the effects of ePRO usage in clinical studies analyzing economic and patient outcome as end points.

Conflicts of Interest

MG is the founder and stakeholder of ABATON GmbH. MK, PK, JM, DV, AK, PS, and JK declare no conflicts of interest.

References

1. Challa D, Kvrgic Z, Cheville A, Crowson C, Bongartz T, Mason T, et al. Patient-provider discordance between global assessments of disease activity in rheumatoid arthritis: a comprehensive clinical evaluation. Arthritis Res Ther 2017 Sep 26;19(1):212 [FREE Full text] [doi: 10.1186/s13075-017-1419-5] [Medline: 28950896]
2. Fautrel B, Alten R, Kirkham B, de la Torre I, Durand F, Barry J, et al. Call for action: how to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. Rheumatol Int 2018 Jun;38(6):935-947 [FREE Full text] [doi: 10.1007/s00296-018-4005-5] [Medline: 29564549]

3. Hendrikx J, de Jonge MJ, Fransen J, Kievit W, van Riel PL. Systematic review of patient-reported outcome measures (PROMs) for assessing disease activity in rheumatoid arthritis. RMD Open 2016 Aug 18;2(2):e000202 [FREE Full text] [doi: 10.1136/rmdopen-2015-000202] [Medline: 27651921]
4. Gossec L, Dougados M, Dixon W. Patient-reported outcomes as end points in clinical trials in rheumatoid arthritis. RMD Open 2015;1(1):e000019 [FREE Full text] [doi: 10.1136/rmdopen-2014-000019] [Medline: 26509052]
5. Bingham CO, Gaich CL, DeLozier AM, Engstrom KD, Naegeli AN, de Bono S, et al. Use of daily electronic patient-reported outcome (PRO) diaries in randomized controlled trials for rheumatoid arthritis: rationale and implementation. Trials 2019 Mar 22;20(1):182 [FREE Full text] [doi: 10.1186/s13063-019-3272-0] [Medline: 30902094]
6. Hiligsmann M, Rademacher S, Kaal KJ, Bansback N, Harrison M. The use of routinely collected patient-reported outcome measures in rheumatoid arthritis. Semin Arthritis Rheum 2018 Dec;48(3):357-366. [doi: 10.1016/j.semarthrit.2018.03.006] [Medline: 29709290]
7. Fransen J, Langenegger T, Michel B, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. Rheumatology (Oxford) 2000 Mar;39(3):321-327. [doi: 10.1093/rheumatology/39.3.321] [Medline: 10788543]
8. Bruce B, Fries J. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003 Jan;30(1):167-178. [Medline: 12508408]
9. Alexanderson H, Del Grande M, Bingham CO, Orbai A, Sarver C, Clegg-Smith K, et al. Patient-reported outcomes and adult patients' disease experience in the idiopathic inflammatory myopathies. report from the OMERACT 11 Myositis Special Interest Group. J Rheumatol 2014 Mar 15;41(3):581-592 [FREE Full text] [doi: 10.3899/jrheum.131247] [Medline: 24429182]
10. Mahajan TD, Mikuls TR. Recent advances in the treatment of rheumatoid arthritis. Curr Opin Rheumatol 2018 May;30(3):231-237 [FREE Full text] [doi: 10.1097/BOR.0000000000000496] [Medline: 29461286]
11. Köhler, Günther, Kaudewitz, Lorenz. Current Therapeutic Options in the Treatment of Rheumatoid Arthritis. J Clin Med 2019 Jun 28;8(7) [FREE Full text] [doi: 10.3390/jcm8070938] [Medline: 31261785]
12. Quinn MA, Emery P. Window of opportunity in early rheumatoid arthritis: possibility of altering the disease process with early intervention. Clin Exp Rheumatol 2003;21(5 Suppl 31):S154-S157. [Medline: 14969068]
13. Smolen JS, Schöls M, Braun J, Dougados M, FitzGerald O, Gladman DD, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis 2018 Jan;77(1):3-17 [FREE Full text] [doi: 10.1136/annrheumdis-2017-211734] [Medline: 28684559]
14. Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Álvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017 Jun;76(6):948-959. [doi: 10.1136/annrheumdis-2016-210602] [Medline: 27979873]
15. Battafarano D, Ditmyer M, Bolster M, Fitzgerald J, Deal C, Bass A, et al. 2015 American College of Rheumatology Workforce Study: Supply and Demand Projections of Adult Rheumatology Workforce, 2015-2030. Arthritis Care Res (Hoboken) 2018 Apr;70(4):617-626 [FREE Full text] [doi: 10.1002/acr.23518] [Medline: 29400009]
16. Yu Z, Lu B, Agosti J, Bitton A, Corrigan C, Fraenkel L, et al. Implementation of Treat-to-Target for Rheumatoid Arthritis in the US: Analysis of Baseline Data From a Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2018 May;70(5):801-806 [FREE Full text] [doi: 10.1002/acr.23343] [Medline: 28834390]
17. Gazitt T, Oren S, Reitblat T, Lidar M, Gurman AB, Rosner I, et al. Treat-to-target concept implementation for evaluating rheumatoid arthritis patients in daily practice. Eur J Rheumatol 2019 Jul;6(3):136-141 [FREE Full text] [doi: 10.5152/eurjrheum.2019.18195] [Medline: 31329541]
18. Zink A, Braun J, Gromnica-Ihle E, Krause D, Lakomek HJ, Mau W, et al. [Memorandum of the German Society for Rheumatology on the quality of treatment in rheumatology - Update 2016]. Z Rheumatol 2017 Apr;76(3):195-207. [doi: 10.1007/s00393-017-0297-1] [Medline: 28364218]
19. Krusche M, Sewerin P, Kleyer A, Mucke J, Vossen D, Morf H, Rheumadocs und Arbeitskreis Junge Rheumatologie (AGJR). [Specialist training quo vadis?]. Z Rheumatol 2019 Oct;78(8):692-697. [doi: 10.1007/s00393-019-00690-5] [Medline: 31468164]
20. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, Arbeitskreis Junge Rheumatologen. [Use of medical apps and online platforms among German rheumatologists : Results of the 2016 and 2018 DGRh conference surveys and research conducted by rheumadocs]. Z Rheumatol 2019 Nov;78(9):839-846. [doi: 10.1007/s00393-018-0578-3] [Medline: 30542914]
21. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
22. van Riel P, Alten R, Combe B, Abdulganieva D, Bousquet P, Courtenay M, et al. Improving inflammatory arthritis management through tighter monitoring of patients and the use of innovative electronic tools. RMD Open 2016;2(2):e000302 [FREE Full text] [doi: 10.1136/rmdopen-2016-000302] [Medline: 27933206]

23. Schoels M, Aletaha D, Smolen JS, Bijlsma JWJ, Burmester G, Breedveld FC, et al. Follow-up standards and treatment targets in rheumatoid arthritis: results of a questionnaire at the EULAR 2008. Ann Rheum Dis 2010 Mar;69(3):575-578 [FREE Full text] [doi: 10.1136/ard.2009.108472] [Medline: 19389716]
24. Singh JA, Saag KG, Bridges SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol 2016 Jan;68(1):1-26 [FREE Full text] [doi: 10.1002/art.39480] [Medline: 26545940]
25. Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Álvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017 Jun;76(6):948-959. [doi: 10.1136/annrheumdis-2016-210602] [Medline: 27979873]
26. van Vollenhoven R. Treat-to-target in rheumatoid arthritis - are we there yet? Nat Rev Rheumatol 2019 Mar;15(3):180-186. [doi: 10.1038/s41584-019-0170-5] [Medline: 30700865]
27. Knitza J, Raab C, Lambrecht A, Simon D, Hagen M, Bayat S, et al. The Urge for Mobile Apps in Rheumatology – a German Patient Perspective [abstract]. Arthritis Rheumatol 2019 Nov 08;71(suppl 10) [FREE Full text] [doi: 10.2196/preprints.19661]
28. Ibfelt E, Jensen D, Hetland M. The Danish nationwide clinical register for patients with rheumatoid arthritis: DANBIO. Clin Epidemiol 2016;8:737-742 [FREE Full text] [doi: 10.2147/CLEP.S99490] [Medline: 27822121]
29. Gandrup J, Yazdany J. Using Health Information Technology to Support Use of Patient-Reported Outcomes in Rheumatology. Rheum Dis Clin North Am 2019 May;45(2):257-273 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.007] [Medline: 30952397]
30. Ciurea A, Scherer A, Exer P, Bernhard J, Dudler J, Beyeler B, Rheumatologists of the Swiss Clinical Quality Management Program for Axial Spondyloarthritis. Tumor necrosis factor α inhibition in radiographic and nonradiographic axial spondyloarthritis: results from a large observational cohort. Arthritis Rheum 2013 Dec 27;65(12):3096-3106 [FREE Full text] [doi: 10.1002/art.38140] [Medline: 23983141]
31. Curtis JR, Chen L, Danila MI, Saag KG, Parham KL, Cush JJ. Routine Use of Quantitative Disease Activity Measurements among US Rheumatologists: Implications for Treat-to-target Management Strategies in Rheumatoid Arthritis. J Rheumatol 2018 Jan 15;45(1):40-44 [FREE Full text] [doi: 10.3899/jrheum.170548] [Medline: 29142029]
32. Mathijssen E, Vriezekolk J, Popa C, van den Bemt BJF. Shared decision making in routine clinical care of patients with rheumatoid arthritis: an assessment of audio-recorded consultations. Ann Rheum Dis 2020 Feb;79(2):170-175. [doi: 10.1136/annrheumdis-2019-216137] [Medline: 31662320]
33. Falzer PR. Treat-to-target and shared decision making in rheumatoid arthritis treatment: Is it feasible? Int J Rheum Dis 2019 Sep;22(9):1706-1713. [doi: 10.1111/1756-185X.13664] [Medline: 31359630]
34. Marengo MF, Suarez-Almazor ME. Improving treatment adherence in patients with rheumatoid arthritis: what are the options? Int J Clin Rheumtol 2015 Oct 01;10(5):345-356 [FREE Full text] [doi: 10.2217/ijr.15.39] [Medline: 27087857]
35. World Health Organization. Adherence to Long-term Therapies: Evidence for Action. WHO. Geneva, Switzerland: World Health Organization; 2003. URL: https://www.who.int/chp/knowledge/publications/adherence_full_report.pdf?ua=1 [accessed 2020-06-24]
36. Navarro-Millán I, Zinski A, Shurbaji S, Johnson B, Fraenkel L, Willig J, et al. Perspectives of Rheumatoid Arthritis Patients on Electronic Communication and Patient-Reported Outcome Data Collection: A Qualitative Study. Arthritis Care Res (Hoboken) 2019 Jan 27;71(1):80-87 [FREE Full text] [doi: 10.1002/acr.23580] [Medline: 29669191]
37. Secher A, Glintborg B, Gudbergsen H, Krogh N, Sørensen IJ, Jensen D, et al. Comparing patient-reported outcomes entered at home versus at hospital, and testing touch screens for initial recruitment to scientific trials in arthritis patients. Scand J Rheumatol 2019 May 16;48(3):178-184. [doi: 10.1080/03009742.2018.1522666] [Medline: 30444168]
38. Walker U, Mueller R, Jaeger V, Theiler R, Forster A, Dufner P, et al. Disease activity dynamics in rheumatoid arthritis: patients' self-assessment of disease activity via WebApp. Rheumatology (Oxford) 2017 Oct 01;56(10):1707-1712 [FREE Full text] [doi: 10.1093/rheumatology/kex229] [Medline: 28957553]
39. Joe J, Demiris G. Older adults and mobile phones for health: a review. J Biomed Inform 2013 Oct;46(5):947-954 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.008] [Medline: 23810858]
40. Nowell WB, Curtis JR, Nolot SK, Curtis D, Venkatachalam S, Owensby JK, et al. Digital Tracking of Rheumatoid Arthritis Longitudinally (DIGITAL) Using Biosensor and Patient-Reported Outcome Data: Protocol for a Real-World Study. JMIR Res Protoc 2019 Sep 26;8(9):e14665 [FREE Full text] [doi: 10.2196/14665] [Medline: 31573949]
41. Hendrikx J, Fransen J, van Riel PLCM. Monitoring rheumatoid arthritis using an algorithm based on patient-reported outcome measures: a first step towards personalised healthcare. RMD Open 2015;1(1):e000114 [FREE Full text] [doi: 10.1136/rmdopen-2015-000114] [Medline: 26629364]

Abbreviations

AGJR: Arbeitsgemeinschaft Junge Rheumatologie
ePRO: electronic patient-reported outcome
FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue
HAQ-DI: Health Assessment Questionnaire Disability Index
PRO: patient-reported outcome

RA: rheumatoid arthritis
SDM: shared decision making
SF-36: Short Form-36

Edited by C Dias; submitted 04.02.20; peer-reviewed by T Hügle, F Dittrich, P Iyer, G Murdaca, X Ji; comments to author 21.03.20; revised version received 22.04.20; accepted 09.05.20; published 20.07.20.

Please cite as:
Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, Sewerin P, Knitza J
Acceptance, Usage, and Barriers of Electronic Patient-Reported Outcomes Among German Rheumatologists: Survey Study
JMIR Mhealth Uhealth 2020;8(7):e18117
URL: http://mhealth.jmir.org/2020/7/e18117/
doi:10.2196/18117
PMID:32390592

Original Paper

A Novel Device for Smartphone-Based Fundus Imaging and Documentation in Clinical Practice: Comparative Image Analysis Study

Maximilian W M Wintergerst[1], MD; Linus G Jansen[1]; Frank G Holz[1], MD; Robert P Finger[1], MD, PhD

Department of Ophthalmology, University Hospital Bonn, Bonn, Germany

Corresponding Author:
Maximilian W M Wintergerst, MD
Department of Ophthalmology
University Hospital Bonn
Ernst-Abbe-Strasse 2
Bonn, 53127
Germany
Phone: 49 228287 ext 15505
Email: maximilian.wintergerst@ukbonn.de

Abstract

Background: Smartphone-based fundus imaging allows for mobile and inexpensive fundus examination with the potential to revolutionize eye care, particularly in lower-resource settings. However, most smartphone-based fundus imaging adapters convey image quality not comparable to conventional fundus imaging.

Objective: The purpose of this study was to evaluate a novel smartphone-based fundus imaging device for documentation of a variety of retinal/vitreous pathologies in a patient sample with wide refraction and age ranges.

Methods: Participants' eyes were dilated and imaged with the iC2 funduscope (HEINE Optotechnik) using an Apple iPhone 6 in single-image acquisition (image resolution of 2448 × 3264 pixels) or video mode (1248 × 1664 pixels) and a subgroup of participants was also examined by conventional fundus imaging (Zeiss VISUCAM 500). Smartphone-based image quality was compared to conventional fundus imaging in terms of sharpness (focus), reflex artifacts, contrast, and illumination on semiquantitative scales.

Results: A total of 47 eyes from 32 participants (age: mean 62.3, SD 19.8 years; range 7-93; spherical equivalent: mean –0.78, SD 3.21 D; range: –7.88 to +7.0 D) were included in the study. Mean (SD) visual acuity (logMAR) was 0.48 (0.66; range 0-2.3); 30% (14/47) of the eyes were pseudophakic. Image quality was sufficient in all eyes irrespective of refraction. Images acquired with conventional fundus imaging were sharper and had less reflex artifacts, and there was no significant difference in contrast and illumination ($P<.001$, $P=.03$, and $P=.10$, respectively). When comparing image quality at the posterior pole, the mid periphery, and the far periphery, glare increased as images were acquired from a more peripheral part of the retina. Reflex artifacts were more frequent in pseudophakic eyes. Image acquisition was also possible in children. Documentation of deep optic nerve cups in video mode conveyed a mock 3D impression.

Conclusions: Image quality of conventional fundus imaging was superior to that of smartphone-based fundus imaging, although this novel smartphone-based fundus imaging device achieved image quality high enough to document various fundus pathologies including only subtle findings. High-quality smartphone-based fundus imaging might represent a mobile alternative for fundus documentation in clinical practice.

(JMIR Mhealth Uhealth 2020;8(7):e17480) doi:10.2196/17480

KEYWORDS

smartphone-based fundus imaging; smartphone-based funduscopy; smartphone; retinal imaging; mHealth; mobile phone; smartphone imaging; smartphone funduscopy; smartphone ophthalmoscope

Introduction

Imaging of the eye using smartphones has become increasingly popular and allows for an inexpensive and mobile fundus examination and documentation [1,2]. Although initially the lens needed to be held in the optical path manually [1-3], different proprietary adapters are available now [4-7]. Smartphone-based fundus imaging can be performed based on both direct [4,5] and indirect ophthalmoscopy [6,8-10]. Given the low costs and great mobility, smartphone-based fundus imaging has the potential to revolutionize eye care, especially in lower-resource settings [11,12]. Furthermore, smartphones allow for continuous connectivity, and hence, smartphone-based fundus imaging might pave the way for telemedicine in ophthalmology [13].

With the advent of smartphone-based fundus imaging, multiple applications in ophthalmology including smartphone-based diabetic retinopathy screening [14-20] and smartphone-based optic nerve head evaluation [5,21,22] have become available. In addition, smartphone-based fundus imaging has successfully been used for documentation of hypertensive retinopathy [23] and its utility for fundus documentation in a pediatric emergency department has been evaluated [24]. Furthermore, direct smartphone-based fundus imaging might have merits in teaching medical students and other health care trainees fundus examination as it seems much easier to understand and master than conventional direct funduscopy [25,26].

Here, we evaluated a novel, high-quality optics adapter for smartphone-based fundus imaging in single-image acquisition or video mode for documentation of a variety of retinal/vitreous pathologies in a patient sample with wide refraction and age ranges.

Methods

Subject Recruitment

Patients were consecutively recruited from the retina outpatient clinic at the Department of Ophthalmology of the University of Bonn, Germany. Ethical approval was obtained from the Ethics Committee of the University of Bonn (approval ID 209/16), and informed consent was obtained from all study participants or their legal guardians prior to study initiation. The Declaration of Helsinki was followed. Exclusion criteria were severe media opacities or any contraindications for pupil dilation.

Image Acquisition

Eyes were dilated with tropicamide (5.0 mg/mL) and phenylephrine (100 mg/mL) and imaged with the iC2 funduscope (HEINE Optotechnik GmbH & Co. KG; Figure 1) using an iPhone 6 (Apple Inc.). The smartphone-based fundus imaging device allows for a retinal field of view of up to 34 °, has a CE sign, and is classified as an ophthalmologic instrument group 1 according to the international DIN EN ISO 15004-2 standard. The weight of this handheld smartphone-based fundus imaging device is 300 g and it can be operated by one hand (focus adjustment and trigger for image acquisition), while the other hand placed on the forehead of the patient can be used to stabilize the device. Focus can be adjusted via a 2-step system: (1) a wheel at the front of the casing (Figure 1A) allows for a first manual compensation of the approximate refraction (±15 D); and (2) the autofocus of the smartphone camera achieves a fine adjustment of the focus (±3 D) after slightly pushing the image-acquisition trigger below the refraction wheel to its first stage. The image is then acquired by pushing the trigger to its second stage. Supported smartphones include the iPhone 8, 7, 6, 6s, 5s, and 5se. Images can be exported from the smartphone for further use. The app allows for a quick examination for immediate imaging without the need of entering patient data beforehand. Fundus illumination is adjustable via three brightness levels (50%, 75%, and 100% illumination). The device comes with an iOS app and the iPhone and smartphone-based fundus imaging device are connected via Bluetooth. Smartphone-based fundus imaging was performed by the same examiner (MWMW) in all patients in a darkened room. No scleral indentation was used for peripheral examination. Smartphone-based fundus imaging was performed in single-image acquisition (round images, image resolution of 2448 × 3264 pixels) or video mode (slightly narrowed images, image resolution of 1248 × 1664 pixels). In addition, a subgroup of the eyes were dilated and imaged with conventional fundus imaging (VISUCAM 500; Carl Zeiss Meditec). In case fundus images were optimized for brightness, the postprocessing was equally performed in the whole image, with no selective adjustments in parts of the image made.

Figure 1. Optical setup of the smartphone-based funduscopy device evaluated in this study. The (A) smartphone-based fundus imaging device (iC2 funduscope; HEINE Optotechnik GmbH & Co. KG) evaluated in this study has a separate optical path for the illumination system and the imaging system. (B) Optical paths in the plane of the pupil (I=illumination's light path, II=imaging system's light path) with approximate scale. (C) The optical setup consists of four aspheric lenses and a polarized illumination source with a 3000-K light-emitting diode. The device allows for a field-of-view of up to 34 °. Lens 3 is adjustable for compensations of refractions up to ±15 D.

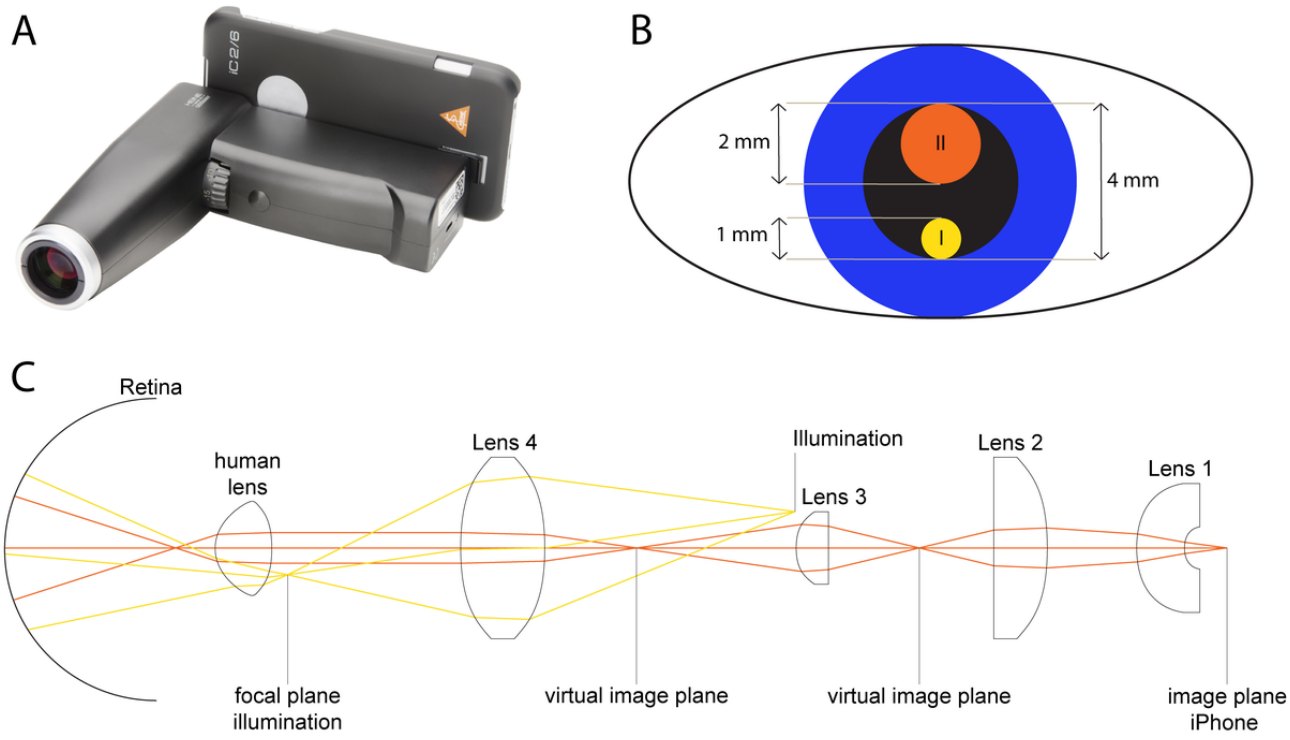


Image Analysis

Images were qualitatively analyzed for sharpness as previously reported [22]; reflex artifacts (on a scale from 0 to 3; 3=reflex artifacts in <10% of the field of view; 2=reflex artifacts in >10% of the field of view; 1=reflex artifacts in >30% of the field of view; and 0=reflex artifacts in >50% of the field of view); and contrast and illumination (on a scale from 1 to 3; 3=all vessels contrast well against fundus pigmentation; 2=most vessels contrast against fundus pigmentation; and 1=only few retinal vessels visible). Usable field of view, which was not blocked by image artifacts, was assessed. The Kruskal–Wallis test was used for multiple comparison between groups for nonparametric data.

Results

A total of 47 eyes of 32 participants with healthy eyes or the following vitreous/retinal disease were recruited (see Table 1 for characteristics of the sample): healthy (8 eyes), exudative age-related macular degeneration (6 eyes), retinal detachment (4 eyes), adult-onset vitelliform macular dystrophy (4 eyes), intermediate-stage age-related macular degeneration (3 eyes), status post retinal surgery (3 eyes), dome-shaped maculopathy (2 eyes), choroidal nevus (2 eyes), uveal coloboma (2 eyes), central serous chorioretinopathy (2 eyes), central retinal artery occlusion (1 eye), retrohyaloidal hemorrhage (1 eye), Coats disease (1 eye), status post blunt ocular trauma with choroidal rupture (1 eye), posterior uveitis (1 eye), branch retinal vein occlusion (1 eye), retinoschisis with retinal detachment (1 eyes), status post retinal gunshot injury (1 eye), asteroid hyalosis (1 eye), peripheral retinal hole (1 eye), and endogenous endophthalmitis (1 eye). Visualization of the fundus using smartphone-based fundus imaging was possible in all eyes following dilation, irrespective of refraction. Image quality was sufficient to document various specific fundus pathologies including pathologies with only subtle findings as in macular dystrophy, as demonstrated in the exemplary images of the posterior pole in Figure 2. A total of 14 eyes were imaged by both smartphone-based fundus imaging and conventional fundus imaging (see Figure 3 for exemplary images). Images acquired with conventional fundus imaging were sharper and had less reflex artifacts, compared with images acquired by smartphone-based fundus imaging (P<.001 and P=.03; Figure 4). There was no significant difference in contrast and illumination (P=.10; Figure 4), and usable field of view was not different between smartphone-based fundus imaging and conventional fundus imaging—0.99% (0.03%) and 100% (0%), respectively (P=.07).

Adjustable depth of focus allowed for documentation of retinal detachments in varying focal planes (Figure 5). Applicability in children was demonstrated by imaging a 7-year-old child with Coats disease (Figure 6). We found image quality to be limited in the far periphery, but sufficient to detect the Coats-specific pathological alterations (Figures 6C and 6D). In contrast to the far periphery, the midperiphery could be imaged with only slightly reduced image quality as compared with the posterior pole (Figure 7).

Documentation of pathology was more easily achieved using the video mode than the single-image acquisition mode. Multimedia Appendix 1 shows an exemplary smartphone-based

fundus examination in video mode where first the healthy right eye and then the left eye exhibiting vitreous synchysis were imaged. Furthermore, smartphone-based fundus imaging in video mode with slightly altering camera directions conveys a 3D impression of structures such as deepened optic nerve cups (Multimedia Appendix 2). We termed this mock 3D impression *pseudo-biomicroscopic effect.*

Glare of the image periphery (Figures 2F and 7B) and minor reflexes (Figures 5A and 5B) were image artifacts that we encountered during both single-image acquisition and video mode, with the latter artifact being predominantly present in pseudophakic eyes. Glare increased as images were acquired from more peripheral parts of the retina (Figures 6C and 6D). Smartphone-based fundus imaging with this device was also possible without pupil dilation but we found it to be more difficult and time consuming (>2 minutes for one image of the posterior pole); furthermore, reduced field of view due to increased glare limits its application (see Figure 8 for an exemplary image acquired without pupil dilation).

Table 1. Characteristics of the study sample (N=47).

Characteristic	Value
Age (years), mean (SD); range	62.3 (19.8); 7 to 93
Sex, n (%)	
Male	29 (62)
Female	18 (38)
Lens status, n (%)	
Phakic, clear lens	33 (70)
Pseudophakic	14 (30)
Visual acuity (logMAR), mean (SD); range	0.48 (0.66); 0 to 2.3
Refraction (spherical equivalent; D), mean (SD); range	–0.78 (3.21); –7.88 to 7.0

Figure 2. Exemplary smartphone-based fundus images of the posterior pole. (A) Healthy eye with slight peripapillary atrophy, (B) central retinal artery occlusion with a cherry red spot, (C) multifocal preretinal subhyaloidal hemorrhage, (D) adult vitelliform macular dystrophy, (E) massive subretinal combined with subhyaloidal hemorrhage, and (F) traumatic peripapillary choroidal rupture.

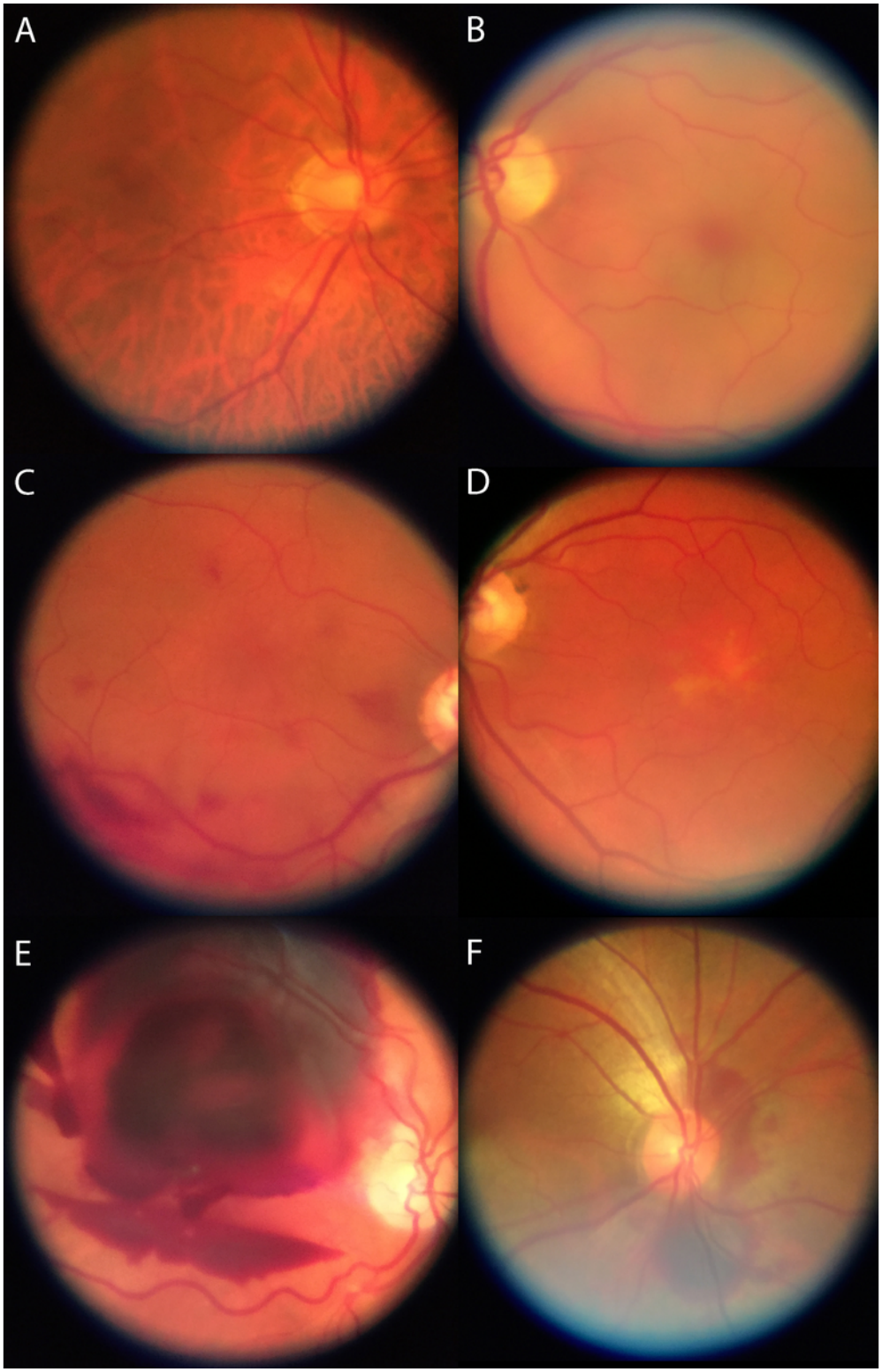


Figure 3. Exemplary comparison of smartphone-based fundus imaging and conventional fundus imaging. Exemplary comparison of two eyes (A+B and C+D) which were both imaged with smartphone-based fundus imaging (A and C) and conventional fundus imaging (B and D).

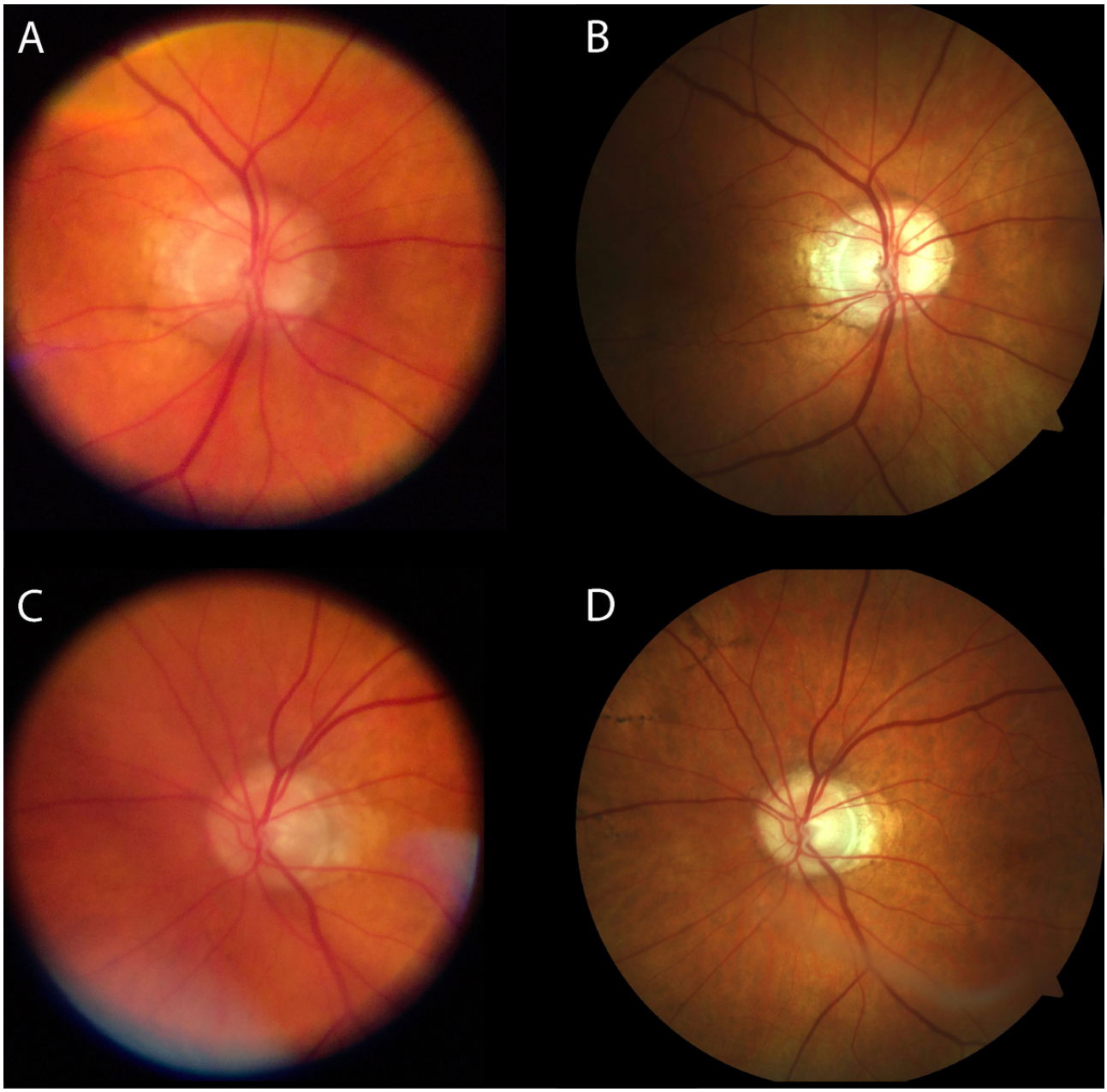


Figure 4. Comparison of image quality between smartphone and conventional fundus imaging. Image quality in terms of sharpness (left; semiquantitative scale from 1 to 5 as previously reported [22]), reflex artifacts (middle; scale from 0 to 3; 3=reflex artifacts in <10% of the field of view; 2=reflex artifacts in >10% of the field of view; 1=reflex artifacts in >30% of the field of view; 0=reflex artifacts in >50% of the field of view), and contrast and illumination (right; scale from 1 to 3; 3=all vessels contrast well against fundus pigmentation; 2=most vessels contrast against fundus pigmentation; 1=only few retinal vessels visible) is shown for smartphone-based fundus imaging and conventional fundus imaging.

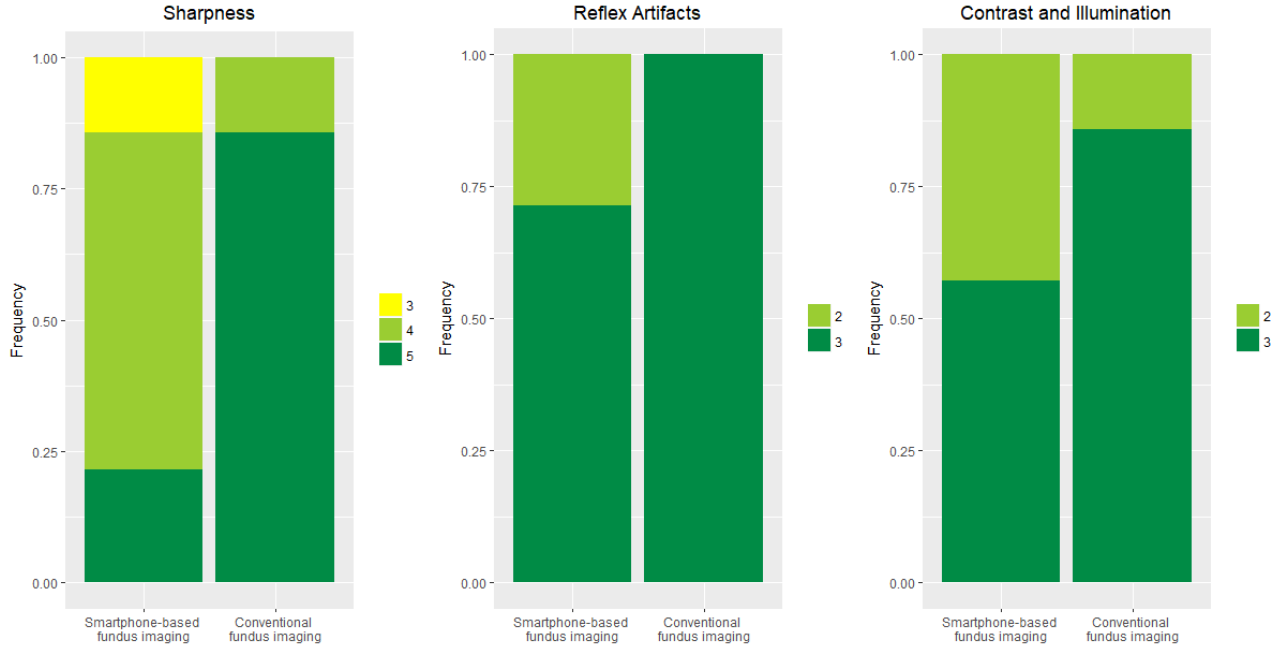


Figure 5. Documentation of different focal planes with smartphone-based fundus imaging in a total retinal detachment. With varying focus, different focal planes (indicated by black arrows) from (A) the optic disc to (B) nasally detached retina to (C) temporal bulla of detached retina can be documented. Note the characteristic reflex artifact in (A) and (B) which can predominantly be seen in pseudophakic eyes.

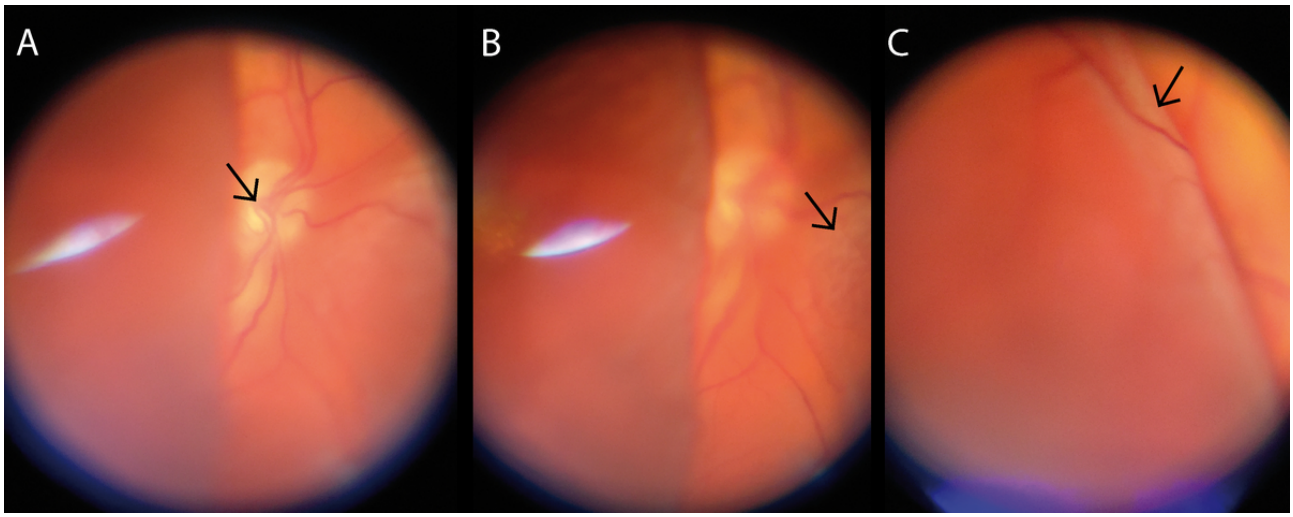


Figure 6. Smartphone-based fundus images from a 7-year-old child with Coats disease. (A) and (B) exudates superior from the upper arcade; (C) and (D) massive exudation and telangiectasia of retinal vessels in the far superior-temporal periphery, which were also funduscopically difficult to see.

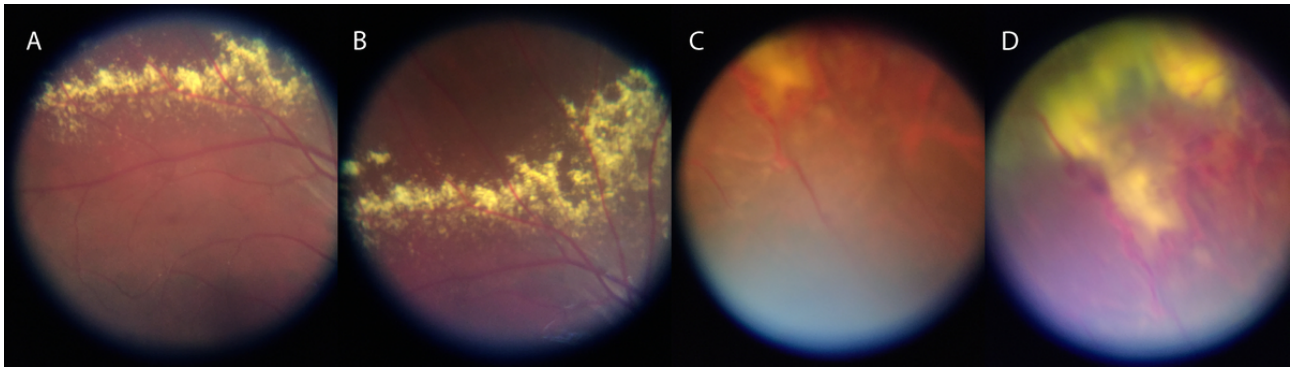


Figure 7. Exemplary smartphone-based fundus images from the mid periphery. (A) and (B) Choroidal nevi, (C) choroidal detachment (the out-of-focus area indicated by the black arrow), and (D) a subretinal demarcation line (high water mark) in a case of longstanding retinal detachment.

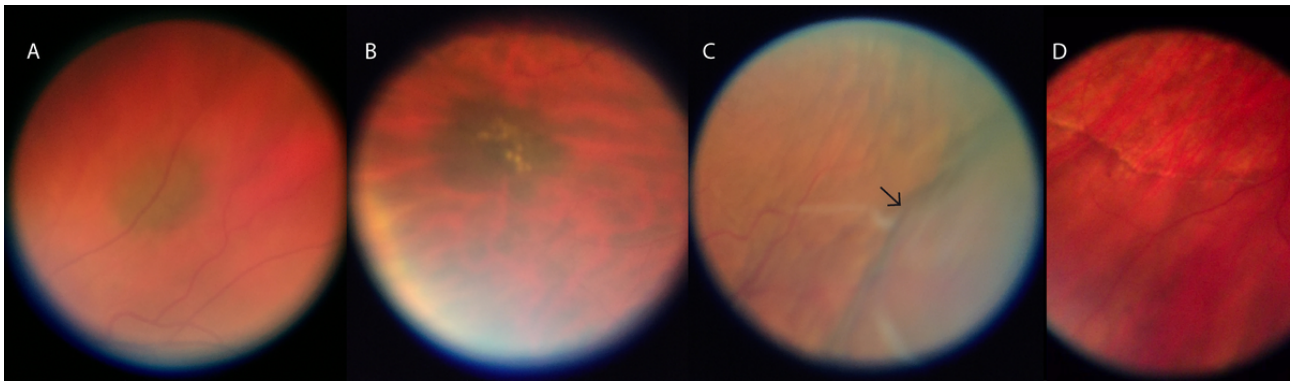
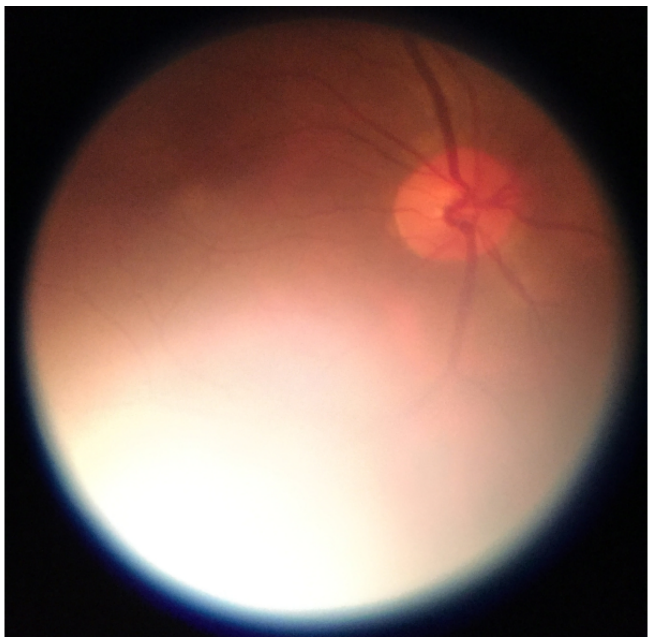


Figure 8. Smartphone-based fundus image of the posterior pole acquired without pupil dilation. Compared to imaging with pupil dilation, field of view is constrained due to increased glare from the image periphery.



Discussion

Smartphone-based fundus imaging with this device allows for retinal imaging of a diverse group of retinal pathologies in adults and children. Best image quality is achieved at the posterior pole and decreases with more peripheral imaging. Use of the video mode had several advantages over single-image acquisition mode but comes with a higher data volume, reduced image resolution, and the need for postprocessing (ie, selection of relevant parts of the video).

Our study further supports existing data showing that retinal imaging using smartphone-based fundus imaging can be successfully applied to a variety of ophthalmologic diseases and might be implemented in health care systems

[5,14-19,21-24]. Smartphone-based fundus imaging might be especially applicable in children as they are often familiar with and interested in smartphones. For example, it has been shown that smartphone-based fundus imaging can be used in a pediatric emergency department setting for screening for fundus pathologies [24].

Image quality of conventional fundus imaging was superior to that of smartphone-based fundus imaging, yet smartphone-based image quality was sufficient to document various fundus pathologies, including only subtle findings. Imaging the peripheral retina is possible; however, due to optical limitations—also related to the eye not being a perfect optical system [27]—image quality is reduced. Glare increased with more peripheral imaging and was particularly pronounced without pupil dilation. This might indicate a correlation of image glare with the effective aperture in the optical path. As the viewing angle increases in imaging of the mid and especially far retinal periphery, the effective aperture at the level of the pupil gets smaller and exhibits asymmetric distortions [28]; therefore, among other imaging artifacts, light scatter increases, which leads to increased glare in the image periphery. In addition, we observed characteristic small reflex artifacts which were predominantly seen in pseudophakic eyes, albeit not reducing overall image quality. Cataract might be a further confounder of image quality in smartphone-based fundus imaging; however, there were no major lens opacities in our study population.

Smartphone-based fundus imaging without pupil dilation might appear promising as no topical application of drugs is needed and examination can be performed directly without any waiting time. Although this device was not specifically built for imaging without pupil dilation, we showed that it is in principle possible, but not feasible due to reduced image quality and increased examination time. Yet, there are other approaches for smartphone-based fundus imaging which are capable of image acquisition without pupil dilation [4,24,29]. However, pupil dilation is still recommended to achieve best image quality in smartphone-based fundus imaging [22].

Smartphone-based fundus imaging using this device allows for retinal imaging of a diverse group of retinal pathologies in adults and children. Using the video mode has several advantages over single-image acquisition mode but comes with a higher data volume, reduced image resolution, and the need for postprocessing. We showed that video documentation of deepened optic disc cups conveys a mock 3D impression, which we termed *pseudo-biomicroscopic effect*. Although smartphone-based fundus imaging using video mode has already been reported [30], this is the first time this mock 3D impression is described. The pseudo-biomicroscopic effect might aid evaluation of optic nerve heads, for example, in the setting of a lower-resource glaucoma screening.

The strengths of our study are a comparison with conventional fundus imaging, semiquantitative analyses of image quality, a detailed description of the device, an application to a variety of different retinal diseases, wide refraction and age ranges of participants, and a novel approach for video-based documentation of the 3D structure of the optic nerve cup. However, we did not compare this device with other smartphone-based fundus imaging adapters, had a relatively small sample size, and did not evaluate applicability in eyes with clinically relevant lens opacities.

Image quality of conventional fundus imaging was superior to that of smartphone-based fundus imaging in terms of sharpness and reflex artifacts, although this novel smartphone-based fundus imaging device achieved image quality high enough to document various fundus pathologies, including only subtle findings. Hence, smartphone-based fundus imaging with this device might represent a mobile alternative for high-quality fundus documentation, for example, in immobilized intensive care patients. An image quality comparison between different smartphone-based fundus imaging approaches and further studies on the applicability of this device are warranted.

Authors' Contributions

MWMW was responsible for the literature review. MWMW and LGJ analyzed the data. MWMW wrote the manuscript and prepared the figures. RPF supervised this research project. All authors critically reviewed the manuscript.

Conflicts of Interest

Else Kröner-Fresenius-Stiftung/German Scholars Organization (EKFS/GSO 16) provided funding to RPF and the BONFOR GEROK Program, Faculty of Medicine, University of Bonn (Grant No. O-137.0028) provided funding to MWMW. Heine Optotechnik GmbH provided funding and the device used in this study to MWMW. MWMW received a travel grant from DigiSight Technologies, was a consultant for and received a grant from Heine Optotechnik, and received imaging devices from D-EYE. FGH was a consultant for Heidelberg Engineering, received a grant from Novartis, Bayer, GeneTech, Acucela, Alcon, Allergan, and Optos, personal fees from Bayer, GeneTech, Acucela, Boehringer Ingelheim, Alcon, Allergan, and Optos, and nonfinancial support from Optos and Carl Zeiss Meditec Inc.; RPF was a consultant for Bayer, Novartis, Santen, Opthea, Novelion, Retina Implant, and Oxford Innovation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Multimedia Appendix 1

Exemplary smartphone-based fundus examination of a healthy eye and an eye with vitreous synchysis. Both eyes are from the same patient and first the healthy right eye and then the left eye exhibiting vitreous body synchysis are documented. The second eye is examined from 27 seconds onward.

[MP4 File (MP4 Video), 68335 KB - mhealth_v8i7e17480_app1.mp4]

Multimedia Appendix 2
Pseudo-biomicroscopic effect. An eye with a deepened optic nerve cup was examined with smartphone-based fundus imaging in video mode. Examination from slightly altered directions generates a mock 3D impression.
[MP4 File (MP4 Video), 22088 KB - mhealth_v8i7e17480_app2.mp4]

References

1. Lord RK, Shah VA, San Filippo AN, Krishna R. Novel uses of smartphones in ophthalmology. Ophthalmology 2010 Jun;117(6):1274.e1-1274.e3. [doi: 10.1016/j.ophtha.2010.01.001]
2. Ryan ME, Rajalakshmi R, Prathiba V, Anjana RM, Ranjani H, Narayan KV, et al. Comparison Among Methods of Retinopathy Assessment (CAMRA) Study. Ophthalmology 2015 Oct;122(10):2038-2043. [doi: 10.1016/j.ophtha.2015.06.011]
3. Bastawrous A. Smartphone fundoscopy. Ophthalmology 2012 Feb;119(2):432-433.e2 [FREE Full text] [doi: 10.1016/j.ophtha.2011.11.014] [Medline: 22305322]
4. Russo A, Morescalchi F, Costagliola C, Delcassi L, Semeraro F. A novel device to exploit the smartphone camera for fundus photography. J Ophthalmol 2015;2015:1-5. [doi: 10.1155/2015/823139]
5. Bastawrous A, Giardini ME, Bolster NM, Peto T, Shah N, Livingstone IAT, et al. Clinical validation of a smartphone-based adapter for optic disc imaging in Kenya. JAMA Ophthalmol 2016 Feb 01;134(2):151. [doi: 10.1001/jamaophthalmol.2015.4625]
6. Sharma A, Subramaniam SD, Ramachandran K, Lakshmikanthan C, Krishna S, Sundaramoorthy SK. Smartphone-Based Fundus Camera Device (MII Ret Cam) and Technique with Ability to Image Peripheral Retina. European Journal of Ophthalmology 2015 May 08;26(2):142-144. [doi: 10.5301/ejo.5000663]
7. DeBuc DC. The role of retinal imaging and portable screening devices in tele-ophthalmology applications for diabetic retinopathy management. Curr Diab Rep 2016 Nov 14;16(12):132. [doi: 10.1007/s11892-016-0827-2] [Medline: 27841014]
8. Maamari RN, Keenan JD, Fletcher DA, Margolis TP. A mobile phone-based retinal camera for portable wide field imaging. Br J Ophthalmol 2014 Apr;98(4):438-441. [doi: 10.1136/bjophthalmol-2013-303797] [Medline: 24344230]
9. Hong SC. 3D printable retinal imaging adapter for smartphones could go global. Graefes Arch Clin Exp Ophthalmol 2015 Apr 22;253(10):1831-1833. [doi: 10.1007/s00417-015-3017-z] [Medline: 25900816]
10. Ludwig C, Murthy S, Pappuru R, Jais A, Myung D, Chang R. A novel smartphone ophthalmic imaging adapter: user feasibility studies in Hyderabad, India. Indian J Ophthalmol 2016;64(3):191-200. [doi: 10.4103/0301-4738.181742] [Medline: 27146928]
11. Garg SJ. Applicability of smartphone-based screening programs. JAMA Ophthalmol 2016 Feb 01;134(2):158-159. [doi: 10.1001/jamaophthalmol.2015.4823] [Medline: 26605498]
12. Lodhia V, Karanja S, Lees S, Bastawrous A. Acceptability, usability, and views on deployment of Peek, a mobile phone mHealth intervention for eye care in Kenya: qualitative study. JMIR Mhealth Uhealth 2016 May 09;4(2):e30 [FREE Full text] [doi: 10.2196/mhealth.4746] [Medline: 27160779]
13. Jin K, Lu H, Su Z, Cheng C, Ye J, Qian D. Telemedicine screening of retinal diseases with a handheld portable non-mydriatic fundus camera. BMC Ophthalmol 2017 Jun 13;17(1):89. [doi: 10.1186/s12886-017-0484-5] [Medline: 28610611]
14. Russo A, Morescalchi F, Costagliola C, Delcassi L, Semeraro F. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy. Am J Ophthalmol 2015 Feb;159(2):360-364.e1. [doi: 10.1016/j.ajo.2014.11.008] [Medline: 25447109]
15. Bolster NM, Giardini ME, Bastawrous A. The diabetic retinopathy screening workflow: potential for smartphone imaging. J Diabetes Sci Technol 2015 Nov 23;10(2):318-324. [doi: 10.1177/1932296815617969] [Medline: 26596630]
16. Rajalakshmi R, Arulmalar S, Usha M, Prathiba V, Kareemuddin KS, Anjana RM, et al. Validation of smartphone based retinal photography for diabetic retinopathy screening. PLoS One 2015 Sep 24;10(9):e0138285. [doi: 10.1371/journal.pone.0138285] [Medline: 26401839]
17. Micheletti JM, Hendrick AM, Khan FN, Ziemer DC, Pasquel FJ. Current and next generation portable screening devices for diabetic retinopathy. J Diabetes Sci Technol 2016 Feb 16;10(2):295-300 [FREE Full text] [doi: 10.1177/1932296816629158] [Medline: 26888973]
18. Toy BC, Myung DJ, He L, Pan CK, Chang RT, Polkinhorne A, et al. Smartphone-based dilated fundus photography and near visual acuity testing as inexpensive screening tools to detect referral warranted diabetic eye disease. Retina 2016;36(5):1000-1008. [doi: 10.1097/iae.0000000000000955] [Medline: 26807627]
19. Bhat S, Bhaskaranand M, Ramachandra C, Margolis T, Fletcher D, Solanki K. Fully-automated diabetic retinopathy screening using cellphone-based cameras. In: Invest Ophthalmol Vis Sci. 2015 Presented at: ARVO Annual Meeting; 05/03/2015-05/07/2015; Denver, Colorado p. 1428.
20. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond) 2018 Jun;32(6):1138-1144 [FREE Full text] [doi: 10.1038/s41433-018-0064-9] [Medline: 29520050]

21. Russo A, Mapham W, Turano R, Costagliola C, Morescalchi F, Scaroni N, et al. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading vertical cup-to-disc ratio. J Glaucoma 2016 Sep;25(9):e777-e781. [doi: 10.1097/IJG.0000000000000499] [Medline: 27513903]
22. Wintergerst MWM, Brinkmann CK, Holz FG, Finger RP. Undilated versus dilated monoscopic smartphone-based fundus photography for optic nerve head evaluation. Sci Rep 2018 Jul 06;8(1):10228 [FREE Full text] [doi: 10.1038/s41598-018-28585-6] [Medline: 29980724]
23. Muiesan ML, Salvetti M, Paini A, Riviera M, Pintossi C, Bertacchini F, et al. Ocular fundus photography with a smartphone device in acute hypertension. J Hypertens 2017 Aug;35(8):1660-1665. [doi: 10.1097/HJH.0000000000001354] [Medline: 28306635]
24. Day LM, Wang SX, Huang CJ. Nonmydriatic fundoscopic imaging using the pan optic iExaminer system in the pediatric emergency department. Acad Emerg Med 2017 May;24(5):587-594 [FREE Full text] [doi: 10.1111/acem.13128] [Medline: 27801997]
25. Hakimi AA, Lalehzarian SP, Lalehzarian AS, Dallalzadeh LO, Boodaie BD, Oggoian R, et al. The utility of a smartphone-enabled ophthalmoscope in pre-clinical fundoscopy training. Acta Ophthalmol 2019 Mar;97(2):e327-e328 [FREE Full text] [doi: 10.1111/aos.13934] [Medline: 30302917]
26. Mamtora S, Sandinha MT, Ajith A, Song A, Steel DHW. Smart phone ophthalmoscopy: a potential replacement for the direct ophthalmoscope. Eye (Lond) 2018 Nov;32(11):1766-1771 [FREE Full text] [doi: 10.1038/s41433-018-0177-1] [Medline: 30042410]
27. Gubisch RW. Optical performance of the human eye. J. Opt. Soc. Am 1967 Mar 01;57(3):407. [doi: 10.1364/josa.57.000407]
28. Fedtke C, Manns F, Ho A. The entrance pupil of the human eye: a three-dimensional model as a function of viewing angle. Opt. Express 2010 Oct 07;18(21):22364-22376. [doi: 10.1364/oe.18.022364] [Medline: 20941137]
29. Toslak D, Thapa D, Chen Y, Erol MK, Paul Chan RV, Yao X. Trans-palpebral illumination: an approach for wide-angle fundus photography without the need for pupil dilation. Opt Lett 2016 Jun 15;41(12):2688-2691 [FREE Full text] [doi: 10.1364/OL.41.002688] [Medline: 27304264]
30. Toslak D, Ayata A, Liu C, Erol MK, Yao X. Wide-field smartphone fundus video camera based on miniaturized indirect ophthalmoscopy. Retina 2018 Feb;38(2):438-441 [FREE Full text] [doi: 10.1097/IAE.0000000000001888] [Medline: 29095361]

Edited by G Eysenbach; submitted 16.12.19; peer-reviewed by H Philippin, C Perera; comments to author 10.03.20; revised version received 17.03.20; accepted 10.04.20; published 29.07.20.

Please cite as:
Wintergerst MWM, Jansen LG, Holz FG, Finger RP
A Novel Device for Smartphone-Based Fundus Imaging and Documentation in Clinical Practice: Comparative Image Analysis Study
JMIR Mhealth Uhealth 2020;8(7):e17480
URL: https://mhealth.jmir.org/2020/7/e17480
doi:10.2196/17480
PMID:32723717

Original Paper

Co-designing a Self-Management App Prototype to Support People With Spinal Cord Injury in the Prevention of Pressure Injuries: Mixed Methods Study

Julia Amann[1,2], PhD; Maddalena Fiordelli[1,3], PhD; Mirjam Brach[1,4], MPH, MBA; Sue Bertschy[1,4], PhD; Anke Scheel-Sailer[5], MD; Sara Rubinelli[1,4], PhD

[1]Swiss Paraplegic Research, Nottwil, Switzerland

[2]Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[3]Institute of Communication and Health, University of Lugano, Lugano, Switzerland

[4]Department of Health Sciences and Medicine, University of Lucerne and Swiss Paraplegic Research, Lucerne, Switzerland

[5]Swiss Paraplegic Center, Nottwil, Switzerland

Corresponding Author:
Julia Amann, PhD
Swiss Paraplegic Research
Guido A Zäch Strasse 4
Nottwil, 6207
Switzerland
Phone: 41 41939 ext 6567
Email: julia.amann@hest.ethz.ch

Abstract

Background: Spinal cord injury is a complex chronic health condition that requires individuals to actively self-manage. Therefore, an evidence-based, self-management app would be of value to support individuals with spinal cord injury in the prevention of pressure injuries.

Objective: The main objectives of this study were to (1) establish a co-design approach for developing a high-fidelity prototype app for the self-management of individuals with spinal cord injury, (2) design the prototype that resulted from this process, and (3) conduct the first usability assessment of the prototype app.

Methods: We adopted a co-design approach to develop an evidence-based app prototype. Starting from a preliminary content model (based on clinical guidelines for the prevention of pressure injuries) and three research-based user personas, we conducted an ideation workshop involving individuals with spinal cord injury and health care professionals. The ideation workshop formed the basis for two consecutive design sprints. The result of this co-design phase was an interactive app prototype. The prototype was evaluated in two rounds of usability testing (N=4 and N=15, respectively) using a combination of qualitative and quantitative methods.

Results: The co-design process resulted in a high-fidelity prototype with two key components: a self-management component and a communication component. The final prototype included a combination of features to support individuals with spinal cord injury in the prevention of pressure injuries, namely a smart camera, pressure injury diary, expert consultation, reminders, and knowledge repository. Findings of the usability testing showed that most participants navigated the app fluently with little back and forth navigation and were able to successfully complete a set of assigned tasks. These positive results are supported by the average system usability score achieved (78.5/100; range 47.5-95.0) and our qualitative analysis of the semistructured interviews. Despite an overall positive evaluation of the app prototype, we identified areas for improvement (eg, inclusion of a search function).

Conclusions: Individuals with spinal cord injury often need to navigate competing interests and priorities, paired with uncertainty about the accuracy and relevance of clinical recommendations. Understanding what matters to individuals with spinal cord injury can help guide the design of behavioral interventions that are useful and acceptable to these individuals in their daily lives. This study shows that involving individuals with spinal cord injury and health care professionals in co-designing a self-management app can foster knowledge cocreation at the intersection of lived experience, medical expertise, and technical solutions.

(JMIR Mhealth Uhealth 2020;8(7):e18018) doi:10.2196/18018

KEYWORDS

co-design; mHealth; eHealth; self-management; spinal cord injury; pressure injury

Introduction

Spinal cord injury is a complex chronic health condition that also makes individuals prone to several secondary complications, including pressure injuries [1-3]. In fact, pressure injuries are one of the most common secondary complications affecting community-dwelling individuals with spinal cord injury [3-5]. In addition to having a considerable impact on a person's health, quality of life, and well-being, there are also high economic costs associated with the treatment of pressure injuries [4,6]. Several risk factors associated with pressure injuries have been recognized, including sociodemographic, neurological, functional, clinical, biological, and medical care management [7]. Some of these factors such as age or lesion level are nonmodifiable. For example, natural skin aging increases the risk of developing a pressure injury.

In this context, some researchers have highlighted the key role that self-management plays in the prevention of pressure injuries [2,8]. Indeed, several guidelines and recommendations exist to guide individuals on how to prevent pressure injuries [9]. Yet, research indicates that individuals with spinal cord injury may not always follow evidence-based recommendations regarding physical activity, diet, and other preventive measures (eg, regular skin checks and pressure relief) once they return to their daily routine after the first rehabilitation [10-12]. These findings may be explained by considering that individuals with spinal cord injury adopt different prevention styles, characterized by different preventive behaviors and attitudes toward pressure injuries and prevention more generally [13]. Understanding what matters to individuals is essential to ensure that interventions are acceptable and rooted in the reality of the individuals' daily lives [14]. In light of this, it has become evident that self-management support needs to be tailored to individual prevention styles to motivate and engage individuals.

Given that self-management programs have long focused on equipping individuals with the knowledge and skills they need to manage chronic health conditions rather than seeking to understand how to best accommodate different life situations, it comes as no surprise that there is no conclusive evidence on how to best support community-dwelling individuals with spinal cord injury in the prevention of pressure injuries [15,16]. Self-management can be broadly defined as "the tasks that individuals must undertake to live with one or more chronic conditions. These tasks include having the confidence to deal with medical management, role management and emotional management of their conditions" [17]. In this paper, we consider self-management as a set of activities and behaviors that an individual actively performs or avoids so as to prevent or alleviate symptoms related to pressure injuries.

In recognizing the complexity and fragmentation of evidence-based and preference-sensitive information that is relevant for a person to effectively self-manage, it becomes clear that a conventional top-down approach to self-management support may no longer be adequate. Rather, there is a need to engage different stakeholder groups in the development of self-management programs to ensure that solutions meet the requirements for successful implementation and favorable health outcomes [14]. Accordingly, there has been a significant increase in participatory approaches to design and improve self-management programs for individuals with chronic health conditions [18-21]. A particular area of research that has developed from this trend for participatory approaches is the co-design of mobile health (mHealth) apps [22-25]. Several recently published studies present different approaches to co-designing self-management apps for health conditions, including asthma [26], cystic fibrosis [27], sickle cell disease [23], and spinal cord injury [28]. Arguably, co-design has intuitive appeal and is promoted on a political level. The normative assumption is that the outcome of solutions that are co-designed by users and professionals ought to be better [29]. Yet, there is limited research on how to effectively engage medically fragile populations in generating user specifications for mHealth tools to foster self-management [30,31].

Guided by this literature, we aimed to develop an evidence-based, self-management app to support individuals with spinal cord injury in the prevention of pressure injuries. In this paper, we describe the co-design approach used, present the app prototype that resulted from this process, and report the findings of a first usability assessment. We conclude by critically reflecting on co-design as an approach to enrich the development of self-management apps for individuals with disabilities more generally. This study constitutes part of a larger project on the prevention of pressure injuries in community-dwelling individuals with spinal cord injury in Switzerland.

Methods

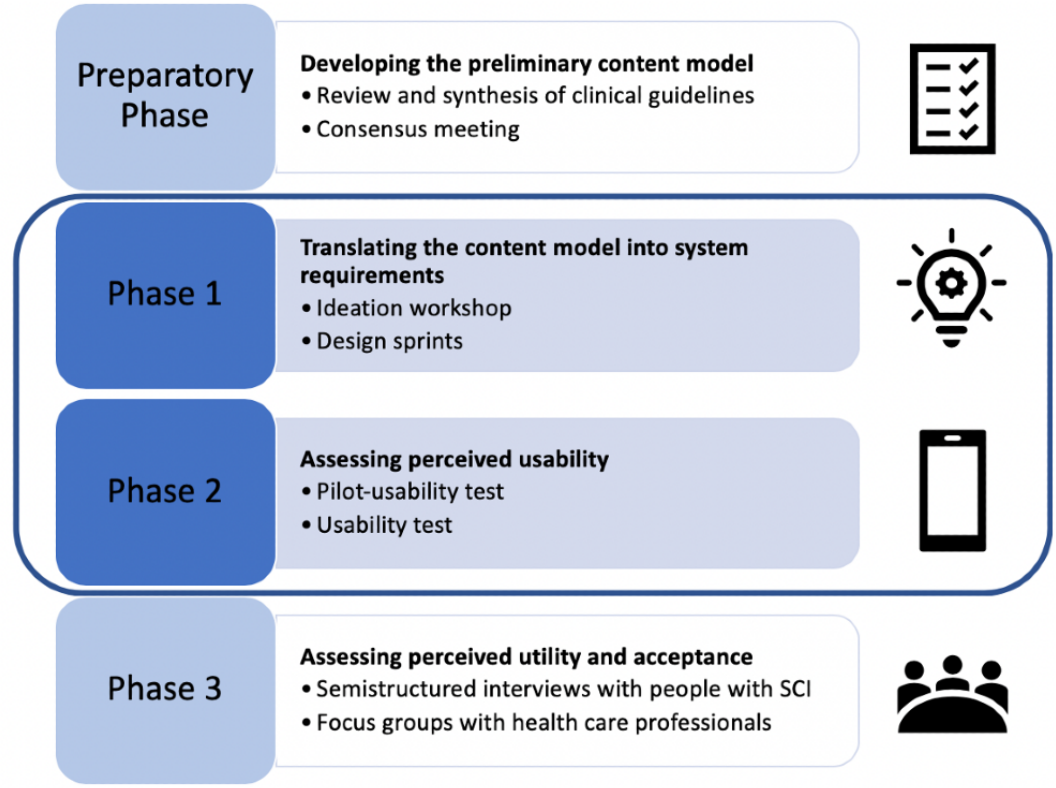
Study Design

We followed a co-design and development approach similar to that described by Gray et al [30], incorporating qualitative research methods (ie, semistructured interviews) into user-centered design approaches (ie, ideation workshop, design sprints, usability tests). The study was approved by the Cantonal Ethics Commission (EKNZ 2017-01787).

In a preparatory phase, we involved different stakeholders in assessing available clinical guidelines for the prevention of pressure injuries in people with spinal cord injury. Based on the findings of this preparatory phase, we developed a preliminary content model for the app prototype. In Phase 1, this preliminary content model was translated into concrete functional and nonfunctional requirements of the app prototype. In Phase 2, we performed a usability test to assess the perceived usability of the app prototype. In Phase 3, we conducted semistructured interviews and focus groups with health care professionals to assess the utility of the app prototype and to

determine whether it could be implemented in practice. An overview of our methodological approach is displayed in Figure 1. The present paper describes the outcomes of Phase 1 and Phase 2. Methods and findings of the preparatory phase have been published elsewhere [9].

Figure 1. Co-design and development approach. SCI: spinal cord injury.



Preparatory Phase: Developing a Preliminary Content Model

To identify the most relevant guidelines for the prevention of pressure injuries that community-dwelling individuals with spinal cord injury can and should perform, we held a consensus meeting involving 15 health care professionals specialized in spinal cord injury. In addition, two individuals with spinal cord injury and an insurance representative participated in the meeting. The consensus meeting resulted in a set of 98 guidelines spread across 12 categories: (A) Support surface, (B) Repositioning, (C) Nutrition, (D) Skincare, (E) Skin assessment, (F) Exercising, (G) Collaboration with health professionals/caregivers, (H) Transfers, (I) Clothing, (J) Body function and structure, (K) Personal factors, and (L) General. This selection of clinical guidelines formed the preliminary content model of the app (for detailed results see [9]). In other words, these guidelines served as building blocks that would guide the development of the evidence-based app content.

Participants (Phase 1 and Phase 2)

Through the different phases of the project, we aimed for balanced participant samples in terms of age (>18 years), gender, and lesion level, and health care professionals' expertise. Individuals with spinal cord injury and health care professionals were recruited with the support of health professionals from the Swiss Paraplegic Center (in particular members of the interdisciplinary Decubitus-Care team), participants of the consensus meeting described above, as well as through support of the Swiss Paraplegic Association and Parahelp, a home care service provider specialized in spinal cord injury care. This recruitment strategy was complemented by an online call for participants published through the Paraplegie Community [32], an online community for individuals with spinal cord injury. People who expressed interest to participate in the study were contacted by email or telephone. All participants received detailed study information and were asked to sign a consent form. We did not offer any financial incentives. Table 1 presents information on the study populations of Phase 1 and Phase 2.

Table 1. Study participants in Phase 1 and Phase 2.

Phase	N
Ideation workshop (Phase 1)	
Project team	2
Parahelp home care provider	2
Wound specialist	1
Nutritionist	1
People with SCI[a]	5
User experience designer	2
Design sprints (Phase 1)	
Project team	4
User experience designer	2
Usability test (Phase 2)	
Pilot test	4
Usability test	15

[a]SCI: spinal cord injury.

Phase 1: Translating the Content Model Into System Requirements

The key challenge was then to translate the content model into concrete functional and nonfunctional requirements of an app prototype that would be perceived as useful and acceptable by prospective users. To this end, we collaborated closely with individuals with spinal cord injury, health care professionals specialized in spinal cord injury, as well as user experience designers. The translation process consisted of two main activities.

First, we held a 1-day ideation workshop with health care professionals, researchers, people with spinal cord injury, and user experience designers. The workshop pursued two specific aims: (1) to translate the preliminary content model into concrete functions, and (2) to foster stakeholder engagement and commitment to the project. For this purpose, we sought to gain an in-depth understanding of the challenges faced by the different stakeholders in the prevention of pressure injuries and to collect ideas on how an app could address these challenges.

The ideation workshop consisted of a set of individual and group activities commonly used in user experience design [33] and was moderated by a trained user experience designer. We aimed to encourage participants to draw on their own experiences and expertise but also to reflect on those of the other participants. Activities incorporated the evidence-based content model developed in the preparatory phase [9] and 3 research-based user personas [34,35] to stimulate discussion. User personas are fictitious characters to represent different prospective user types who are characterized by specific goals and behaviors [35]. The user personas adopted for this study described the three different prevention types that had been identified in earlier work: the thoughtful, the selective, and the delegator [13]. Each prevention type is characterized by different preventive behaviors, knowledge, and attitudes toward the prevention of pressure injuries; collaboration with health care professionals; and attitudes toward spinal cord injury in general [13]. For example, in one of the exercises, participants worked in smaller groups to create a list of app features and functionalities that would be perceived as “cool” or “not cool” by the respective type their group had been assigned. In another activity, participants were again divided into smaller groups to brainstorm how specific guidelines might be translated into concrete contents and functions (ie, identify opportunities and challenges). As individual activities, participants had to write a “love letter” and a “one-star review,” respectively, to describe what they liked and disliked about the (at this point) fictional app. Moreover, we asked participants to sketch their ideas for user interfaces to activate their creative thinking and problem-solving skills. Impressions from the ideation workshop are presented in Multimedia Appendix 1.

Following the ideation workshop, we conducted two consecutive design sprints [36]. During these design sprints, the research team and user experience designers aimed to synthesize, condense, and prioritize the ideas that were generated during the ideation workshop. For this purpose, we reviewed the material collected during the ideation workshop individually and then as a group. We revisited the user journey and potential use cases. Finally, we used a feasibility-impact matrix [37] to guide priority setting in selecting the functions to be implemented in the first iteration of the app prototype. More specifically, we mapped the different features and functionalities (written on Post-It notes) onto a physical easy/hard–low/high impact matrix to determine those that were high impact (ie, with great potential benefit for the user) and feasible (ie, easy to implement). Features that were deemed high impact but difficult to implement for technical reasons (eg, integrated ruler in the smart camera to indicate the size of the pressure injury) were not included in the first iteration of the app prototype. Following this process, the first prototype in the form of a clickable user interface was developed.

Data that were collected during the ideation workshop and design sprints included: (1) written Post-It notes, (2) participants' drawings, and (3) field notes taken by the research team. We used an online project management tool (RealTime Board) to collaboratively collate, synthesize, and analyze the materials in a comparative process akin to thematic analysis [38], together with user experience designers. In our analysis, we focused on capturing participants' needs and desires, as well as ideas for concrete functions and features. We then followed a narrative approach to group similar ideas and concepts into overarching, inductively derived categories that would ultimately form the structural model of the app prototype.

Phase 2: Assessing Perceived Usability

The central objective of Phase 2 was to examine the perceived usability and usefulness of the app from the perspective of potential service users. In this phase, we also aimed to collect suggestions for improvement and ideas for designing the functions, content, and navigation of the app. A pilot usability test was carried out with 4 individuals with spinal cord injury using a first iteration of the app prototype. Based on the findings of the pilot test, a second and more refined version of the prototype was developed and then tested by 15 individuals with spinal cord injury. Table 2 presents an overview of the participants' characteristics.

Table 2. Characteristics of study participants in Phase 2.

Characteristic	Pilot test (N=4)	Usability test (N=15)
Age (years), mean (range)	55.5 (48-64)	40.8 (28-58)
Gender, n (%)		
Male	3 (75)	11 (73)
Female	1 (25)	4 (27)
Lesion level, n (%)		
Paraplegic	2 (50)	7 (47)
Quadriplegic	2 (50)	8 (53)

Both usability tests were carried out in a laboratory setting (with one exception that was carried out at the person's home) following the same procedure (Textbox 1), which was informed by earlier work on mHealth usability studies [39]. In particular, we combined two common usability-testing approaches, task completion [40] and think-aloud technique [41,42], with additional observational data on navigation fluidity and navigation challenges. In doing so, we aimed to determine whether the app prototype was designed in such a way that users were able to fluently navigate it. We complemented these assessments with a semistructured interview and a self-administered questionnaire to gain further insights into how participants perceived the usability of the app prototype. Upon completion of the pilot test, the app prototype was also presented to the individuals with spinal cord injury and health care professionals that had taken part in the ideation workshop to share with them the outcome and to collect informal feedback.

Textbox 1. Procedure for usability testing.

Usability Test

As part of the usability test, participants were asked to complete a series of 11 tasks using the app prototype. The 11 tasks used were chosen as they represent typical use cases (eg, "Set a reminder for a mattress check for January 30, 2019"). While completing the tasks, participants were instructed to verbalize their navigation behavior ("think-aloud technique") [43] (eg, where they look for certain information or where they would expect to find it). Participants with limited hand functioning were provided with an assistive pen to facilitate navigation. A structured observation sheet was developed by the research team through internal pretesting and expert recommendation provided by the user experience designer. Prior to data collection, a research assistant received detailed coding instructions. During the usability test, the following data were collected: (1) task completion (ie, did the participant successfully complete the task, rated as concluded, partially concluded, not concluded); (2) task completion time (ie, time required to complete the task successfully, recorded as time in seconds); (3) navigation fluidity (ie, how fluently participants were navigating the app prototype, rated on a 5-point scale from "irregular"=1 to "fluent"=5); and (4) challenges related to completion of the task (observed and stated by participants, and noted down by a research assistant).

In addition, participants' navigation behavior and commentary were recorded using a screen camera, screen-capturing software, and audio-recording device (see Multimedia Appendix 2). After data collection was complete, a second coder (JA) went back to the coding to ensure intercoder reliability.

Semistructured interview

Following the usability test, participants took part in a semistructured interview guided by the technology acceptance model [44], which aimed to capture participants' views regarding usability and perceived usefulness of the app prototype, as well as their attitudes and intentions toward using the app (see Multimedia Appendix 3). Moreover, we aimed to elicit ideas and suggestions for improvement. All interviews were transcribed verbatim and analyzed using inductive thematic analysis [38].

Questionnaire

Finally, the study participants were asked to complete a short questionnaire on system usability. The questionnaire consisted of 10 items to be rated on the system usability scale (SUS), which is a 10-point Likert scale [26]. The SUS score was calculated for each participant.

Results

Phase 1: Content and Functionalities of the App Prototype

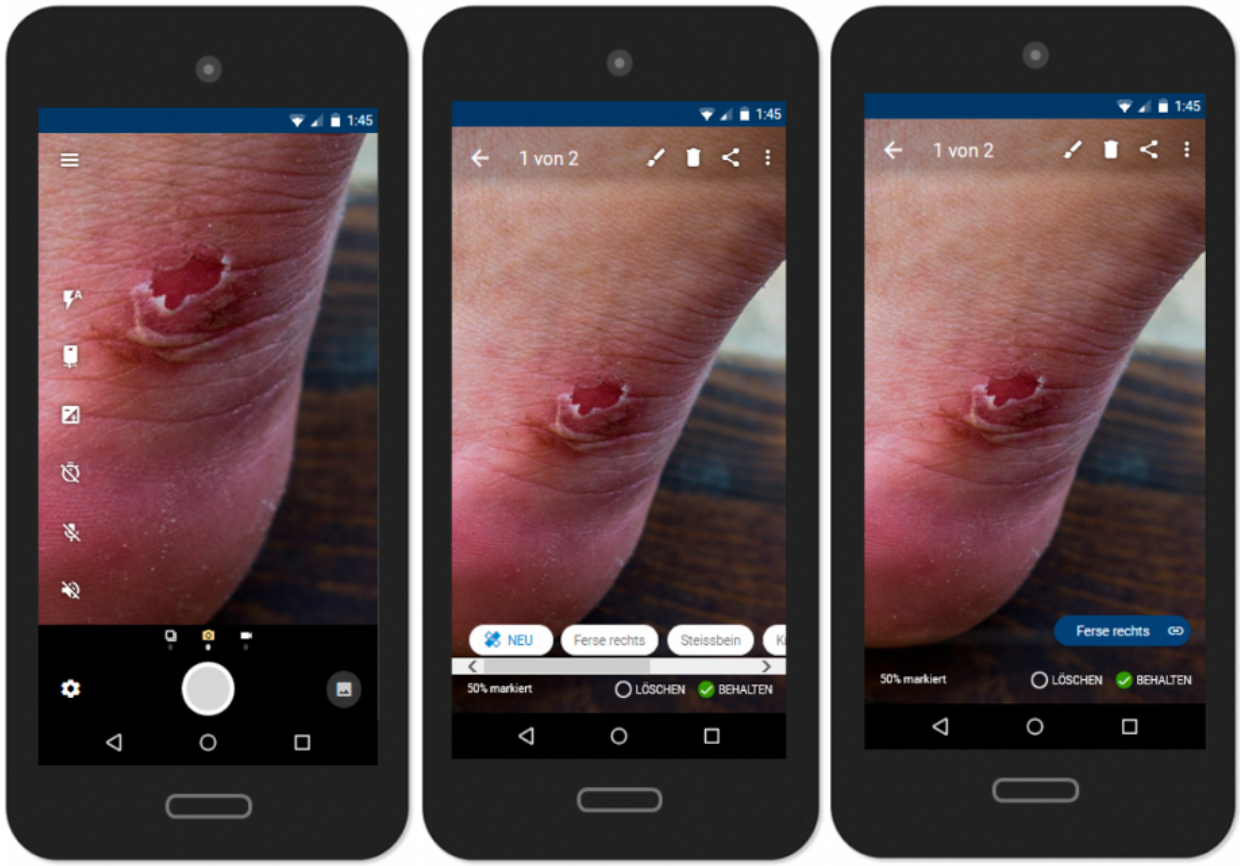
Within the scope of the co-design process, we identified two key components of the app. The first is a self-management component that would need to support individuals in documenting their pressure injuries, including visual and written information (disease monitoring); finding relevant, evidence-based information on pressure injury prevention and related topics (disease knowledge); and receiving personalized and actionable recommendations based on their preferences and needs (motivational support). The second is a communication component that would allow users to safely and quickly get in touch with a health care provider to inquire about pressure injury-related topics (eg, nutrition) and send pictures and information on pressure injuries to receive feedback.

In the subsections below, we describe the key functionalities of the app prototype that we identified as necessary to achieve the goals of these two components. An overview is provided in Multimedia Appendix 4. Note that the prototype presented here is the iteration used for the usability test in Phase 2.

Smart Camera

A central function of the app is the smart camera (Figure 2). Pictures and videos taken with the smart camera can be directly added to the pressure injury diary, the documentation function of the app, or forwarded to a health care provider.

Figure 2. Smart camera.



The need for a smart camera resulted from a vivid discussion around the poor quality of pictures often taken by people with spinal cord injuries. Particularly, health care professionals emphasized that they need high-quality pictures to be able to make reasonable judgments about the condition and state of a pressure injury when providing advice remotely. Based on these discussions, several additional features envisioned by the workshop participants emerged, including a voice-operated shutter release function, which would be particularly useful for people with limited hand functioning; a multi-shot mode; and tips and tricks for taking good pictures (eg, paying attention to light conditions, angles, and image sharpness). Despite the desire of participants to have an integrated ruler to indicate the size of the pressure injury, this function could not be implemented due to technical constraints. It was thus suggested to instead provide recommendations on how size can be inferred, for example by placing a coin next to the pressure injury or by using self-adhesive ruler tape.

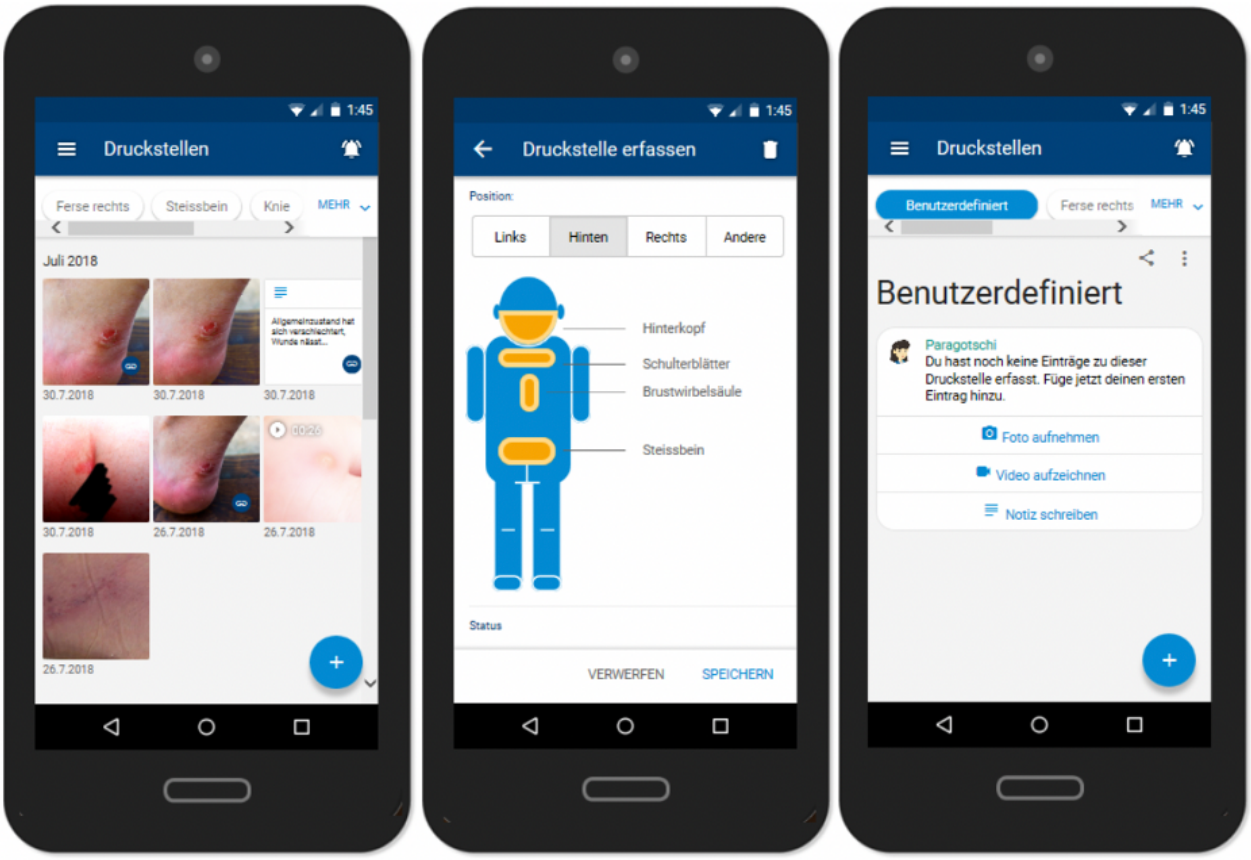
Even though health care professionals considered a function that would help users take better pictures as highly valuable, they also emphasized that pictures cannot replace a physical exam but rather provide some indication of the urgency of the situation. This was also a warning message they would like to see integrated into the final smart camera. An additional camera feature that people with spinal cord injury desired for privacy reasons was the “censoring function,” which is a tool that allows users to easily edit pictures using a black marker, which might

be desired to hide intimate body areas as an example. The individuals with spinal cord injury also emphasized that sometimes they may just want to check their skin using their smartphone camera as a mirror, rather than saving or sharing pictures. This is why pictures and videos are not saved automatically into the pressure injury diary.

Pressure Injury Diary

The pressure injury diary, as the documentation function of the app, allows users to store and manage their pictures, videos, and other relevant information such as date, time, and location of the pressure injury. The documentation function is intended to support users in the early detection and observation of conspicuous skin areas. New entries can be entered directly using the camera function (as shown in Figure 2) or in the "My pressure injuries" section (Figure 3). In addition to pictures, personal notes can be added. When adding a picture, participants are also prompted to add additional information, including the location of the pressure injury and the presumed cause of the injury, by selecting from a drop-down menu.

Figure 3. Pressure injury diary.

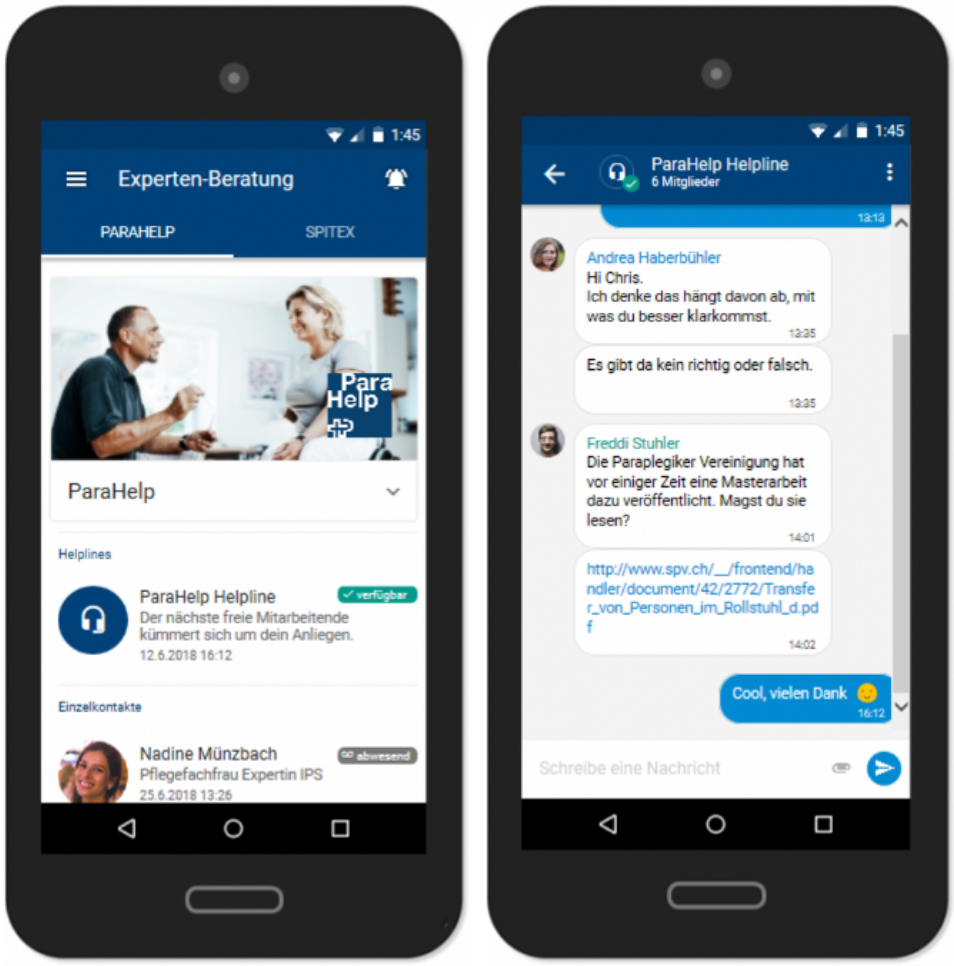


As desired by the workshop participants, a filter function was implemented to allow users to easily search for and filter entries. Being able to review, edit, share, or delete entries was underlined as very important, particularly from the service users' perspective. Another important aspect highlighted by the workshop participants was the need for "hidden picture storage," which prevents pictures from being directly transferred to the user's phone gallery. In other words, pictures taken within the app would remain within the app and would not appear anywhere else. This hidden storage was considered to serve both to simplify documentation (all photos are in one place, making it easier to find and compare pictures) and to protect the individuals' privacy. Several workshop participants mentioned how easily it could happen that an unpleasant and possibly embarrassing picture may pop up on the screen when showing a friend holiday pictures on one's phone.

Expert Consultation

Workshop participants agreed that expert consultations constitute an integral part of the app (Figure 4). They suggested that being able to record and send audio messages would be particularly helpful for people with limited hand functioning. In terms of design and functionalities, participants referred to WhatsApp as a solution that most people are fairly familiar and comfortable with. WhatsApp was described as the current status quo medium for sharing and receiving support requests, including pictures of pressure injuries. It was thus argued that to replace this well-established practice, the app would need to provide additional benefits to users.

Figure 4. Expert consultation.

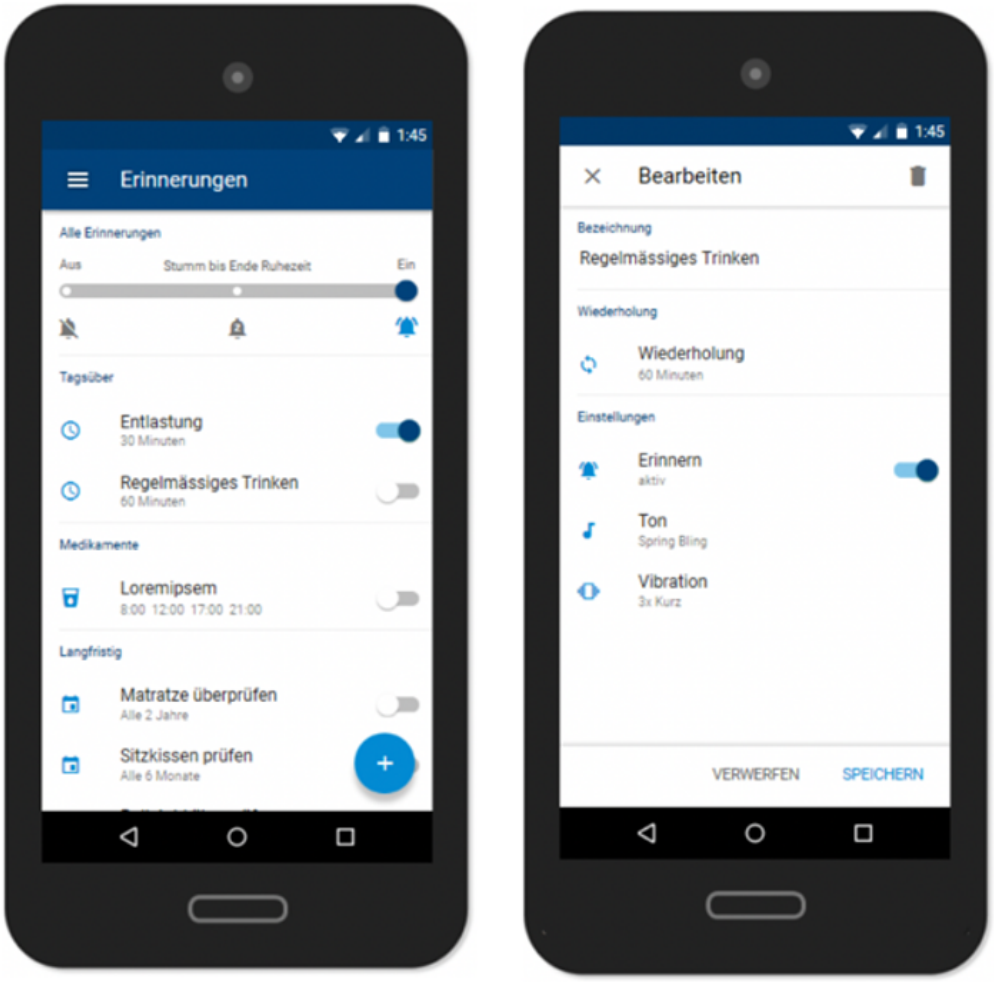


Individuals with spinal cord injury recognized the key benefit in the fact that all communication with health care professionals would be secure, thereby not jeopardizing their privacy. Health care professionals confirmed that a system that would allow them to display their absences would help to avoid common problems such as pictures not being received on time due to an employee's holiday absence. A controversial aspect discussed in this context was the level of desired privacy. Although some service users argued that they would prefer to simply send an anonymous request as a way of maintaining their autonomy and decision power, health care professionals emphasized the significance of knowing who they are providing advice to. They argued that they would first need to verify whether the person contacting them is entitled to use their service, in line with their service mandate (limited to individuals with spinal cord injury). In addition, they would need access to the person's medical history to be able to give them the best possible recommendation. However, precisely what such a registration process can and should look like was not further discussed within the scope of the workshops.

Reminders

A somewhat more debated feature of the app was the reminder functionality (Figure 5). Workshop participants agreed that many of the guidelines for the prevention of pressure injuries could be transferred into the app in the form of reminders (eg, for regular pressure relief). Although some workshop participants identified reminders as a useful support tool, including for informal caregivers, others perceived the reminders to be patronizing and annoying, especially when presented to users in a generic form. In this sense, participants described recommendations such as "You should engage in pressure relief x times a day" as poorly tailored and likely to be ineffective. Participants agreed that, ideally, the app would include sufficient information to provide them with tailored, timely interventions (ie, provide recommendations in the moment when needed rather than at predefined times) such as "At the moment pressure on your right heel is critical, please check your skin and relieve immediately to avoid a pressure injury." However, there was no consensus as to how this could be achieved in practice.

Figure 5. Reminders.



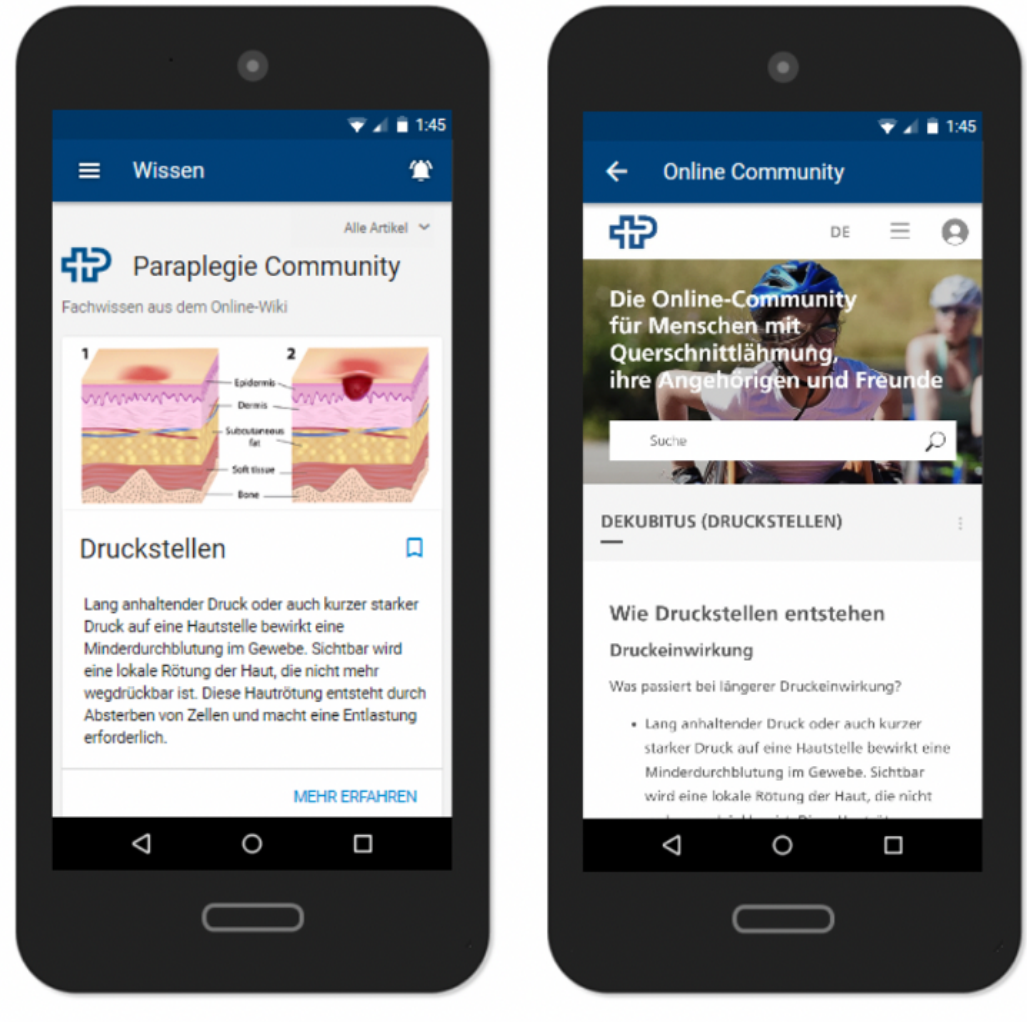
As a result of these discussions, the current app prototype provides users with a selection of reminder templates relating to everyday activities (eg, regular drinking, repositioning) and more long-term events (eg, check the seat cushion or mattress). Users can modify the frequency/date of each reminder. Users can also add new reminders and edit or delete them at any time. Workshop participants agreed that ideally these reminder settings should be completed together with a health care professional. In addition, the specification of resting periods during which signal tones and vibration functions are deactivated was designed to prevent users from receiving reminders unintentionally, for example during the night.

Knowledge Repository

In addition to specific functionalities that would translate the guidelines identified in the preparatory phase more indirectly, participants also acknowledged the importance of having a knowledge repository for finding relevant information to foster awareness of different preventive measures. It was also suggested that caregivers may benefit from being able to read up information in an easily accessible format. Workshop participants emphasized the importance of visual materials, including both images and videos. They also advised against presenting large chunks of text as users may feel overwhelmed.

In conceptualizing the knowledge repository (Figure 6), we drew on existing resources, namely the knowledge repository of the Paraplegie Community [32], which already provides a wide range of patient education material. The content is available in four languages (German, French, Italian, and English) and is regularly updated by experts in the field of spinal cord injury. Workshop participants agreed that it would be useful to directly incorporate the knowledge repository of the Paraplegie Community into the app prototype as it was both a fitting solution and a pragmatic choice.

Figure 6. Knowledge repository.



Phase 2: Perceived Usability of the App Prototype

Analysis of the task completion exercise (11 tasks) indicated a high level of usability. As shown in Table 3, most of the participants managed to complete the majority of tasks, at least partially. In cases where participants did not complete tasks, this was often caused by technical limitations of the app prototype or a misinterpretation of the task at hand. Given that participants were asked to verbalize their navigation behavior while completing the tasks (think-aloud technique), task completion times were not representative of the actual time it took participants to complete specific tasks. In some cases, participants would pause midway when performing a task to ask questions or to comment on design or navigation elements, leading to inaccurate task completion times. As a consequence, we did not analyze task completion times in-depth. Yet, our analysis of participants' navigation recordings showed that most participants navigated the app very fluently with little back and forth navigation (Table 3). These positive results are mirrored in the average system usability score we calculated from the individual participants' ratings displayed in Table 3. The app prototype received an average rating of 78.5/100 (range 47.5-95.0).

Table 3. Individual assessments of the 11 tasks.

Participant ID	Number of tasks completed	Number of tasks partially completed	Number of tasks not completed	Average navigation fluency (1-5)	SUS[a] score (0-100)
AM0708	7	2	2	4	80
ChC1308	8	1	2	4	75
FB2908	9	2	0	4.2	87.5
FH0709	9	2	0	4.6	72.5
FR0608	9	1	1	4.4	90
GB0608	8	2	1	4.5	95
JL1409	5	4	2	4.5	75
ML2908	3	3	5	4	85
MS0908	10	0	1	4.8	80
NH0908	5	2	4	4.7	77.5
SIM1709	3	4	3	4.1	90
SM1308	6	3	2	4	75
TF0708	9	1	1	4.6	70
TH0908	6	2	3	4.4	47.5
TNS2108	7	2	2	4	77.5

[a]SUS: system usability scale.

Our analysis indicated that reoccurring challenges to the completion of tasks were primarily related to the following aspects: (1) the testing device itself (Android, as iPhone users were unfamiliar with the device, its icons, and navigation system); (2) misunderstandings or misinterpretations regarding the task (eg, participant set a reminder for a different time than requested by the task); (3) confusion caused by labeling/word choice of the app prototype (eg, the term “chat” led to confusion); (4) preconceived behaviors and preconceptions (eg, in one of the tasks, participants were asked to send a picture to a health care professional. However, instead of sending the picture through the app, some participants tried to export the picture to email); and (5) malfunctioning/limited functionality of the app prototype (eg, participant would click the “share” icon, which was not activated in the prototype). No structural differences in task completion and navigation were observed in the study population.

Findings of the task completion exercise were supported by our qualitative analysis of the semistructured interviews. Despite the limited functionality of the app prototype, most participants rated the prototype’s usability quite high. They appreciated the logical, intuitive structure and the simplicity of the app prototype, as well as the familiar functionalities.

> *No, super logical. Even if you come from a different corner, I go over the camera and take a new picture, or I go over to pressure injuries and take a new picture that works, that's cool.* [MS0908]

> *Well, I like it. Also, that it has some sort of chat history. That's really something most people are familiar with. These are the functions people know. It's easy. It's clear.* [FB2908]

Although most participants rated the usability of the app prototype as high, some of its aspects were also criticized. For example, several respondents pointed out that some content was depicted in a font size that was too small, making it difficult to read, and that it would be good to have an overview of all the things the app provides rather than having to look for it. Further criticism related to the lack of a search function to identify relevant content. Participants also noted that navigation difficulties could arise, especially for people who were not yet familiar with the app's structure.

> *It might be good to see the possibilities the app offers right at the beginning. To show you what you can do, otherwise, you have to click through. I think it's fine if you are familiar with it, but when you open the app for the first time you have to search for everything.* [AM0708]

> *Exactly, so it should be clearer and that you don't have to scroll too much because you lose a lot of time and you can't really read it well if you skip something.* [TH0908]

Similar to the findings of the usability test, our qualitative analysis of the interviews indicated that some of the navigation difficulties related to the testing device itself rather than to the navigation structure of the app prototype. One of the study participants also noted that elderly users may have difficulties navigating.

> *For me, the difficulty was that I don't have a Samsung. But if you have the app, it’s on your own phone and you know how [to use it]. Of course, for elderly people, it is a bit more difficult.* [GB0806]

> *Well, it was more of a fight with the Samsung [phone]. I know the iPhone by heart. If I could've worked with the iPhone, it would've probably been faster.* [JL1409]

The navigation issue was, however, described as minimally problematic by some of the participants, claiming that it was part of a natural learning process that one goes through whenever installing a new app. In this way, they suggested that users would explore the app through trial and error to get to know and understand it better.

> *Well, the first time you access it, it's just a little bit of trying things out, and eventually you'll get how it works.* [FB2908]

> *Then you just press something, you see what happens - what is it that I'm looking for exactly. It's a bit like a natural flow guiding you […] It's somehow automatic, you just try things out as you go along.* [ML2908]

Discussion

Principal Findings

Using spinal cord injury as a case in point, this study details the process and result of engaging different stakeholder groups in the development of a self-management app. Our findings illustrate how self-management needs can be translated into meaningful technical solutions by involving the relevant stakeholders in an open dialog and creative exercises. The fact that all activities were guided by the previously developed evidence-based content model (preparatory phase) helped to ensure that the app prototype would draw on the latest scientific evidence [9]. Given that many apps that address specific disability conditions are informational and provide only limited functionality, findings of this study are of great relevance to advancing the development of mHealth solutions for people with disabilities [45].

Co-designing mHealth Solutions: Lessons Learned

In the following, we critically reflect on co-design as an approach to enrich the development of self-management apps more generally. In particular, we would like to highlight three key lessons learned from this project that are in line with recent work in the field of co-designing health services [30,46], and for specifically designing self-management support for people with spinal cord injury [28].

A first point relates to what is commonly referred to as needs assessment. Most scholars and practitioners will agree that mHealth apps should be tailored to the needs of their intended users as the end beneficiaries. For this purpose, needs assessments are carried out. Unlike other forms of needs assessments (eg, through surveys or interviews), the co-design approach we adopted in this study allowed us to understand different needs in context [31]. In other words, different stakeholders' needs were enacted, contextualized, and put into perspective through an interactive exchange among the different stakeholders involved in the prevention and treatment of pressure injuries. Our experiences reflect those of earlier work [47], highlighting that user personas can be particularly helpful in this process as they allowed participants to think beyond their personal experiences and consider different scenarios.

A second point relates to translating needs into supporting materials. In traditional, top-down patient education, experts will—based on the needs they identified—devise educational materials to address the needs of the target population. We adopted a different approach to translating needs into support materials. More specifically, we engaged the prospective users of the app both in the identification of needs *and* in the identification of technical solutions that could help to address them. In doing so, we were able to access a previously untapped source of ideas and knowledge, resulting in a rich catalog of desirable features and functionalities of the app prototype. The user experience designer greatly facilitated the co-design process by illustrating how different technical solutions could address specific requirements or issues raised by the participants (eg, relating to data protection and privacy). As demonstrated by earlier work [48], we also observed how creativity was unleashed during the ideation workshop. Creativity-focused exercises such as drawing mock-ups of user interfaces not only stimulated discussion but also led to instances of knowledge cocreation at the intersection of lived experience, medical expertise, and technical solutions, as illustrated by the description of the individual functionalities of the app prototype.

A third point relates to our learning process as a research team. While moving along the co-design process, we as a research team were continuously confronted with our own scientifically grounded assumptions about the self-management of pressure injuries. When initially conceptualizing the app prototype, for example, we placed much greater emphasis on the self-management component of the app, neglecting the importance of communication aspects, which turned out to be an essential component. Being open to this learning process, despite being challenging at times, was indispensable and allowed us to collaborate with our study participants on eye level. During the co-design process, it also became evident that there are conflicting concepts and desires, not only between the different stakeholder groups but also within the same groups. The prime example to mention here is that people with spinal cord injury requested timely interventions (rather than arbitrary, generic advice), which is difficult if not impossible to reconcile with their wish for as much autonomy and privacy as possible. In addition, the conflicting nature of some evidence-based guidelines was identified as a key challenge [9]. In trying to manage these tensions and possible points of conflict, we also experienced the dark sides of co-design [29,49]. We had to accept and deal with the fact that consensus among all stakeholders was not always attainable nor compatible with the financial and organizational constraints we faced [50,51]. This involved negotiation, mediation, and, most importantly, managing participants' and other stakeholders', including funders', expectations [52]. In line with earlier work [53], we also had to face the fact that our co-design and development approach had failed to adequately consider the need for an implementation plan and a business model.

Our values, knowledge, and experiences as researchers inevitably shape our work in one way or another [54,55]. Verbalizing our assumptions and expectations regarding the

co-design process and its outcomes from the very beginning can help us to establish a baseline. Against this baseline, we can then consider the added value of co-design by asking ourselves: What specifically did we gain from involving different stakeholders at eye level? What are the aspects we would have likely failed to consider? Could some of our taken-for-granted assumptions be refuted? Moreover, being conscious and reflective about the underpinnings of our work can help us to maintain our research integrity and minimize the risk of tokenism when co-designing health solutions with different stakeholders. Critical reflection can also promote out-of-the-box thinking, a much-needed skill when addressing design problems that could be tackled by a myriad of potential solutions [56,57]. Without such critical reflection, research teams may fail to explore alternative solutions and instead fixate on preset notions and ideas, which may result in suboptimal design and development choices [58]. To promote such out-of-the-box thinking, we chose to prioritize stakeholders' lived experience and clinical guidelines instead of relying on existing theoretical models to inform the design and development of the app prototype. At a next stage, we plan to link our findings to existing behavior change theories [59] to devise detailed intervention content for a pilot trial. This will then allow us to perform a rigorous evaluation of our self-management app and help us to identify its most effective components [60].

Strengths and Limitations

It needs to be acknowledged that the individuals agreeing to take part in this study may have been more technology savvy and may have thus held more positive attitudes toward mHealth solutions as compared to those individuals that did not take part. As a consequence, the app prototype may fail to account for the specific needs of less tech-savvy target audiences. To counteract this limitation, we incorporated user personas in the co-design process to allow and encourage participants to think beyond their personal expectations and experiences with technology. Moreover, there are also some limitations related to the functionality of the app prototype used to carry out the usability tests. Given that the app prototype was web-based, functionality was limited and sometimes impeded, which may have led to poorer evaluations. To account for this, all participants were made aware of these inherent limitations of the technology before starting the usability test. From a methodological point of view, combining the task completion exercise with the think-aloud technique during usability testing presented some limitations. More specifically, it prevented us from assessing the actual time it would take a participant to complete a specific task. Last but not least, we cannot be sure how truly participatory our co-design approach was. Although we aimed to involve the different stakeholder groups as equal partners, we cannot know with certainty whether this is how they experienced the co-design process. We thus recommend researchers using co-design methodologies to incorporate process evaluations into their research to gain a better understanding of how different stakeholders experience their participation.

Conclusions

Spinal cord injury is a complex chronic health condition, requiring those affected to navigate competing interests and priorities, paired with uncertainty about the accuracy and relevance of clinical recommendations. This study shows that involving individuals with spinal cord injury and health care professionals in co-designing a self-management app is both a feasible and enriching exercise in that it fosters knowledge cocreation at the intersection of lived experience, medical expertise, and technical solutions. In light of a current dearth of mHealth solutions tailored to the needs of community-dwelling individuals with spinal cord injury, this study makes an important contribution by advancing our knowledge on how to design interventions that can motivate behavior change, specifically regarding the prevention of pressure injuries. However, co-designing self-management solutions is a time and resource-intensive endeavor. Future research is needed to evaluate the impact of a co-designed self-management app and to demonstrate its additional value over conventional top down–designed solutions.

Acknowledgments

The authors would like to thank Suva, the Swiss Accident Insurance Fund, for providing funding for this project. We would also like to thank all study participants for their time, trust, and valuable contributions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ideation Workshop.
[PDF File (Adobe PDF File), 522 KB - mhealth_v8i7e18018_app1.pdf]

Multimedia Appendix 2
Usability Test.
[PDF File (Adobe PDF File), 198 KB - mhealth_v8i7e18018_app2.pdf]

Multimedia Appendix 3

Interview Guide.
[PDF File (Adobe PDF File), 122 KB - mhealth_v8i7e18018_app3.pdf]

Multimedia Appendix 4
Core Themes and matching Functions.
[PDF File (Adobe PDF File), 108 KB - mhealth_v8i7e18018_app4.pdf]

References

1. Regan MA, Teasell RW, Wolfe DL, Keast D, Mortenson WB, Aubut JL, Spinal Cord Injury Rehabilitation Evidence Research Team. A systematic review of therapeutic interventions for pressure ulcers after spinal cord injury. Arch Phys Med Rehabil 2009 Feb;90(2):213-231 [FREE Full text] [doi: 10.1016/j.apmr.2008.08.212] [Medline: 19236976]
2. Scheel-Sailer A, Wyss A, Boldt C, Post MW, Lay V. Prevalence, location, grade of pressure ulcers and association with specific patient characteristics in adult spinal cord injury patients during the hospital stay: a prospective cohort study. Spinal Cord 2013 Nov 3;51(11):828-833. [doi: 10.1038/sc.2013.91] [Medline: 23999106]
3. Brienza D, Krishnan S, Karg P, Sowa G, Allegretti AL. Predictors of pressure ulcer incidence following traumatic spinal cord injury: a secondary analysis of a prospective longitudinal study. Spinal Cord 2018 Jan 12;56(1):28-34. [doi: 10.1038/sc.2017.96] [Medline: 28895575]
4. Gabbe BJ, Nunn A. Profile and costs of secondary conditions resulting in emergency department presentations and readmission to hospital following traumatic spinal cord injury. Injury 2016 Aug;47(8):1847-1855. [doi: 10.1016/j.injury.2016.06.012] [Medline: 27343134]
5. Carlson M, Vigen CLP, Rubayi S, Blanche EI, Blanchard J, Atkins M, et al. Lifestyle intervention for adults with spinal cord injury: Results of the USC-RLANRC Pressure Ulcer Prevention Study. J Spinal Cord Med 2019 Jan;42(1):2-19 [FREE Full text] [doi: 10.1080/10790268.2017.1313931] [Medline: 28414254]
6. White BA, Dea N, Street JT, Cheng CL, Rivers CS, Attabib N, et al. The Economic Burden of Urinary Tract Infection and Pressure Ulceration in Acute Traumatic Spinal Cord Injury Admissions: Evidence for Comparative Economics and Decision Analytics from a Matched Case-Control Study. J Neurotrauma 2017 Oct 15;34(20):2892-2900. [doi: 10.1089/neu.2016.4934] [Medline: 28791891]
7. Marin J, Nixon J, Gorecki C. A systematic review of risk factors for the development and recurrence of pressure ulcers in people with spinal cord injuries. Spinal Cord 2013 Jul 16;51(7):522-527. [doi: 10.1038/sc.2013.29] [Medline: 23588570]
8. Kruger EA, Pires M, Ngann Y, Sterling M, Rubayi S. Comprehensive management of pressure ulcers in spinal cord injury: current concepts and future trends. J Spinal Cord Med 2013 Nov;36(6):572-585 [FREE Full text] [doi: 10.1179/2045772313Y.0000000093] [Medline: 24090179]
9. Fiordelli M, Zanini C, Amann J, Scheel-Sailer A, Brach M, Stucki G, et al. Selecting Evidence-Based Content for Inclusion in Self-Management Apps for Pressure Injuries in Individuals With Spinal Cord Injury: Participatory Design Study. JMIR Mhealth Uhealth 2020 May 20;8(5):e15818 [FREE Full text] [doi: 10.2196/15818] [Medline: 32432559]
10. Fogelberg DJ, Powell JM, Clark FA. The role of habit in recurrent pressure ulcers following spinal cord injury. Scand J Occup Ther 2016 Nov 14;23(6):467-476. [doi: 10.3109/11038128.2015.1130170] [Medline: 26763555]
11. Olney CM, Vos-Draper T, Egginton J, Ferguson J, Goldish G, Eddy B, et al. Development of a comprehensive mobile assessment of pressure (CMAP) system for pressure injury prevention for veterans with spinal cord injury. J Spinal Cord Med 2019 Nov 31;42(6):685-694. [doi: 10.1080/10790268.2019.1570437] [Medline: 30702395]
12. Garber SL, Rintala DH. Pressure ulcers in veterans with spinal cord injury: a retrospective study. J Rehabil Res Dev 2003;40(5):433-441 [FREE Full text] [doi: 10.1682/jrrd.2003.09.0433] [Medline: 15080228]
13. Zanini C, Brach M, Lustenberger N, Scheel-Sailer A, Koch HG, Stucki G, et al. Engaging in the prevention of pressure injuries in spinal cord injury: A qualitative study of community-dwelling individuals' different styles of prevention in Switzerland. J Spinal Cord Med 2020 Mar 12;43(2):247-256 [FREE Full text] [doi: 10.1080/10790268.2018.1543094] [Medline: 30540555]
14. Louise L, Annette B. Drawing straight lines along blurred boundaries: qualitative research, patient and public involvement in medical research, co-production and co-design. Evid Policy 2019 Aug 01;15(3):409-421. [doi: 10.1332/174426419x15552999451313]
15. Baron JS, Sullivan KJ, Swaine JM, Aspinall A, Jaglal S, Presseau J, et al. Self-management interventions for skin care in people with a spinal cord injury: part 1-a systematic review of intervention content and effectiveness. Spinal Cord 2018 Sep 25;56(9):823-836 [FREE Full text] [doi: 10.1038/s41393-018-0138-3] [Medline: 29802393]
16. Baron JS, Sullivan KJ, Swaine JM, Aspinall A, Jaglal S, Presseau J, et al. Self-management interventions for skin care in people with a spinal cord injury: part 2-a systematic review of use of theory and quality of intervention reporting. Spinal Cord 2018 Sep 25;56(9):837-846 [FREE Full text] [doi: 10.1038/s41393-018-0136-5] [Medline: 29795415]
17. Corrigan JM, Greiner AC, Adams K, editors. The 1st Annual Crossing the Quality Chasm Summit: A Focus on Communities: Report of a Summit. Washington DC: National Academies Press; 2004.

18. Allin S, Shepherd J, Tomasone J, Munce S, Linassi G, Hossain SN, et al. Participatory Design of an Online Self-Management Tool for Users With Spinal Cord Injury: Qualitative Study. JMIR Rehabil Assist Technol 2018 Mar 21;5(1):e6 [FREE Full text] [doi: 10.2196/rehab.8158] [Medline: 29563075]
19. Fletcher K, Foley F, Murray G. Web-Based Self-Management Programs for Bipolar Disorder: Insights From the Online, Recovery-Oriented Bipolar Individualised Tool Project. J Med Internet Res 2018 Oct 24;20(10):e11160 [FREE Full text] [doi: 10.2196/11160] [Medline: 30355553]
20. Mirkovic J, Jessen S, Kristjansdottir OB, Krogseth T, Koricho AT, Ruland CM. Developing Technology to Mobilize Personal Strengths in People with Chronic Illness: Positive Codesign Approach. JMIR Form Res 2018 Jun 05;2(1):e10774 [FREE Full text] [doi: 10.2196/10774] [Medline: 30684404]
21. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a Mental Health eClinic to Improve Access to and Quality of Mental Health Care for Young People: Using Participatory Design as Research Methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
22. Revenäs Å, Opava CH, Martin C, Demmelmaier I, Keller C, Åsenlöf P. Development of a web-based and mobile app to support physical activity in individuals with rheumatoid arthritis: results from the second step of a co-design process. JMIR Res Protoc 2015 Feb 09;4(1):e22 [FREE Full text] [doi: 10.2196/resprot.3795] [Medline: 25665589]
23. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: A self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan 30;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
24. Marent B, Henwood F, Darking M, EmERGE Consortium. Development of an mHealth platform for HIV Care: Gathering User Perspectives Through Co-Design Workshops and Interviews. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e184 [FREE Full text] [doi: 10.2196/mhealth.9856] [Medline: 30339132]
25. Jessen S, Mirkovic J, Ruland CM. Creating Gameful Design in mHealth: A Participatory Co-Design Approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
26. Bangor A, Kortum P, Miller J. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776] [Medline: 19236976]
27. Floch J, Zettl A, Fricke L, Weisser T, Grut L, Vilarinho T, et al. User Needs in the Development of a Health App Ecosystem for Self-Management of Cystic Fibrosis: User-Centered Development Approach. JMIR Mhealth Uhealth 2018 May 08;6(5):e113 [FREE Full text] [doi: 10.2196/mhealth.8236] [Medline: 29739742]
28. Mortenson WB, Singh G, MacGillivray M, Sadeghi M, Mills P, Adams J, et al. Development of a Self-Management App for People with Spinal Cord Injury. J Med Syst 2019 Apr 22;43(6):145. [doi: 10.1007/s10916-019-1273-x] [Medline: 31011881]
29. Dudau A, Glennon R, Verschuere B. Following the yellow brick road? (Dis)enchantment with co-design, co-production and value co-creation in public services. Public Manage Rev 2019 Sep 01;21(11):1577-1594. [doi: 10.1080/14719037.2019.1653604]
30. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving Patient Experience and Primary Care Quality for Patients With Complex Chronic Disease Using the Electronic Patient-Reported Outcomes Tool: Adopting Qualitative Methods Into a User-Centered Design Approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
31. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth Delivered Interventions: A Systematic Review to Assess Key Methods and Processes. Curr Nutr Rep 2016 Jul 4;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
32. Paraplegie Community. URL: https://community.paraplegie.ch/de/ [accessed 2020-02-05]
33. Gibbons S. UX mapping methods compared: A cheat sheet. Nielsen Norman Group. 2017. URL: https://www.nngroup.com/articles/ux-mapping-cheat-sheet/ [accessed 2020-02-05]
34. Turner AM, Reeder B, Ramey J. Scenarios, personas and user stories: user-centered evidence-based design representations of communicable disease investigations. J Biomed Inform 2013 Aug;46(4):575-584 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.006] [Medline: 23618996]
35. Sripathi V, Sandru V. Effective Usability Testing?Knowledge of User Centered Design is a Key Requirement. Int J Emerg Technol Adv Eng 2013;3(1):627-635 [FREE Full text]
36. Keijzer-Broers WJW, de Reuver M. Applying agile design sprint methods in action design research: prototyping a health and wellbeing platform. In: Parsons J, Tuunanen T, Venable J, Donnellan B, Helfert M, Kenneally J, editors. Tackling Society's Grand Challenges with Design Science. DESRIST 2016. Lecture Notes in Computer Science. Switzerland: Springer; 2016.
37. Faste H, Bergamasco M. A strategic map for high-impact virtual experience design. In: McDowall IE, Dolinsky M, editors. The Engineering Reality of Virtual Reality. Bellingham, WA: SPIE Press; 2009.
38. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

39. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb 20;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
40. 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 01;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
41. Backman C, Harley A, Peyton L, Kuziemsky C, Mercer J, Monahan MA, et al. Development of a Path to Home Mobile App for the Geriatric Rehabilitation Program at Bruyère Continuing Care: Protocol for User-Centered Design and Feasibility Testing Studies. JMIR Res Protoc 2018 Sep 24;7(9):e11031 [FREE Full text] [doi: 10.2196/11031] [Medline: 30249591]
42. Marchak JG, Cherven B, Williamson Lewis R, Edwards P, Meacham LR, Palgon M, et al. User-centered design and enhancement of an electronic personal health record to support survivors of pediatric cancers. Support Care Cancer 2019 Dec 18. [doi: 10.1007/s00520-019-05199-w] [Medline: 31853699]
43. Borenstein J, Chiou C, Henning JM, Wilson A, Hohlbauch AA, Richards MS, et al. Physician attitudes toward strategies to promote the adoption of medical evidence into clinical practice. Am J Manag Care 2003 Mar;9(3):225-234 [FREE Full text] [Medline: 12643340]
44. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
45. Jones M, Morris J, Deruyter F. Mobile Healthcare and People with Disabilities: Current State and Future Needs. Int J Environ Res Public Health 2018 Mar 14;15(3):515 [FREE Full text] [doi: 10.3390/ijerph15030515] [Medline: 29538292]
46. Palmer VJ, Weavell W, Callander R, Piper D, Richard L, Maher L, et al. The Participatory Zeitgeist: an explanatory theoretical model of change in an era of coproduction and codesign in healthcare improvement. Med Humanit 2019 Sep 28;45(3):247-257 [FREE Full text] [doi: 10.1136/medhum-2017-011398] [Medline: 29954854]
47. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
48. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: Using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
49. Steen T, Brandsen T, Verschuere B. The dark side of co-creation and co-production: Seven evils. In: Brandsen T, Steen T, Verschuere B, editors. Co-Production and Co-Creation: Engaging Citizens in Public Services. New York: Routledge; 2018:284-293.
50. Bains W. Leadership and innovation. How consensus management blocks genuine innovation. Biosci Hypoth 2009 Jan;2(5):277-281. [doi: 10.1016/j.bihy.2009.08.002]
51. McAfee A. When too much IT knowledge is a dangerous thing. MIT Sloan management review. 2003. URL: https://sloanreview.mit.edu/article/when-too-much-it-knowledge-is-a-dangerous-thing/ [accessed 2020-02-05]
52. Rajapakse R, Brereton M, Sitbon L, Roe P. A Collaborative approach to design individualized technologies with people with a disability. New York: Association for Computing Machinery; 2015 Presented at: Annual Meeting of the Australian Special Interest Group for Computer Human Interaction; December 2015; Parkville, Australia.
53. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that Matters via a Multidisciplinary Requirements Development Approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
54. Thomas K. The structure of scientific revolutions. Chicago and London: University of Chicago Press; 1962.
55. Mesel T. The necessary distinction between methodology and philosophical assumptions in healthcare research. Scand J Caring Sci 2013 Sep;27(3):750-756. [doi: 10.1111/j.1471-6712.2012.01070.x] [Medline: 22935081]
56. Hennipman EJ, Oppelaar EJR, van der Veer GC, Bongers B. Rapid and rich prototyping: proof of concepts for experience. In: The ergonomics of cool interaction. New York: Association for Computing Machinery; 2008 Presented at: 15th European conference on cognitive ergonomics; 2008; Funchal, Portugal p. 1-6.
57. McBride M. The software architect: essence, intuition, and guiding principles. In: Companion to the 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications. New York: Association for Computing Machinery; 2004 Presented at: 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications; 2004; Vancouver, Canada p. 230-235. [doi: 10.1145/1028664.1028764]
58. Dow SP, Glassco A, Kass J, Schwarz M, Schwartz DL, Klemmer SR. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM Trans Comput Hum Interact 2010 Dec;17(4):1-24. [doi: 10.1145/1879831.1879836]
59. 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]
60. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 29.01.20; peer-reviewed by K Fitzner, R Cruz Martínez; comments to author 20.02.20; revised version received 03.04.20; accepted 19.04.20; published 09.07.20.

Please cite as:
Amann J, Fiordelli M, Brach M, Bertschy S, Scheel-Sailer A, Rubinelli S
Co-designing a Self-Management App Prototype to Support People With Spinal Cord Injury in the Prevention of Pressure Injuries: Mixed Methods Study
JMIR Mhealth Uhealth 2020;8(7):e18018
URL: https://mhealth.jmir.org/2020/7/e18018
doi:10.2196/18018
PMID:32673241

Original Paper

Co-Design in the Development of a Mobile Health App for the Management of Knee Osteoarthritis by Patients and Physicians: Qualitative Study

Kelly J Mrklas[1,2], BSc, MSc; Tanya Barber[3], MA; Denise Campbell-Scherer[4], MD, PhD; Lee A Green[3], MD, MPH; Linda C Li[5,6], PT, PhD; Nancy Marlett[7,8], PhD; Jean Miller[7], PhD; Brittany Shewchuk[2], BSc; Sylvia Teare[7], MED; Tracy Wasylak[1,9], MSc, CHE; Deborah A Marshall[2], PhD

[1]Strategic Clinical Networks, Provincial Clinical Excellence, Alberta Health Services, Calgary, AB, Canada

[2]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]Enhancing Alberta Primary Care Research Networks, Department of Family Medicine, University of Alberta, Edmonton, AB, Canada

[4]Department of Family Medicine, University of Alberta, Edmonton, AB, Canada

[5]Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada

[6]Arthritis Research Canada, Richmond, BC, Canada

[7]O'Brien Institute for Public Health, University of Calgary, Calgary, AB, Canada

[8]Community Rehabilitation and Disability Studies, Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[9]Faculty of Nursing, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Deborah A Marshall, PhD
Department of Community Health Sciences
Cumming School of Medicine
University of Calgary
Health Research Innovation Centre (HRIC)
3C58 - 3280 Hospital Drive NW
Calgary, AB, T2N4Z6
Canada
Phone: 1 403 210 6377
Email: damarsha@ucalgary.ca

Abstract

Background: Despite a doubling of osteoarthritis-targeted mobile health (mHealth) apps and high user interest and demand for health apps, their impact on patients, patient outcomes, and providers has not met expectations. Most health and medical apps fail to retain users longer than 90 days, and their potential for facilitating disease management, data sharing, and patient-provider communication is untapped. An important, recurrent criticism of app technology development is low user integration design. User integration ensures user needs, desires, functional requirements, and app aesthetics are responsive and reflect target user preferences.

Objective: This study aims to describe the co-design process for developing a knee osteoarthritis minimum viable product (MVP) mHealth app with patients, family physicians, and researchers that facilitates guided, evidence-based self-management and patient-physician communication.

Methods: Our qualitative co-design approach involved focus groups, prioritization activities, and a pre-post quality and satisfaction Kano survey. Study participants included family physicians, patient researchers and patients with knee osteoarthritis (including previous participants of related collaborative research), researchers, key stakeholders, and industry partners. The study setting was an academic health center in Southern Alberta.

Results: Distinct differences exist between what patients, physicians, and researchers perceive are the most important, convenient, desirable, and actionable app functional requirements. Despite differences, study participants agreed that the MVP should be electronic, should track patient symptoms and activities, and include features customized for patient- and physician-identified factors and international guideline-based self-management strategies. Through the research process, participants negotiated

consensus on their respective priority functional requirements. The highest priorities were a visual symptom graph, setting goals, exercise planning and daily tracking, and self-management strategies. The structured co-design with patients, physicians, and researchers established multiple collaborative processes, grounded in shared concepts, language, power, rationale, mutual learning, and respect for diversity and differing opinions. These shared team principles fostered an open and inclusive environment that allowed for effective conceptualization, negotiation, and group reflection, aided by the provision of tangible and ongoing support throughout the research process, which encouraged team members to question conventional thinking. Group-, subgroup-, and individual-level data helped the team reveal how and for whom perspectives about individual functional requirements changed or remained stable over the course of the study. This provided valuable insight into how and why consensus emerged, despite the presence of multiple and differing underlying rationales for functional requirement prioritization.

Conclusions: It is feasible to preserve the diversity of perspectives while negotiating a consensus on the core functional requirements of an mHealth prototype app for knee osteoarthritis management. Our study sample was purposely constructed to facilitate high co-design interactivity. This study revealed important differences between the patient, physician, and researcher preferences for functional requirements of an mHealth app that did not preclude the development of consensus.

(JMIR Mhealth Uhealth 2020;8(7):e17893) doi:10.2196/17893

KEYWORDS

health services research; app; knee osteoarthritis; community-based participatory research

Introduction

Background

Knee osteoarthritis is an inflammatory condition affecting over 25% of middle-aged to older adults [1], causing significant disability and reduced health-related quality of life. Knee osteoarthritis is a significant, growing economic and health burden to afflicted individuals and the society at large [2,3] and is one of the most prevalent diseases worldwide [4].

Fortunately, mobile health (mHealth) technology, defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices," including apps [4], holds great promise for advancing the treatment and management of chronic diseases, such as knee osteoarthritis. Mobile technology is the fastest spreading technology in modern history, expanding more than 28-fold, from an average of 3.7 mobile cellular subscriptions per 100 individuals in 1997 to 103 per 100 individuals in 2017 [5]. With thousands of apps released daily [6], mobile technology has been ascribed *limitless potential* to enhance patient and provider access to evidence-based, effective health care resources, at a lower cost [7].

Importantly, recent osteoarthritis treatment and management guidelines endorse patient self-management as a means to increase self-awareness of symptoms and better prepare patients to assume active roles in shared medical decision making [8]. mHealth can enable patient self-management through the use of wearable aides for activity monitoring and behavioral change using personalized, real-time feedback [9], through acquisition of new knowledge, and by enhancing patient-provider communication [10,11].

Despite the doubling of osteoarthritis-targeted mHealth apps over the last decade, both app volume and app research focused on knee osteoarthritis are incongruent with its high global prevalence [12]. Most current apps fail to meet patient and provider expectations for disease management, data sharing, and enhancing communication [8,13,14]. Furthermore, although users demonstrate high interest and demand for health and medical apps [14], the vast majority fail to retain even a third of users after 90 days because of infrequent use, high user turnover, and low loyalty [15].

Unfortunately, research on knee osteoarthritis app development, assessment, and effectiveness [4,8,14] is generally lacking [13], as is a fulsome understanding of the documented or novel barriers that lead to app discontinuity to inform meaningful design improvements. Multiple existing studies document key issues, including functionality, content and content personalization, interactivity, behavior change theory integration, and sustained use [13-17].

Calls to adopt more participatory, coproduction approaches to technology design have emerged over the last decade in many disciplines [18-20]; lack of co-design is a recurring criticism of app development and design. User involvement is crucial to ensure that needs, preferences, functional requirements, and app aesthetics are responsive, reflect, and meet end user needs [18,19]. Empirically tested mHealth apps oriented to and built specifically with and for the needs, preferences, and activities of patients with knee osteoarthritis and their providers [8] are needed.

The call for apps that focus specifically on knee osteoarthritis self-management, patient-clinician decision support, and shared decision making [8] has not yet been answered; the anticipated benefits of mHealth have not been fully realized.

Objectives

This study aimed to describe the co-design process of knee osteoarthritis minimum viable product (MVP) app development with and for patients and family physicians. An MVP (versus a more comprehensive design product) was chosen because of its specific focus on establishing key end user requirements, meeting early adopter needs, and supporting ongoing iterative research–focused app development, at optimized time and cost [21]. The intent of MVP app development was to facilitate guided self-management, provide evidence-based information to patients and physicians, and facilitate communication while addressing patient needs and challenges [22]. We sought to understand whether and how patient and provider preferences

for the functional requirements of a knee osteoarthritis app differ, and whether patient-provider consensus was possible, in the development of an app supporting knee osteoarthritis management [14].

Methods

Engaging People With Lived Experience as Research Team Members

This participatory research project engaged people with lived experience of knee osteoarthritis from idea inception through data collection, analysis, and dissemination. People with lived experience provided crucial direction in co-design approach development with team members trained as patient and community engagement researchers (PACERs) [23], informed by the Guidance for Reporting Involvement of Patients and the Public 2.0 checklist [24].

Research Coalition

Patients, patient researchers, physicians, researchers, decision makers, industry partners, and trainees were engaged to co-design an MVP by combining 3 stakeholder-specific collaborative research approaches. This included a patient-to-patient approach [23] developed by and for patients through PACER in which qualitatively trained patients co-design using a 3-step grounded theory-based research process: (1) *set* to clarify and explore the topic, (2) *collect* to interview, and (3) *reflect* through collaborative analysis. The second approach mobilized tacit health care provider knowledge and was developed by the Enhancing Alberta Primary Care Research Networks (ENACT) team [25]. ENACT supports practice-based research networks and academic and community practitioners conducting research in primary care. The third approach was iteratively developed by team researchers, informed by integrated knowledge translation [26] and collaborative participatory design approach principles [18].

Conceptual Framework

Conceptual underpinnings are derived from the synthesis-based mHealth design framework for osteoarthritis self-management by Choi et al [8], which highlights evidence-based decision support, osteoarthritis assessment, shared decision making, self-management (such as education, physical activity, feedback, symptom/movement, and joint function monitoring), and data visualization for patients and providers.

Recruitment

Patient participants, previously recruited through the media, the Arthritis Society, and posters for participation in several preceding PACER studies [22,23,27] were purposively recruited and reinvited to continue their research involvement in this study. Patients were eligible if they reported knee pain "on most days of the month at any time in the past and any pain in the past 12 months" [28]. Family physicians who had either participated or showed interest but were unable to participate in the preceding study [22] were invited by email and asked to nominate a colleague if they could not attend. Physicians were *early majority* [29] practitioners (ie, the first sizable segment of providers to adopt an innovation after seeing others try it) and were recruited to avoid designing around unrepresentative perspectives of *early adopters*. Physicians had a diverse range of experiences in practice settings in Alberta.

Key stakeholders, comprising decision makers from Alberta Health Services, the main provincial health care service provider, and the Alberta Bone and Joint Research Institute, were invited to join the study from inception. Our industry partner was invited to join the third interactive co-design session.

Study Design

We planned 3 full day, co-design sessions involving semistructured focus groups, ranking and prioritization activities, and a presession and postsession quality and satisfaction (ie, convenience and importance) Kano survey [30]. Sessions were full day, face-to-face meetings hosted at an academic health campus in Calgary, Alberta, Canada (Table 1). Sessions were informed by previous studies [22,23,27,28] and iteratively by findings arising from each co-design session. This study was reviewed and approved by the Conjoint Health Research Ethics Board at the University of Calgary (REB161372).

Table 1. Co-design session objectives, methods, and outputs.

Event	Objectives	Methods (analysis)	Outputs
Session 1 (March 27, 2017)	• Establish MVP[a] symptom and quality-of-life measures for patients and physicians • 2. Establish parameters for MVP use as communication and self-management tool	• Semistructured focus groups (thematic analysis of flip chart data, notes, transcribed notes)	• Tool category features: symptoms and activity, red flags/triggers, and guided self-management strategies • Summaries, executive summaries
Session 2 (April 18, 2017)	• Condense potential functional requirements • Determine relative importance and define functionality of MVP requirements • Explore how functional requirement use by patients and physicians, to improve patient outcomes	• Semistructured focus groups (member check, initial theming and thematic analysis of flip chart data, notes, transcribed notes) • Provisional dot voting (frequency counts used as a provisional prioritization criteria for each group)	• Categorized tool features • Inputs: goal setting, context, symptom tracking, activity tracking, plans/strategies, prognosis prediction (input) • Interaction reminders: daily, event-based, periodic outputs; and feedback to patients, physician summary, red flags, prognosis prediction (output) • Summaries, executive summaries
Kano presurvey and postsurveys (October 1, 2017 and October 5, 2018)	• Determine how stakeholders (patients, physicians, researchers, and decision makers) rated functional requirements by importance and convenience before and after group introduction and review of MVP	• Mean (SD) importance score by participant group and all respondents, frequency count by category and participant group) • Convenience scores (reported by participant group, frequency count by category and participant group)	• Quantified importance/convenience scores for functional requirements • Thematic analysis of qualitative comments by group (if required)
Session 3 (October 3, 2018)	• To review MVP appearance (wireframes) and mock function, and provide feedback on functional requirements for design iteration with development team and gather a definitive prioritization and ranking of functional requirements for inclusion in the final MVP using dot voting	• Semistructured focus group discussion (member check, initial theming and thematic analysis of notes: main takeaways) • Dot voting (frequency counts/range on task 1: must-have, won't-have prioritization and task 2: desirability and actionability prioritization, reported by participant group and all respondents)	• Must-have, won't-have dot voting results by participant group and for all respondents, for each functional requirement • Desirability and actionability dot voting results, by participant group, all respondents, for each functional requirement

[a]MVP: minimum viable product.

Co-Design Sessions 1 to 3

A total of 7 *a priori* objectives defined the problem, needs, and scope and functions of the proposed MVP development (Table 1). Co-design participants were separated into 2 multistakeholder groups (two groups at S1 and S2, only one multistakeholder group at S3), comprising the patients, physicians, researchers, and decision makers, led by research staff and 2 note takers each. This same format and procedures were applied to all sessions.

S1 and S2 guiding questions and activities sought participant perspectives on symptoms and MVP functionality, including how physicians and patients measure and identify red flag symptoms that trigger a family physician visit, how perceived quality of life is affected by different symptoms, patient and provider experience, and observed symptom variation. The team investigated symptom prioritization and rating by both patients and physicians using the Western Ontario and McMaster Universities Arthritis Index (WOMAC) Numerical Rating Scale 3.1 index [31]. The WOMAC is a validated, 24-item, self-administered knee and hip osteoarthritis index that assesses 3 dimensions: pain, disability, and joint stiffness. These data established broad MVP use parameters, including supporting patient-physician communication and guided self-management, exploring MVP benefits/limitations, and potential capability for research-guided self-management.

The findings generated at earlier co-design and dot voting/prioritization activities [32,33] were reconsidered by participants during each session, and they were iteratively used to help identify and define key functional requirements. This iterative process was also used to surface discrepant views for individual and group reflections, and subsequent discussion [34].

Dot voting prioritization involved participants applying 2 each of *must-have* (green) and *won't-have* (red) stickers to their priority requirements, in any desired configuration. Similarly, S3 participants were given 10 each of desirability (blue) and actionability (yellow) stickers to rank applicable requirements. Votes were tallied by group (patients, physicians, and researchers) and combined with other findings for prioritization. Both sessions helped define the functional requirement characteristics.

In the 6 months preceding S3, our industry partner applied early findings to technical analysis and data consolidation and generated recommendations for appropriate technologies and

basic functional requirements. MVP mood boards (visual guide of skeletal framework) were developed for S3 to illustrate page layout, content arrangement, function range and prioritization, display rules, and the effects of different scenarios on app display. Boards were used to collect feedback and generate definitive MVP functional requirement prioritization.

A Kano survey, conceptually grounded in the two-factor motivational theory by Herzberg et al [30,35], assessed satisfaction with proposed functionalities, and the degree to which functionalities were required for satisfaction [36]. The Kano survey is based on the theory of attractive quality, often used to assess and relate customer satisfaction with specific quality attributes [37]. In health care, it is used for function-satisfaction interface [38] assessments, including patient perceptions of service quality and quality expectations, quality elements, patient-provider relationships, satisfaction, and assessing how expectations vary with increased awareness to inform appropriate and aligned service requirement design [38].

A web-enabled Kano survey, comprising 10 three-part question clusters, was emailed to all co-design participants 3 days preceding and immediately following S3. The survey was accompanied by a link to click-through MVP static wireframes for independent review by participants (Multimedia Appendix 1) [36,39].

Data Collection and Analysis

Table 1 itemizes key data collection events, objectives, methods, and planned outputs. S1 and S2 discussions were audio recorded and key points transcribed by a research assistant, supplemented by findings from the note takers. Text was analyzed line by line, and important patterns pertaining to session objectives and research questions were identified [40]. Action-oriented themes were framed as *take-aways* and refined for coherence [40]. Summaries were generated to inform iterative co-design.

The Kano survey results were collated, frequencies tabulated, and visualized graphically by requirement and stakeholder group (ie, patients, physicians, and researchers). Functional requirements were ranked based on the combined study findings. Mean importance (SD) was calculated for each requirement and for each participant group, and an adjusted mean importance (SD) was calculated across all respondents. Summaries describing and quantifying the co-design process and its outputs were generated for each phase. Survey comments were collated and analyzed for important patterns [40].

Dot voting [32,33] was used to identify, rank, and prioritize functional requirements on convenience dimensions (ie, must-have, won't-have, desirable, and actionable). Frequencies (range) were presented for each dimension by the participant group (ie, patients, physicians, researchers, and all participants) and summarized.

Results

Co-Design Participants

A total of 28 unique co-design participants (13 males and 15 females) took part in at least one session. Overall, 4 patients, 5 physicians, 12 researchers (including 7 team, 3 PACERs, and 2 ENACT researchers), 3 trainees, and 2 decision makers took part in the co-design process to refine concepts, functional requirements, and their relative priority (Table 2). Two industry partner team members observed and interacted at the S3 discussion, and neither completed Kano surveys and dot voting.

Table 2. Co-design participant demographics (sessions 1-3).

Session and participants by group	Male, n	Female, n	Participants by session, n
Session 1			
Patients	2	2	4
Physicians	1	1	2
UCalgary[a]/academic researchers	2	4	6
ENACT[b] researchers	1	1	2
PACER[c]	0	3	3
Decision makers	1	1	2
Trainees	0	0	0
Total Session 1	7	12	19
Session 2			
Patient	2	0	2
Physicians	0	2	2
UCalgary/academic researchers	2	4	6
ENACT researchers	1	1	2
PACER	0	2	2
Decision makers	1	1	2
Trainees[d]	3	0	3
Total Session 2	9	10	19
Session 3			
Patients	2	2	4
Physicians	1	1	2
UCalgary/academic researchers	1	3	4
ENACT researchers	1	1	2
PACER	0	3	3
Decision makers	0	0	0
Trainees	0	0	0
Industry partners	2	0	2
Total Session 3	7	10	17

[a]UCalgary: University of Calgary.

[b]ENACT: Enhancing Alberta Primary Care Research Networks team.

[c]PACER: patient and community engagement researchers.

[d]Computer Science trainees, University of Calgary.

Co-Design Sessions 1 and 2: Focus Group Findings

S1 discussions generated a unanimous agreement on basic MVP features and purpose, scope, patient/physician preferences and rationale, design action items, and evolving refinements of app functionalities (Textbox 1). Overall, the group thought that the MVP should have the following:

- Be electronic (ie, cell phone or web-enabled).
- Help patients manage symptoms and activity.
- Include customized red flags/triggers.
- Include evidence-based guided self-management strategies (eg, output structured by the Osteoarthritis Research Society International guidelines), including pain, weight management and aids, scheduling for self-management strategies, track progression, and report history (eg, activity, symptoms, red flags), and self-management plan for physician visits [31,41].

Textbox 1. Sessions 1 and 2 summary: participant-selected inputs, outputs and interactions/reminders.

Goal setting, context (input)

- Patient-customized goals defined on first app use (symptoms, quality of life, activity and linked to custom activities, reminders). Tracking of activities as output. Patient customizable comorbidities; visualized (homunculus), symptoms, activity history, previous plans, strategies

Symptom tracking (input)

- Dimensions: pain, stiffness, function, others (swelling, warmth, and inflammation) by validated, evidence-based tools. Patient-guided entry: threshold approach (eg, visual analog scale 0-10 provided, if higher than predetermined threshold, prompts location, duration, and intensity). Journal for situation-specific symptom record. Symptom history as output

Activity tracking (input)

- Patient-customized goal setting and exercises (evidence-based physiotherapy exercises). Exercise resource links per patients' needs

Plans, strategies (input)

- Sliders: My Exercise Plan, My Diet, My Medication (prescription, topical, and over the counter), and My Assistive Devices/Supports (aids, accessories, and other therapies). Categories to include attempted therapies with evidence-based *drop down* list and customizability. Ranking function to gather customized patient data about utility of strategies

Prognosis, prediction (input)

- Algorithm to capture patient symptoms, other information, and prediction of osteoarthritis severity. Patients prompted to enter information, context, and demographic information on first app use with customizable symptom tracking (time and frequency). Prognosis prediction as output, graphic visualization for patients/providers. Minimum input, thresholds required.

Feedback to patients (output)

- Report historical summary of goals, symptoms, activity, plans/strategies on patient dashboard. Provider summary graph as a separate dashboard with one chart containing symptoms and activity.

Red flags (output)

- Automated, predefined rules to capture symptoms, activity. Data considered red flags or signs stimulate action (eg, physician visit/emergency department). Patient button to journal red flags as events to share with providers

Prognosis, prediction [42] (output)

- On the basis of symptom inputted by patient, graph displays symptom fluctuation; sharable with providers

Daily, event-based, and periodic (interactions and reminders)

- Customized input for daily, event-based, periodic reminders. Reminders seen as small red sign with numbers. Custom daily input used for symptom/activity tracking. Periodic input (other than initial patient input) for goal setting, plans, strategies. Automated reminders aligned with exercises, goals, pain, etc

S1 discussions generated additional insights considered during S2 dot voting/prioritization activities. To facilitate further refinement and discussion, participants decided to use S2 dot voting to help generate consensus during their discussions rather than dot voting to prioritize as initially planned (Textbox 1).

When discrepancies surfaced, these were highlighted during and at the end of agenda sections at each session, and then revisited for further exploration during report back times and group discussions [34]. Participants eventually agreed to and provided details on 12 functional requirements (*inputs* [n=5]: goal setting and context, symptom tracking, activity tracking, plans or strategies, and prognosis prediction; *interaction reminders* [n=3]: daily, event-based and periodic reminders; and *outputs* [n=4]: feedback to patients, physician summary, red flags, and prognosis prediction). Participants advocated for validated instruments and high-quality evidence sources for medically relevant requirements identified by researchers. S1 and S2 findings were consolidated and requirements refined with the industry partner to inform S3 findings.

Session 3 Findings

Overall, 17 individuals (7 males and 10 females) attended session 3, including 4 patients and 4 providers, 2 physicians and 2 ENACT researchers, 3 PACERs, and 2 industry partner facilitators. S3 was preceded and followed by Kano survey assessments and involved in-depth consideration of the refined functionalities that emerged from S2. Cumulative findings were reviewed and checked by participants at the start and high-priority requirements refined iteratively through S3 assessments, prioritization activities, and group discussions.

Kano Survey: Importance

The Kano surveys helped qualitatively describe, enumerate, and reveal evolving co-design processes and outputs (Tables 3 and 4).

Importance assessments revealed the *top 6* functional requirements: track pain symptoms, visual graph of symptoms, self-management strategies, setting goals and follow through, track functional impairment, and plan exercises and daily tracking. Although differences in perceived importance arose between participant groups, the *top 6* requirements remained the same pre- and post-Kano survey, summarized in Table 5.

Patients scored self-management as highly important and patients referred to the MVP as a source of motivation, control, planning, and a means of encouraging positive behavior. Patients appreciated having knee osteoarthritis management functions in a single spot, using the app to facilitate a physician-patient interaction, and customizing and tracking progress over time.

Table 3. Presession Kano survey: importance by participant group (n=13).

App feature	Importance (9-point Likert scale: 1=not at all important to 9=extremely important)						Mean adjusted importance (n=13)	
	Patients (n=6)		Physicians (n=3)		Researchers (n=3)			
	Mean (SD)	Rank[a]	Mean (SD)	Rank[a]	Mean (SD)	Rank[a]	Mean (SD)	*Overall rank* [a,b]
If the app could show you a graph of your 7-year osteoarthritis severity prediction, how do you feel?	7.3 (1.03)	6	4.3 (2.08)	8	6.0[c] (1.41)	9	6.27 (1.85)	*9*
If the app could help you to set goals and follow through, how do you feel?	8.2 (0.98)	2	8.3 (0.58)	2	8.7 (0.58)	1	8.33 (0.78)	*1*
If the app could help you set a plan with various exercises and track them daily, how do you feel?	6.7 (2.94)	7	8.7 (0.58)	1	7.7 (1.53)	5	7.42 (2.27)	*6*
If the app could allow you to track your pain symptoms over time, how do you feel?	7.8 (1.33)	4	7.7 (1.53)	4	8.3 (0.58)	4	7.92 (1.16)	*4*
If the app could allow you to track your stiffness symptoms over time, how do you feel?	7.5 (1.22)	5	6.0 (2.65)	7	6.3 (1.15)	8	6.83 (1.64)	*7*
If the app could allow you to track your functional impairment symptoms over time, how do you feel?	8.0 (1.26)	3	8.3 (1.15)	2	8.7 (0.58)	2	8.25 (1.06)	*2*
If the app could show you a graph of your symptoms over time, how do you feel?	8.5 (0.84)	1	6.7 (2.08)	6	8.3 (0.58)	3	8.00 (1.35)	*3*
If the app could give you strategies to help you self-manage your arthritis, how do you feel?	7.3 (1.21)	6	8.0 (0)	3	7.3 (1.15)	7	7.50 (1.00)	*5*
If the app could let you flag certain days where arthritis impacted your plans, how do you feel?	5.2 (2.56)	9	4.0 (2.0)	9	7.7 (0.58)	6	5.50 (2.39)	*10*
If the app could give you reminders to update your information (symptoms, exercise, goal tracking), how do you feel?	6.3 (2.16)	8	7.0 (1.0)	5	5.7 (0.58)	10	6.33 (1.61)	*8*

[a]Rank subjectively assessed based on a combination of mean scores and overall mean adjusted scores, 1=highest rank, 10=lowest rank.

[b]Italics emphasize the overall rank for each functional requirement.

[c]2 responses only.

Table 4. Postsession Kano survey: importance by participant group (n=12).

App feature	Importance (9-point Likert scale 1=not at all important to 9=extremely important)						Mean adjusted importance (n=12)	
	Patients (n=7)		Physicians (n=2)		Researchers (n=3)			
	Mean (SD)	Rank[a]	Mean (SD)	Rank[a]	Mean (SD)	Rank[a]	Mean (SD)	*Overall rank* [a,b]
If the app could show you a graph of your 7-year osteoarthritis severity prediction, how do you feel?	5.90 (2.19)	7	5.50 (2.12)	8	6.70 (1.15)	6	6.00 (1.86)	*8*
If the app could help you to set goals and follow through, how do you feel?	7.00 (2.52)	4	9.00 (0)	1	6.00 (1.73)	7	7.10 (2.23)	*4*
If the app could help you set a plan with various exercises and track them daily, how do you feel?	6.60 (1.81)	5	8.50 (0.71)	3	6.00 (2.83)	8	6.80 (1.89)	*6*
If the app could allow you to track your pain symptoms over time, how do you feel?	7.40 (1.27)	2	7.50 (0.71)	6	8.30 (0.58)	3	7.70 (1.07)	*1*
If the app could allow you to track your stiffness symptoms over time, how do you feel?	5.60 (2.23)	9	3.00 (1.41)	9	8.00 (1.00)	4	5.80 (2.42)	*9*
If the app could allow you to track your functional impairment symptoms over time, how do you feel?	5.90 (2.34)	8	8.50 (0.71)	2	8.30 (0.58)	2	6.90 (2.19)	*5*
If the app could show you a graph of your symptoms over time, how do you feel?	7.00 (1.83)	3	7.50 (2.12)	4	8.70 (0.58)	1	7.50 (1.68)	*2*
If the app could give you strategies to help you self-manage your arthritis, how do you feel?	7.70 (1.11)	1	7.50 (2.12)	5	7.00 (1.00)	5	7.50 (1.17)	*3*
If the app could let you flag certain days where arthritis impacted your plans, how do you feel?	5.10 (1.77)	10	2.50 (0.71)	10	5.70 (4.16)	10	4.80 (2.48)	*10*
If the app could give you reminders to update your information (symptoms, exercise, goal tracking), how do you feel?	6.30 (1.89)	6	6.00 (2.83)	7	6.00 (2.00)	9	6.20 (1.85)	*7*

[a]Subjectively assessed rank based on mean scores and overall mean adjusted scores, 1=highest rank, 10=lowest rank.

[b]Italics emphasize the overall rank for each functional requirement.

Table 5. Functional requirement importance ranking: presession and postsession 3 Kano survey.

Presession Kano survey results	Rank[a]	Postsession Kano survey results	Rank[a]
Set goals and follow through	1	Track pain symptoms	1
Track functional impairment symptoms	2	Visual graph of symptoms	2
Visual graph of symptoms	3	Self-management strategies	3
Track pain symptoms	4	Set goals and follow through	4
Self-management strategies	5	Track functional impairment symptoms	5
Plan exercises and daily tracking	6	Plan exercises and daily tracking	6
Track stiffness symptoms	7	Reminders to update info	7
Reminders to update info	8	7-year osteoarthritis severity prediction	8
7-year osteoarthritis severity prediction	9	Track stiffness symptoms	9
Flag days	10	Flag days	10

[a]1=highest rank, 10=lowest rank.

Kano Survey: Convenience

There were clear alignments and differences in how patients, physicians, and researchers assessed the convenience of MVP requirements (Multimedia Appendix 2).

Patients' presession *must-have* features overlapped with that of the physicians (eg, planning exercises and daily tracking) and expanded in number postsurvey. Setting a plan with exercises and daily tracking remained *must-haves* for patients throughout. Patients scored a 7-year osteoarthritis severity prediction as attractive presession and postsession; researchers and physicians

were indifferent. Researchers reported few *must-haves* and aligned with physicians (ie, tracking functional impairment, graphing symptoms, and the ability to plan exercises and track daily).

The *top 6* functional requirements generated by convenience assessments differed from importance assessments in only 2 ways: the inclusion of reminders and the shifting of self-management strategies to a slightly lower rank. Stiffness symptom tracking and flags remained the lowest.

On the basis of the cumulative S1 to S3 findings and discussions, requirements were relabeled into 7 categories, as shown in Table 6.

Table 6. Session 3 dot voting results: revised functional requirement categories and summary of must-have and won't-have features.

Functional requirements	Must-have features				Won't-have features			
	Patients, n	Physicians, n	Researchers[a], n	Total, n	Patients, n	Physicians, n	Researchers[a], n	Total, n
Symptoms graph and summary (charts, diagrams to visualize symptoms, goal achievement, context, and communication)	4	1	3	8	0	0	0	0
Severity prediction (7-year osteoarthritis severity prediction tool [42], WOMAC[b])	0	0	0	0	3	1	0	4
Setting goals (shared goal setting including work, chores, sports, and hobbies)	0	1	2	3	1	0	0	1
Tracking activity (for events and outcomes, including activities, pain swelling, function, mood, fatigue and interventions, plans, and activities)	0	0	0	0	0	0	0	0
Reminders (reminders to update customized patient information)	0	0	0	0	0	0	3	3
Flags (flags for identifying arthritis burdensome days)	0	0	0	0	2	1	2	5
Information and strategies (self-management strategies including exercises, other conditions, medications, red flags, local resources)	2	0	0	2	0	0	0	0

[a]Missing data (n=1).

[b]WOMAC: Western Ontario and McMaster Universities Arthritis Index.

Session 3 Dot Voting Findings: Must-Have, Won't-Have, Desirability, Actionability

There was a strong preference for a symptom graph and summary (Table 6), consistent with previous assessments. Although physicians and researchers rated setting goals as a *must-have* feature, patients favored information and strategies.

Flags was perceived as the highest *won't-have* requirement (n=5). Patients (n=3) and physicians (n=1) *did not want* severity prediction, which was consistent with physician importance, but contrary to patient importance findings (Tables 3 and 4). Researchers did not want reminders (n=3).

The most *desirable* MVP features were: symptoms graph and summary, setting goals, information and strategies, severity prediction, tracking activity, and reminders, with preference variability for severity prediction and reminders, consistent with importance and convenience assessments (Multimedia Appendix 3). The least *desirable* requirement was *flags*, which was unchanged from other assessment findings.

Actionable requirements by descending frequency were symptoms graph and summary, tracking activity, setting goals, severity prediction, information and strategies, reminders, and flags (Multimedia Appendix 4). With the exception of severity prediction, the findings were highly consistent.

Discussion

Principal Findings

Co-design participants considered and prioritized MVP functional requirements using iterative, qualitative co-design [26] methods over 3 interactive sessions. Overall, the highest priority requirements included the following: (1) symptoms graph and summary; (2) setting goals; (3) tracking activity; and (4) information and strategies.

Clear differences in preferences existed between stakeholders and were documented throughout the research process. These findings are consistent with other studies examining the priorities and needs of patients and physicians in the management of knee osteoarthritis [22], yet diverse perspectives

among patients and physicians did not preclude consensus on MVP functional requirements.

The continuous involvement of participants from inception ensured that the understanding, language, and goals were negotiated. We supported diversity of thought with shared governance and decision making, role clarity, and by maintaining a safe, respectful, transparent research environment.

Limitations

The pattern of our findings and their consistency with related literature suggest that we captured important alignments and differences between participants through co-design; however, further study to reveal co-design dynamics is required.

It is possible that the patient participants in this study are not representative of *typical* patients with knee osteoarthritis. Individuals actively participating in research over longer periods are more likely to be motivated and have different needs and priorities than those who are less engaged in their own self-management [43-45]. Further validation of these findings within a broader knee osteoarthritis patient population is necessary.

This study was limited in scope to MVP development and purposely involved a smaller user group; however, the research was nested within a larger collaborative research program. Study results were immediately integrated into subsequent co-design and *alpha* testing with a representative sample of patients and physicians (personal communication by DA Marshall, November 2019).

Comparison With Prior Work

Early S1 and S2 findings revealed high-level agreement on functional requirements among co-design partners. These findings align well with findings from mixed group co-design research by Revenas et al [16,46,47]. However, from the point of initial agreement forward, we observed different, evolving thought patterns from patient, physician, and researcher perspectives. Fluctuating findings reflected emergent mutual understanding and thought diversity [22] among co-design partners. For example, group-level findings varied as a shared understanding of severity prediction emerged, relative to other functional requirements. The discussion revealed a newly developed understanding that the extent and timing of future physical knee osteoarthritis decline was not valued by patients relative to their proximal, preventative needs. Although severity prediction was initially important for patients and physicians, and patients thought it desirable/actionable, it became a clear *won't-have* feature for patients, and rated less important and convenient for both groups, by the end of S3.

Persistent thought diversity observed throughout our study contrasts with reports of thought homogeneity [16,17,46,47], or what Revenas et al describe as *participant convergence*. *Participant convergence* is the blending of participant perspectives [16,17,46,47], resulting in the inability to discern the diversity of patient and provider voices in the process [17]. As Pilemalm and Timpka report [26], the application of a rigorous co-design approach helped us establish a transparent, safe, and supportive environment for participants to freely express and consider diverse viewpoints [34], and helped preserve thought diversity. Voices were integrated by iteratively raising, airing, and reconciling conflicting interests [34]. These findings also align with a case study by Craven et al [48], in that all participants shared responsibility for identifying discrepancies and contributed to their exploration and reconciliation. In summary, we avoided loss of voice by involving multiple representative co-design partners (ie, patient, physician, and researcher) at each session, using reflective, responsive group processes, supporting open, transparent dialogue and power sharing, developing common language, and actively fostering a culture of mutual respect for the differing, yet equally valued contextual expertise of participants [49,50].

Our findings were highly consistent with key requirements identified by mixed users co-designing rheumatoid arthritis self-management tools, including customization, self-regulation, and exercise planning or follow-up [47]. They were also consistent with general design features of effective electronic health interventions, including social context and support, contacts with intervention, tailoring, and self-management [12,13,51,52].

The continued involvement of participants through each research phase was key [22]. In doing so, we may have avoided common difficulties such as establishing a shared starting point, rationale, and purpose. These and other commonly documented design process challenges (eg, maligned goals and tasks or difficulties turning ideas into concrete app features) [48] did not arise.

From inception, the team discussed expectations for participation and engaged in negotiation and clarification of roles [34]. Shared governance and decision-making principles were operationalized, as evidenced by the groups' spontaneous repurposing of S2 dot voting methods. Participants openly expressed reservations and negotiated this modification with ease. It is possible we avoided commonly documented partnership challenges [17,48,52] by adopting structural and process components, ensuring adequate resources, and time, by actively engaging our stakeholders in revealing and reconciling multiple, diverging perspectives [34], and by matching participants with research phases [26,53].

The findings are promising; however, systematic assessment and quantification of co-design processes, outcomes, and impact is needed to validate these findings and reveal co-design mechanisms [54].

Finally, there are mixed effectiveness findings and documented challenges associated with mHealth development to support and manage patients with chronic conditions [13,51] such as knee osteoarthritis. These include a lack of sustained app use, diminished product relevance, low daily patient routinization [48,55], high turnover, low app use, and disloyalty [14]. These challenges, coupled with the low reported likelihood of successful app development (an estimated 1 in 10,000 in 2018) [56], necessitate the use of deliberate strategies to optimize interactivity and app relevance for target users.

Our prioritized functional requirements address documented gaps [8], effective design features [12,52], and core mHealth characteristics [57]. For patients, prioritized requirements

addressed the inability of patients to track/assess symptoms (eg, pain), a lack of apps to support shared decision making with providers and support more informed patient self-management, and broadened focus beyond education [12,13,51]. For providers, prioritized requirements addressed the ability of an app to facilitate joint function measurement and enhanced decision support.

By reviewing the high-priority, midrange, and low-priority MVP functional requirements, we co-designed an MVP that addresses important, documented barriers for patients and providers in their use of mHealth [51] to manage knee osteoarthritis. The research was carried out in a way that was inclusive of diverse perspectives, yet facilitated consensus.

Conclusions

In conclusion, this research represents an important intermediate step in an interactive, ongoing dialogue with knee osteoarthritis patients, their providers, and the health research community about mHealth use to support knee osteoarthritis management [22,23,58]. This study offers other researchers tangible rationale for and an example of tailored co-design. The findings reveal how structural and process aspects can facilitate the presence and authenticity of patient, provider, and researcher voices while optimizing an MVP for future research phases.

Acknowledgments

The team would like to acknowledge Towards Optimized Practice, now the Accelerating Change Transformation Team, for their assistance with physician recruitment. We acknowledge collaborative insights from Chris Smith (Alberta Bone and Joint Health Institute) and the administrative assistance of Ms Cassandra McLaughlin for project team and team meeting support. We are grateful for the assistance provided by Chris Hawk, Rachel Quapp, Ali Kamran, and Spencer Briere, supervised by Christian Jacob (Computer Science 599, Department of Computer Science, University of Calgary). We acknowledge the assistance of our industry partner, Tactica Interactive, including Kevin Glasier (Interactive Producer), Michael Coutts (Account Manager), Bryce Macewan (Project Manager), Robin Hillman (Business Analyst), and Chris Lacey (Creative Director), for their contributions to the collaborative co-design process, including assistance with S3, concepts/mood board wireframes, technical analysis, and prototyping.

This study was funded by the Canadian Institutes for Health Research Knowledge to Action Grant KAL147547. The funding agency was not involved in any aspect of project design, nor in the collection, analysis, or interpretation of the data and its dissemination through the development of this manuscript.

DAM is funded through a Canada Research Chair (2008-2018) and the Arthur JE Child Chair in Rheumatology Research. KJM is a doctoral candidate in the Department of Community Health Sciences, University of Calgary, supported by the Integrated Knowledge Translation Research Network, Canadian Institutes of Health Research Foundation Grant (FDN #143237). ENACT is funded by the Alberta Innovates Translational Health Chair Award (Grant# 201300234).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Session 3 co-design participant surveys.
[PDF File (Adobe PDF File), 196 KB - mhealth_v8i7e17893_app1.pdf]

Multimedia Appendix 2
Kano survey findings: summary of convenience response frequencies pre- and post-session 3, by functional requirement.
[PDF File (Adobe PDF File), 54 KB - mhealth_v8i7e17893_app2.pdf]

Multimedia Appendix 3
Session 3 dot voting results: desirability for each functional requirement by participant type.
[PNG File , 81 KB - mhealth_v8i7e17893_app3.png]

Multimedia Appendix 4
Session 3 dot voting results: actionability for each functional requirement by participant type.
[PNG File , 66 KB - mhealth_v8i7e17893_app4.png]

References

1. Allen KD, Golightly YM. Epidemiology of osteoarthritis: state of the evidence. Curr Opin Rheumatol 2015 May;27(3):276-283 [FREE Full text] [doi: 10.1097/BOR.0000000000000161] [Medline: 25775186]

2. Kiadaliri AA, Lohmander LS, Moradi-Lakeh M, Petersson IF, Englund M. High and rising burden of hip and knee osteoarthritis in the Nordic region, 1990-2015: findings from the global burden of disease study. Acta Orthop 2018 Apr;89(2):177-183 [FREE Full text] [doi: 10.1080/17453674.2017.1404791] [Medline: 29160139]
3. Maetzel A, Li LC, Pencharz J, Tomlinson G, Bombardier C, Community Hypertension and Arthritis Project Study Team. The economic burden associated with osteoarthritis, rheumatoid arthritis, and hypertension: a comparative study. Ann Rheum Dis 2004 Apr;63(4):395-401 [FREE Full text] [doi: 10.1136/ard.2003.006031] [Medline: 15020333]
4. 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 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
5. Mobile Cellular Subscriptions (Per 100 People). World Bank. 2019. URL: https://data.worldbank.org/indicator/IT.CEL.SETS.P2 [accessed 2019-08-02]
6. Clement J. Average Number of New Android App Releases Per Day From 3rd Quarter 2016 to 1st Quarter 2018. Statista. 2019. URL: https://www.statista.com/statistics/276703/android-app-releases-worldwide/ [accessed 2019-08-02]
7. World Health Organization. MHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2020-04-28]
8. Choi W, Zheng H, Franklin P, Tulu B. mHealth technologies for osteoarthritis self-management and treatment: a systematic review. Health Informatics J 2019 Sep;25(3):984-1003 [FREE Full text] [doi: 10.1177/1460458217735676] [Medline: 29090628]
9. 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 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
10. Arthritis: Self-Management Education Workshops. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/arthritis/interventions/self_manage.htm [accessed 2019-08-02]
11. Sorting Through the Evidence for the Arthritis Self-Management Program and the Chronic Disease Self-Management Program: Executive Summary of ASMP/CDSMP Meta-Analyses. Centre for Disease Control and Prevention. 2011. URL: https://www.cdc.gov/arthritis/docs/ASMP-executive-summary.pdf [accessed 2020-04-28]
12. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile health apps for self-management of rheumatic and musculoskeletal diseases: systematic literature review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730 [FREE Full text] [doi: 10.2196/14730] [Medline: 31769758]
13. Azevedo AR, de Sousa HM, Monteiro JA, Lima AR. Future perspectives of Smartphone applications for rheumatic diseases self-management. Rheumatol Int 2015 Mar;35(3):419-431. [doi: 10.1007/s00296-014-3117-9] [Medline: 25168866]
14. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
15. Farago P. App Engagement: The Matrix Reloaded. Flurry. 2012. URL: https://www.flurry.com/post/113379517625/app-engagement-the-matrix-reloaded [accessed 2019-08-02]
16. Revenäs A, Opava CH, Martin C, Demmelmaier I, Keller C, Åsenlöf P. Development of a web-based and mobile app to support physical activity in individuals with rheumatoid arthritis: results from the second step of a co-design process. JMIR Res Protoc 2015 Feb 9;4(1):e22 [FREE Full text] [doi: 10.2196/resprot.3795] [Medline: 25665589]
17. Revenas A. Co-Designing a Mobile Internet Service for Self-Management of Physical Activity in Rheumatoid Arthritis. Stockholm, Sweden: The Department of Neurobiology, Care Sciences and Society, Division of Physiotherapy, Karolinska Institute; 2016.
18. Scariot CA, Heemann A, Padovani S. Understanding the collaborative-participatory design. Work 2012;41(Suppl 1):2701-2705. [doi: 10.3233/WOR-2012-0656-2701] [Medline: 22317129]
19. Giacomin J. What is human centred design? Design J 2015 Apr 28;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
20. Graham I, Tetroe J, Maclean R. Some basics of integrated knowledge translation research. In: Turning Knowledge Into Action: Practical Guidance on How to Do Integrated Knowledge Translation Research. Adelaide, Australia: Lippincott-JBI; 2014.
21. What is the Minimum Viable Product? Venture Hacks. 2009. URL: http://venturehacks.com/articles/minimum-viable-product [accessed 2020-04-28]
22. Barber T, Sharif B, Teare S, Miller J, Shewchuk B, Green LA, et al. Qualitative study to elicit patients' and primary care physicians' perspectives on the use of a self-management mobile health application for knee osteoarthritis. BMJ Open 2019 Feb 1;9(1):e024016 [FREE Full text] [doi: 10.1136/bmjopen-2018-024016] [Medline: 30782723]
23. Miller JL, Teare SR, Marlett N, Shklarov S, Marshall DA. Support for living a meaningful life with osteoarthritis: a patient-to-patient research study. Patient 2016 Oct;9(5):457-464. [doi: 10.1007/s40271-016-0169-9] [Medline: 27085710]
24. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. Br Med J 2017 Aug 2;358:j3453 [FREE Full text] [doi: 10.1136/bmj.j3453] [Medline: 28768629]
25. Scaling Up Integration In Alberta: Pilot Study. EnACt. 2019. URL: https://primarycareresearch.ca [accessed 2019-01-08]

26. Graham ID, Kothari A, McCutcheon C, Integrated Knowledge Translation Research Network Project Leads. Moving knowledge into action for more effective practice, programmes and policy: protocol for a research programme on integrated knowledge translation. Implement Sci 2018 Feb 2;13(1):22 [FREE Full text] [doi: 10.1186/s13012-017-0700-y] [Medline: 29394932]
27. Shklarov S, Marshall DA, Wasylak T, Marlett NJ. 'Part of the team': mapping the outcomes of training patients for new roles in health research and planning. Health Expect 2017 Dec;20(6):1428-1436 [FREE Full text] [doi: 10.1111/hex.12591] [Medline: 28660732]
28. Cibere J, Sayre E, Guermazi A, Nicolaou S, Kopec J, Esdaile J, et al. Natural history of cartilage damage and osteoarthritis progression on magnetic resonance imaging in a population-based cohort with knee pain. Osteoarthritis Cartilage 2011 Jun;19(6):683-688 [FREE Full text] [doi: 10.1016/j.joca.2011.02.008] [Medline: 21329760]
29. Rogers E. Diffusion of Innovations. Fifth Edition. New York, USA: Free Press; 2003.
30. Kano N, Seraku N, Takahashi F, Tsjui S. Attractive quality and must-be quality. J Jpn Soc Qual Contr 1984;14(2):39-48.
31. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014 Mar;22(3):363-388 [FREE Full text] [doi: 10.1016/j.joca.2014.01.003] [Medline: 24462672]
32. Dot Voting. Government Of Ontario. 2018 Mar 02. URL: https://www.ontario.ca/page/dot-voting [accessed 2020-04-28]
33. Design Method Toolkit: Dot Voting. Digital Society School. 2015. URL: https://toolkits.dss.cloud/design/method-card/dot-voting/ [accessed 2020-04-28]
34. Pilemalm S, Timpka T. Third generation participatory design in health informatics--making user participation applicable to large-scale information system projects. J Biomed Inform 2008 Apr;41(2):327-339 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.004] [Medline: 17981514]
35. Herzberg F, Mausner B, Peterson R, Capwell D. Job Attitudes: Review of Research and Opinion. Pittsburg, PA: Psychological Service of Pittsburg; 1957.
36. Zacharias D. The Complete Guide to the Kano Model: Prioritizing Customer Satisfaction and Delight. Folding Burritos. 2018. URL: https://foldingburritos.com/kano-model/ [accessed 2020-04-28]
37. Fundin A, Nilsson L. Using Kano's theory of attractive quality to better understand customer experiences with e‐services. Asian J Qual 2003 Aug 21;4(2):32-49. [doi: 10.1108/15982688200300018]
38. Materla T. Analyzing Factors Affecting Patient Satisfaction Using the Kano Model. Scholars' Mine. 2018. URL: https://scholarsmine.mst.edu/doctoral_dissertations/2681/ [accessed 2020-06-03]
39. Moorman J. Leveraging the Kano Model for Optimal Results. Ux Magazine. 2012. URL: https://uxmag.com/articles/leveraging-the-kano-model-for-optimal-results [accessed 2020-04-28]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthr Cartilage 2007 Sep;15(9):981-1000 [FREE Full text] [doi: 10.1016/j.joca.2007.06.014] [Medline: 17719803]
42. Sharif B, Marshall D, Faris P, Kopec J, Guermazi A, Sayre E, et al. Development of predictive models for 7-year whole-joint multi-feature MRI score among a population-based early knee osteoarthritis cohort: results from the Valseko study. Osteoarthr Cartilage 2017 Apr;25(Suppl 1):S294. [doi: 10.1016/j.joca.2017.02.497] [Medline: 17719803]
43. Maguire K, Britten N. 'How can anybody be representative for those kind of people?' Forms of patient representation in health research, and why it is always contestable. Soc Sci Med 2017 Jun;183:62-69. [doi: 10.1016/j.socscimed.2017.04.049] [Medline: 28463721]
44. Martin GP. 'Ordinary people only': knowledge, representativeness, and the publics of public participation in healthcare. Sociol Health Illn 2008 Jan;30(1):35-54 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01027.x] [Medline: 18254832]
45. Oehrlein E. Tackling Representativeness: A Roadmap and Rubric. National Health Council. 2018. URL: http://www.nationalhealthcouncil.org/sites/default/files/Oehrlein_HTAi_Representativeness_final.pdf [accessed 2020-04-28]
46. Revenas A. SP0135 co-designing a mobile internet service for self-management of physical activity in rheumatoid arthritis. Ann Rheum Dis 2016;75(Suppl 2):32.3-3233 [FREE Full text] [doi: 10.1136/annrheumdis-2016-eular.6208]
47. Revenäs A, Opava CH, Åsenlöf P. Lead users' ideas on core features to support physical activity in rheumatoid arthritis: a first step in the development of an internet service using participatory design. BMC Med Inform Decis 2014 Mar 22;14:21 [FREE Full text] [doi: 10.1186/1472-6947-14-21] [Medline: 24655757]
48. Craven M, Lang A, Martin J. Developing mHealth apps with researchers: multi-stakeholder design considerations. In: Marcus A, editor. Design, User Experience, and Usability: User Experience Design for Everyday Life Applications and Services. New York, USA: Springer; 2014:15-24.
49. Salsberg J. Integrated knowledge translation and participatory research. In: Graham ID, Tetroe J, Pearson A, editors. Turning Knowledge Into Action: Practical Guidance on How to Do Integrated Knowledge Translation Research. Adelaide, Australia: Lippincott-JBI; 2014.

50. Salsberg J, Parry D, Pluye P, Macridis S, Herbert CP, Macaulay AC. Successful strategies to engage research partners for translating evidence into action in community health: a critical review. J Environ Public Health 2015;2015:191856 [FREE Full text] [doi: 10.1155/2015/191856] [Medline: 25815016]
51. Pellegrini CA, Steglitz J, Hoffman SA. E-health intervention development: a synopsis and comment on 'what design features are used in effective e-health interventions? A review using techniques from critical interpretive synthesis'. Transl Behav Med 2014 Dec;4(4):342-345 [FREE Full text] [doi: 10.1007/s13142-014-0283-y] [Medline: 25584082]
52. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from critical interpretive synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
53. Kothari A, McCutcheon C, Graham ID. Defining integrated knowledge translation and moving forward: a response to recent commentaries. Int J Health Policy Manag 2017 May 1;6(5):299-300 [FREE Full text] [doi: 10.15171/ijhpm.2017.15] [Medline: 28812820]
54. Hoekstra F, Mrklas KJ, Sibley KM, Nguyen T, Vis-Dunbar M, Neilson CJ, et al. A review protocol on research partnerships: a coordinated multicenter team approach. Syst Rev 2018 Nov 30;7(1):217 [FREE Full text] [doi: 10.1186/s13643-018-0879-2] [Medline: 30497527]
55. Novak LL, Unertl KM, Holden RJ. Realizing the potential of patient engagement: designing IT to support health in everyday life. Stud Health Technol Inform 2016;222:237-247 [FREE Full text] [Medline: 27198106]
56. Wilson N, Prentice B, Van Huizen HG, Baker V, Kyte A, Smith D. Predicts 2014: Application Development. Gartner Research. 2013. URL: https://www.gartner.com/en/documents/2625820/predicts-2014-application-development [accessed 2020-04-28]
57. Davis TL, DiClemente R, Prietula M. Taking mHealth forward: examining the core characteristics. JMIR Mhealth Uhealth 2016 Aug 10;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
58. Marshall D, Barber T, Shewchuk B, Mrklas K, Miller J, Teare S, et al. PP115 A mobile health app to improve knee osteoarthritis self-management. Int J Technol Assess Health Care 2019 Jan 3;34(S1):109-110. [doi: 10.1017/s0266462318002532]

Abbreviations

ENACT: Enhancing Alberta Primary Care Research Networks
mHealth: mobile health
MVP: minimum viable product
PACER: patient and community engagement research
WOMAC: Western Ontario and McMaster Universities Arthritis Index

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by R Dodd, C Cheng; comments to author 28.03.20; revised version received 29.04.20; accepted 13.05.20; published 10.07.20.

Please cite as:
Mrklas KJ, Barber T, Campbell-Scherer D, Green LA, Li LC, Marlett N, Miller J, Shewchuk B, Teare S, Wasylak T, Marshall DA
Co-Design in the Development of a Mobile Health App for the Management of Knee Osteoarthritis by Patients and Physicians: Qualitative Study
JMIR Mhealth Uhealth 2020;8(7):e17893
URL: http://mhealth.jmir.org/2020/7/e17893/
doi:10.2196/17893
PMID:32673245

Original Paper

Smartphone App to Self-Monitor Nausea During Pediatric Chemotherapy Treatment: User-Centered Design Process

Astrid Eliasen[1,2,3]; Mikkel Kramme Abildtoft[4]; Niels Steen Krogh[4]; Catherine Rechnitzer[2]; Jesper Sune Brok[2]; René Mathiasen[2]; Kjeld Schmiegelow[2,3]; Kim Peder Dalhoff[1,3]

[1]Department of Clinical Pharmacology, Bispebjerg and Frederiksberg University Hospital, Copenhagen, Denmark

[2]Department of Paediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark

[3]Institute of Clinical Medicine, Faculty of Medicine, University of Copenhagen, Copenhagen, Denmark

[4]ZiteLab ApS, Copenhagen, Denmark

Corresponding Author:
Astrid Eliasen
Department of Clinical Pharmacology
Bispebjerg and Frederiksberg University Hospital
Bispebjerg Bakke 23, indgang 20C 2. sal
Copenhagen, 2400
Denmark
Phone: 45 38635102
Email: astrid.eliasen@regionh.dk

Abstract

Background: Nausea and vomiting are common and distressing side effects for children receiving chemotherapy. Limited evidence is available to guide antiemetic recommendations; therefore, prospective and reliable evaluation of antiemetic efficacy is needed. Smartphone apps can be used to effortlessly and precisely collect patient-reported outcomes in real time.

Objective: Our objective was to develop a smartphone app to monitor nausea and vomiting episodes in pediatric cancer patients aged 0 to 18 years and to test its usability and adherence to its use.

Methods: We used a user-centered design process and the evolutionary prototype model to develop and evaluate the app. Multidisciplinary group discussions and several rounds of patient feedback and modification were conducted. We translated the validated Pediatric Nausea Assessment Tool to assess nausea severity in children aged 4 to 18 years. The child's own term for nausea was interactively incorporated in the nausea severity question, with response options expressed as 4 illustrative faces. Parent-reported outcomes were used for children aged 0 to 3 years. Reminders were sent using push notifications in order to ensure high response rates. Children aged 0 to 18 years who were undergoing chemotherapy were recruited from the Department of Pediatric Oncology at Copenhagen University Hospital Rigshospitalet to evaluate the app.

Results: The app's most important function was to record nausea severity in children. After assistance from a researcher, children aged 4 to 18 years were able to report their symptoms in the app, and parents were able to report symptoms for their children aged 0 to 3 years. Children (n=20, aged 2.0-17.5 years) and their parents evaluated the app prospectively during a collective total of 60 chemotherapy cycles. They expressed that the app was user-friendly, intuitive, and that the time spent on data entry was fair. The response rates were on average 92%, 93%, and 80% for the day before, the first day of, and the next 3 days after chemotherapy, respectively. Researchers and clinicians were able to obtain an overview of the patient's chemotherapy dates and responses through a secure and encrypted web-based administrative portal. Data could be downloaded for further analysis.

Conclusions: The user-friendly app could be used to facilitate future pediatric antiemetic trials and to refine antiemetic treatment during chemotherapy.

(JMIR Mhealth Uhealth 2020;8(7):e18564) doi:10.2196/18564

KEYWORDS

mobile applications; patient-reported outcome measures; patient compliance; neoplasms; antiemetics; nausea; vomiting; cancer; children; app

Introduction

Background

Chemotherapy-induced nausea and vomiting are among the most common and burdensome side effects that are experienced by children who are undergoing cancer treatment [1,2]. The symptoms are often accompanied by profound physical and psychological consequences, including dehydration, electrolyte imbalance, weight loss, and disruption of normal childhood activities [1,2]. The prevention of chemotherapy-induced nausea and vomiting has been improved over the years with novel antiemetic drug combinations [3-6], but approximately 50% of pediatric cancer patients still suffer from chemotherapy-induced nausea and vomiting [7-9].

Correct assessment is an essential first step to manage nausea experienced by children with cancer, but it is difficult to assess subjective symptoms such as nausea in children. Clinicians and children express and interpret nausea differently, which makes it a challenge to score the symptoms with high validity and reliability; therefore, antiemetic efficacy in pediatric trials is often solely based on vomiting (an objective outcome), while nausea (a subjective outcome) is rarely included as a clinical outcome [7,8,10]. Nausea is an important outcome measure because it occurs more often than vomiting and is one of the most distressing side effects for children with cancer [2]. Clinical trials should, therefore, assess antiemetic efficacy in children receiving chemotherapy using both nausea and vomiting as clinical outcomes.

The side effects of chemotherapy are often underdocumented by clinicians during prospective clinical trials [11-13]. Patient-reported outcomes can give more valid and reliable information on the side effects of chemotherapy [14]. Patient-reported outcomes are defined as symptoms that are directly reported by the patient, without interpretation by a clinician or anyone else and are used as clinical trial outcomes [14]. One barrier to accurately reporting the side effects experienced by patients as patient-reported outcomes is that the accuracy can be affected by the patient's memory [14]. It is extremely important to record patient-reported outcomes at assigned times during pediatric clinical trials as recall bias may be more pronounced among children than among adults [15]. Also, pediatric cancer patients must often commute frequently between hospital and home, and some children live alternately with different parents or caregivers; in these situations, paper diaries may get lost or forgotten.

Prior Work

Electronic diaries can collect patient-reported outcomes with more precision and effectiveness than paper diaries [16], and the increased number of smartphone users allows an opportunity to develop smartphone apps that collect real time patient-reported outcomes [17,18]. Even though pediatric patient-reported outcomes are advocated, parent-reported outcomes may be needed when the child is too young to complete a patient-reported outcome questionnaire [19].

There are several smartphone apps for adult patients with cancer that can monitor real time patient-reported outcomes and that are used in clinical trials and to support patients as they manage the side effects of chemotherapy [17,20,21]. In contrast, only a few smartphone apps have been developed to monitor patient-reported outcomes in children and adolescents with cancer [22]. Wang et al [23] developed a smartphone app where patients aged 8 years or older were able to record several symptoms such as pain, fatigue, depression, and anxiety. A user-centered design process has also been applied in the development of smartphone apps to assess and manage pain in children and adolescents [24,25]; Stinson et al [24] developed a game-based smartphone app to assess pain in adolescents with cancer [24]. The game-based nature and built-in reward system was appealing to users and promoted high adherence rates with a mean of 81% adolescents completing 100% of the tasks within a 2-week period [24].

Study Goals

In this study, we aimed to develop and evaluate a smartphone app to monitor nausea severity in children from 0 to 18 years of age undergoing chemotherapy. We used a user-centered design process where we engaged children who were undergoing chemotherapy and their parents during the design, development, and evaluation of the app. The evaluation process included an investigation of the usability of the app and of adherence to the app's use.

Methods

Overview

This study was conducted in Copenhagen, Denmark and was part of a prospective observational study. The study was approved by the Danish Data Protection Agency (RH-2016-219, I-Suite: 04804) and the Danish National Research Ethics Committee (H-18020267). The app was approved by the regulatory authorities in *Center for It, Medico og Telefoni* in the Capital Region of Denmark.

We used a user-centered design process and the evolutionary prototype model to develop and evaluate the app [26,27], and we involved pediatric cancer patients and their parents throughout the process (Figure 1 and Figure 2). Medical researchers and software engineers worked synergistically in the development of the app. Pediatric oncologists were consulted throughout the process because they have an explicit understanding of the users (pediatric cancer patients). The app was developed in 4 phases: (1) The context of use was specified (ie, the users, the tasks that the users would perform, and the environment in which the users would use the app were characterized), (2) requirements were specified (ie, requirements and user goals that must be met for the app to be successful were identified), (3) design solutions were produced (ie, an app prototype was developed), and (4) the app was evaluated (ie, for data completeness and to further refine the app).

Figure 1. The app was developed and evaluated with the user-centered design process (adapted from ISO 9241-210:2019).

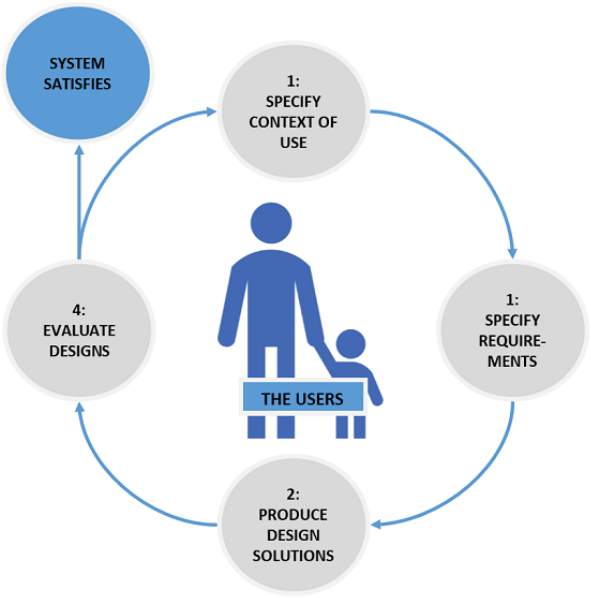
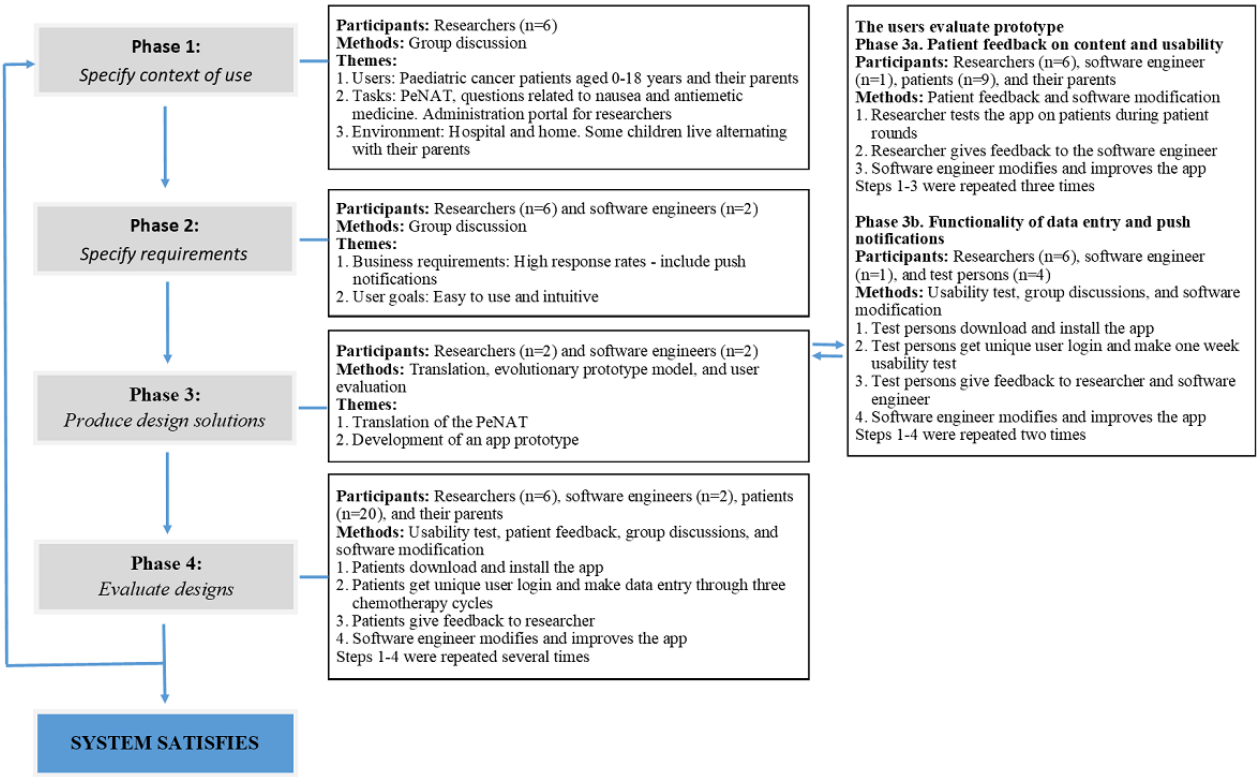


Figure 2. Workflow of the development and evaluation process according to the user-centered design process and the evolutionary prototype model.



Specify Context of Use (Phase 1)

Researchers including pediatric oncologists (n=4), a clinical pharmacologist (n=1), and a medical researcher (n=1) participated in the first discussions. The advantages of developing an app and its content, as well as strategies to obtain complete data sets were discussed. The discussion was based on previous experience with pediatric antiemetic trials and a literature review that was used to identify the best validated tool for pediatric patient-reported nausea severity.

Specify Requirements (Phase 2)

The requirements were discussed with 2 software engineers from ZiteLab in Copenhagen, Denmark [28]. A requirement was that the app should include a customized notification system to ensure high response rates. We aimed for a user-friendly and intuitive app, and the usability was refined during several rounds of patient feedback and modification (phases 3a and 3b).

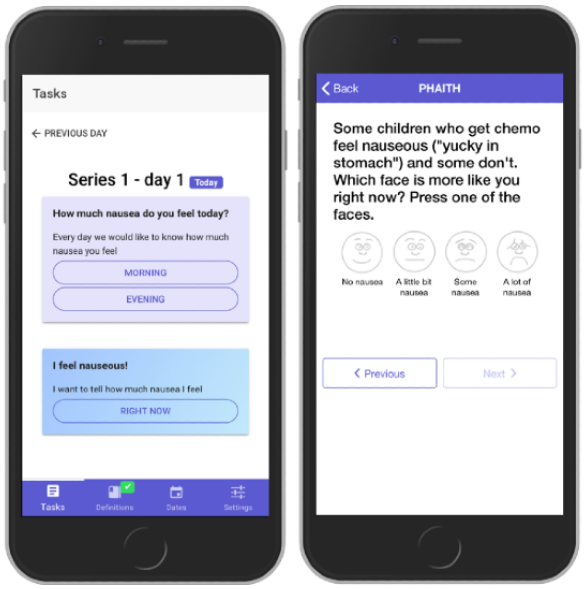
Produce Design Solutions (Phase 3)

Overall Structure

The app was developed with hybrid web-based technology to allow a single code base to be deployed for both iPhone and Android operating systems. Feedback was received from the users (phases 3a and 3b), and subsequent prototypes were produced, each with additional functionality or improvements until a fully functional app was developed [27].

The users were able to get an overview of the day's tasks when the app was opened (Figure 3). Each morning, the user reported nausea severity. Each evening, the user reported nausea severity, retching and vomiting episodes, and antiemetic medications. In addition, users could report nausea severity at any time they (or in the case of parents for children from 0 to 3 years of age, their child) felt nauseated.

Figure 3. The app with overview of the day's task (left) and the patient's expression of nausea severity (right).



Data Entry on Nausea Severity

We translated the Pediatric Nausea Assessment Tool (PeNAT), a valid and reliable tool to assess nausea severity among children who are pediatric oncology patients (from 4 to 18 years of age) [29], from English to Danish. A questionnaire in the PeNAT ensured that the child understood the concept of nausea, and it focused the child's attention to their feeling of nausea. Four faces in the PeNAT corresponded to the nausea severity encountered by the child—no nausea, mild nausea, moderate nausea, and severe nausea (Figure 3). Children older than 8 years of age were presented with the 4 faces simultaneously, while children aged 8 years or younger were presented with 2 faces simultaneously, and the pair presented depended on whether the child said they felt no nausea or some nausea.

We translated the PeNAT according to The International Medical Interpreters Association Guide on Medical Translation [30]. The developers of the PeNAT were contacted for an explanation of the concepts before the translation process. A pediatric oncologist (CR) and a medical researcher (AE), both with substantial knowledge of the subject terminology, independently translated the tool from English to Danish. The translations were compared and merged into a single final translation, and if necessary, a third researcher (MKA) was consulted.

The app had a definition module where the child was able to define the term used for nausea as they understood it, in accordance with the PeNAT [29]. The child could then express nausea severity by tapping one of 4 faces that were presented horizontally and that represented increasing nausea severity from left to right; the child's own terminology for nausea was incorporated in the question (Figure 3).

The app was also developed to address the need for parent-reported outcomes for children aged 0 to 3 years. For example, one question in the PeNAT is "Which face is more like you right now?"; in the parent-reported tool, this question is worded as "Which face is more like your child right now?"

Accordingly, 3 versions of the app were developed: one for children aged 0 to 3 years (parent-reported outcome), one for children aged 4 to 8 years (2 faces presented simultaneously), and one for children aged 9 to 18 years (4 faces presented simultaneously).

Data Entry on Questions Related to Nausea

The patients were asked about vomiting and retching episodes each day. These questions were answered with 4 single bullet options: no; yes, 1 to 2 times; yes, 3 to 5 times; or yes, 6 or more times. An episode was defined as a vomit or retch that was separated from another vomit or retch by at least one minute.

Push Notifications

Push notifications were sent at 9 AM and 7 PM if the questionnaires had not been answered each morning and evening, respectively.

Web-Based Administration Portal

The web-based administrative portal was created as a secure and encrypted home page. Data were anonymized (with

pseudonyms) and data storage and transmission followed general data protection regulations. The portal consisted of a front end (presentation layer for researchers) and a back end (data access layer for software engineers).

The researchers could create unique log-ins for patients including username, password, age group (0 to 3 years, 4 to 8 years, or 9 to 18 years), and scheduled chemotherapy dates. The dates could be changed by the researchers, and the users could forward changes to the researchers with a simple tap in the app, if the dates were incorrect. The researchers could get an overview of the patient's scheduled chemotherapy dates and view the patient's responses and response rates. The data could be downloaded for further analysis.

Patient Feedback on Content and Usability (Phase 3a)

The pilot assessment feedback came from children who were pediatric patients and their parents, all of whom were recruited at the Department of Pediatric Oncology, Copenhagen University Hospital Rigshospitalet, during regular patient rounds in November 2018 and December 2018. We used a qualitative usability testing approach with 3 cycles of patient feedback and modification. Children (n=9) and their parents participated (3 in each of 3 cycles). AE introduced the app to the test participants who were asked to comment on words, to interpret phrases in the questionnaire, and to suggest improvements to the interface. A software engineer (MKA) modified the app according to the feedback. The app prototype was modified until no further changes were suggested.

Functionality of Data Entry and Push Notifications (Phase 3b)

Clinical researchers (n=4) from the Department of Clinical Pharmacology, Bispebjerg and Frederiksberg University Hospital tested the functionalities of log-in, data entry, and push notifications during 2 test periods, each lasting one week. Test persons (n=4), 2 with an Android and 2 with an iPhone smartphone, downloaded the app and were given unique log-in information. The test person registered if they were able to log-in, if they received push notifications, and if they could enter data in the app. MKA and AE performed group discussions with these individuals after each test period and discussed the functionality of the app. MKA modified the app according to the feedback.

Evaluate Designs (Phase 4)

Following the development of a fully functional app, we included patients to test usability of the app and adherence to its use. Patients (n=20) were recruited at the Department of Pediatric Oncology, Copenhagen University Hospital Rigshospitalet during March 2019 and April 2019. All parents provided written informed consent for their child to participate in the study. Eligible patients were children aged 0 to 18 years who would be receiving at least 3 cycles of chemotherapy. The child and at least one parent was required to be able to communicate fluently in Danish. At least one parent needed to have an Android or iPhone smartphone. Parents of children aged 0 to 3 years needed to be able to describe the nausea severity experienced by their child.

One or both parents downloaded and installed the app from Google Play or App Store and were given unique log-in information. It was emphasized that the data collected were to be used for research purposes, and that the parents should also inform their pediatric oncologist or nurse about their child's symptoms. The questions in the app were delivered to each child by AE who also taught the parents to administer the app. The children (with assistance from their parents) were asked to record their own nausea severity in the app, and AE noted if they were able to naturally navigate through the app's features. The parents could then assist the child to enter data about nausea severity for the day before and for the first 4 days after the start of chemotherapy during 3 chemotherapy cycles. The children and their parents were contacted in person before each cycle to promote adherence to use of the app. A semistructured interview was conducted at that time, and the participants were asked what they liked or disliked about the app, if the app was easy to use, and if the time spent on data entry was fair. AE provided continuous feedback to MKA who modified the app according to the feedback. This was continued until there were no further recommendations for changes in the app.

The web-based administrative portal was fully developed in this phase. AE tested if log-ins for new patients could be created and if the portal provided a user-friendly interface and overview of data completeness. MKA tested whether the back end received the entered data accurately and if data could be downloaded for further analysis.

We subsequently interviewed parents of patients aged 0 to 3 years in order to determine if the parents could grade the nausea severity experienced by their child.

A paired, two-tailed *t* test was used to determine if the response rates differed between morning and evening and if the response rates differed between chemotherapy cycles. $P<.05$ was deemed significant.

Results

Context of Use, Requirements, and Design Solutions (Phases 1, 2, and 3)

During phases 1 and 2, we specified in which context the app would be used, and we agreed on the overall requirements of the app (Figure 1 and Figure 2). Based on the functions and requirements that were described by the researchers, in phase 3, a beta-version smartphone app was created by software engineers from ZiteLab in Copenhagen, Denmark [28]. These functions and requirements were refined and modified according to feedback from children who were pediatric cancer patients, parents, and participants (phase 3a and phase 3b).

Participant Characteristics

The demographic and disease characteristics of the children who were pediatric cancer patients (phases 3a and 4) and the demographic characteristics of clinical researchers (phase 3b) who participated in the user-centered testing are shown in Table 1.

Table 1. Baseline demographics and disease characteristics of participants who were included in the development and evaluation of the app.

Characteristics	Phase 3a (n=9)	Phase 3b (n=4)	Phase 4 (n=20)
Age in years[a], mean (range)	7.2 (2.9-16.5)	32 (30-36)	7.4 (2.0-17.5)
Gender, n (%)			
Female	4 (44)	4 (100)	8 (40)
Male	5 (56)	0 (0)	12 (60)
Primary diagnosis, n (%)			
Acute lymphoblastic leukemia	5 (56)	N/A[b]	9 (45)
Brain tumor	1 (11)	N/A	3 (15)
Hodgkin lymphoma	1 (11)	N/A	3 (15)
Kidney tumor	0 (0)	N/A	1 (5)
Non-Hodgkin lymphoma	1 (11)	N/A	1 (5)
Osteosarcoma	1 (11)	N/A	2 (10)
Rhabdomyosarcoma	0 (0)	N/A	1 (5)
Duration of illness in days[a], mean (range)	111 (7-251)	N/A	110 (1-299)

[a]At the time of study recruitment.

[b]N/A: not applicable.

Patient Feedback on Content and Usability (Phase 3a)

The questions in the PeNAT were clear and easy to understand, and the patients and their parents had no comments relating to the meaning of words or phrases in the translation. The parents of one newly diagnosed patient (less than one week) reported that it would be demanding to enter data, which was in contrast with the feedback from parents of patients who had been through several rounds of chemotherapy, who expressed that they could easily overcome to enter data in the app.

One suggestion for the interface was to include emojis as pictorial expressions of the questions. Emojis were not included in the PeNAT part of the app to be true to the original PeNAT paper version.

Through interviews with parents, we found that the actual antiemetics that were given did not always correspond with the prescribed medications in the medical record, especially when the child was discharged from hospital. To capture the actual medication use, a medicine module was developed to let parents fill in administered antiemetics, numbers of administrations, and whether the patient used antiemetic rescue medications on any given day. The medicine module included an autofill function where the parents could copy answers from the previous day with a single tap and modify answers that had changed.

Functionality of Data Entry and Push Notifications (Phase 3b)

During the first test period, problems with the log-in function required that participants restarted the app. All test participants received push notifications informing them to enter data in the app, and they were subsequently able to do so.

Design Evaluation (Phase 4)

The app was evaluated prospectively in patients (n=20) during a total of 60 chemotherapy cycles.

Usability

Overall, the patients and their parents expressed that the app was user-friendly, intuitive, and that time spent on data entry was fair. Satisfaction with the app was illustrated with the following quotes: A male participant aged 12 years said, “It is a really cool app, and the questions are very relevant.” A parent of a 4-year old female said that, “she really likes to use the app, and she also uses the ad hoc function to tell how she is feeling.”

The semistructured interviews revealed suggestions for changes in the app. First, the log-in session was expanded from 2 hours to 3 months, so that the users did not need provide log-in information at every new task. Second, 15% (3/20) of patients were so nauseated that they could not overcome the nausea to answer the questions on that specific day; therefore, the parents requested a solution enabling them to go back and help the child answer questions from the previous day. This option was then included in the app allowing users the possibility to return to answer questions from within the previous 2 days. Third, the days with questions were individualized, so they continued for 5 days after the last day with chemotherapy in the cycle. This feature was added because 15% (3/20) of the children who received chemotherapy over several days felt nauseated longer than four days after the first chemotherapy administration.

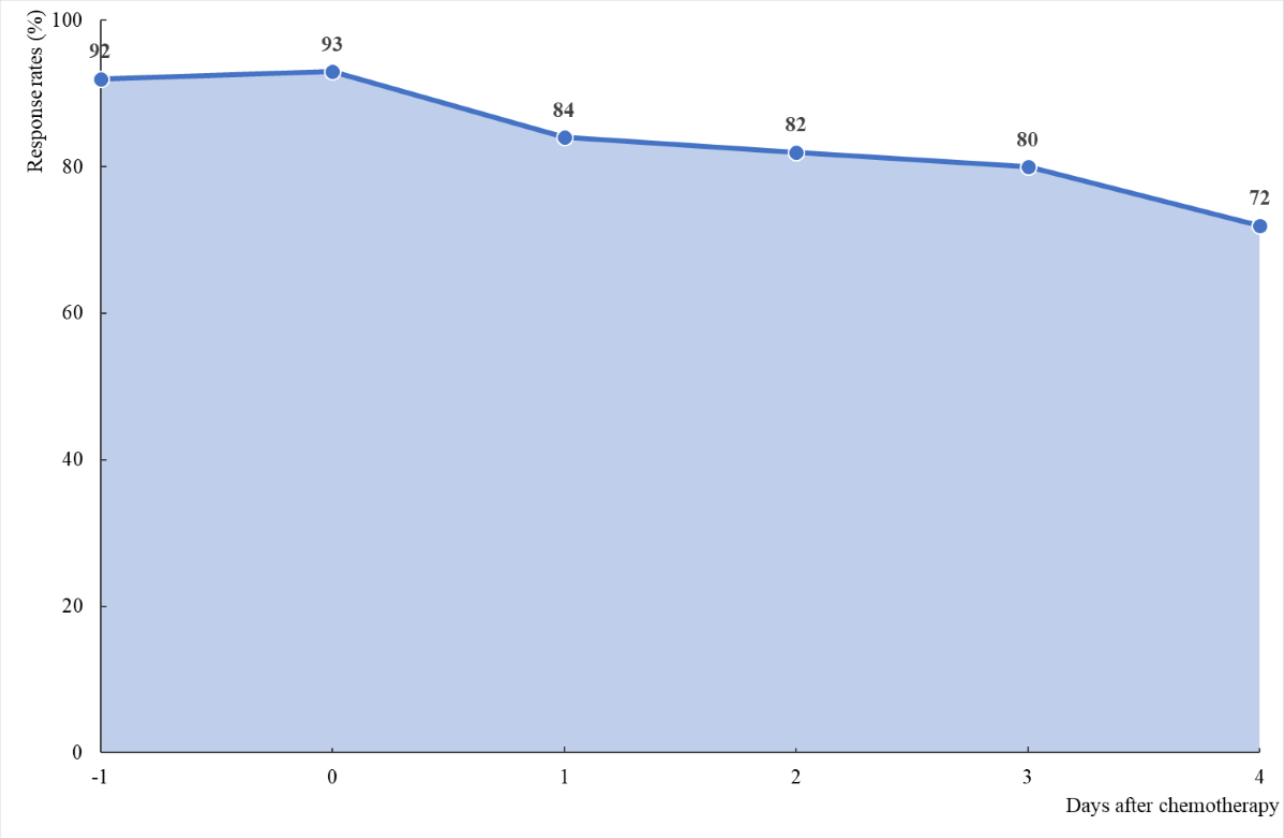
Two adolescent patients suggested that the app should include a module with an overview of their answers. It is our intention to include this feature in a future version of the app.

Adherence to Use

All patients and their parents expressed that push notifications were a convenient and efficient method to ensure that they completed their tasks. The response rates were on average 92%, 93%, and 80% for the day before, the first day, and the next 3 days after chemotherapy, respectively (Figure 4). Individual response rates of 100%, 50%, or 0% were given each day if the

patient answered both morning and evening, answered either morning or evening, or did not answer, respectively.

Figure 4. Data completeness for 20 patients using the app during 60 chemotherapy cycles.



The personal contact between the researcher and the participants before each chemotherapy cycle promoted high response rates the day before and the day of chemotherapy administration. The next 3 days after chemotherapy, 15% (3/20) of the children did not respond, each with different reasons—one patient did not receive the push notifications, one patient had difficulties with the log-in function, and one patient had a limited internet connection at home. One adolescent (1/20, 5%) downloaded the app on her own smartphone but did not complete the questions due to severe nausea. Her parents then downloaded the app and assisted her.

There were no significant changes in response rates when response rates for cycles 1 to 3 were analyzed separately (t_5=2.57, P=.45 for comparing response rates for cycle 1 with those of cycle 3). The response rates in the morning were slightly higher than those in the evening, but the difference was insignificant (t_5=2.57, P=.11).

Additional Data on Usability for Patients Aged 0 to 3 Years

The interviews showed that most parents (21/22, 95%) of the children aged 0 to 3 years could explain how their child reacted when feeling nauseous, and they could use the app to express their child's nausea severity. Only the parents of one child (1/22, 5%) did not know how their child reacted when feeling nauseous.

Discussion

Principal Results

We described the process of developing and evaluating a smartphone app to collect pediatric patient-reported outcomes of nausea severity and episodes of vomiting and retching. Patients were able to self-report in the app the severity of nausea that they experienced with assistance from their parents, and clinicians and researchers had a good historical overview of their symptoms in the web-based administrative portal. Multidisciplinary group discussions ensured that the app and the web-based administrative portal met both medical and technical demands. It was essential that the app had a high usability, so patients and their parents were involved early in the development to get insight into their preferences.

Generally, patients prefer mobile data collection over more traditional methods such as paper diaries, because it is easier and faster to record data into a mobile device than it is to record in a paper diary [31]. Some children thought it was interesting to answer questions in the app because they liked to press the faces reflecting nausea severity. They easily understood the interactive component because the child's own term for nausea was incorporated into the questions. These features are not possible in a paper diary or a web-based survey.

Usability of the Smartphone App

Nausea has a higher incidence than vomiting during administration of chemotherapy, but it is a challenge to

recognize subjective symptoms such as nausea in children. Furthermore, children under 4 years of age are often excluded from studies that investigate the burden of chemotherapy-induced nausea and vomiting in children [32-34]. The target users of this app are children aged 0 to 18 years (and their parents) undergoing cancer treatment. Our subjective impression was that children aged 4 to 18 years were able to express their nausea severity in the app. One 17-year-old patient expressed (about nausea) that "it hurts in my stomach and I have no appetite," whereas a 7-year-old patient expressed that nausea was "like an elevator moving up through the stomach and all the way to the throat." Children aged 0 to 3 years are too young to complete a patient-reported outcome instrument, and the PeNAT is not validated for this age group; however, parents of patients aged 0 to 3 years expressed that they could easily grade the nausea severity experienced by their child. By including this age group, researchers were able to obtain data about parent-reported nausea severity and vomiting episodes for children aged 0 to 3 years.

Patient Adherence With Paper-Based and Electronic Diaries

It was important that the patients adhered to the use of the app, including patients with severe nausea who likely had reduced energy to perform extra tasks during chemotherapy. In a paper version of the PeNAT, 17% of the patients did not use the questionnaire for three or more days out of seven days [32]. The high response rates in our study were attributed to the individual patient reminders. Also, data were captured in real time, and we could identify and respond to nonadherence early by following the patient responses in the web-based administrative portal. The reasons for nonadherence were mainly technical difficulty–based (mainly in the first version of the app) and not that the tasks were too demanding. The response rates did not decrease over three chemotherapy cycles; therefore, the app has the advantage that it could be used in antiemetic trials using a crossover design, where patients undergo two chemotherapy cycles.

Real time patient-reported outcomes reduce the risk of recall bias and improve data quality. After requests from patients and parents, patients were allowed to return to answer questions from within the previous 2 days. This allowed patients with severe symptoms on one day to respond the following day. The survey time was recorded in the administration system which permitted analysis of data accuracy and the ability to exclude data that was not registered in real time. This is in contrast to a paper-based approach, where the researchers are not aware of whether patients are back filling or completing the paper diary just prior to returning it to the researchers [24].

Challenges in Using Apps to Collect Patient-Reported Outcomes

The digital revolution offers exciting new tools to support research, but along with these new opportunities come some challenges. We conducted a preliminary assessment of the suitability of apps in the target population, and every participant in this study had a smartphone. Other target populations may not have access to digital technology, and the target population should, therefore, be assessed to determine whether digital data collection is appropriate.

The development of an app is time consuming compared to the time required to develop a paper diary or a simple electronic web-based survey. Simple electronic diaries for use on smartphones can be created at online research platforms without extensive programming skills [35,36]. More advanced options, such as individual patient reminders and interactive components, require assistance from experienced software engineers. Also, data storage in these platforms may not adhere to general data protection regulations. The cost of developing an app depends on the complexity of the interface and functions, and additional costs must be considered for data storage and distribution in Google Play and App Store; however, electronic data collection reduces the time used for data handling because researchers do not need to manually record data from paper diaries [31]. This more precise and efficient data collection method could also outweigh the time and price spent on preparing the trial [16-18].

Technical support needs to be available to maintain and upgrade the app over time in clinical trials using apps for data collection. Technical problems mainly arose when the software was updated or when a new version of the app was launched. The major technical problems were that patients could not log-in to the app and that push notifications were not sent. Other barriers that could prevent efficient data entry are limited access to internet or if a patient signs out and forgets the username and password. These problems require that investigators of the clinical trial pay attention to both the function of the app and to feedback from patients throughout the clinical trial so that technical problems can be solved immediately. In our study, one researcher was signed into the app with a test user log-in to capture technical difficulties which were then directly forwarded to a software engineer. A researcher also directly transcribed patient-reported outcomes from the patients at any time they were not able to report patient-reported outcomes in the app.

The data entered by the patients will be used to determine patient-related risk factors of chemotherapy-induced nausea and vomiting, and the app will be used to collect pediatric patient-reported outcomes in upcoming antiemetic trials. The overview of the patient-reported outcomes can be used in the future by clinicians to better refine antiemetic treatment for forthcoming chemotherapy.

Conclusions

The app and the web-based administrative portal demonstrated good usability for patients and researchers, and pediatric patient-reported outcomes of nausea severity were collected efficiently. The app can be used to facilitate future research regarding antiemetic efficacy in pediatric cancer patients and to better refine antiemetic treatment in upcoming chemotherapy cycles.

Acknowledgments

We sincerely appreciate the generosity of the children and parents who participated in this study. This study was funded by the Danish Childhood Cancer Foundation and the Dagmar Marshall Foundation.

Conflicts of Interest

KS has received personal fees from Jazz Pharmaceuticals and Servier for work other than that presented herein. The authors are also the developers of the smartphone app.

References

1. Phillips RS, Friend AJ, Gibson F, Houghton E, Gopaul S, Craig JV, et al. Antiemetic medication for prevention and treatment of chemotherapy-induced nausea and vomiting in childhood. Cochrane Database Syst Rev 2016 Feb 02;2:CD007786 [FREE Full text] [doi: 10.1002/14651858.CD007786.pub3] [Medline: 26836199]
2. Hedström M, Haglund K, Skolin I, von Essen L. Distressing events for children and adolescents with cancer: child, parent, and nurse perceptions. J Pediatr Oncol Nurs 2003;20(3):120-132. [doi: 10.1053/jpon.2003.76] [Medline: 12776260]
3. Patel P, Robinson PD, Thackray J, Flank J, Holdsworth MT, Gibson P, et al. Guideline for the prevention of acute chemotherapy-induced nausea and vomiting in pediatric cancer patients: A focused update. Pediatr Blood Cancer 2017 Oct;64(10). [doi: 10.1002/pbc.26542] [Medline: 28453189]
4. Dupuis LL, Sung L, Molassiotis A, Orsey AD, Tissing W, van de Wetering M. 2016 updated MASCC/ESMO consensus recommendations: prevention of acute chemotherapy-induced nausea and vomiting in children. Support Care Cancer 2017 Jan;25(1):323-331. [doi: 10.1007/s00520-016-3384-y] [Medline: 27565788]
5. Hesketh PJ, Kris MG, Basch E, Bohlke K, Barbour SY, Clark-Snow RA, et al. Antiemetics: American Society of Clinical Oncology Clinical Practice Guideline Update. J Clin Oncol 2017 Oct 01;35(28):3240-3261. [doi: 10.1200/JCO.2017.74.4789] [Medline: 28759346]
6. Eliasen A, Dalhoff K, Mathiasen R, Schmiegelow K, Rechnitzer C, Schelde AB, et al. Pharmacogenetics of antiemetics for chemotherapy-induced nausea and vomiting: a systematic review and meta-analysis. Crit Rev Oncol Hematol 2020 May;149:102939. [doi: 10.1016/j.critrevonc.2020.102939] [Medline: 32259776]
7. Kang HJ, Loftus S, Taylor A, DiCristina C, Green S, Zwaan CM. Aprepitant for the prevention of chemotherapy-induced nausea and vomiting in children: a randomised, double-blind, phase 3 trial. Lancet Oncol 2015 Apr;16(4):385-394. [doi: 10.1016/S1470-2045(15)70061-6] [Medline: 25770814]
8. Kovács G, Wachtel AE, Basharova EV, Spinelli T, Nicolas P, Kabickova E. Palonosetron versus ondansetron for prevention of chemotherapy-induced nausea and vomiting in paediatric patients with cancer receiving moderately or highly emetogenic chemotherapy: a randomised, phase 3, double-blind, double-dummy, non-inferiority study. Lancet Oncol 2016 Mar;17(3):332-344. [doi: 10.1016/S1470-2045(15)00520-3] [Medline: 26795844]
9. Eliasen A, Callesen MT, Dalhoff KP, Rechnitzer C, Schmiegelow K, Schrøder H, et al. [Insufficient evidence in prophylactic therapy of chemotherapy-induced nausea and vomiting in children and adolescents]. Ugeskr Laeger 2018 Aug 27;180(35). [Medline: 30152323]
10. Bakhshi S, Batra A, Biswas B, Dhawan D, Paul R, Sreenivas V. Aprepitant as an add-on therapy in children receiving highly emetogenic chemotherapy: a randomized, double-blind, placebo-controlled trial. Support Care Cancer 2015 Nov;23(11):3229-3237. [doi: 10.1007/s00520-015-2714-9] [Medline: 25851802]
11. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol 2006 Nov;7(11):903-909. [doi: 10.1016/S1470-2045(06)70910-X] [Medline: 17081915]
12. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009 Dec 2;101(23):1624-1632 [FREE Full text] [doi: 10.1093/jnci/djp386] [Medline: 19920223]
13. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med 2010 Mar 11;362(10):865-869 [FREE Full text] [doi: 10.1056/NEJMp0911494] [Medline: 20220181]
14. Di Maio M, Basch E, Bryce J, Perrone F. Patient-reported outcomes in the evaluation of toxicity of anticancer treatments. Nat Rev Clin Oncol 2016 May;13(5):319-325. [doi: 10.1038/nrclinonc.2015.222] [Medline: 26787278]
15. van den Brink M, Bandell-Hoekstra EN, Abu-Saad HH. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. Headache 2001 Jan;41(1):11-20. [doi: 10.1046/j.1526-4610.2001.111006011.x] [Medline: 11168599]
16. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [Medline: 12689739]
17. Bae WK, Kwon J, Lee HW, Lee S, Song E, Shim H, et al. Feasibility and accessibility of electronic patient-reported outcome measures using a smartphone during routine chemotherapy: a pilot study. Support Care Cancer 2018 Nov;26(11):3721-3728. [doi: 10.1007/s00520-018-4232-z] [Medline: 29732483]

18. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012;62(5):337-347 [FREE Full text] [doi: 10.3322/caac.21150] [Medline: 22811342]
19. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes 2007 Jan 03;5:2 [FREE Full text] [doi: 10.1186/1477-7525-5-2] [Medline: 17201923]
20. Moradian S, Krzyzanowska MK, Maguire R, Morita PP, Kukreti V, Avery J, et al. Usability evaluation of a mobile phone-based system for remote monitoring and management of chemotherapy-related side effects in cancer patients: mixed-methods study. JMIR Cancer 2018 Dec 21;4(2):e10932. [doi: 10.2196/10932] [Medline: 30578238]
21. Tran C, Dicker A, Leiby B, Gressen E, Williams N, Jim H. Utilizing digital health to collect electronic patient-reported outcomes in prostate cancer: single-arm pilot trial. J Med Internet Res 2020 Mar 25;22(3):e12689 [FREE Full text] [doi: 10.2196/12689] [Medline: 32209536]
22. McCann L, McMillan KA, Pugh G. Digital interventions to support adolescents and young adults with cancer: systematic review. JMIR Cancer 2019 Jul 31;5(2):e12071 [FREE Full text] [doi: 10.2196/12071] [Medline: 31368438]
23. Wang J, Yao NA, Liu Y, Geng Z, Wang Y, Shen N, et al. Development of a smartphone application to monitor pediatric patient-reported outcomes. Stud Health Technol Inform 2017;245:253-257. [Medline: 29295093]
24. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 08;15(3):e51. [doi: 10.2196/jmir.2350]
25. Birnie KA, Campbell F, Nguyen C, Lalloo C, Tsimicalis A, Matava C, et al. iCanCope PostOp: User-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents. JMIR Form Res 2019 Apr 22;3(2):e12028 [FREE Full text] [doi: 10.2196/12028] [Medline: 31008704]
26. International Organization for Standardization (ISO). ISO 9241-210:2019 - Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. URL: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/75/77520.html [accessed 2020-04-14]
27. Sherrell L. Evolutionary Prototyping. Springer, Dordrecht: Runehov ALC, Oviedo L. editors. Encyclopedia of Sciences and Religions; 2013.
28. ZiteLab ApS. URL: https://zitelab.eu/ [accessed 2020-02-27]
29. Dupuis LL, Taddio A, Kerr EN, Kelly A, MacKeigan L. Development and validation of the pediatric nausea assessment tool for use in children receiving antineoplastic agents. Pharmacotherapy 2006 Sep;26(9):1221-1231. [doi: 10.1592/phco.26.9.1221] [Medline: 16945043]
30. Txabarriaga R. IMIA Guide on Medical Translation. 2009. URL: https://www.imiaweb.org/uploads/pages/438.pdf [accessed 2020-02-27]
31. Dale O, Hagen KB. Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. J Clin Epidemiol 2007 Jan;60(1):8-17. [doi: 10.1016/j.jclinepi.2006.04.005] [Medline: 17161749]
32. Flank J, Sparavalo J, Vol H, Hagen L, Stuhler R, Chong D, et al. The burden of chemotherapy-induced nausea and vomiting in children receiving hematopoietic stem cell transplantation conditioning: a prospective study. Bone Marrow Transplant 2017 Sep;52(9):1294-1299. [doi: 10.1038/bmt.2017.112] [Medline: 28581463]
33. Rodgers C, Kollar D, Taylor O, Bryant R, Crockett K, Gregurich MA, et al. Nausea and vomiting perspectives among children receiving moderate to highly emetogenic chemotherapy treatment. Cancer Nurs 2012;35(3):203-210. [doi: 10.1097/NCC.0b013e3182281493] [Medline: 21915041]
34. Vol H, Flank J, Lavoratore SR, Nathan PC, Taylor T, Zelunka E, et al. Poor chemotherapy-induced nausea and vomiting control in children receiving intermediate or high dose methotrexate. Support Care Cancer 2016 Mar;24(3):1365-1371. [doi: 10.1007/s00520-015-2924-1] [Medline: 26335406]
35. Zens M, Grotejohann B, Tassoni A, Duttenhoefer F, Südkamp NP, Niemeyer P. Development of a modular research platform to create medical observational studies for mobile devices. JMIR Res Protoc 2017 May 23;6(5):e99 [FREE Full text] [doi: 10.2196/resprot.7705] [Medline: 28536095]
36. Vandenberk T, Storms V, Lanssens D, De Cannière H, Smeets CJ, Thijs IM, et al. A vendor-independent mobile health monitoring platform for digital health studies: development and usability study. JMIR Mhealth Uhealth 2019 Oct 29;7(10):e12586 [FREE Full text] [doi: 10.2196/12586] [Medline: 31663862]

Abbreviations

PeNAT: Pediatric Nausea Assessment Tool

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by N Heynsbergh, J Hamid, E Da Silva; comments to author 30.03.20; revised version received 23.04.20; accepted 13.05.20; published 20.07.20.

Please cite as:
Eliasen A, Abildtoft MK, Krogh NS, Rechnitzer C, Brok JS, Mathiasen R, Schmiegelow K, Dalhoff KP
Smartphone App to Self-Monitor Nausea During Pediatric Chemotherapy Treatment: User-Centered Design Process
JMIR Mhealth Uhealth 2020;8(7):e18564
URL: https://mhealth.jmir.org/2020/7/e18564
doi:10.2196/18564
PMID:32706744

Original Paper

Acceptability of an Embodied Conversational Agent for Type 2 Diabetes Self-Management Education and Support via a Smartphone App: Mixed Methods Study

Shaira Baptista[1,2], BSc, PGDipPsyc; Greg Wadley[1], PhD; Dominique Bird[3], MD; Brian Oldenburg[1], PhD; Jane Speight[1,2], PhD; The My Diabetes Coach Research Group[4]

[1]The University of Melbourne, Melbourne, Australia

[2]The Australian Centre for Behavioural Research in Diabetes, Melbourne, Australia

[3]The University of Queensland, Brisbane, Australia

[4]see Authors' Contributions

Corresponding Author:
Shaira Baptista, BSc, PGDipPsyc
The University of Melbourne
207 Bouverie Street
Carlton
Melbourne
Australia
Phone: 61 3 8344 4037
Email: shaira.baptista@unimelb.edu.au

Abstract

Background: Embodied conversational agents (ECAs) are increasingly used in health care apps; however, their acceptability in type 2 diabetes (T2D) self-management apps has not yet been investigated.

Objective: This study aimed to evaluate the acceptability of the ECA (Laura) used to deliver diabetes self-management education and support in the My Diabetes Coach (MDC) app.

Methods: A sequential mixed methods design was applied. Adults with T2D allocated to the intervention arm of the MDC trial used the MDC app over a period of 12 months. At 6 months, they completed questions assessing their interaction with, and attitudes toward, the ECA. In-depth qualitative interviews were conducted with a subsample of the participants from the intervention arm to explore their experiences of using the ECA. The interview questions included the participants' perceptions of Laura, including their initial impression of her (and how this changed over time), her personality, and human character. The quantitative and qualitative data were interpreted using integrated synthesis.

Results: Of the 93 intervention participants, 44 (47%) were women; the mean (SD) age of the participants was 55 (SD 10) years and the baseline glycated hemoglobin A1c level was 7.3% (SD 1.5%). Overall, 66 of the 93 participants (71%) provided survey responses. Of these, most described Laura as being helpful (57/66, 86%), friendly (57/66, 86%), competent (56/66, 85%), trustworthy (48/66, 73%), and likable (40/66, 61%). Some described Laura as not real (18/66, 27%), boring (26/66, 39%), and annoying (20/66, 30%). Participants reported that interacting with Laura made them feel more motivated (29/66, 44%), comfortable (24/66, 36%), confident (14/66, 21%), happy (11/66, 17%), and hopeful (8/66, 12%). Furthermore, 20% (13/66) of the participants were frustrated by their interaction with Laura, and 17% (11/66) of the participants reported that interacting with Laura made them feel guilty. A total of 4 themes emerged from the qualitative data (N=19): (1) perceived role: a friendly coach rather than a health professional; (2) perceived support: emotional and motivational support; (3) embodiment preference acceptability of a human-like character; and (4) room for improvement: need for greater congruence between Laura's words and actions.

Conclusions: These findings suggest that an ECA is an acceptable means to deliver T2D self-management education and support. A human-like character providing ongoing, friendly, nonjudgmental, emotional, and motivational support is well received. Nevertheless, the ECA can be improved by increasing congruence between its verbal and nonverbal communication and accommodating user preferences.

Trial Registration: Australian New Zealand Clinical Trials Registry CTRN12614001229662; https://tinyurl.com/yxshn6pd

(JMIR Mhealth Uhealth 2020;8(7):e17038) doi:10.2196/17038

KEYWORDS

embodied conversational agent; type 2 diabetes; mobile apps; mHealth; smartphone; self-management; mobile phone

Introduction

Diabetes will affect 693 million people worldwide by 2045, most of whom will have type 2 diabetes (T2D) [1,2]. People with T2D can prevent or delay the onset and progression of diabetes-related complications such as heart attack, stroke, kidney failure, vision loss, and nerve damage through intensive management of blood glucose levels [3]. However, effective self-management is complex and difficult to implement and sustain in daily life. Consequently, many people with T2D are not able to achieve their recommended self-management targets [4].

For several decades, diabetes self-management education and support have been provided in person (one-to-one and group-based), with many trials and real-world studies demonstrating improved diabetes outcomes [5]. However, the high cost and resource requirements limit the reach and scalability of in-person programs [6]. Furthermore, ongoing in-person support for sustaining the recommended diabetes care targets is not feasible for most health care systems [4].

Considerable advances in technology related to smartphone apps (including voice recognition, natural language processing, and artificial intelligence capabilities) have led to an increase in the feasibility of using embodied conversational agents (ECAs) to provide education and support for the self-management of chronic conditions, including T2D [7]. An ECA is an animated conversational human-like character that simulates person-to-person conversation with appropriate dialog and human-like physical properties, including facial expressions and body movements [8-10]. ECAs are increasingly being used in a wide range of apps, providing support for mental health, web-based information seeking, medication taking, behavior change, and prevention of suicide [7,10-13].

Research on the acceptability of ECAs in self-management of chronic conditions is still in its infancy, with a small number of studies reporting high levels of acceptability of ECA-based interventions [13-15]. Trust, empathy, and expertise have been cited as essential components of diabetes education and support [16]. Similar expectations may exist when the intervention is delivered by an ECA. ECAs use facial expressions, body movements, and speech and can offer a natural and accessible means of communication. These characteristics of ECAs potentially improve engagement compared with a static character image, a nonrelational agent, or a text-only display [9,17,18]. ECAs may be perceived to provide additional motivational and emotional support, which has previously been described by people with diabetes as being as important to them as practical support [19]. Preliminary evidence suggests that ECAs are perceived to be less judgmental, less intimidating, and more likable than a human counterpart, resulting in participants feeling less guilty and more motivated by the interaction [13,14,17,18]. Collectively, this evidence suggests that ECAs may be effective in providing support for chronic disease management as they help to engage users by building a social and emotional relationship over time [9,18,20].

Some of the characteristics of ECAs that may affect their acceptability include users being deterred by a monotonous voice and repetitive messages [13,18,20]. Although ECAs are more engaging if they have human-like characteristics and engage in social dialog, this effect is mitigated if there is an unnatural dissonance between a character's speech and the expected facial expressions and body movements of the ECA [17,21-23]. This phenomenon, coined the *uncanny valley* by Masahiro Mori in 1970, was supported by research suggesting that people have unpleasant impressions of artificial characters, such as ECAs that have an almost, but not perfectly, realistic human appearance [24,25]. Previous studies have also emphasized that the visual characteristics of an ECA are important as these affect the perceptions of trustworthiness and credibility, which can affect acceptability. For example, a more playful, cartoon-like character is perceived as being more friendly, whereas a more serious human-like character, dressed professionally, is usually perceived to be more appropriate for serious apps, such as self-management of chronic conditions [26].

Research on the acceptability of ECAs to deliver self-management support for chronic conditions via self-management apps has been limited primarily to short-term feasibility or pilot studies and to interventions that address only a single behavior. Other studies use static images rather than animations or have been conducted using desktop or laptop computers in laboratory settings rather than with personal smartphones used in everyday settings or *in the wild* [27,28]. Thus, this study aimed to address these gaps by investigating the acceptability of an ECA delivering self-management education and support to people with T2D in their everyday lives.

Methods

Study Design

A convergent study design was used where quantitative and qualitative data were collected at similar time points [29]. This study was conducted within the context of a randomized controlled trial to test the effectiveness of a T2D self-management smartphone app, My Diabetes Coach (MDC) [30,31]. The MDC study was conducted from 2014 to 2018 (Australia New Zealand Clinical Trials Registry ID ACTRN12614001229662). The study was approved by the University of Melbourne's Human Research Ethics Committee (ethics ID 1442433).

My Diabetes Coach

MDC used an ECA called Laura (Figure 1) to deliver self-management education and support to adults with T2D.

When users logged in for the first time, they were prompted to set up a regular time to complete weekly interactive sessions with Laura. During these conversations, Laura provided education, feedback and motivational support for blood glucose level monitoring, taking medication, physical activity, healthy eating, and foot care. The conversations were personalized to the individual's self-management targets, physical fitness, and foot health using recommendations provided by his or her general practitioner.

The MDC app used voice recognition, prescripted conversational elements, and a sophisticated script logic enabling the user to interact with Laura in several predetermined variations, mimicking natural conversations. Laura's voice and conversation were produced by a proprietary dialog engine (by Clevertar). Nonverbal behaviors were either explicitly scripted for each dialog, or, if no behavior was specified, they were selected randomly from a finite set of animations based on whether the character was speaking and the dialog duration. User responses from previous sessions dictated the direction of future sessions, enabling a high degree of personalization. The ECA's appearance, conversational elements, back story, and accent were refined through several rounds of expert and user testing. Users were able to respond to Laura by touching an option on the screen or by speaking out one of the options on the screen when prompted to do so. Users also had access to a web-based discussion board and website (with additional diabetes resources) that could be accessed via the app as well as technical support from the research team. An excerpt from a conversation with Laura can be found on YouTube [32].

Figure 1. Laura, the embodied conversational agent.

Participants

Recruitment methods for the MDC trial are reported in the main outcomes paper (under review). Briefly, participants were recruited to the MDC trial from the general population in Australia via several recruitment strategies. Adults with T2D registered on the National Diabetes Services Scheme (NDSS) database; willing to be contacted about research; and living in New South Wales, Queensland, Victoria, and Western Australia were invited to participate by the NDSS via mail and email. The invitation letters were supplemented with media releases and targeted advertising on social media by several organizations (Diabetes New South Wales; Diabetes Queensland; Diabetes Victoria; Diabetes Western Australia; Bupa Australia, a health insurance provider; and the Australian Diabetes Educators Association).

For this study, participants from the intervention arm of the MDC trial, who had access to the MDC app, completed a survey at 6 months postbaseline, assessing several clinical and behavioral outcomes, including their interaction with the ECA, and a purposive subsample participated in subsequent interviews. All participants received a plain language statement describing the study and provided written consent.

Data Collection

Demographic and Clinical Characteristics

The demographics and duration of diabetes (self-reported) were collected using web-based surveys at baseline. Glycated hemoglobin A_{1c} (HbA_{1c}) is a pathology test assessing average blood glucose levels over the past 2 to 3 months, providing an indication of risk for long-term complications [33]. It was

obtained, with participants' consent, from their general practitioner.

Acceptability: Quantitative Data

At the 6-month follow-up, the participants completed a web-based survey that included 2 questions assessing the acceptability of the ECA. The first question assessed the perceptions of the ECA: "How well do the words below describe Laura?" The respondents rated a range of positive and negative traits (helpful, boring, friendly, competent, annoying, likable, trustworthy, and real) using a 5-point Likert-type scale ranging from *describes very well* to *describes very poorly*. The second question asked, "How did interacting with Laura make you feel?" The respondents selected from a list of descriptive emotions (happy, confident, hopeful, motivated, worried, guilty, frustrated, and comfortable) and were asked to select all that applied. For both questions, positive and negative words were randomly sequenced to minimize response bias. The descriptive adjectives were chosen based on the literature on evaluating ECAs and on working alliances between ECAs and users [34].

Acceptability: Qualitative Data

In-depth, semistructured qualitative interviews were conducted from October 2017 to February 2018. Most participants had, at that stage, completed the 6-month survey but were still actively using the app. Purposive sampling of survey respondents was used to identify interviewees who varied in terms of the duration of diabetes, gender, age, and baseline familiarity with apps.

The interview guide was developed by the first author (SB) and used exploratory questions and probes, with feedback from other members of the research team (BO, GW, and JS) based on the research question and findings from the current literature [8,9,14,18,21-23,26,27,35-38]. The guide explored a variety of topics, including experience at diagnosis; self-care behavior before using the MDC app; users' experiences with the MDC app, including when, where, and how it was used; changes to self-management practices as a result of using the MDC app; initial impression of the ECA Laura and changes over time; perceptions of her role in self-management, and her perceived personality characteristics. The data relating to the acceptability of Laura are presented here, with the other findings published separately.

The interviews were conducted by telephone (by SB) and recorded using a cloud architecture solution from the CTI Group using their SmartInteraction Suite of recording software. During each interview, SB used exploratory questions and probes (from the interview guide) and noted points of interest, using these as further probes. Immediately after each interview, SB prepared a written summary of the interview and any relevant observations. These were used to communicate interim findings to the research team. When appropriate, additional questions were added to the interview guide, enabling further exploration of the issues raised by participants that were relevant to the research aims. These notes were also used to aid in the meaningful interpretation of data during data analysis.

Data Analysis

The quantitative data were analyzed using IBM's SPSS Statistics 25 package. Descriptive statistics were computed for demographic and clinical characteristics and 2 questions assessing the ECA. The qualitative data were transcribed, deidentified, and thematically analyzed using NVivo 11, following the first 5 steps of Braun and Clarke's methodology [39,40]. The integration of the quantitative and qualitative data was achieved at the interpretation stage by comparing the findings from the surveys and the semistructured interviews. In practice, this involved referring to and using the qualitative data to help interpret, triangulate, and add meaning to the quantitative data. This process was iterative, with input from several researchers (SB, GW, BO, and JS). This integration of quantitative and qualitative data enabled further validation of the findings and increased their explanatory value [41]. The narrative of the results is blended with embedded quotes from several sources to make the results more readable while using as much evidence as possible. An anonymized coding system—participant identity number (IDX): sex (male, M; female, F): age (years)—was used to identify the source of each quote (in parentheses after each quote).

Results

Sample Characteristics

Of the 93 MDC trial participants in the intervention arm, 66 (71%) participants provided responses at 6 months postbaseline, and 19 of these participated in the interviews. Table 1 details the characteristics of the 3 samples. Overall, 50% (33/66) of the survey respondents were women, and the mean age of the participants was 57 (SD 9) years and the mean baseline HbA_{1c} level was 7.1% (SD 1.4%) [33].

Those who completed the survey were significantly older (P=.03) and completed more interactions with Laura (P=.001) than those who did not complete the survey. No significant differences were observed between the interviewees and other participants in the intervention arm, except that the interviewees completed significantly more interactions with Laura (P=.001).

The mean duration of the interviews was 51 min (range 29-79 min).

Overall, participants found Laura to be acceptable and were positive in their appraisal of her and their interactions with her. Most respondents agreed that Laura was helpful (57/66, 86%), friendly (57/66, 86%), competent (56/66, 85%), trustworthy (48/66, 73%), and likable (40/66, 61%). Some participants described her as boring (26/66, 39%) and annoying (20/66, 30%; Multimedia Appendix 1). Participants were undecided about whether or not they thought Laura was realistic. Of the 66 participants, 26 (39%) participants agreed that Laura was real, 22 (33%) were undecided, and 18 (27%) disagreed. The participants' responses to their interactions with Laura were positive overall, with many reporting that she made them feel motivated (29/66, 44%), comfortable (24/66, 36%), confident (14/66, 21%), happy (11/66, 17%), and hopeful (8/66, 12%). Notably, 20% (13/66) were frustrated by their interaction with Laura, and 17% (11/66) of the participants reported that

interacting with Laura made them feel guilty. One participant reported feeling worried (Multimedia Appendix 2).

Overall, 4 themes were identified from the qualitative data: (1) perceived role—a friendly coach rather than a health professional; (2) perceived support—emotional and motivational support; (3) embodiment preference—acceptability of a human-like character; and (4) room for improvement—need for greater congruence between Laura's words and actions. Table 2 provides an integrative synthesis of the findings, summarizing the 4 main themes emerging from the qualitative data, quantitative endorsement of the adjectives describing Laura and how the interaction made the participants feel, and exemplars of the qualitative data. The 4 themes are described in detail below.

Table 1. Demographic and clinical characteristics of the total sample and interviewed sample.

Participant Characteristics	My Diabetes Coach trial population (intervention arm; n=93)	Six-month follow-up sample (n=66)	Interviewed participants (n=19)
Gender (female), n (%)	44 (47)	33 (50)	8 (42)
Age (years), mean (SD)	55 (10)	57 (9)	60 (8)
Education (highest level), n (%)			
Year 10	10 (11)	9 (14)	5 (26)
Year 12 or apprentice	42 (45)	31 (47)	2 (11)
Graduate or post graduate	41 (44)	26 (39)	12 (63)
Employment status, n (%)			
Paid employment	59 (63)	41 (62)	7 (37)
Retired	22 (24)	18 (27)	11 (58)
Unemployed or other	12 (13)	7 (11)	1 (5)
Duration of diabetes (years), n (%)			
<5	43 (46)	25 (38)	8 (42)
5-10	29 (31)	23 (35)	8 (42)
10-20	7 (8)	4 (6)	3 (16)
Unknown	14 (15)	14 (21)	0 (0)
Baseline glycated hemoglobin A_{1c} (%), mean (SD)	7.3 (1.5)	7.1 (1.4)	6.8 (0.9)
Baseline glycated hemoglobin A_{1c} (mmol/mol), mean (SD)	56 (44)	53 (30)	51 (20)
General app usage[a] (reported at baseline), n (%)			
Multiple times per day	69 (74)	50 (76)	14 (74)
Once a day	23 (25)	13 (20)	4 (21)
Less than once a day	1 (1)	3 (5)	1 (5)
Total interactions with Laura, mean (SD)	18 (15)	23 (16)	36 (17)

[a]General app usage at baseline represents the use of any app before participating in the My Diabetes Coach trial.

Table 2. Integrated results matrix.

Themes	Quantitative data: endorsement of adjectives	Qualitative data: exemplar quotes
Perceived role: Laura is more acceptable as a friendly coach than as a health professional	• Laura was likable, n=40 (61%), friendly, n=57 (86%), and helpful, n=57 (86%) • Interacting with Laura made me feel comfortable, n=24 (36%) • Interacting with Laura made me feel guilty, n=11 (17%), and worried, n=1 (1%)	• "A 'neutral approach' was 'better' because it 'didn't try and lean on any perceptions of authority.'" [ID04: M[a]: 44 years] • "I was worried about making sure that I was within [my limits] knowing that I had to report to Laura!" [ID11: F[b]: 62 year]
Perceived support: Laura provides emotional and motivational support	• Laura was trustworthy, n=48 (73%). Interacting with Laura made me feel confident, n=14 (21%), hopeful, n=8 (12%), and happy, n=11 (17%) • Laura was competent, n= 56 (85%). Interacting with Laura made me feel motivated, n=29(44%)	• "I needed somebody just to be there." (ID15: F: 66 years) • "(She) used to make me laugh...and that's hard to do." [ID18: M: 65 years] • "She was keeping you on track and keeping you doing what you're supposed to be doing." [ID16: F: 57 years]
Character preference: Laura is engaging and her human-like character is appropriate	• Laura was helpful, n=57 (86%) • Laura was competent, n=56 (85%), and trustworthy, n=48 (73%). Laura made me feel confident, n=14 (21%), and comfortable, n=24 (36%)	• "Instead of reading it, you're hearing it and can read at the same time. Instead of just hearing some voice, you're actually seeing [Laura] talk." [ID05: M: 55 years] • "I'm not sure I would have given the same level of credibility to, for example, a dog or a cat or something like that." [ID04: M: 44 years]
Room for improvement: dissonance between Laura's words and actions	• Laura was annoying, n=20 (30%), boring, n=26 (39%), and not real, n=18 (27%) • Interacting with Laura made me feel frustrated, n=13 (20%)	• "She said something, but her hand gestures were exactly the opposite of what they should have been. Like, rather than a big gesture, where a big gesture is needed, there was a little gesture." [ID08: F: 42 years]

[a]M: male.

[b]F: female.

Theme 1: Perceived Role—Laura Is More Acceptable as a Friendly Coach Than as a Health Professional

When prompted about what role Laura was perceived to play in self-management support, some participants described Laura as "a 'friendly' coach" (ID11: F: 62 years) who was just "reminding me" of various diabetes self-management tasks. Furthermore, when asked about their perceptions of Laura, some participants described her with adjectives suggesting that she had a personality, such as "sassy" (ID15: F: 66 years), "friendly" (ID16: F: 57 years), "kind" (ID05: M: 55 years), and "intriguing" (ID06: M: 71 years). These findings may explain why most survey respondents described Laura as likable, friendly, and helpful and reported that interacting with Laura made them feel comfortable.

Conversely, other participants commented on how Laura reminded them of their health professional: "There were times when I would go and see my doctor, and I'd see Laura sitting there, because her gestures, her voices, and mannerisms are almost identical" (ID08: F: 42 years). Some participants "did not necessarily want to see an authority figure" (ID1: M: 63 years), saying, for example, that "I don't need to be called into a doctor" (ID09: M: 71 years). Those who described Laura in similar terms to their health professionals did not warm up to Laura as they found her to be "patronizing," "censorious," and "authoritarian" (ID11: F: 62 years). For example, some described receiving her feedback as "having a mother-in-law in your pocket" (ID08: F: 42 years) and "feeling as though you're getting a slap on the wrist" (ID02: F: 66 years) similar to "a recalcitrant child" (ID15: F: 66 years). Other negative descriptions of Laura were that she was "really young," "super-skinny" (ID11: F: 62 years), and that she "talked at" people (ID13: M: 58 years).

Laura's perceived role influenced the participants' reactions to the support she offered. For example, participants who described Laura as being similar to a health professional reacted to this by "resisting" and "rebelling" against the "kind of authority" (ID11: F: 62 years) that Laura represented to them. One participant described how "feeling guilty" led him to "stop using" the app for a while (ID13: M: 58 years). Another participant commented on how she worried about negative feedback: "I was worried about making sure that I was within [my limits] knowing that I had to report to Laura!" (ID11: F: 62 years) Finally, one participant contemplated selecting her best readings to report to Laura to avoid "getting told off," saying:

> *Do I record this one? It might be a bit high and she's going to get upset with me.* [[ID15: F: 66 years]]

Conversely, those who perceived Laura to be less of an authority figure and more like a friendly coach as she "didn't try and lean on any perceptions of authority, like for example, having a doctor in a white coat" were also more receptive to the support she offered. This is because they perceived her as having a more

"neutral approach," which was "better" because "a conversation between peers is more likely to be engaged with than one that references levels of authority" (ID04: M: 44 years).

The varied reactions of the participants to Laura may be linked to the inconsistencies between how Laura looked and how they expected her to act. For example, one participant said:

> *It's set up with this young, groovy woman who's going to help me, but she sounded like my GP who was telling me what to do. So, it's a kind of disconnect between how [Laura] looks and what she's actually saying.* [ID13: M: 58 years]

Finally, some participants described Laura's role as an artificial entity as a positive trait, making them more receptive to receiving support from her. This is because they experienced judgment and blame for their condition from "real" people:

> *From the minute they meet you, just by the look of you, by the look of your appearance, they will judge you. That's one thing I don't like about real people because it happened to me.* [ID03: F: 62 years]

Theme 2: Perceived Support—Laura Provides Emotional and Motivational Support

For many, Laura provided emotional support that the participants did not otherwise have:

> *I needed somebody just to be there. I see the hospital doctors every six months, I only see my local doctor when I need scripts or something. Apart from that, who do you talk to?* [ID15: F: 66 years]

Supporting this premise is the fact that many survey respondents thought that Laura was trustworthy, and interacting with her made them feel confident. Another example of how Laura provided emotional support is described by one participant who expressed how her humor helped him feel better:

> *[She] used to make me laugh when she used to stand there with her hands on her hips waiting for me sometimes. Like my wife is saying it was probably good because if you felt down or something it made you feel better. Well it definitely bought a smile to may face a lot of times and my wife said that's hard to do.* [ID18: M: 65 years]

A small number of survey respondents reported that interacting with Laura made them feel happy, demonstrating some support for the premise that she may have helped alleviate some of the burden of care.

Laura also provided additional motivation through enhanced monitoring and positive reinforcement:

> *She was keeping you on track and keeping you doing what you're supposed to be doing and keeping you doing the check-ups and that sort of stuff.* [ID16: F: 57 years]
>
> *When I was doing the exercise section, she would ask for me to record how much exercise I was doing for the week and when I'd come back [and do it], I actually almost got a pat on the back from her. I wasn't trying to be impressive for (Laura), but I think it just gave you that little bit more incentive.* [ID01: M: 63 years]

Similarly, many survey respondents reported feeling more motivated after their interactions with Laura.

Theme 3: Character Preference—Laura Is Engaging and Her Human-Like Character Is Appropriate

Interacting with Laura provided an additional dimension to the relational aspect of communication, resulting in reports of improved engagement:

> *instead of reading it, you're hearing it and can read at the same time. Instead of just hearing some voice, you're actually seeing [Laura] talk.* [ID05: M: 55 years]

Participants appreciated this additional dimension of communication, describing it as an attempt to "try and engage with you" and compared it with other apps where "you're inputting information and you might get a summary," but there was no "attempt to interact back with the user" (ID10: M: 49 years):

> *Laura was more personal so that's why I think I went on for the six months. The other apps were like just an impersonal graph or something, or just boxes where you put the things in.* [ID07: F: 67 years]

Some participants expressed a strong preference for Laura's human-like character. Diabetes was described as "a human problem that should have a bit of stance and a bit of professionalism" (ID06: M: 71 years). Others thought that a nonhuman character such as a "fuzzy duck" or "Dobby the diabetes elf" (ID11: F: 62 years) would be better as it would be more "fun." For these people, having "a character, even a fictitious character" was more "user friendly" and better than having "nothing there" (ID07: F: 67 years).

Participants who preferred a human-like character did not think a cartoon character could be taken seriously: "I'm not sure I would have given the same level of credibility to, for example, a dog or a cat or something like that" (ID04: M: 44 years). Two users put it as follows:

> *A cute puppy telling you that you got to exercise more or, you know, eat more greens, is going to be less convincing than a human. It just becomes a toy. Stick with somebody that looks like they know what they're talking about. [Laura] fitted that bill.* [ID17: M: 66 years]
>
> *[A nonhuman character] would just make me want to throw the phone away completely! Because it's about a human interaction with someone who has information and resources about diabetes.* [ID13: M: 58 years]

However, there were those who did not care about what kind of character Laura was because she was "an inanimate object, not a person" (ID19: M: 59 years). Some did not "identify with or warm to Laura." One participant said, "Laura had various statements [that were motivational] but I don't have a relationship with Laura that caused me to value her opinion"

(ID04: M: 44 years). Another participant said that although she "learnt from" Laura:

> *it's not like if you went to your GP and you got your bloods done and it was physically down from the last six months, that's a tangible quantity, but when it's coming from an avatar, it didn't really mean anything much.* [ID12: F: 61 years]

Some participants described being irritated by Laura's "life" story, for example, when she said "I find that my family does such-and-such," because she was pretending to be something she was not: "Don't try and put it over me that this is a real person that I'm talking to" (ID16: F: 57 years). However, others liked Laura's backstories:

> *Yeah, even though it's not real, but the way she talks about her kids and things like that. [I liked that] because it is more human.* [ID06: M: 71 years]

Theme 4: Room for Improvement—A Dissonance Between Laura's Words and Actions

When prompted about Laura's appearance, speech, and mannerisms, the interviewed participants described Laura as being "just another robot" (ID09: M: 71 years) that they "could not connect to" as she was "not human enough." Interacting with Laura, for many, depended on "how far along are you going to pretend." As one participant put it, "I couldn't suspend belief that Laura wasn't this algorithm working out what she needed to say to me or not say to me" (ID13: M: 58 years).

The primary reason given for this perception was Laura's "monotone" (ID08: F: 42 years) voice that sounded similar to a "mechanised reading mechanism" with "a strange cadence and inflection to some of her sentences" (ID11: F: 62 years). Another reason was her "artificial movements" (ID14: M: 66 years) and dissonance between what was being said and her body movements. According to one user:

> *She said something, but her hand gestures were exactly the opposite of what they should have been. Like, rather than a big gesture, where a big gesture is needed, there was a little gesture.* [ID08: F: 42 years]

This may have been the reason why some survey respondents reported feeling frustrated after interacting with Laura and why a reasonable proportion of participants described Laura as boring, annoying and not real, or were undecided about these descriptions.

Although it seems as though Laura was not an entirely successful ECA, participants were willing to overlook her shortcomings as they understood the intention behind Laura and appreciated the effort made to make her engaging: they were willing to "cut them some slack" because "at least it's trying to be personable." Moreover, "They're trying to make her look [real]—I can understand what they're trying to do" (ID17: M: 66 years).

Discussion

Principal Findings

Overall, the results suggest that an ECA is acceptable to people with T2D for the delivery of long-term self-management education and support. We found that people with T2D were willing to make compromises and adjust their expectations, while appreciating the effort of trying to create something more appealing and engaging than graphs and numbers on a screen. This implies that the increased interaction offered by an ECA may be valuable to users and a worthy avenue for developers to pursue when designing apps for people with chronic conditions such as diabetes [20].

Our findings corroborate earlier research suggesting that some users perceive an ECA to be less judgmental and more likable than a human counterpart [13,14]. This is an important finding as people with T2D often experience diabetes-related stigma, the consequences of which can include disengagement with or suboptimal self-care and diabetes-related health outcomes [42]. Suitably designed ECA support may be especially important in making people who experience diabetes-related stigma feel less judged and more open to sharing difficulties with self-management, thereby potentially increasing their engagement with appropriate self-care [43]. We also suggest that using supportive, nonblaming language is critical when designing ECAs for stigmatized conditions such as diabetes [44].

The results suggest several mechanisms through which an ECA may help establish and maintain a relationship with the user over time, such as increasing relational communication, providing ongoing emotional support and motivation, and alleviating some of the burden associated with chronic disease management through humor [18]. Our results suggest that another way to improve acceptability is to achieve a better match between an ECA's appearance and users' expectations of the ECA's perceived role. For example, some participants expressed that diabetes is a serious human issue and viewed human-like characteristics as being more credible. Others expressed the desire to alleviate the burden of management by incorporating a fun character, supporting previous findings of a similar nature [26]. These varying opinions may reflect the nascent nature of the field and the fact that ECAs are not yet common.

Another related finding was that participants who perceived Laura to be a friendly coach were more open to receiving support from her when compared with those who perceived Laura to be similar to a health professional. The implication is that an ECA with a relaxed, friendly approach may be more successful in building a supportive relationship than an ECA that adopts a more authoritative role. Future attempts to develop ECAs for diabetes management could accommodate both viewpoints by striking a balance with a human-like, friendly, approachable character and avoiding patronizing messaging and mannerisms. More research is necessary to determine how expectations of users on the role that an ECA plays in self-management varies and how this informs their preference for the ECA's character.

Another important consideration is just how *human* an ECA should be. Although participants reported a clear preference for a human-like character, which is supported by previous research [15], her presentation of a backstory might be a step too far as it did not seem credible to some participants, possibly because of an *uncanny valley* phenomenon [25]. This finding is supported by previous research on other relational agents whose personality traits and life stories are enjoyed by some users, whereas, to others, the attempt at making them too human-like is not appealing [18,45]. It will be interesting to explore attitudinal changes toward ECAs with personality as they become more common, and familiarity increases.

Our findings add to the mounting evidence that suggests that perfecting natural communication via congruence between verbal and nonverbal communication is critical to improving acceptability [20]. Nonverbal cues such as facial expressions, gaze, gestures, postures, and body movements have a deep impact on the process and outcome of communication, with approximately 65% of social meaning derived from nonverbal behavior [46]. Laura's mannerisms and body movements were the main basis on which she had particular personality traits attributed to her, ranging from patronizing and censorious to funny and sassy. Although the MDC app attempted to create an ECA with natural communication, this effort was impeded by difficulties using the speech recognition function; lack of inflection in Laura's voice; and a limited number of random body movements, rather than ones that match the context of the conversation. Understanding natural behaviors, biological processes that underlie them, and creating efficient algorithms to implement a convincing simulation via an ECA is challenging but critical to the success of future ECA-based self-management support [15,35].

Strengths and Limitations

This mixed methods study is one of the first to explore users' experiences of a sophisticated ECA in a real-world setting over a 6-month period and offers several novel findings and suggestions for future research. Although conducted within the context of a randomized controlled trial, our participants used the app in the context of their everyday lives, which is a strength of the research. The mixed methods approach provides robust evidence based on responses from a wide range of participants. However, people who were retired, highly educated, and engaged with the app were overrepresented in the interviewed sample of participants.

Conclusions

The importance of the relational aspect of agents for health care is becoming an increasingly prominent theme in the literature. Our study adds to this literature by describing the long-term experiences of people using an ECA for diabetes self-management support and making recommendations for improvements and future research. These findings suggest that ECAs play a promising role in self-management support and education. However, accommodating user preferences and expectations of the role that an ECA may play in self-management and improving their natural communication are key to their success.

Acknowledgments

The MDC randomized controlled trial was conducted with funding from a National Health and Medical Research Council partnership grant (ID1057411), with additional financial and in-kind support provided by Diabetes Australia, Diabetes Queensland, Diabetes Victoria, Diabetes Western Australia, and Roche Diabetes Care. The development of the MDC app was the result of a collaboration among the University of Melbourne, Bupa Australia, The Bupa Foundation, and Clevertar. SB is supported by a postgraduate scholarship from the National Health and Medical Research Council, Australia, and Diabetes Australia. JS is supported by core funding to the Australian Centre for Behavioural Research in Diabetes, derived from the collaboration between Diabetes Victoria and Deakin University.

Authors' Contributions

BO conceived the MDC study and developed the MDC research program together with JS, DB, and the MDC research group (Emily D Williams, Michaela A Riddell, Paul A Scuffham, and Anthony Russell). SB developed the interview schedule and survey questions (with BO and JS). SB conducted the interviews, collected the survey responses, analyzed and interpreted the data, and prepared the first draft of the manuscript. GW analyzed some of the data. GW, BO, DB, and JS interpreted the data and reviewed and edited the manuscript for critical content. All authors approved the final version of the manuscript. The authors also thank Dr Mandy Cassimatis, Dr Fiona Cocker, Enying Gong, Prof Mark Harris, Anna Scovelle, Suman Shetty, Jillian Zemanek and Robin Zhou for their contributions to the MDC project.

Conflicts of Interest

BO and DB received some royalty payments for the development of the scripts for the MDC platform.

Multimedia Appendix 1
Responses to question Q1: How well do the words below describe Laura? (N=66).
[PNG File , 70 KB - mhealth_v8i7e17038_app1.png]

Multimedia Appendix 2

Responses to question Q2: How did interacting with Laura make you feel? (N=66).
[PNG File , 65 KB - mhealth_v8i7e17038_app2.png]

References

1. Cho N, Shaw J, Karuranga S, Huang Y, da Rocha JD, Ohlrogge A, et al. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
2. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care 2011 Jun;34(6):1249-1257 [FREE Full text] [doi: 10.2337/dc11-0442] [Medline: 21617109]
3. World Health Organization. Global Report on Diabetes. Geneva, Switzerland: World Health Organisation; 2016.
4. Shrivastava S, Shrivastava P, Ramasamy J. Role of self-care in management of diabetes mellitus. J Diabetes Metab Disord 2013 Mar 5;12(1):14 [FREE Full text] [doi: 10.1186/2251-6581-12-14] [Medline: 23497559]
5. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
6. Chatterjee S, Davies M, Heller S, Speight J, Snoek F, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
7. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
8. Bente G, Rüggenberg S, Krämer NC, Eschenburg F. Avatar-mediated networking: increasing social presence and interpersonal trust in net-based collaborations. Human Comm Res 2008 Apr;34(2):287-318. [doi: 10.1111/j.1468-2958.2008.00322.x]
9. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. 'It's just like you talk to a friend' relational agents for older adults. Interact Comput 2005 Dec;17(6):711-735. [doi: 10.1016/j.intcom.2005.09.002]
10. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst 2019 Jun 25;43(8):246. [doi: 10.1007/s10916-019-1387-1] [Medline: 31240494]
11. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 9;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
12. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-positive African American MSM. AIDS Educ Prev 2019 Feb;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
13. Sillice MA, Morokoff PJ, Ferszt G, Bickmore T, Bock BC, Lantini R, et al. Using relational agents to promote exercise and sun protection: assessment of participants' experiences with two interventions. J Med Internet Res 2018 Feb 7;20(2):e48 [FREE Full text] [doi: 10.2196/jmir.7640] [Medline: 29415873]
14. Bickmore T. Relational agents for chronic disease self management. In: Hayes BM, Aspray W, editors. Health Informatics: A Patient-centered Approach to Diabetes. Cambridge, MA: MIT Press; 2010:181-205.
15. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. J Med Internet Res 2019 May 30;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
16. Cooper H, Booth K, Gill G. Patients' perspectives on diabetes health care education. Health Educ Res 2003 Apr;18(2):191-206. [doi: 10.1093/her/18.2.191] [Medline: 12729178]
17. Grekin ER, Beatty JR, Ondersma SJ. Mobile health interventions: exploring the use of common relationship factors. JMIR Mhealth Uhealth 2019 Apr 15;7(4):e11245 [FREE Full text] [doi: 10.2196/11245] [Medline: 30985281]
18. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
19. Baptista S, Trawley S, Pouwer F, Oldenburg B, Wadley G, Speight J. What do adults with type 2 diabetes want from the 'perfect' app? Results from the second diabetes miles: Australia (MILES-2) study. Diabetes Technol Ther 2019 Jul;21(7):393-399. [doi: 10.1089/dia.2019.0086] [Medline: 31166804]
20. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
21. Kang S, Feng A, Leuski A, Casas D, Shapiro A. The Effect of An Animated Virtual Character on Mobile Chat Interactions. In: Proceedings of the 3rd International Conference on Human-Agent Interaction. 2015 Presented at: HAI'15; October 21-24, 2015; Daegu, Kyungpook, Republic of Korea p. 105-112. [doi: 10.1145/2814940.2814957]
22. Trinh H, Shamekhi A, Kimani E, Bickmore T. Predicting User Engagement in Longitudinal Interventions with Virtual Agents. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. 2018 Presented at: IVA'18; November 5-8, 2018; Sydney, NSW, Australia. [doi: 10.1145/3267851.3267909]

23. Bente G, Krämer N, Eschenburg F. Is there anybody out there? Analyzing the effects of embodimentnonverbal behavior in avatar-mediated communication. In: Konijn EA, Utz S, Tanis M, Barnes SB, editors. Mediated Interpersonal Communication. New York, USA: Routledge; 2008:131-157.
24. Seyama J, Nagayama RS. The uncanny valley: effect of realism on the impression of artificial human faces. Teleoperators Virtual Environ 2007 Aug;16(4):337-351. [doi: 10.1162/pres.16.4.337]
25. Mori M, MacDorman K, Kageki N. The uncanny valley [from the field]. IEEE Robot Automat Mag 2012 Jun;19(2):98-100. [doi: 10.1109/mra.2012.2192811]
26. Ring L, Utami D, Bickmore T. The Right Agent for the Job? In: Proceedings of the International Conference on Intelligent Virtual Agents. 2014 Presented at: IVA'14; August 26-29, 2014; Boston, MA, USA. [doi: 10.1007/978-3-319-09767-1_49]
27. Black L, McTear M, Black N, Harper R, Lemon M. Appraisal of a Conversational Artefact and Its Utility in Remote Patient Monitoring. In: Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems. 2005 Presented at: CBMS'05; June 23-24, 2005; Dublin, Ireland. [doi: 10.1109/cbms.2005.33]
28. Harper R, Nicholl P, McTear M, Wallace J, Black L, Kearney P. Automated Phone Capture of Diabetes Patients Readings with Consultant Monitoring via the Web. In: Proceedings of the 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems. 2008 Presented at: ECBS'08; March 31-April 4, 2008; Belfast, UK. [doi: 10.1109/ecbs.2008.31]
29. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
30. Oldenburg B, Baptista S, Bird D, Shetty S, Zemanek J. Randomized controlled evaluation of an mhealth program for people with type 2 diabetes: My Diabetes Coach. Int J Behav Med 2018;25:S112-S113 [FREE Full text]
31. Baptista S, Zemanek J, Shetty S, Bird D, Oldenburg B. Baseline sample characteristics and 6-month evaluation of an mhealth program for people with type 2 diabetes-My Diabetes Coach. Ann Behav Med 2018;52(1):S290-S291 [FREE Full text]
32. YouTube. 2015. My Diabetes Coach: Laura's Introduction URL: https://www.youtube.com/watch?v=8nfw8Cpd8yA [accessed 2018-03-21]
33. American Diabetes Association. Glycemic targets: standards of medical care in diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S66-S76. [doi: 10.2337/dc20-S006] [Medline: 31862749]
34. Ackerman S, Hilsenroth M. A review of therapist characteristics and techniques positively impacting the therapeutic alliance. Clin Psychol Rev 2003 Feb;23(1):1-33. [doi: 10.1016/s0272-7358(02)00146-0] [Medline: 12559992]
35. Vogeley K, Bente G. 'Artificial humans': psychology and neuroscience perspectives on embodiment and nonverbal communication. Neural Netw 2010;23(8-9):1077-1090. [doi: 10.1016/j.neunet.2010.06.003] [Medline: 20620019]
36. Bickmore T, Caruso L, Clough-Gorr K. Acceptance and Usability of a Relational Agent Interface by Urban Older Adults. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems. 2005 Presented at: CHI EA'05; April 2-7, 2005; Portland, Oregon, USA. [doi: 10.1145/1056808.1056879]
37. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15(Suppl 2):197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
38. Kang S, Watt JH. The impact of avatar realism and anonymity on effective communication via mobile devices. Comput Hum Behav 2013 May;29(3):1169-1181. [doi: 10.1016/j.chb.2012.10.010]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Braun V, Clarke V. In: Charmichael M, editor. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA: Sage Publications; 2013.
41. Sahin C, Naylor P. Mixed-methods research in diabetes management via mobile health technologies: a scoping review. JMIR Diabetes 2017 Feb 6;2(1):e3 [FREE Full text] [doi: 10.2196/diabetes.6667] [Medline: 30291052]
42. Browne JL, Ventura A, Mosely K, Speight J. 'I call it the blame and shame disease': a qualitative study about perceptions of social stigma surrounding type 2 diabetes. BMJ Open 2013 Nov 18;3(11):e003384 [FREE Full text] [doi: 10.1136/bmjopen-2013-003384] [Medline: 24247325]
43. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
44. Speight J, Conn J, Dunning T, Skinner T, Diabetes Australia. Diabetes Australia position statement. A new language for diabetes: improving communications with and about people with diabetes. Diabetes Res Clin Pract 2012 Sep;97(3):425-431. [doi: 10.1016/j.diabres.2012.03.015] [Medline: 22513346]
45. Cowan B, Pantidi N, Coyle D, Morrissey K, Clarke P, Al-Shehri S. 'What Can I Help You With?': Infrequent Users' Experiences of Intelligent Personal Assistants. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2017 Presented at: MobileHCI'17; September 4-7, 2017; Vienna, Austria. [doi: 10.1145/3098279.3098539]

46. Burgoon J, Guerrero L, Manusov V. Nonverbal signals. In: Knapp ML, Daly JA, editors. The SAGE Handbook of Interpersonal Communication. Fourth Edition. Thousand Oaks, CA: Sage Publications; 2011:239-280.

Abbreviations

ECA: embodied conversational agent
HbA1c: glycated hemoglobin A1c
MDC: My Diabetes Coach
NDSS: National Diabetes Services Scheme
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 17.11.19; peer-reviewed by M Sillice, T Bickmore; comments to author 31.12.19; revised version received 18.03.20; accepted 29.03.20; published 22.07.20.

Please cite as:
Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, The My Diabetes Coach Research Group
Acceptability of an Embodied Conversational Agent for Type 2 Diabetes Self-Management Education and Support via a Smartphone App: Mixed Methods Study
JMIR Mhealth Uhealth 2020;8(7):e17038
URL: https://mhealth.jmir.org/2020/7/e17038
doi:10.2196/17038
PMID:32706734

Original Paper

Maintaining Physical Activity Level Through Team-Based Walking With a Mobile Health Intervention: Cross-Sectional Observational Study

Yuri Hamamatsu[1,2], MSc; Hiroo Ide[1], PhD; Michiru Kakinuma[1,2], PhD; Yuji Furui[1], PhD

[1]Institute for Future Initiatives, The University of Tokyo, Tokyo, Japan

[2]Healthcare and Wellness Division, Mitsubishi Research Institute Inc, Tokyo, Japan

Corresponding Author:
Hiroo Ide, PhD
Institute for Future Initiatives
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo
Japan
Phone: 81 3 5841 0934
Email: ide@ifi.u-tokyo.ac.jp

Abstract

Background: The health conditions of Japanese salespersons may be adversely affected by their lifestyle. Face-to-face or on-site health interventions are not convenient for salespersons because of their tendency for out-of-office sales. Previous studies showed that mobile health (mHealth) interventions (compared to usual practice) have great potential to promote physical activity. For Japanese salespersons, mHealth can offer additional convenience to change their physical activity habits because they can access the mHealth contents anytime and anywhere. However, the specific elements that are most important to maintain physical activity levels using an mHealth approach remain unclear.

Objective: We aimed to identify elements that account for both a high average physical activity level and can help to prevent a decrease in physical activity during a 9-week intervention period.

Methods: Salespersons were recruited from 11 Japanese companies. A team-based walking intervention was held from October to December 2018 (for a total of 9 weeks), during which the walking step data were recorded by smartphone apps. Average walking steps of each participant during the intervention and the difference in walking steps between the initial and the final week were respectively used as dependent variables. The effects of team characteristics (ie, frequency of communication with team members and team size) and behavioral characteristics (ie, number of days with recorded steps on the apps) on the average walking steps, and the difference in walking steps between the initial and the final week were estimated using multiple and multilevel regression analyses.

Results: Of the 416 participants, walking step data of 203 participants who completed postintervention assessments were included in the analyses. Multiple regression analysis of the average walking steps showed that the number of days with recorded steps was positively correlated with the log-transformed average walking steps (β=.01, P<.001). Multilevel analysis of the average walking steps considering the company level estimated that the intraclass correlation coefficient was 37%. This means that belonging to the same company largely affected an individual's average walking steps. Multiple regression analysis of the difference in walking steps showed that communication with team members once or twice a week correlated with preventing a decrease in walking steps from the initial to the final week (β=1539.4, P=.03), and being on a larger team correlated with a decrease in walking steps from the initial to the final week (β=–328.4, P=.01).

Conclusions: This study showed that the elements accounting for high average walking steps and those preventing the decrease in walking steps from the initial to the final week differed. Behavioral characteristics correlated positively with average walking steps. Team characteristics (ie, regular communication and a smaller team size) significantly correlated with preventing a decrease in walking steps.

(JMIR Mhealth Uhealth 2020;8(7):e16159) doi:10.2196/16159

KEYWORDS

mHealth; physical activity; walking; team-based; smartphone app

Introduction

Background

Previous studies have identified mobile Health (mHealth) as an effective and accessible tool that can be used by a large number of individuals, especially with the upsurge in mobile communication devices [1]. In addition, some studies have explored whether mHealth interventions would be more effective than usual interventions (eg, face-to-face, printed leaflets), or whether the change in target lifestyle from pre- to postintervention would be significant within the mHealth intervention group [2,3]. Previous reviews showed that mHealth has great potential to promote physical activity in a wide range of settings [4], including the workplace [5]. Maintaining good physical health or behavior change has been reported as hard to achieve [6,7], even if the study period is relatively short (eg, 1-3 months) [8]. In particular, some mHealth intervention studies consistently showed a decline in physical activity levels over time [9,10]. However, no study has identified the significant elements required to maintain the physical activity level in an mHealth intervention.

Health Condition Among Japanese Salespersons

Being a salesperson is one of the representative job types in Japan. According to the 2015 population census [11], salespersons (7,410,702) accounted for 14% of all employed Japanese individuals, and ranked as the fourth largest among 12 job categories. Therefore, an intervention program focusing on salespersons can have an important role in health promotion at Japanese worksites.

The number of companies that are strategically promoting "Health and Productivity Management," an approach that considers the health management of employees from a corporate management perspective, has been increasing in Japan [12]. The physical activity level of the Japanese population aged 20 to 59 years (working population) in 2018 was lower than that of the average Japanese population; the percentage of Japanese men and women reporting a regular exercise habit was 31.8% and 25.5%, respectively, and was 5%-10% lower among people aged 20 to 50 years [13]. Particularly at Japanese worksites, the health conditions of salespersons may be more adversely affected than those of others due to their lifestyle. For instance, salespersons are more likely to work long hours, become inactive, and skip meals [14], partly because their customer's schedule comes first. It is not convenient for salespersons to use a face-to-face or an on-site health intervention such as providing healthy company cafeteria meals, because of their tendency toward out-of-office sales. Therefore, mHealth can offer additional convenience to change the physical activity habits of Japanese salespersons because they can access the mHealth contents anytime and anywhere. However, to the best of our knowledge, there has been no mHealth research that has focused specifically on Japanese salespersons.

Team-Based and Competitive Intervention Program

Previous studies have suggested that a team-based competitive approach would be effective to improve individual health-related behaviors and health outcomes. Shape Up Rhode Island, a statewide team-based campaign to promote weight loss and walking, showed that changes in weight were similar among teammates, and being on a more active team was associated with a greater increase in activity for individual members [15,16]. Accordingly, we expect that being a part of a team can improve the individual physical activity of team members through supportive interactions that encourage healthy behaviors [17,18]. In addition, it can be anticipated that people in a competitive environment use rankings as motivation to work toward their individual goals [18]. Therefore, we expect that an intervention program based on a team-based competitive approach would also be effective in targeting salespersons.

In this study, we conducted a team-based competitive walking intervention for Japanese salespersons via a smartphone app. The participants were required to record their daily step counts via the smartphone app. The participants could also check their individual daily step counts and their team's weekly ranking via the smartphone app.

Goal of this Study

This observational study aimed to identify elements that account for maintenance of physical activity during the intervention. Two different approaches that could be used to estimate the extent to which participants maintain their physical activity levels are calculating the average physical activity level and the amount of change of physical activity from the initial to the final period during the intervention. Therefore, the purpose of our study was two-fold, namely: (1) to estimate the effect of team and individual behavioral characteristics on the average walking steps of each participant, and (2) to examine the effect of team and individual behavioral characteristics on the difference in walking steps between the initial and final week of the program.

Methods

Contents of the "Walking Intervention"

We held the "Walking Intervention" to promote walking activity via smartphone apps at 11 companies representing the financial, manufacturing, and pharmaceutical industries.

To recruit participants, we distributed the original leaflets to sales departments at each company. We did not provide any participant incentives for applicants at the recruitment stage of this study. Applicants who were regular employees, working in the sales department of each company, and who could carry smartphones with them during working hours to record their walking steps were eligible to participate in this intervention. The decision to include applicants under medical treatment in this intervention was officially determined at the discretion of a physician. We registered 416 participants from September to October 2018. A total of 85 teams of 3 to 8 members each were

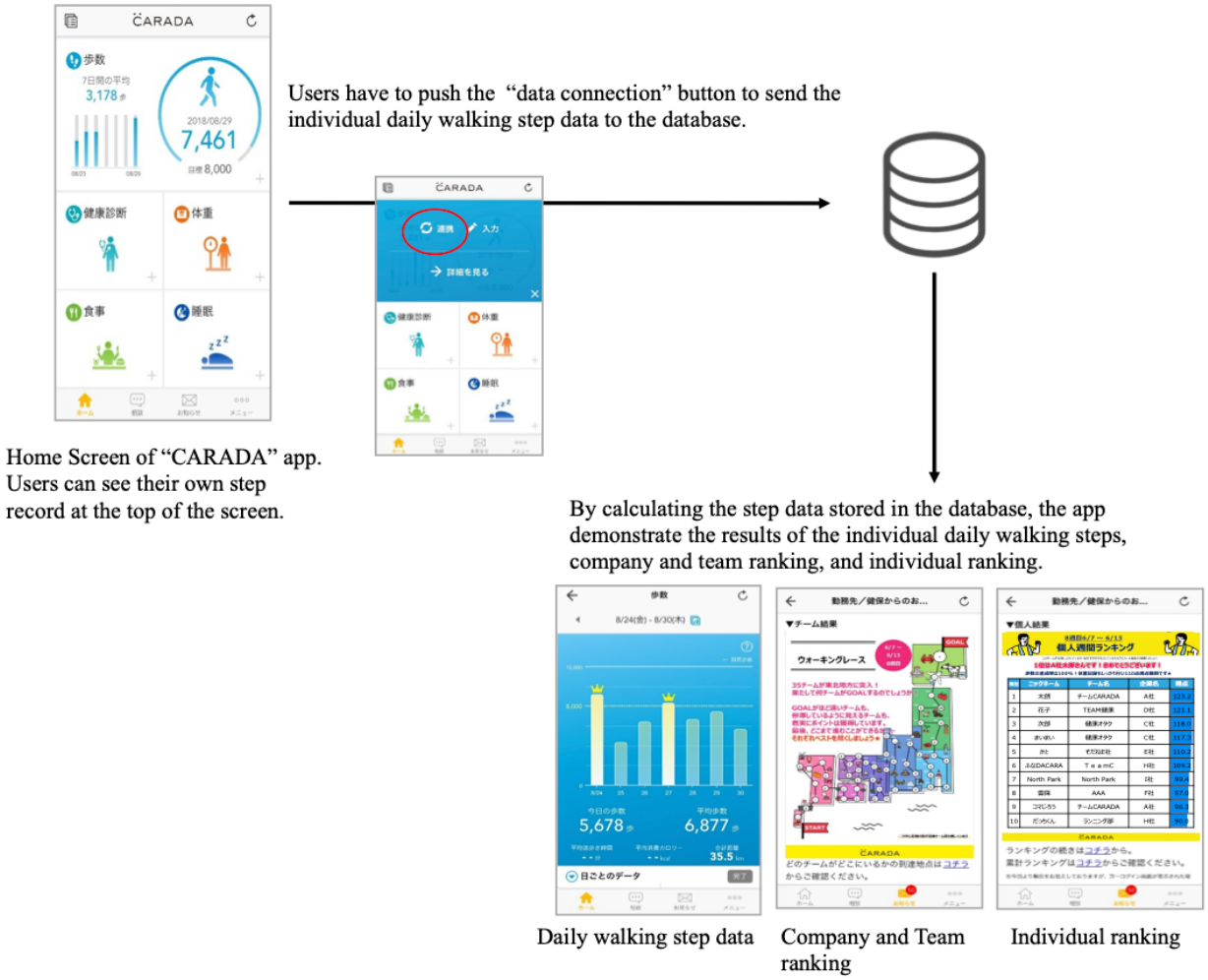
formed. The team was formed by self-selected members or by an intervention manager at each company who recruited participants from their own worksites.

Before the Walking Intervention started, we explained to each participant the health benefits of walking, but the participants did not have to set individual daily walking goals (eg, 8000 steps a day). Since the participants' worksites and residences geographically varied, we did not have any joint event to promote the participants' walking behavior. In addition, we did not officially provide any messages to the participants recommending them to walk more.

Since salespersons are likely to be out of the office for sales and need an easy-to-participate program that can be conducted anytime and anywhere, we held the intervention via smartphone apps. The participants downloaded the app CARADA (MTI Ltd, Tokyo, Japan) to allow them to check their daily walking steps. Individual and team rankings were calculated weekly by the walking step counts displayed in the app.

Figure 1 demonstrates the features of the CARADA app. The app can automatically record the participants' daily walking steps and shows their step records at the top of the home screen. To send their daily walking step data to the database, the participants have to push the "data connection" button manually. The app displays the results of the individual daily walking steps, along with the company, team, and individual rankings by calculating the step data stored in the database. Based on previous studies indicating the effectiveness of social comparison [18], we expected that the participants would be motivated by checking the company, team, and individual ranking results.

Figure 1. Features of the CARADA app.



We held a seminar at each company before starting the walking intervention during which the intervention manager explained the purpose of the intervention, the rules, and how to download the CARADA app. Seminars were held at a total of 7 companies; the other 4 companies did not have the chance to gather the participants together before the intervention. Therefore, for the companies that did not have a seminar in advance, we provided materials that included the same information as provided in the seminar: the purpose of the intervention, the rules, and how to download the CARADA app.

The intervention was held from October 17 to December 17, 2018 (total of 9 weeks or 62 days). One company had a 2-week delay from October 31 to December 31, 2018.

Postintervention Assessment

We administered questionnaires to the participants after the intervention (January 2019) via the CARADA app to assess the frequency of communication between team members, daily means of transportation during working hours, the reason for participating in this intervention, and the applicability of an

app-based walking intervention. According to previous studies [15-18], we hypothesized that some factors (eg, good team relationship) would be essential for the success of a worksite health intervention. Therefore, we developed the questionnaires based on the hypothesis and through discussion with the intervention managers (see Multimedia Appendix 1). To reduce the burden on the participants, we did not insist that they answer the questionnaires both prior to and after the intervention.

Statistical Analyses

Because the distribution of average walking steps was right-skewed, we could not apply a parametric analysis that requires a normal distribution. Therefore, the Wilcoxon rank-sum test or Kruskal-Wallis rank-sum test was used to compare the median of average walking steps between the explanatory categorical variables. One-way analysis of variance (ANOVA) was used to compare the average difference in walking steps between the initial and final week between the explanatory categorical variables.

We performed analyses to estimate the effects of team and behavior characteristics on average walking steps and the difference of walking steps between the initial and the final week. The following team characteristics were included: team size (number of participants in a team, ranging from 3 to 8 members) and the frequency of communication with team members (rarely=reference, once or twice a month, once or twice a week, three or four times a week, every day). As a behavioral characteristic, the number of days with recorded steps of each participant was examined.

We first performed multiple regression and multilevel analyses to examine the effect of team and individual behavioral characteristics on the average walking steps, and then performed an additional multiple regression analysis to examine the effect of team and individual behavioral characteristics on the difference in walking activity between the initial and the final week throughout the program.

For the first analyses, the average walking steps of each participant during the intervention was the dependent variable. Multiple regression analysis was performed to examine the relationship of team and behavioral characteristics with average walking steps. As the degree of health consciousness and motivation for physical activity were considered to likely differ among the 11 companies, we subsequently applied the random intercept model to consider the random effect of the company to which each participant belonged. In addition, we used the business category (finance, manufacturing, and pharmaceutical) as a confounding variable because the lifestyle and work style among salespersons could vary according to business category. The average walking steps of each participant was log-transformed as a dependent variable, because the values were always above zero and the distribution was right-skewed.

For the second analysis, the difference in walking steps between the initial (ie, first and second week) and the final week (ie, ninth week) for each participant was the dependent variable. The difference in walking steps between the initial and the final week was calculated by subtracting the average walking steps for the final week from that of the initial week. Multilevel analysis was also used because the data had a clustered structure (ie, step count data of participants within companies). However, because the eligible data for the analysis of difference in walking steps included that of a company with step counts from only one participant, we did not apply the multilevel analysis for the difference in walking steps.

We used the following factors as confounding variables: sex (male=1, female=reference), age (years), daily means of transportation during working hours (1=use of a public transportation system, not using public transportation=reference) or walking (walking=1, not walking=reference), the reason for deciding to participate in the intervention (voluntary participation=1, invitation from their colleagues or boss=reference), the business category to which the participants belonged (finance=reference, manufacturing, pharmaceutical), and whether the company to which the participants belonged had a seminar before the intervention (having a seminar=1, not having a seminar=reference). We examined the correlation coefficients of all explanatory variables. All correlation coefficients were between –0.5 and 0.5, and there were no strong correlations observed among the explanatory variables.

Statistical analyses were performed using R 3.5.1 software with the dplyr [19], ggplot2 [20], and lme4 [21] packages. A P value <.05 was considered statistically significant.

Ethical Approval

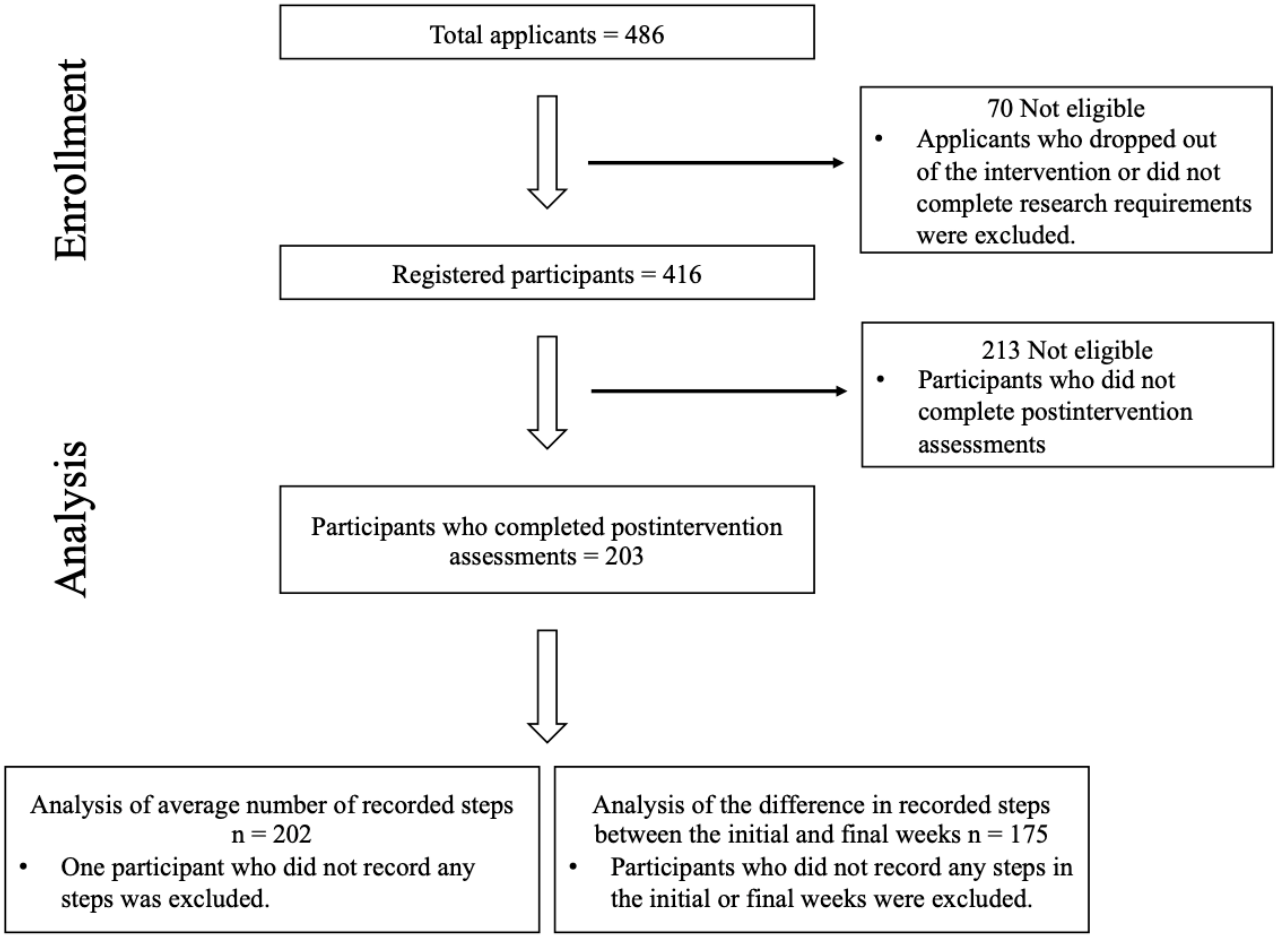
This study was conducted with payouts from the Ministry of Health, Labor, and Welfare, Japan. The University of Tokyo Human Research Ethics Committee approved this study (18-213). All participants received a written and/or verbal explanation of the intervention requirements before providing consent. We provided an informed consent form when the participants downloaded the CARADA app.

Results

Recruitment and Participants

Of the 486 people who applied to participate in the walking intervention, 70 were not eligible for official participation because they either dropped out of the intervention or did not comply with the research requirements. Therefore, we officially registered 416 participants for the walking intervention. We included the walking step data of 203 participants who completed the assessments after the intervention in the final analyses (Figure 2). We did not include any participants with incomplete data in the analysis (n=213) who did not complete the final assessment, because some essential explanatory variables were obtained only from the final assessment. For example, it was not possible to ascertain the exact subjective team evaluation (eg, frequency of communication with team members) from the noncompleters. Therefore, if a team was composed of only noncompleters, we could not analyze that team in this study.

Figure 2. Study flow diagram.



Characteristics of Participants

Table 1 summarizes the characteristics of the final 202 participants included in the analyses on the average walking steps during the intervention. We excluded one of the 203 participants who completed the assessments after the intervention from this analysis because they failed to record any steps during the intervention. Table 2 shows the characteristics of 175 participants included in the analyses with respect to the difference in walking steps between the initial and the final week (28 of 203 participants who did not record any steps at the initial or during the final week were excluded).

The sample predominantly comprised men for both the analysis on the average walking steps and the analysis on the difference of walking steps. The majority of participants communicated with their team members once or twice a week among the five communication categories. The average team size was 5.8. The number of recorded step days was 54.7 for the analysis on the average walking steps and was 58.5 days for the analysis on the difference of walking steps (Table 1 and Table 2).

Table 1. Characteristics of participants for the analysis on average walking steps during the intervention (N=202[a]).

Characteristic	Value
Sex (male), n (%)	159 (78.7)
Age (years), mean(SD)	41.3 (8.91)
Daily means of transportation, n (%)	
Public transportation system=no	131 (64.9)
Public transportation system=yes	71 (35.1)
Walking=no	137 (67.8)
Walking=yes	65 (32.2)
Voluntary participation, n (%)	141 (70.1)
Business category, n (%)	
Finance	90 (44.6)
Manufacturing	22 (10.9)
Pharmaceutical	90 (44.6)
Having a seminar before the intervention, n (%)	154 (76.2)
Frequency of communication with team members, n (%)	
Rarely	31 (15.4)
Once or twice a month	36 (17.9)
Once or twice a week	73 (36.3)
Three or four times a week	28 (13.9)
Every day	33 (16.4)
Did not answer	1 (0.5)
Team size, mean (SD)	5.8 (1.62)
Number of days with recorded steps, mean (SD)	54.7 (12.44)

[a]One participant who did not record any steps at all was excluded from the total 203 participants.

Table 2. Characteristics of participants for the analysis on the difference of walking steps between the initial and the final week (N=175[a]).

Characteristic	Value
Sex (male), n (%)	136 (77.7)
Age (years), mean (SD)	41.9 (9.04)
Daily means of transportation, n (%)	
Public transportation system=no	108 (61.7)
Public transportation system=yes	67 (38.3)
Walking=no	115 (65.7)
Walking=yes	60 (34.3)
Voluntary participation, n (%)	122 (70.1)
Business category, n (%)	
Finance	78 (44.6)
Manufacturing	17 (9.7)
Pharmaceutical	80 (45.7)
Having a seminar before the intervention (%)	132 (75.4)
Frequency of communication with team members (%)	
Rarely	22 (12.6)
Once or twice a month	30 (17.2)
Once or twice a week	66 (37.9)
Three or four times a week	26 (14.9)
Every day	30 (17.2)
Did not answer	1 (0.6)
Team size, mean (SD)	5.8 (1.60)
Number of days with recorded steps, mean (SD)	58.5 (6.49)

[a]Twenty-eight participants who did not record any steps in the initial or final week were excluded from the total 203 participants.

Multimedia Appendix 2 shows the results of the comparison between the completers (n=203), who completed the final assessment, and the noncompleters (n=213), who did not complete the final assessment, in terms of age, sex, and the average recorded walking steps. The completers were significantly older, walked more, and the proportion of female completers was higher than that of noncompleters.

Descriptive Statistics

Figure 3 shows the distributions of the average walking steps and difference in walking steps between the initial and the final week. The distribution of the average walking steps was right-skewed, and the distribution for the difference in walking steps tended to be normal.

The mean and median (IQR) average walking steps was 8024 and 7908 (5710-9819), respectively, with a range of 1724 to 22,965. The Shapiro-Wilk normality test showed a significant effect (W=0.96, P<.001). The mean and median (IQR) difference in walking steps between the initial and the final week was –898.0 and –760.4 (–2462.6 to 878.8), respectively, with a range of –7908.5 to 6429.2. The Shapiro-Wilk normality test was not significant (W=0.99, P=.19).

Figure 3. Distribution of the average walking steps and difference in walking steps between the initial and the final week for each participant.

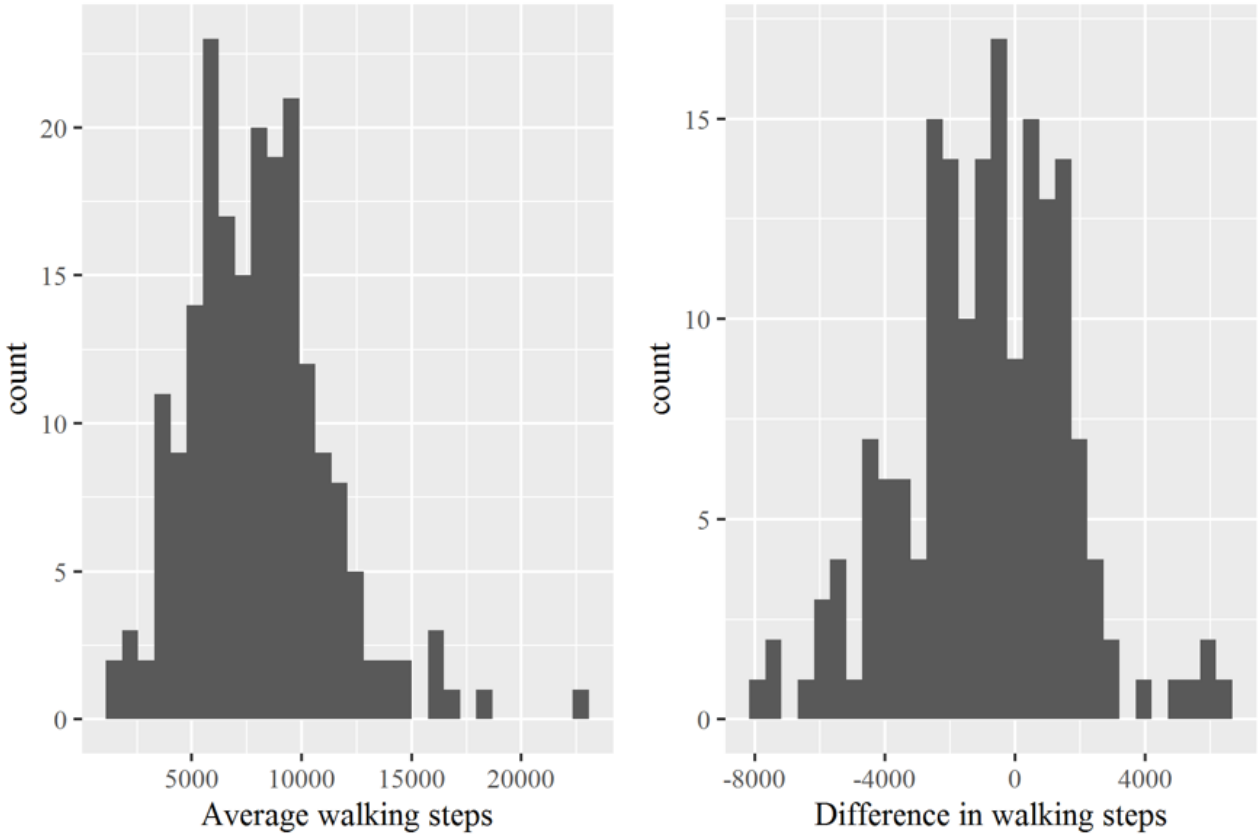


Table 3 shows the Wilcoxon or Kruskal-Wallis rank-sum test results for the average walking steps of each participant for each categorical variable. The average walking steps differed significantly between the following variables: using the public transportation system or walking, the company's business category to which the participants belonged, and having a seminar before the intervention. The average walking steps and age were significantly correlated (Pearson correlation coefficient 0.23, *P*=.001). The average walking steps was also significantly correlated with the number of days with recorded steps (Pearson correlation coefficient 0.37, *P*=.001). Team size did not significantly correlate with average walking steps. The participants who did not have a seminar recorded significantly higher average walking steps than those who did receive the seminar in advance (Table 3).

Table 3. Average walking steps for each explanatory variable.

Categorical variables	n	Median (IQR) average steps[a]	*P* value[b]
Sex			.08
Female	43	8404 (6888-10,228)	
Male	159	7623 (5610-9701)	
Use of public transportation system			<.001
No	131	6775 (5350-8618)	
Yes	71	9527 (8170-11,185)	
Walking			<.001
No	137	7025 (5576-9183)	
Yes	65	9149 (7570-11,003)	
Voluntary participation			.86
No	60	7806 (5572-9799)	
Yes	141	7910 (5782-9831)	
Did not answer	1	9333 (N/A[c])	
Business category			.003
Finance	90	8363 (5961-10,370)	
Manufacturing	22	9235 (7864-9936)	
Pharmaceutical	90	6778 (5391-9123)	
Taking a seminar before the intervention			<.001
No	48	8816 (8040-10,346)	
Yes	154	6967 (5391-9468)	
Frequency of communication with team members			.27
Rarely	31	8058 (5667-9626)	
Once or twice a month	36	7154 (4505-9251)	
Once or twice a week	73	7750 (5722-10,158)	
Three or four times a week	28	7385 (5319-9151)	
Every day	33	8529 (6602-10,333)	
Did not answer	1	9333 (N/A)	

[a]The distribution of average walking steps of each participant was right-skewed; therefore, the medians of average walking steps with IQR are shown for each categorical variable.

[b]The *P* values for categorical variables were estimated by the Wilcoxon or Kruskal-Wallis rank-sum test.

[c]N/A: not applicable.

One-way ANOVA results for the difference in walking steps between the initial and the final week for each categorical variable demonstrated that participants who used public transportation systems or walked for their daily transportation recorded a significantly smaller difference in walking steps between the initial and the final week than those who did not use public transportation systems or did not walk for daily transportation. There were no significant correlations between any of the continuous variables and the difference in walking steps (Table 4).

Table 4. Difference in walking steps between the initial and the final week for each explanatory variable.

Categorical variables	n	Mean (SD) of difference	*P* value[a]
Sex			.22
Female	39	–453 (2638)	
Male	136	–1026 (2542)	
Use of public transportation system			.006
No	108	–1315 (2671)	
Yes	67	–226 (2252)	
Walking			.01
No	115	–1241 (2519)	
Yes	60	–240 (2550)	
Voluntary participation			.72
No	52	–1017 (2284)	
Yes	122	–865 (2690)	
Did not answer	1	1216 (N/A[b])	
Business category			.79
Finance	78	–762 (2325)	
Manufacturing	17	–846 (1786)	
Pharmaceutical	80	–1042 (2926)	
Having a seminar before the intervention			.10
No	43	–333 (2186)	
Yes	132	–1082 (2661)	
Frequency of communication with team members			.42
Rarely	22	–1797 (2220)	
Once or twice a month	30	–1082 (3244)	
Once or twice a week	66	–633 (2446)	
Three or four times a week	26	–995 (1751)	
Every day	30	–624 (2891)	
Did not answer	1	1216 (N/A)	

[a]The *P* values for categorical variables were calculated by one-way analysis of variance.

[b]N/A: not applicable.

Multiple Regression and Multilevel Analyses

Table 5 shows the multiple regression analysis results for the average walking steps (multiple R^2=0.36; adjusted R^2=0.31). The behavioral characteristic (number of days with recorded steps) was positively correlated with the log-transformed average walking steps (β=.01, *P*<.001).

Table 5. Multiple regression analysis for factors influencing average walking steps during the intervention[a] (N=192[b]).

Variables	β[c]	SE	*P* value
Intercept	8.05	0.22	<.001
Sex (male=1)	–.02	0.07	.80
Age (years)	.01	0.003	.02
Daily means of transportation (using public transportation system=1)	.18	0.07	<.001
Daily means of transportation (by walking=1)	.04	0.06	.50
Voluntary participation (yes=1)	.04	0.07	.54
Business category			
Finance (reference)	N/A[d]	N/A	N/A
Manufacturing	.09	0.1	.36
Pharmaceutical	–.09	0.08	.27
Having a seminar before the intervention (yes = 1)	–.07	0.08	.41
Frequency of communication with team members			
Rarely (reference)	N/A	N/A	N/A
Once or twice a month	–.08	0.09	.38
Once or twice a week	–.02	0.08	.80
Three or four times a week	–.03	0.1	.77
Every day	.03	0.1	.75
Team size	–.01	0.02	.48
Number of days with recorded steps	.01	0.002	<.001

[a]The participants' average steps were log-transformed.

[b]Participants with missing data for explanatory variables were excluded from the analysis.

[c]Partial regression coefficients are reported.

[d]N/A: not applicable.

Table 6 shows the multilevel analysis results for the average walking steps according to the participants' company level. The intraclass correlation coefficient (ICC) was 37% in the null model (0.07 [variance of company]/0.12 [variance of residual]+0.07 [variance of the company]), indicating that belonging to the same company largely affected an individual's average walking steps. The Akaike Information Criterion value for the null model was 178.93 and was 207.3 for Model 1. In this multilevel analysis, the number of days with recorded steps still showed a significant positive correlation with the log-transformed average walking steps.

Table 6. Multilevel analysis results for average walking steps during the intervention[a] (N=192[b]).

Variables	β[c]	SE	*P* value
(Intercept)[d]	8.21	0.28	<.001
Sex (male=1)	.12	0.07	.07
Age (years)	.002	0.003	.43
Daily means of transportation (using public transportation system=1)	.14	0.07	.04
Daily means of transportation (walking=1)	–.01	0.06	.82
Voluntary participation (yes=1)	.06	0.07	.38
Business category			
Finance (reference)	N/A[e]	N/A	N/A
Manufacturing	.05	0.20	.82
Pharmaceutical	–.16	0.21	.48
Having a seminar before the intervention (yes=1)	–.10	0.20	.63
Frequency of communication with team members			
Rarely (reference)	N/A	N/A	N/A
Once or twice a month	–.05	0.08	.56
Once or twice a week	–.01	0.08	.90
Three or four times a week	–.08	0.09	.38
Every day	.06	0.09	.52
Team size	–.0008	0.02	.96
Number of days with recorded steps	.01	0.002	<.001

[a]The participants' average steps during the intervention were log-transformed.

[b]Participants with missing explanatory variable data were excluded from the analysis.

[c]Partial regression coefficients are reported.

[d]Intercept of null model=9.02 (–0.09)

[e]N/A: not applicable.

Table 7 shows the multiple regression analysis results for the difference in walking steps between the initial and the final week (multiple R^2=0.13, adjusted R^2=0.04). For the team characteristics, significant correlations were observed between having communication with team members once or twice a week and preventing a decrease in walking steps from the initial to the final week. Being on a larger team was also correlated with a decrease in walking steps from the initial to the final week.

Table 7. Multiple regression model results for the difference in walking steps between the initial and the final week (N=167[a]).

Variables	β[b]	SE	*P* value
(Intercept)	2810.2	2331.2	.23
Sex (male=1)	–659.5	526.9	.21
Age (years)	–11.2	23.8	.64
Daily means of transportation (using public transportation system=1)	1051.7	515.1	.04
Daily means of transportation (walking=1)	613.1	487.9	.21
Voluntary participation (yes=1)	–316.8	558.7	.57
Business category			
Finance (reference)	N/A[c]	N/A	N/A
Manufacturing	–11.9	767.7	.99
Pharmaceutical	688.2	599	.25
Having a seminar before the intervention (yes=1)	–453.6	597.5	.45
Frequency of communication with team members			
Rarely (reference)	N/A	N/A	N/A
Once or twice a month	1041.4	741.9	.16
Once or twice a week	1539.4	683.4	.03
Three or four times a week	1142.6	775.4	.14
Every day	1143.8	777.7	.14
Team size	–328.4	130.8	.01
Number of days with recorded steps	–39.4	32.3	.23

[a]The participants with missing explanatory variable data were excluded from the analysis.

[b]Partial regression coefficients are reported.

[c]N/A: not applicable.

Discussion

Principal Results

The average walking steps had a significant positive correlation with the individual behavioral characteristic (number of days with recorded steps). However, preventing a decrease in walking steps from the initial to the final week had significant correlations with two team characteristics (having communication with team members once or twice a week and being on a smaller sized team).

Previous studies have suggested that a positive change in physical activity was most likely to occur at the beginning of the program, and the physical activity level tended to decrease from the beginning to later periods [9,10]. The present study showed that the elements contributing to recording a higher number of average walking steps and those contributing to preventing a decrease in walking steps from the initial to the final week were different. The intervention for individual behavior (eg, offering an easy activity tracking tool) was effective in enhancing the basic physical activity level. The intervention for team relationships (eg, providing support for team members via some communication tools) was effective in preventing a decrease in physical activity level during the 9-week intervention period.

Average Walking Steps

The number of days with recorded steps was positively correlated with the average walking steps throughout the intervention. Recording step counts via smartphone apps can offer an easy way of monitoring the steps, and thus provide direct feedback on daily walking patterns. Previous studies have reported that use of a self-monitoring technique was associated with a positive change in health-related behavior and health outcomes [22,23]. For example, self-weighing led to success with weight loss; furthermore, with a greater number of days that participants monitored their own weight, there was a greater decrease in weight [23]. Intervention programs using self-monitoring methods also showed great potential to increase physical activity [22]. Feedback tools can be effective for increasing physical activity; Kang et al [24] showed that using a pedometer with a feedback tool significantly increased the participants' motivation to become more active.

Smartphone apps can provide more effective self-monitoring and feedback tools than traditional pedometers [25]. For example, smartphone apps can be equipped with visible feedback content, which is not possible with traditional pedometers. It may be more convenient to record and monitor daily step counts via smartphone apps than traditional pedometers, as people usually carry smartphones with them throughout the day. In addition, smartphone apps can offer more

accessible self-monitoring tools than a website-based program [26]. Therefore, it is speculated that the number of days with recorded steps in this study might have been amplified by using smartphone apps with strong self-monitoring and feedback features.

In the multilevel analysis on the average walking steps, the ICC was 37% and being in the same company largely explained the individual variance in average walking steps. It is possible that a company that employed more healthy salespersons happened to have a healthy culture and more motivation toward physical activity as a whole. However, considering the results of previous studies showing that community culture has a significant impact on individual health-related behaviors and health outcomes [27-29], the causal relationship between health outcomes of individual employees and the degree of health consciousness at the company level could be in the opposite direction. This study showed that participants who did not have a seminar before the intervention recorded a higher average number of walking steps. One possible reason is that companies that consider their employees to not have a healthy lifestyle and promote Healthy and Productivity Management [12] might have encouraged the participants to take a seminar to further motivate them. However, the seminar was not effective in improving the employees' behavior. We can assume that interventions to raise awareness about health that are administered company-wide can result in greater success in the change in individual employees' health outcomes.

Difference in Walking Steps

In the analyses on the difference of walking steps between the initial and the final week, a smaller team size correlated with preventing a decrease in walking steps throughout the intervention. It has been suggested that a social network and social support can contribute to increased physical activity levels [30,31]. We assumed that the participants in smaller teams would more easily be able to communicate with each other, and that the positive social support effect was more likely to occur in smaller rather than in larger teams.

Having communication with team members once or twice a week also correlated with preventing a decrease in walking steps from the initial to the final week. Previous studies showed that having a social interaction could reduce participants' attrition [32]. It is possible that having regular communication built social support among team members, which contributed to preventing a decrease in walking steps from the initial to the final week. However, there was no dose-response relationship between the frequency of communication and prevention of a decrease in walking steps. Having communication more than once or twice a week did not correlate with preventing a decrease in walking steps. One possible explanation is that the effect of social support on physical activity performance was influenced by personality type [33]. Some participants may feel encouraged to do more walking, whereas others may feel pressured from social support. Therefore, this finding may indicate that a moderate level of communication is needed to promote physical activity.

Limitations

One of the potential limitations of this study is that we included participants who voluntarily decided to participate in this walking intervention; thus, the sample population might have been comparatively more health-conscious than the general population. We excluded 213 participants from the analyses who did not complete the assessment. In addition, we conducted the assessment via a smartphone app; therefore, it is possible that participants who were unfamiliar with the use of smartphone apps could not properly respond to the assessment. There is a need for further examination on whether the effectiveness of self-monitoring and feedback via smartphone apps that was observed in this study can also be applied to people who are unfamiliar with the use of smartphone apps. The assessment was conducted after the intervention, raising the possibility that only the people who were motivated by the walking intervention might have answered the assessment questionnaire. The people who were highly motivated may have been the participants who were more satisfied with this intervention because of their team relationship or their success with walking behavior. In fact, the average number of recorded walking steps among the completers was higher than that of noncompleters. It is possible that only the people with a high number of recorded walking steps or that had good communication with team members might have been included in this analysis. The same bias could have occurred among the 28 of the 203 participants who did not record any steps in the initial week or during the final week, and were therefore excluded in the analyses due to a lack of difference in walking steps between the initial and the final week.

Additionally, the sample was predominantly comprised of men. According to the population census of 2015 [11], the proportion of men among salespersons was about 56%. The proportion of men in this study was larger than that indicated from the census result. One of the possible reasons for this discrepancy is that the industries selected to focus on in this study tend to employ more male salespersons than female salespersons; for example, the proportion of male salespersons in the manufacturing industry (including the pharmaceutical industry) was over 80% and that in the financial industry was about 50% [11]. Therefore, a wide range of industries should be targeted to generalize these results in future studies.

The number of days with recorded steps might also reflect the participants' motivation for walking; that is, participants with a more inherent motivation to walk would record a higher number of days with steps. This could have resulted in reduced effectiveness of self-monitoring and feedback using smartphone apps compared to expectation.

Another limitation is that the study period was shorter than that used in other studies focusing on the long-term intervention effect on physical activity. For example, Buckingham et al [5] reviewed the studies regarding an mHealth intervention for physical activity and reported that a significant mHealth effect on physical activity was largely observed from 1 month to 12 months after the beginning of the intervention. Although our study period was only 9 weeks, we expected that not all of the participants would maintain their physical activity level during this period. For example, Boyce et al [8] reported that the

greatest change in mean BMI occurred within the first few weeks (ie, 4 weeks) and then plateaued until the end of intervention (ie, at week 12). Therefore, we expected that the physical activity level would decline in most participants, while some would maintain the increased physical activity level. Our aim was to investigate the specific factors that would most contribute to these differences in physical activity levels. Therefore, follow-up research after the intervention (eg, 12 months) should be conducted to assess the long-term effects similar to previous studies.

This study used a multilevel analysis to consider the company effect of this clustered data structure. We also tried to consider the clustered effect of a team, but as some teams had only one eligible participant for analysis, we could not apply the multilevel analysis considering the team effect. We checked the correlation between all of the explanatory variables (including team variables). All correlation coefficients were between –0.5 and 0.5, and the observed explanatory variables were numerically independent. However, it is possible that the clustered effect of a team was not sufficiently investigated.

As this study used a cross-sectional design, we cannot infer causal relationships between the individual recorded steps and team or behavioral characteristics. We did our best to adjust for possible confounders using statistical analysis, but since this study was not a randomized controlled trial, there are many unmeasured confounders that could also explain the results. Therefore, we cannot simply generalize the results from this study.

Conclusion

This study showed that the elements that accounted for raising the average number of walking steps and those that accounted for preventing a decrease in walking steps from the initial to the final week were different. The behavioral characteristic (number of days with recorded steps via smartphone apps) was positively correlated with the average number of walking steps. Team characteristics (having regular communication once or twice a week and a smaller team size) were significantly correlated with preventing a decrease in walking steps. Considering the above factors, we can develop more effective mHealth programs focusing on salespersons, which can be used for maintenance of high physical activity levels.

Acknowledgments

This research was funded with payouts from the Ministry of Health, Labor, and Welfare, Japan: Subsidy for elderly medical care managements at FY 2018 (in Japanese).

Authors' Contributions

YF and MK developed the study concept. YH conducted the recruitment, data collection, analysis, and drafting of the manuscript. HI provided technical advice. All authors reviewed, edited, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaires for assessing motivations and experience after the walking intervention.
[DOCX File , 32 KB - mhealth_v8i7e16159_app1.docx]

Multimedia Appendix 2
Comparison of characteristics between completers and noncompleters in terms of age, sex, and average recorded walking steps.
[DOCX File , 19 KB - mhealth_v8i7e16159_app2.docx]

References

1. 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]
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. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]
4. 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]
5. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: A systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]

6. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
7. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
8. Boyce RW, Dyer EA, Willett TK, Figueroa JL, Jones GR. Police weight-loss competition in the United States with gender, age, and job comparisons with health promotion recommendations. Int J Health Promot Educ 2015 Feb 10;53(4):204-215. [doi: 10.1080/14635240.2015.1005321]
9. Boekhout JM, Berendsen BAJ, Peels DA, Bolman CAW, Lechner L. Evaluation of a Computer-Tailored Healthy Ageing Intervention to Promote Physical Activity among Single Older Adults with a Chronic Disease. Int J Environ Res Public Health 2018 Feb 15;15(2):346 [FREE Full text] [doi: 10.3390/ijerph15020346] [Medline: 29462862]
10. Kolt GS, Rosenkranz RR, Vandelanotte C, Caperchione CM, Maeder AJ, Tague R, et al. Using Web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
11. National Census 2015. Statistics Bureau, Ministry of Internal Affairs and Communication. URL: https://www.stat.go.jp/data/kokusei/2015/kekka.html [accessed 2019-06-19]
12. Kenkou Keiei No Suishin (Japanese). Ministry of Economy, Trade and Industry. URL: https://www.meti.go.jp/policy/mono_info_service/healthcare/kenko_keiei.html [accessed 2019-06-19]
13. National health and nutrition examination survey at FY 2018 (Japanese). Ministry of Health Labour and Welfare. URL: https://www.mhlw.go.jp/content/000615325.pdf [accessed 2020-06-09]
14. Yamaguchi M, Shimizu K, Nakazawa Y, Ikeda H, Ohfusa H. The relationship between occupations and lifestyle-related disease. J Japan Soc Human Dry Dock 2003 Sep 30;18(2):191-194. [doi: 10.11320/ningendock1986.18.191]
15. Leahey TM, Crane MM, Pinto AM, Weinberg B, Kumar R, Wing RR. Effect of teammates on changes in physical activity in a statewide campaign. Prev Med 2010 Jul;51(1):45-49 [FREE Full text] [doi: 10.1016/j.ypmed.2010.04.004] [Medline: 20394768]
16. Leahey TM, Kumar R, Weinberg BM, Wing RR. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418. [doi: 10.1038/oby.2012.18] [Medline: 22310234]
17. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
18. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
19. Wickham H, François R, Henry L, Müller K. dplyr: A grammar of data manipulation. R package version 0. URL: https://CRAN.R-project.org/package=dplyr [accessed 2020-06-09]
20. Wickham H. Ggplot2: Elegant Graphics For Data Analysis. New York: Springer International Publishing; Jul 30, 2010:2016.
21. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft 2015;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
22. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key Elements of mHealth Interventions to Successfully Increase Physical Activity: Meta-Regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
23. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265. [doi: 10.1002/oby.20946] [Medline: 25521523]
24. Kang M, Marshall SJ, Barreira TV, Lee J. Effect of pedometer-based physical activity interventions: a meta-analysis. Res Q Exerc Sport 2009 Sep;80(3):648-655. [doi: 10.1080/02701367.2009.10599604] [Medline: 19791652]
25. Fong SS, Ng SS, Cheng YT, Zhang J, Chung LM, Chow GC, et al. Comparison between smartphone pedometer applications and traditional pedometers for improving physical activity and body mass index in community-dwelling older adults. J Phys Ther Sci 2016 May;28(5):1651-1656 [FREE Full text] [doi: 10.1589/jpts.28.1651] [Medline: 27313391]
26. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012 Apr 20;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
27. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
28. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]
29. Leahey TM, Gokee LaRose J, Fava JL, Wing RR. Social influences are associated with BMI and weight loss intentions in young adults. Obesity (Silver Spring) 2011 Jun;19(6):1157-1162. [doi: 10.1038/oby.2010.301] [Medline: 21164501]

30. Lindsay Smith G, Banting L, Eime R, O'Sullivan G, van Uffelen JGZ. The association between social support and physical activity in older adults: a systematic review. Int J Behav Nutr Phys Act 2017 Apr 27;14(1):56 [FREE Full text] [doi: 10.1186/s12966-017-0509-8] [Medline: 28449673]
31. Suzuki E, Takao S, Subramanian SV, Doi H, Kawachi I. Work-based social networks and health status among Japanese employees. J Epidemiol Community Health 2009 Sep;63(9):692-696. [doi: 10.1136/jech.2008.082453] [Medline: 19286690]
32. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
33. Al Ayubi SU, Parmanto B, Branch R, Ding D. A Persuasive and Social mHealth Application for Physical Activity: A Usability and Feasibility Study. JMIR Mhealth Uhealth 2014 May 22;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]

Abbreviations

ANOVA: analysis of variance
ICC: intraclass coefficient
mHealth: mobile health

Edited by G Eysenbach, W Curioso; submitted 06.09.19; peer-reviewed by G Signorelli, C Katie, T Stütz, O Rivera; comments to author 24.02.20; revised version received 07.04.20; accepted 21.04.20; published 03.07.20.

Please cite as:
Hamamatsu Y, Ide H, Kakinuma M, Furui Y
Maintaining Physical Activity Level Through Team-Based Walking With a Mobile Health Intervention: Cross-Sectional Observational Study
JMIR Mhealth Uhealth 2020;8(7):e16159
URL: https://mhealth.jmir.org/2020/7/e16159
doi:10.2196/16159
PMID:32618576

Original Paper

Effect of a Health System–Sponsored Mobile App on Perinatal Health Behaviors: Retrospective Cohort Study

Caroline Cawley[1], MPH; Hannelore Buckenmeyer[1], MBA; Trina Jellison[2], MHA, BSN; Joseph B Rinaldi[3], BS; Keri B Vartanian[3], PhD

[1]Wildflower Health, San Francisco, CA, United States

[2]Women and Children's Institute, Providence St Joseph Health, Renton, WA, United States

[3]Center for Outcomes Research and Education, Providence Portland Medical Center, Portland, OR, United States

Corresponding Author:
Keri B Vartanian, PhD
Center for Outcomes Research and Education
Providence Portland Medical Center
5251 NE Glisan Street
Portland, OR, 97213
United States
Phone: 1 503 215 2594
Email: Keri.Vartanian@providence.org

Abstract

Background: Pregnancy mobile apps are becoming increasingly popular, with parents-to-be seeking information related to their pregnancy and their baby through mobile technology. This increase raises the need for prenatal apps with evidence-based content that is personalized and reliable. Previous studies have looked at whether prenatal apps impact health and behavior outcomes among pregnant and postpartum individuals; however, research has been limited.

Objective: The primary objective of this study is to assess whether the use of a health system–sponsored mobile app—Circle by Providence—aimed at providing personalized and reliable health information on pregnancy, postpartum recovery, and infant care is associated with improved health outcomes and increased healthy behaviors and knowledge among users.

Methods: This observational study compared app users and app nonusers using a self-reported survey and electronic medical records. The study took place over 18 months and was conducted at Providence St. Joseph Health in Portland, Oregon. The sample included patients who received prenatal care at one of seven Providence clinics and had a live birth at a Providence hospital. Recruitment occurred on a rolling basis and only those who completed the survey were included. Survey respondents were separated into app users and app nonusers, and survey responses and clinical outcomes were compared across groups using univariate and adjusted multivariate logistic regression.

Results: A total of 567 participants were enrolled in the study—167 in the app user group and 400 in the nonuser group. We found statistically significant differences between the two groups for certain behavior outcomes: subjects who used the app had 75% greater odds of breastfeeding beyond 6 months postpartum (P=.012), were less likely to miss prenatal appointments (P=.046), and were 50% more likely to exercise 3 or more times a week during pregnancy (P=.04). There were no differences in nutritional measures, including whether they took prenatal vitamins, ate 5 fruits or vegetables a day, or drank caffeine. We found no differences in many of the infant care outcomes; however, there was an increase in awareness of “purple crying.” Finally, there were no significant differences in measured clinical health outcomes, including cesarean births, length of hospital stays (in minutes), low birth weight infants, preterm births, small-for-gestational-age births, large-for-gestational-age births, and neonatal intensive care unit stays.

Conclusions: The use of the Circle app, which provides access to personalized and evidence-based health information, was associated with an increase in certain healthy behaviors and health knowledge, although there was no impact on clinical health outcomes. More research is needed to determine the impact of mobile prenatal apps on healthy pregnancies, clinical health outcomes, and infant care.

(JMIR Mhealth Uhealth 2020;8(7):e17183) doi:10.2196/17183

KEYWORDS

mobile health; perinatal health; health behaviors

Introduction

Demand for prenatal apps has increased over the years, as parents-to-be seek reliable information outside of clinical hours [1]. Lee et al [2] showed that apps related to pregnancy and childcare have become an important resource, particularly among first-time parents. There are currently dozens of available pregnancy apps aimed at providing varying degrees of information to individuals throughout their pregnancy and during the postpartum period.

While mobile pregnancy apps can provide education between prenatal visits, studies have shown that pregnancy apps often lack evidence-based information and localized resources that new parents can trust [2-4]. Other studies have shown that many pregnant individuals are turning to the internet to search for information related to pregnancy symptoms before speaking with their health care provider, sometimes consulting unreliable information [1,5]. Patient-facing technology, including apps, should be tailored to the patient to better engage and educate the patient on key risk factors and other issues related to their pregnancy and birth [6]. Some studies have shown that pregnant patients, especially first-time parents and those at high risk, frequently use mobile apps as a source of information on pregnancy and infant care [2,3,7,8]; however, many of the tools found in pregnancy apps have not been shown to be effective [9]. For example, breastfeeding trackers are increasingly prevalent, but very few have been shown to provide evidence-based information and recommendations [9,10]

To find a way to help pregnant patients outside of traditional prenatal appointments, the Providence St. Joseph Health's Consumer Innovation Team had discussions with 10 providers, including perinatal providers, service line leaders, and educators from the Swedish Medical Center and Providence Health & Services. They also conducted one-on-one interviews with 11 individuals in Seattle, Washington, and 7 individuals in Portland, Oregon, who were either pregnant or new parents. The provider and consumer sessions were facilitated by a 3rd-party user experience research firm. Providers reported that they were limited by time constraints during inpatient visits to address all patient questions, as well as their clinical and nonclinical needs. Pregnant individuals and new parents expressed a desire for information they could trust, that helped them feel "normal," could be personalized to their unique needs, and that connected them to care quickly when they needed additional support. Based on these discussions, the team designed the Circle app with the following goals:

- Provide Providence and its patients and their families with the opportunity to connect and receive personalized health information to promote healthy pregnancies, births, and pediatric care from birth to 18 years of age;
- Give users access to relevant, evidence-based pregnancy and infant care information they can trust;
- Connect users to prenatal and postnatal care and services;
- Provide informative content, to-do lists, and reminders related to prenatal and pediatric care, including tools to track fetal movement, pregnancy weight gain, vaccines, feeding, and diapers that parents-to-be and current parents can use in consultation with their provider;
- Improve patient retention with Providence St. Joseph Health providers;
- Positively impact family well-being;
- Integrate with Providence's MyChart and promote Express Care Virtual, which are online services offered by Providence that are aimed at increasing access to care.

Circle launched in 2016 in Portland, Oregon, and the greater Seattle, Washington, area. In 2018, Circle was acquired by Wildflower Health, a digital health company based in San Francisco. As of September 2018, the app is now available to patients across all Providence St. Joseph Health locations. As of November 2019, there were over 45,000 registered Circle users.

This study examined whether use of the Circle app during the prenatal period was associated with improved health and health behavior outcomes among pregnant and postpartum parents. We hypothesize that use of the Circle app would be associated with improved healthy behaviors and knowledge during and after pregnancy and improved clinical outcomes compared to non-Circle app users.

Methods

Study Overview

This was an observational study that used surveys and electronic medical records to examine a variety of outcomes on health knowledge, healthy behaviors for pregnancy and infant care, and clinical outcomes. The Providence Health and Services Institutional Review Board approved this study.

Data Sources

There were two main data sources for this study: the Providence Birth and Infant Care Survey and electronic medical records. The Providence Birth and Infant Care Survey was developed by Providence St. Joseph Health's Center for Outcomes Research and Education (CORE) in partnership with the Consumer Innovation Team. The survey included validated measures of certain health behaviors and knowledge in the prenatal and postnatal periods as well as infant care. The survey also measured demographics and the use of resources and technology (including the Circle app) to obtain information on pregnancy and infant care.

Electronic medical records from Epic, an electronic health record software application used by Providence St. Joseph Health, included individual-level data on clinic visits, medical procedures, and diagnoses during pregnancy and at birth. These data were used to identify the study sample based on engagement in prenatal care at one of the seven selected clinics and having a live birth at a Providence hospital. This data source

was also used to measure clinical outcomes at birth for the birth parent and infant.

Participants and Recruitment

We used Epic medical health records to select potential study participants based on the following criteria: they had four or more prenatal encounters at one of the seven selected Providence clinics in Portland, were over 18 years of age, and gave birth to a live infant at a Providence hospital in the past 4-6 months. Individuals were excluded if they were an employee of Providence. Recruitment into the study occurred from March 2018 to January 2019. During the study window, a total of 1500 people were identified as potential participants.

Paper surveys were sent out to all potential participants through the mail with a small monetary compensation for their time. Nonresponders received a second survey in the mail as well as an email reminder and electronic versions of the survey. A total of 618 viable surveys were returned for an overall response rate of 41.2%.

Survey responses were used to define the app user group and the comparison nonuser group. The app user group comprised individuals who indicated that they used the Circle app during the prenatal period and the nonuser group comprised individuals who indicated that they did not use the app during the prenatal period or did not use the app at all. Respondents were excluded from the analysis if they did not indicate use of any resources in the prenatal period (n=31) and if their responses concerning Circle use were contradictory (ie, it was unclear if they were, in fact, a Circle user; n=20). After exclusion, the total number of individuals with viable surveys was 567. A total of 167 people qualified for the app user group (those who used Circle in the prenatal period), and 400 people qualified for the comparison nonuser group (those who did not use Circle in the prenatal period).

Outcome Measures

The Providence Birth and Infant Care Survey includes many validated and developed measures of pre- and postnatal health, health care, health behaviors, and health knowledge. We selected three main types of outcome measures for this study:

- Prenatal care and behavior: patient's prenatal and postpartum care, including missed appointments; dental care; exercise, prenatal vitamins, fruit/vegetable and caffeine consumption [11-16]
- Infant care: breastfeeding, preparedness for appointment, awareness of "The Period of Purple Crying" (a time early in an infant's life when there is more crying than normal), and vaccine hesitancy [12,17-20]
- Demographics and socioeconomic status: age, race/ethnicity, gender identity, number of children, marital status, income, education, and insurance type [12]

Clinical outcomes were defined as binary variables (present or absent) using a combination of the International Statistical Classification of Diseases and Related Health Problems-10, Current Procedural Terminology / Healthcare Common Procedure Coding System, and revenue codes. Clinical outcomes included caesarian section, preterm birth (fewer than 37 weeks of gestation), low birth weight (less than 2500 g), small-for-gestational-age birth (weight below the 10th percentile for the gestational age), large-for-gestational-age birth (weight above the 90th percentile for that gestational age), length of stay for a birth event, and presence of a neonatal intensive care unit (NICU) stay.

Analysis

Univariate logistic regressions were used to identify demographic differences between the app user and nonuser groups. Survey outcome measures with Likert-type scales were collapsed into binary responses for positive and negative responses. For responses on a 5-point Likert scale, we collapsed only the clearly positive responses into the positive group with the remaining responses allocated to the negative group. We did not adjust our significance level based on multiple comparisons, which should be considered in the interpretation of results. Analysis of outcomes were conducted using one-step multivariable logistic regressions constructed with adjusting variables based on significant difference across the two groups and understanding of factors that can impact pregnancy outcomes—age (continuous), race, gross household income, number of children, and insurance type. All analyses were conducted in R, version 3.3.3. We considered $P<.05$ to be statistically significant for the purposes of this study.

Results

Demographic and Socioeconomic Profile

The Providence Birth and Infant Care Survey asked questions on demographics and the socioeconomic status of respondents to better understand the profile of individuals using Circle in the prenatal period compared to those who did not. Our results found that the prenatal Circle users and non-Circle users were similar in age and marital status (Table 1). We did find a statistically significant difference in race and ethnicity between users and nonusers ($P=.01$). Circle users were more likely to be white than non-Circle users. Circle users were also more likely to be pregnant with their first child compared to non-Circle users ($P<.001$).

The survey looked at several socioeconomic factors, including self-reported education, household income, and insurance status (Table 2). We found statistically significant differences between the groups in gross household income and insurance type. There were fewer low-income prenatal Circle users than non-Circle users ($P=.001$) and, likewise, fewer Circle users on Medicaid ($P=.02$). Despite this difference in health insurance coverage, there was a similar percentage of people in each group covered by the Providence Health Plan. These differences were adjusted for in our analyses.

Table 1. Demographic profile of study participants (N=567). Italicized *P* values are significant.

Characteristic	Prenatal Circle use (n=167), n (%)	No prenatal Circle use (n=400), n (%)	*P* value
Age group			.81
Under 25	7 (4.2)	26 (6.5)	
25-29	34 (20.4)	84 (21.0)	
30-34	62 (37.1)	133 (35.3)	
35-39	53 (31.7)	119 (29.8)	
40 and older	11 (6.6)	38 (7.5)	
Race/ethnicity			*.01*
White	117 (70.3)	243 (60.7)	
Hispanic	7 (4.2)	39 (9.8)	
Asian	14 (8.5)	60 (15.0)	
Other	29 (17.0)	58 (14.5)	
Number of children			
First child	109 (65.1)	193 (48.4)	*<.001*
More than 1 child	58 (34.9)	207 (51.7)	
Marital status			.65
Married or domestic partnership	145 (86.6)	341 (85.2)	
Single, never married	16 (9.8)	49 (12.3)	
Divorced	2 (1.2)	5 (1.3)	
Something else	4 (2.4)	5 (1.3)	

Table 2. Socioeconomic profile of study participants (N=567). Italicized *P* values are significant.

Characteristic	Prenatal Circle use (n=167), n (%)	No prenatal Circle use (n=400), n (%)	*P* value
Education			.10
High school or less	18 (10.4)	64 (16.1)	
Some college	23 (14.0)	65 (16.4)	
Vocational training or 2-year degree	8 (4.9)	27 (7.2)	
A 4-year college degree or more	118 (70.7)	241 (60.4)	
Gross household income ($ US)			*.001*
$30,000 or less	25 (14.7)	97 (24.3)	
$30,001 to $50,000	14 (8.6)	66 (16.4)	
$50,001 to $70,000	22 (12.9)	52 (13.0)	
$70,001 to $100,000	38 (22.7)	55 (13.8)	
$100,001 to $150,000	33 (19.63)	75 (18.7)	
Insurance type			*.02*
Private coverage	126 (75.2)	253 (63.3)	
Medicaid/Oregon Health Plan	21 (12.7)	87 (21.8)	
I don't have insurance now	4 (2.4)	19 (4.8)	
Other	16 (9.7)	40 (10.0)	
Insured by Providence Health Plan			.45
Yes	70 (41.8)	167 (41.7)	
No	88 (52.7)	199 (49.8)	
I don't know	9 (5.5)	34 (8.4)	

Healthy Behaviors

The study examined prenatal care and healthy habits during pregnancy for the study population (Table 3). We found that very few people in either group felt they were unprepared for their prenatal care appointments and a similar percentage of people in each group received dental care during their pregnancy. We did find that Circle users were significantly less likely to miss prenatal appointments compared to non-Circle users, showing a 45% reduction in the odds of having a missed prenatal care appointment when using Circle (*P*=.046). Survey results also found that Circle users had 50% greater odds of reporting that they exercised three times a week during pregnancy (*P*=.04). All other behaviors measured showed no significant difference.

Table 3. Health behaviors of study participants while pregnant (n=541[a]). Italicized *P* values are significant.

Behavior	Prenatal Circle use (n=162), n (%)	No prenatal Circle use (n=379), n (%)	Logistic regression	
			aOR[b]	*P* value
Prenatal care				
Felt unprepared for appointments	10 (6.3)	16 (4.8)	1.24	.62
Missed prenatal appointments	18 (11.2)	69 (18.1)	0.55	*.046*
Received dental care during pregnancy	103 (63.7)	215 (56.6)	1.10	.64
Health behaviors				
Exercised 3 times a week	82 (50.6)	147 (38.9)	1.50	*.04*
Did not take recommended vitamins	8 (4.9)	23 (6.1)	0.91	.84
Did not eat 5 fruit/vegetables a day	30 (18.8)	79 (20.9)	0.95	.84
Drank more than 2 caffeinated beverages a day	12 (7.4)	31 (8.2)	1.32	.47

[a]Study participants with responses to health behavior questions.

[b]aOR: adjusted odds ratio. Adjusted for age (continuous), race, gross household income, number of children, and insurance type.

Caring for Infants

We assessed several aspects of infant care in the survey (Table 4). The results showed a 75% increase in the odds of breastfeeding for 6 months or more among Circle users (*P*=.012). In addition, there was a statistically significant increase in the awareness of the “Purple Crying” period among prenatal Circle users compared to non-Circle users (*P*=.04). We did not observe any differences in preparedness for infant appointments and vaccine hesitancy between groups.

Table 4. Responses from participants on caring for infants (n=541[a]). Italicized *P* values are significant.

Response	Prenatal Circle use (n=162), n (%)	No prenatal Circle use (n=379), n (%)	Logistic regression	
			aOR[b]	*P* value
Breastfed for more than 6 months	118 (73.1)	226 (59.6)	1.75	*.01*
Unprepared for infant appointments	4 (2.5)	11 (3.0)	0.50	.28
High vaccine hesitancy	9 (5.6)	31 (8.2)	0.73	.44
Never heard of “Purple Crying”	23 (14.1)	105 (27.8)	0.55	*.04*

[a]Study participants with responses to infant care questions.

[b]aOR: adjusted odds ratio. Adjusted for age (continuous), race, gross household income, number of children, and insurance type.

Clinical Health Outcomes

We examined clinical health outcomes for prenatal Circle users and non-Circle users, including cesarean births, length of hospital stays (in minutes), infants with low birth weight, preterm births, small-for-gestational-age births, large-for-gestational-age births, and NICU stays. There were no significant differences between the app user and nonuser groups for all measured clinical outcomes (Table 5).

Table 5. Clinical outcomes for birth parent and infant (n=541[a]). Italicized *P* values are significant.

Outcome	Prenatal Circle use (n=162), n (%)	No prenatal Circle use (n=379), n (%)	Logistic regression		
			aDiff[b]	aOR[c]	*P* value
Maternal outcomes					
Cesarean delivery, n (%)	40 (24.7)	99 (26.1)	—[d]	0.96	0.857
Length of stay (minutes), average (SD)	4256 (2615)	4196 (3591)	284	—	0.369
Infant outcomes, n (%)					
Low birth weight	2 (1.2)	3 (0.8)	—	0.82	0.831
Preterm birth	8 (4.9)	6 (1.6)	—	3.08	0.061
Small for gestational age	7 (4.3)	10 (2.6)	—	2.90	0.077
Large for gestational age	3 (1.9)	6 (1.9)	—	1.00	0.997
NICU[e] stay	8 (4.9)	20 (5.0)	—	0.72	0.452

[a]Study participants with clinical outcomes data.

[b]aDiff: adjusted means difference. Difference adjusted for age (continuous), race, gross household income, number of children, and insurance type.

[c]aOR: adjusted odds ratio. Odds ratio adjusted for age (continuous), race, gross household income, number of children, and insurance type.

[d]Not applicable.

[e]NICU: neonatal intensive care unit.

Discussion

Principal Findings

The Circle app was created with the goal of providing reliable, personalized content to its pregnant patients and new parents. The Providence health system saw the app as a potential tool to better reach patients outside of clinic hours and to help reduce the incidence of negative prenatal and birth outcomes. Data from the self-reported survey demonstrated that the app was associated with improvement in some healthy behaviors—increased exercise during the prenatal period, increased duration of breastfeeding, increased knowledge of the "Purple Crying" period, and a decrease in missed prenatal visits. These findings are in line with research demonstrating the feasibility and suitability of mobile apps in impacting behavioral change in other fields of health [21].

However, this study showed that use of the app was not associated with differences in clinical health outcomes such as premature births, cesarean births, low birth weight babies, and neonatal intensive care unit stays. Previous studies in the mobile health field have also demonstrated the difficulty in impacting clinical health outcomes [10,22,23]. Even in randomized controlled trials, studies that successfully demonstrate the effectiveness of prenatal apps in improving health behavior often fail to find statistically significant differences in neonatal outcomes, delivery, or pregnancy complications [24]. Due to small study populations, many trials may be underpowered to detect clinically and/or statistically significant changes in prenatal health outcomes [25]. More research is needed to better understand how digital health and mobile apps can help move the dial on adverse clinical health outcomes.

Our results contribute to the evolving health care field and the move toward digital care to better engage patients outside of a clinical setting. The use and availability of mobile apps in prenatal care is growing [1,7], and our study shows that the use of prenatal apps that provide reliable, personalized content to the patient can improve some behavior health outcomes among pregnant and new parents.

Limitations

This study has several notable limitations. First, our study population was limited to individuals who received care at 7 Providence clinics and gave birth at Providence hospitals in Portland, Oregon, limiting the number of people who would be eligible for the survey and creating a bias in the sample. Second, we did not use a random sample and could only include individuals who responded to our survey—to ensure they were or were not Circle users—which is another source of potential bias in our sample. Third, this was an observational study, not a randomized controlled trial (ie, respondents were not randomized to the app user and nonuser groups), meaning that there may be underlying differences between the two groups. Fourth, most outcomes were survey-based, which could be subject to reporting bias, such as individuals overreporting positive behaviors and underreporting unhealthy ones. However, since both groups are self-reporting, we can expect to see the same level of inflation in both groups. Findings from the survey on demographics and socioeconomic status suggest that Circle did not reach a diverse racial or socioeconomic population, limiting the generalizability of this study to other patient populations. Finally, while our survey results indicate that some individuals in the nonuser group were using other app-based resources, we did not make any formal comparisons on whether the Circle app led to different outcomes compared to other apps.

Conclusion

Our study demonstrated that using Circle was associated with positive impacts on several key health behavior and knowledge

outcomes but has no significant impact on clinical health outcomes. These health behaviors and knowledge outcomes have important benefits for healthy pregnancies and healthy infant care. Improvements could be made to target a more diverse audience to promote health equity, to expand the impact on health behaviors and knowledge, and to adjust content and other key features to impact clinical health outcomes. While our results demonstrate the potential utility of using apps to promote healthier behaviors and knowledge, more research is needed to determine the efficacy of health system–sponsored, personalized mobile pregnancy apps and their ability to improve health outcomes among pregnant and postpartum parents and their infants.

Acknowledgments

We would like to thank the survey participants who took the time to participate in this study and provide invaluable information. All research was funded by the Providence Consumer Innovation Team at Providence St. Joseph Health. No external funding was received for the purposes of this study.

Authors' Contributions

KBV, JBR, TJ and members of the CORE team contributed to the study design and were responsible for the data analysis. CC wrote the manuscript with revisions and input from HB, KBV, and TJ. All authors read and approved the final manuscript.

Conflicts of Interest

HB holds a position at Wildflower Health, the digital health company that acquired Circle. CC is a paid consultant for Wildflower Health.

References

1. 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]
2. 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]
3. Guerra-Reyes L, Christie VM, Prabhakar A, Harris AL, Siek KA. Postpartum Health Information Seeking Using Mobile Phones: Experiences of Low-Income Mothers. Matern Child Health J 2016 Nov 17;20(Suppl 1):13-21 [FREE Full text] [doi: 10.1007/s10995-016-2185-8] [Medline: 27639571]
4. Bert F, Gualano MR, Brusaferro S, De Vito E, de Waure C, La Torre G, et al. Pregnancy e-health: a multicenter Italian cross-sectional study on Internet use and decision-making among pregnant women. J Epidemiol Community Health 2013 Dec 01;67(12):1013-1018. [doi: 10.1136/jech-2013-202584] [Medline: 24072743]
5. 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 Jun 03;16(6):e147 [FREE Full text] [doi: 10.2196/jmir.3385] [Medline: 24892583]
6. Singh K, Drouin K, Newmark L, Rozenblum R, Lee J, Landman A, et al. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]
7. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
8. 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 26;22(3):608-617. [doi: 10.1177/1460458215574120] [Medline: 25900813]
9. Coughlin SS. The need for research-tested smartphone applications for promoting breastfeeding. Mhealth 2016 May 04;2(5):18-18 [FREE Full text] [doi: 10.21037/mhealth.2016.04.03] [Medline: 27231691]
10. Kennelly MA, Ainscough K, Lindsay KL, O Sullivan E, Gibney ER, McCarthy M, et al. Pregnancy Exercise and Nutrition With Smartphone Application Support: A Randomized Controlled Trial. Obstet Gynecol 2018 May;131(5):818-826. [doi: 10.1097/AOG.0000000000002582] [Medline: 29630009]
11. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012 Nov 26;12(1):1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
12. Centers for Disease Control and Prevention. Pregnancy Risks Assessment Monitoring System (PRAMS). Phase 8 Core Questionnaire. Atlanta, GA: Centers for Disease Control; 2016. URL: https://www.cdc.gov/prams/pdf/questionnaire/Phase-8-Core-Questions-508.pdf [accessed 2020-03-01]
13. Heaman MI, Sword WA, Akhtar-Danesh N, Bradford A, Tough S, Janssen PA, et al. Quality of prenatal care questionnaire: instrument development and testing. BMC Pregnancy Childbirth 2014 Jun 03;14(1):188 [FREE Full text] [doi: 10.1186/1471-2393-14-188] [Medline: 24894497]

14. Lindgren K. Testing the Health Practices in Pregnancy Questionnaire-II. J Obstet Gynecol Neonatal Nurs 2005 Jul;34(4):465-472. [doi: 10.1177/0884217505276308] [Medline: 16020414]
15. McLarty C. Development of a questionnaire to assess knowledge in women with gestational diabetes. Western Australia: Edith Cowan University; 1993. URL: https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1638&context=theses_hons [accessed 2020-03-01]
16. Bland HW, Melton BF, Marshall ES, Nagle JA. Measuring exercise self-efficacy in pregnant women: psychometric properties of the pregnancy-exercise self-efficacy scale (P-ESES). J Nurs Meas 2013 Jan 01;21(3):349-359. [doi: 10.1891/1061-3749.21.3.349] [Medline: 24620510]
17. Ockenden H, Gunnell K, Giles A, Nerenberg K, Goldfield G, Manyanga T, et al. Development and Preliminary Validation of a Comprehensive Questionnaire to Assess Women's Knowledge and Perception of the Current Weight Gain Guidelines during Pregnancy. Int J Environ Res Public Health 2016 Nov 30;13(12):1187 [FREE Full text] [doi: 10.3390/ijerph13121187] [Medline: 27916921]
18. Centers for Disease Control and Prevention. Questionnaires: Breastfeeding and Infant Feeding Practices. 2017. Infant Feeding Practices Study II and Its Year Six Follow-Up URL: https://www.cdc.gov/breastfeeding/data/ifps/index.htm [accessed 2020-03-01]
19. Oladejo O, Allen K, Amin A, Frew PM, Bednarczyk RA, Omer SB. Comparative analysis of the Parent Attitudes about Childhood Vaccines (PACV) short scale and the five categories of vaccine acceptance identified by Gust et al. Vaccine 2016 Sep 22;34(41):4964-4968. [doi: 10.1016/j.vaccine.2016.08.046] [Medline: 27566902]
20. Winstanley A, Gattis M. The Baby Care Questionnaire: a measure of parenting principles and practices during infancy. Infant Behav Dev 2013 Dec;36(4):762-775 [FREE Full text] [doi: 10.1016/j.infbeh.2013.08.004] [Medline: 24050932]
21. 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]
22. Tamrat T, Kachnowski S. Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012 Jul 19;16(5):1092-1101. [doi: 10.1007/s10995-011-0836-3] [Medline: 21688111]
23. Bush J, Barlow DE, Echols J, Wilkerson J, Bellevin K. Impact of a Mobile Health Application on User Engagement and Pregnancy Outcomes Among Wyoming Medicaid Members. Telemed J E Health 2017 Nov;23(11):891-898 [FREE Full text] [doi: 10.1089/tmj.2016.0242] [Medline: 28481167]
24. Daly LM, Horey D, Middleton PF, Boyle FM, Flenady V. The Effect of Mobile App Interventions on Influencing Healthy Maternal Behavior and Improving Perinatal Health Outcomes: Systematic Review. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10012 [FREE Full text] [doi: 10.2196/10012] [Medline: 30093368]
25. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The Usability and Effectiveness of Mobile Health Technology-Based Lifestyle and Medical Intervention Apps Supporting Health Care During Pregnancy: Systematic Review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]

Abbreviations

aDiff: adjusted means difference
aOR: adjusted odds ratio
CORE: Center for Outcomes Research and Education
NICU: neonatal intensive care unit

Edited by G Eysenbach; submitted 24.11.19; peer-reviewed by E Neter, K Marko; comments to author 09.01.20; revised version received 04.03.20; accepted 21.03.20; published 06.07.20.

Please cite as:
Cawley C, Buckenmeyer H, Jellison T, Rinaldi JB, Vartanian KB
Effect of a Health System–Sponsored Mobile App on Perinatal Health Behaviors: Retrospective Cohort Study
JMIR Mhealth Uhealth 2020;8(7):e17183
URL: https://mhealth.jmir.org/2020/7/e17183
doi:10.2196/17183
PMID:32628123

Original Paper

The Effects of Gamification and Oral Self-Care on Oral Hygiene in Children: Systematic Search in App Stores and Evaluation of Apps

Nino Fijačko[1*], MSc; Lucija Gosak[1*], MSc; Leona Cilar[1*], MSc; Alenka Novšak[2*], DMD; Ruth Masterson Creber[3*], RN, MSc, PhD; Pavel Skok[4*], MD, PhD; Gregor Štiglic[1,5*], PhD

[1]Faculty of Health Sciences, University of Maribor, Maribor, Slovenia

[2]Dr. Novsak storitve d.o.o., Ljubljana, Slovenia

[3]Healthcare Policy and Research, Division of Health Informatics, Weill Cornell Medicine, New York, NY, United States

[4]Medical Faculty, University of Maribor, Maribor, Slovenia

[5]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[*]all authors contributed equally

Corresponding Author:
Nino Fijačko, MSc
Faculty of Health Sciences
University of Maribor
Žitna 15
Maribor
Slovenia
Phone: 386 23004764
Email: nino.fijacko@um.si

Abstract

Background: Poor oral hygiene is a great public health problem worldwide. Oral health care education is a public health priority as the maintenance of oral hygiene is integral to overall health. Maintaining optimal oral hygiene among children is challenging and can be supported by using relevant motivational approaches.

Objective: The primary aim of this study was to identify mobile smartphone apps that include gamification features focused on motivating children to learn, perform, and maintain optimal oral hygiene.

Methods: We searched six online app stores using four search terms (“oral hygiene game,” “oral hygiene gamification,” “oral hygiene brush game,” and “oral hygiene brush gamification”). We identified gamification features, identified whether apps were consistent with evidence-based dentistry, performed a quality appraisal with the Mobile App Rating Scale user version (uMARS), and quantified behavior scores (Behavior Change score, uMARS score, and Coventry, Aberdeen, and London-Refined [CALO-RE] score) using three different instruments that measure behavior change.

Results: Of 612 potentially relevant apps included in the analysis, 17 met the inclusion criteria. On average, apps included 6.87 (SD 4.18) out of 31 possible gamification features. The most frequently used gamification features were time pressure (16/17, 94%), virtual characters (14/17, 82%), and fantasy (13/17, 76%). The most common oral hygiene evidence-based recommendation was brushing time (2-3 minutes), which was identified in 94% (16/17) of apps. The overall mean uMARS score for app quality was high (4.30, SD 0.36), with good mean subjective quality (3.79, SD 0.71) and perceived impact (3.58, SD 0.44). Sufficient behavior change techniques based on three taxonomies were detected in each app.

Conclusions: The majority of the analyzed oral hygiene apps included gamification features and behavior change techniques to perform and maintain oral hygiene in children. Overall, the apps contained some educational content consistent with evidence-based dentistry and high-quality background for oral self-care in children; however, there is scope for improvement.

(JMIR Mhealth Uhealth 2020;8(7):e16365) doi:10.2196/16365

KEYWORDS

mobile health; oral health care; gamification; mobile store; evidence-based dentistry; behavior change technique; Mobile Application Rating Scale user version

Introduction

Oral diseases caused by oral hygiene are a major problem worldwide. In some countries, more than 80% of school children are affected [1-3]. Oral diseases can cause severe pain and loss of teeth, both of which affect appearance, dietary intake, and consequently the growth and development of children [4,5]. Poor oral hygiene is associated with poor quality of life and increased morbidity and mortality [1-3,6]. The maintenance of oral hygiene or oral self-care, which includes the use of toothbrushes, dental floss, and other interdental aids for healthy teeth, gums, oral soft tissues, palate, tongue, lips, and salivary glands, is important for quality of life, socialization, overall health, and well-being [7-9].

Brushing at least twice a day with toothpaste containing fluoride is considered an important aspect of the prevention and promotion of good brushing habits at an early age to prevent early childhood dental decay [10-12]. Oral health care providers play a key role in promoting oral health care among children [13]. Maintaining optimal oral hygiene can be challenging owing to different factors contributing to the lack of motivation and the need to initiate initial dental assessments between 6 and 12 months of age [14,15].

As children get older, there are new opportunities to use mobile health (mHealth) apps to support pediatric oral hygiene. Gamification is defined as the use of features to increase target behaviors and engagement. The purpose of gamification is to enable users to perform tasks more effectively while making them more enjoyable [16-18]. Gamification features, including badges, levels, and leader boards [17-19], offer novel approaches in dentistry [20]. Gamification has also been included in asthma apps [21] for similar reasons and has been incorporated into other app reviews for fitness apps [22,23], chronic disease management [24], smoking cessation [25], and health promotion [26]. The aim of our study was to identify apps that include gamification features and evidence-based dentistry (EBD) to support the maintenance of oral self-care in children.

Methods

Systematic Search Criteria and Selection

All apps were searched in June 2019 across six different smartphone app stores (Google Play Store [27], Apple App Store [28], Windows Phone Store [29], Amazon Appstore [30], BlackBerry World [31], and Samsung Galaxy Apps [32]). Our search strategy had two stages. The first evaluation stage was performed in each app store by using the same four search strings ("oral hygiene game," "oral hygiene gamification," "oral hygiene brush game," and "oral hygiene brush gamification"). The PICO criteria [33] were used to define the search criteria as follows: population (children below 13 years), intervention (free and paid apps that contain enlightenment gamification features and that allow users to be part of the interaction), comparison (app contents), and outcome (suitable apps for learning, performing, and maintaining oral health). Apps were excluded if they were duplicate (from multiple search strings and from multiple official web stores), were games, had a non-English interface, and were not related to the oral health care of children above 13 years. The second stage was performed using smartphones. In this stage, the relevant apps were downloaded and evaluated independently by two reviewers (NF and LG). The same reviewers evaluated the eligibility of the apps against preset criteria. Any issues were resolved by discussion among the other members of the study team. We included all free and paid apps that had two or more gamification features in English. Downloaded apps were analyzed on Samsung Galaxy S8 (Samsung, Seoul, South Korea) running Android 9.0 Pie (Google Inc, Mountain View, California, USA) for Android apps and iPhone 7 running iOS 12.3.1 (Apple Inc, Cupertino, California, USA) for iOS apps. If the same app was found for both Android and iOS, we reviewed the app available on Android, as the Google Play Store provides more information about each app than that available on the Apple App Store [34].

Measures and Rating Tools

All ratings and rankings were conducted individually and independently by authors (NF, LG, AN, and LC) with experience, knowledge, and skills in the field of mHealth and experience in different fields of health care (nursing, bioinformatics, and dentistry) to support the synthesis of the search results.

Rating Tool for Gamification Features

Two authors (LC and NF) identified gamification features by using the modified taxonomy of 31 gamification components (Multimedia Appendix 1) [19,35]. The scoring for gamification features was 1 point for the full implementation of gamification features, 0.5 points for partial implementation, and 0 points if the feature was not implemented.

Rating Tool for Oral Hygiene-Related Content Based on Dentistry Evidence

We evaluated EBD in each app using criteria and scored them according to the following four groups: (1) preparing to brush (two items); (2) before brushing (one item); (3) brushing (four items); and (4) after brushing (three items) (Multimedia Appendix 2) [36-40].

Mobile App Rating Scale User Version and Quantifying Behavior Change

Based on the Mobile App Rating Scale user version (uMARS) [41], we performed quality appraisal of apps. Originally, the MARS tool was developed with the aim that researchers could use it to determine whether apps satisfied certain quality criteria instead of relying on a subjective five-star rating system [41,42]. The uMARS is a simplified version of MARS [42], which allows multidimensional measurements (sections A-F) of performance indicators (section A), functionality (section B), esthetics (section C), quality of information (section D), subjective quality of the app (section E), and perceived impact (section F). The uMARS tool was developed to eliminate the need for trained professionals, and according to a study by Stoyanov et al, it has good internal consistency (α=.09), and high reliability [41]. All items were rated on a 5-point Likert scale from 1 (inadequate) to 5 (excellent) and were already used in a similar study [43]. Behavior change was also rated and ranked according to three taxonomies (Behavior Change score [44], uMARS score [41],

and Coventry, Aberdeen, and London-Refined [CALO-RE] score [45]) related to behavior change (Multimedia Appendix 3 and Multimedia Appendix 4). Similar to the rating tool for gamification features, the scoring system allowed 0, 0.5, or 1 point according to implementation. The exception was the uMARS score where items were scored based on a Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Data Analysis

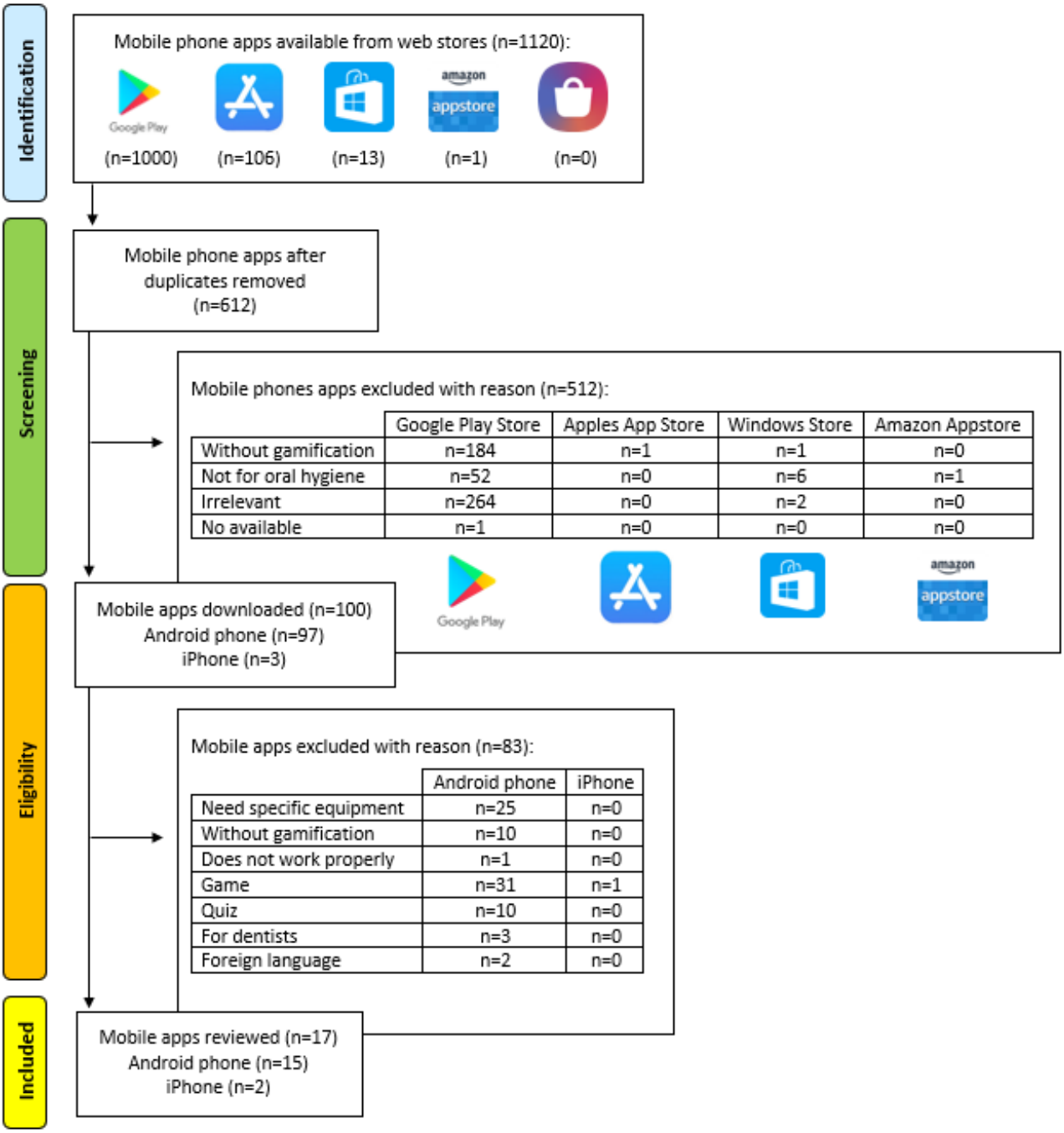
All data analyses and visualizations were prepared using Microsoft Office Professional 2016 (Microsoft Excel 2016; Microsoft Corp, Redmond, Washington, USA) and IBM SPSS Statistics version 25 for Windows (IBM Corp, Armonk, New York, USA). Intraclass correlation coefficients (ICCs) were calculated to provide results on consistency among observational ratings provided by multiple raters [46].

Results

Descriptive Characteristics

We identified 612 potentially relevant apps, and only 17 (3%) met the inclusion criteria. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [47] presents the process of scanning the apps using the inclusion and exclusion criteria. The two largest groups of apps that were excluded in the PRISMA diagram (Figure 1) were apps that required specific equipment or accessories [48] and apps that included gamification but did not include any interaction with the user [49].

Figure 1. PRISMA flow diagram of the selection of the included apps.



The vast majority of apps were free to download (11/17, 65%) and were available in both the Google Play Store and Apple App Store (14/17, 82%). Most apps did not require registration (14/17, 82%) and were updated in the last 2 years (13/17, 76%). Apps were developed in multiple countries, including the United States (6/17, 35%), Japan (2/17, 12%), Australia (1/17, 6%), Germany (1/17, 6%), England (1/17, 6%), Lithuania (1/17, 6%), North Korea (1/17, 6%), South Korea (1/17, 6%), Russia (1/17, 6%), Sweden (1/17, 6%), and Ukraine (1/17, 6%). Apps were developed by small and medium-size enterprises (14/17, 82%), health care–related agencies (2/17, 12%), and an individual (1/17, 6%).

Apps were targeted at different age groups of children. Among the 17 apps, 1 (6%) app was for children aged 5 years or less, 7 (41%) were for children aged 5 to 8 years (41%), 2 (12%) were for children aged 6 to 12 years, and 7 (41%) did not provide information about the target age of children but had a Pan European Game Information 3 (PEGI) certificate.

Table 1 presents all the relevant apps and their basic characteristics [50-80]. We assessed the subjective quality of the apps by answering the following question included in the uMARS framework: “What is your overall (star) rating of the app?” The quality of the apps was rated on a 5-point Likert scale from 1 (the worst app I have used) to 5 (the best app I have used). The mean app score was 3.79 (SD 0.69). The lowest score

was 3, and the highest score was 5. We compared the results with the ratings of users in the smartphone app stores. If the app ratings were available on both smartphone platforms (Android and iOS), we used the mean score. The mean user rating was 4.26 (SD 0.29) (Android: 4.21, SD 0.34; iOS: 4.33, SD 0.54), and the mean number of raters was 4981.79 (SD 12023.77) for Android apps and 359.62 (SD 934.17) for iOS apps. Based on a *t* test, there was a statistically significant difference between uMARS estimates (mean 3.85, SD 0.16) and user ratings (mean 4.26, SD 0.30; t_{32}=−2.27; *P*=.03).

Table 1. Description of the included apps.

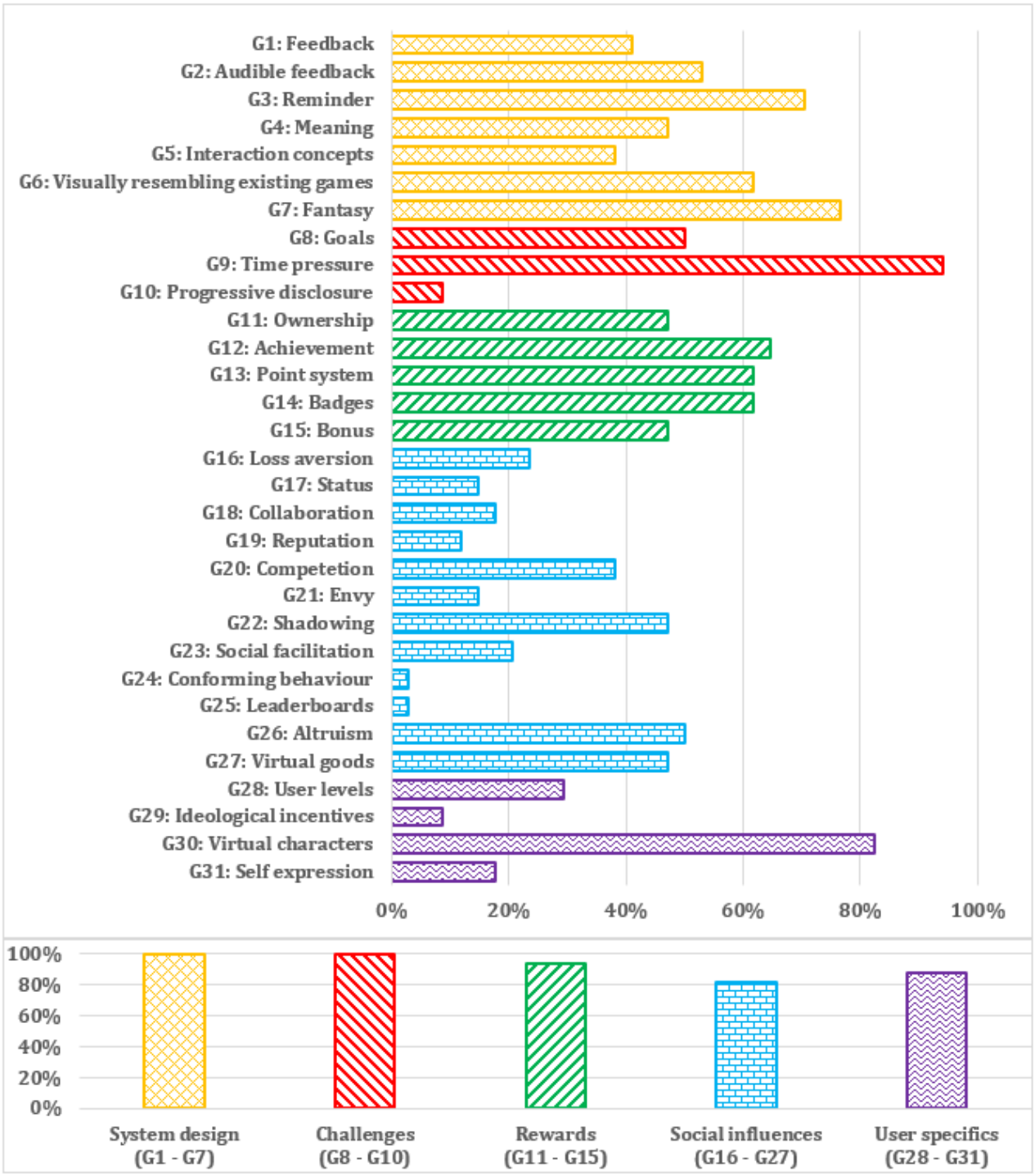
Full app name	Smartphone platform	Rating in the form of stars (number of raters)	
		Google Play Store	Apple App Store
Brush DJ [50,51]	Android & iOS	4.2 (1363)	4.7 (948)
Brush Hero [52,53]	Android & iOS	—[a]	4.2 (24)
Brushing Hero [54,55]	Android & iOS	4.5 (371)	4.2 (24)
Brush Monster [56,57]	Android & iOS	4.4 (199)	5.0 (2)
Brush My Teeth [58,59]	Android & iOS	4.6 (5)	—
Brush'n'Save [60,61]	Android & iOS	4.0 (87)	—
Brush Teeth with The Wiggles [62,63]	Android & iOS	4.3 (132)	4.8 (8)
Brush Up [64,65]	Android & iOS	3.8 (736)	4.1 (64)
Chomper Chums [66,67]	Android & iOS	4.1 (355)	5.0 (7)
Disney Magic Timer by Oral-B [68,69]	Android & iOS	4.0 (36,434)	3.9 (3348)
Mimizavr Clean Teeth [70,71]	Android & iOS	4.0 (14)	—
MyTeeth [72,73]	Android & iOS	5 (1)	3.0 (9)
My Virtual Tooth - Virtual Pet [74,75]	Android & iOS	4.3 (29,945)	4.3 (197)
Timo Kids Routine Timer [76,77]	Android & iOS	4.1 (100)	4.0 (1)
Tooth Hero [78]	Android	3.7 (3)	—
Toothsavers Brushing Game [79]	iOS	—	4.6 (8)
Toothy: Toothbrush Timer [80]	iOS	—	4.5 (34)

[a]Not available.

Gamification Features

On average, apps included 6.87 (SD 4.18) out of 31 possible gamification features. The most frequent gamification features were time pressure (16/17, 94%), virtual characters (14/17, 82%), and fantasy (13/17, 76%). The less frequently represented gamification features were conforming behavior and leaderboards (each 0.5/17, 3%) (Figure 2).

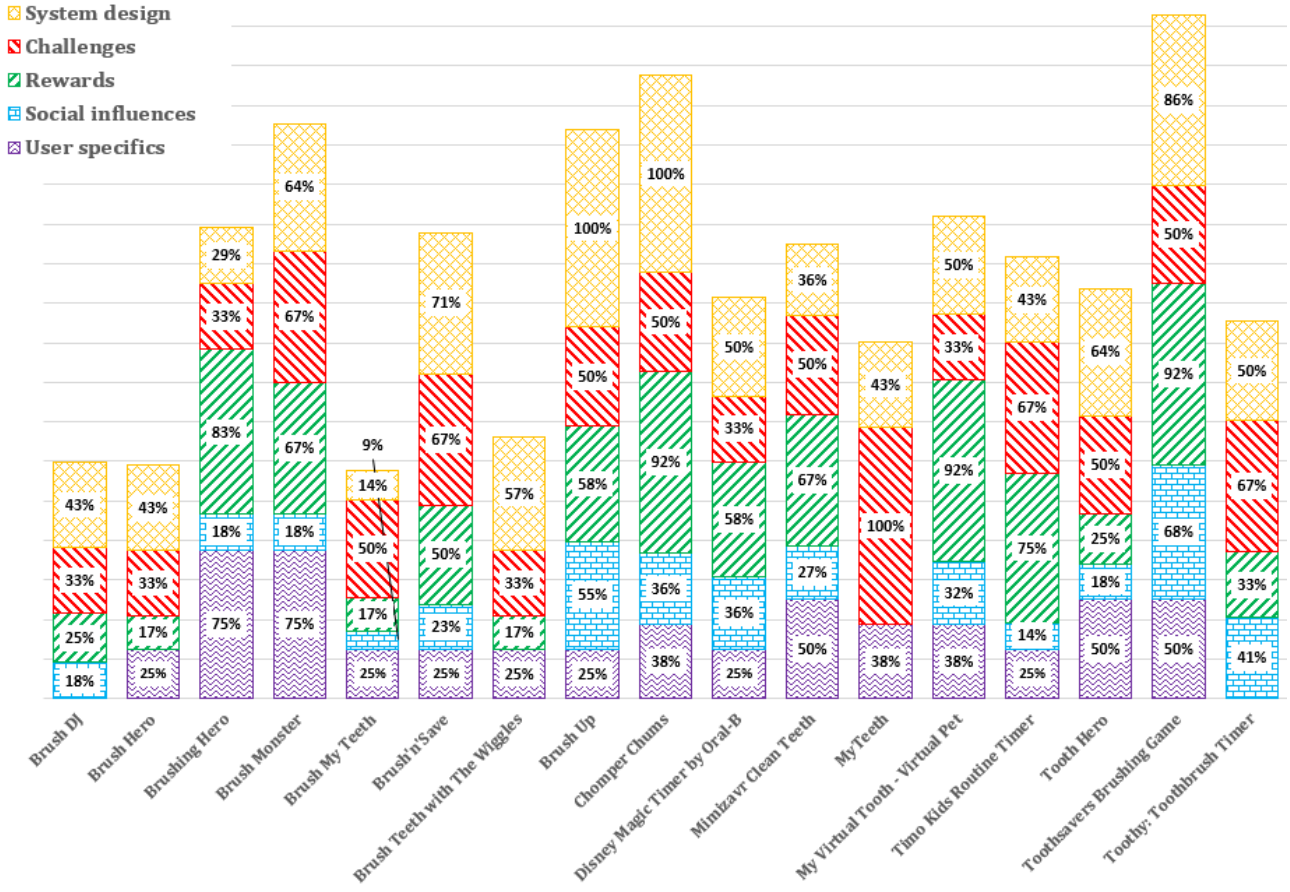
Figure 2. Gamification component categories of gamification features.



We identified all apps based on the gamification features (17/17, 100%) from the system design and challenges part of gamification component categories (Figure 2).

Apps in which we identified the most gamification component categories of gamification features were Toothsavers Brushing Game [79], Chomper Chums [66,67], Brush Monster [56,57], and Brush Up [64,65] (Figure 3).

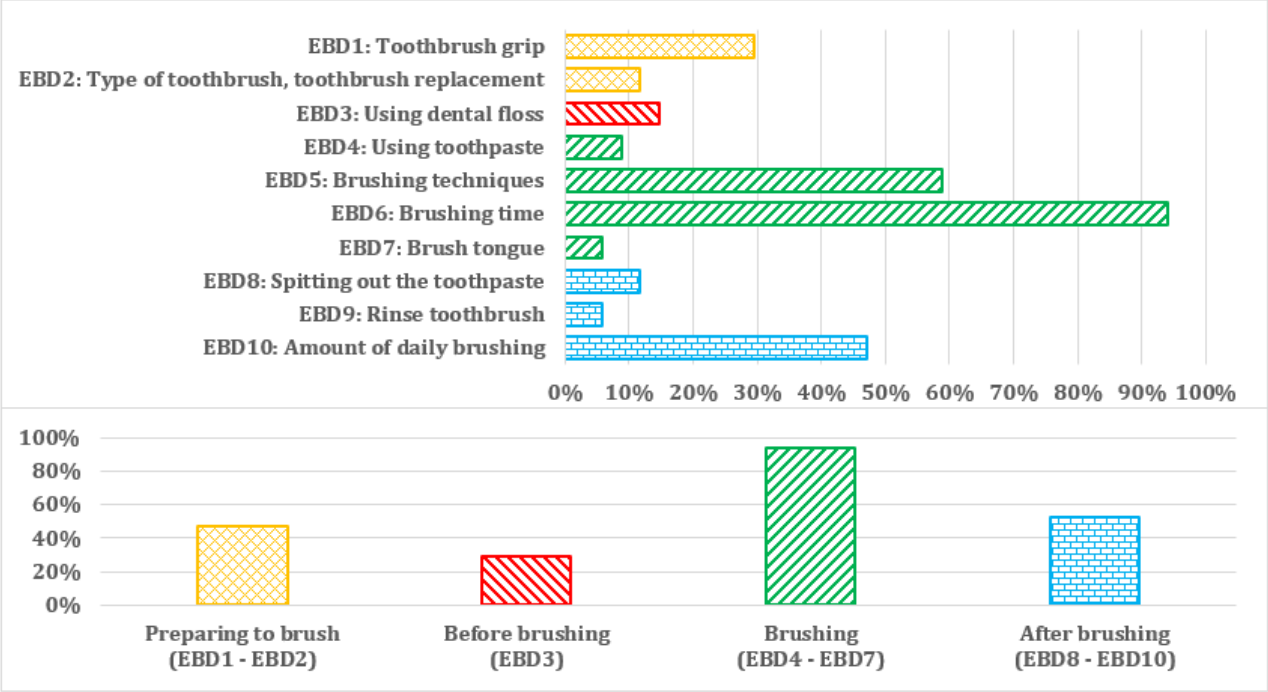
Figure 3. Percentage of gamification component categories for each app.



Oral Hygiene-Related Content Based on Dentistry Evidence

On average, we identified 4.9 (SD 4.98) oral hygiene-related contents in 10 groups based on the level of EBD (EDB1-EDB10). The most common EBD content focused on brushing time (Figure 4) from the EBD group category "Brushing." Most apps (14/17, 82%) had brushing time set to 2 minutes whereas the app Brushing Hero [54,55] had it set to 1 minute. Only two apps (MyTeeth [72,73] and Brush Up [64,65]) had it set to 3 minutes.

Figure 4. Group categories of evidence-based dentistry (EBD) oral hygiene-related content.



The most common brushing technique was the scrub brush technique (5/17, 29%). One app (MyTeeth [72,73]) allowed the user to choose among different brushing techniques (scrub brush technique; circular technique; Bass technique; and vertical technique from red [gums] to white [teeth], or inside, chewing surface, and outside or chewing surface, outside, and inside). Over one-third of the apps did not mention any EBD brushing technique (Figure 5).

On average, the apps included almost 3 out of 10 EBD contents (Figure 6). The MyTeeth app included the most EBD contents.

Figure 5. Brushing techniques used in the apps.

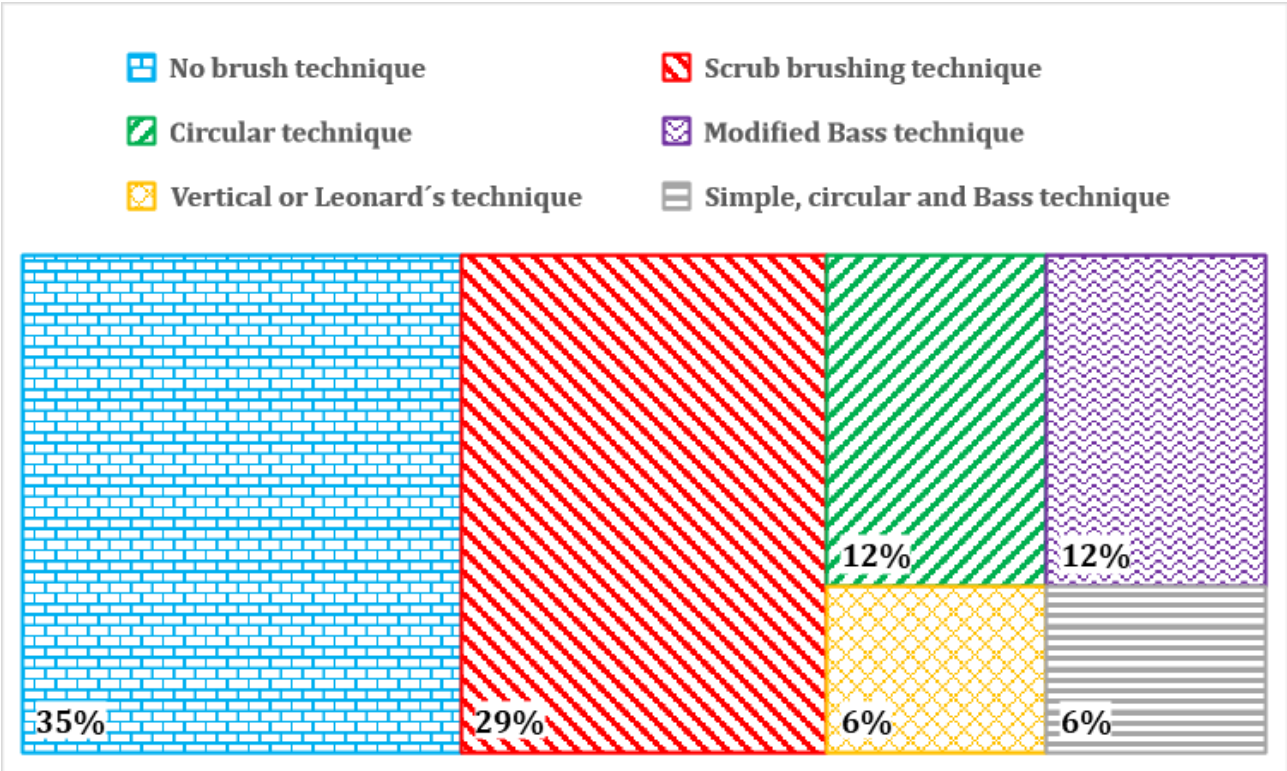
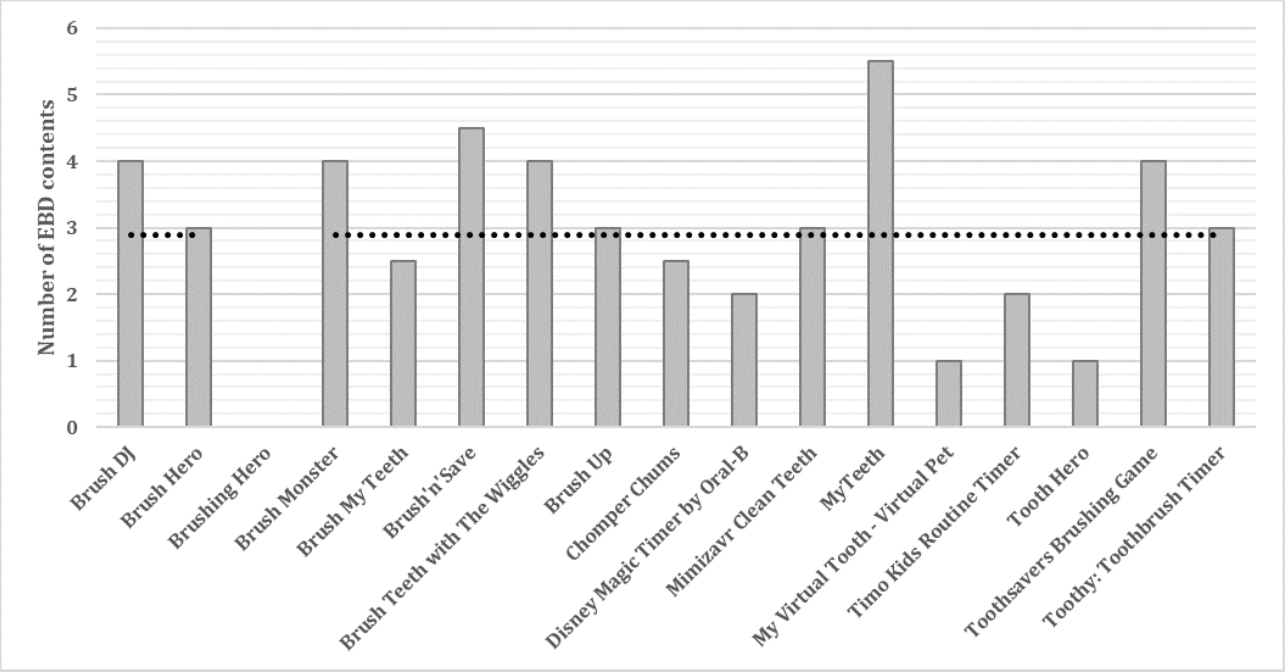


Figure 6. Number of evidence-based dentistry (EBD) contents in the apps.



uMARS App Quality Scores and Oral Hygiene Care Behavior Change Techniques

The total mean score for the section "app quality" was 4.30 (SD 0.36), which is considered a good result. The ICC among app reviewers was high for uMARS ratings (ICC=0.80; 95% CI 0.63-0.92). Toothsavers Brushing Game [79] had the highest score in each section of the uMARS (Table 2).

Table 2. App scores for each section of the Mobile App Rating Scale user version.

Full app name	uMARS[a] section						
	Engagement	Functionality	Esthetics	Information	App quality	App subjective quality	Perceived impact
Brush DJ	3.5	4.5	3.5	4.1	3.9	3.5	3.8
Brush Hero	3.9	3.1	4.7	__[b]	3.9	2.5	2.9
Brushing Hero	4.4	4.8	4.7	—	4.6	4.0	3.3
Brush Monster	5.0	4.8	5.0	4.4	4.8	4.9	4.0
Brush My Teeth	3.7	4.4	4.3	3.1	3.9	3.3	3.2
Brush'n'Save	3.8	5.0	2.5	4.3	3.8	3.9	3.6
Brush Teeth with The Wiggles	4.4	4.8	4.0	—	4.4	3.5	3.2
Brush Up	4.3	4.6	4.3	4.5	4.4	5.0	4.2
Chomper Chums	4.7	4.6	4.2	4.3	4.4	4.1	4.1
Disney Magic Timer by Oral-B	4.3	4.8	4.2	3.0	4.3	3.1	3.2
Mimizavr Clean Teeth	4.5	5.0	5.0	—	4.8	4.3	3.5
MyTeeth	3.2	4.6	4.0	—	3.9	3.3	3.7
My Virtual Tooth - Virtual Pet	4.4	4.6	4.0	—	4.3	3.6	3.1
Timo Kids Routine Timer	4.0	4.9	4.2	—	4.4	3.1	3.6
Tooth Hero	4.4	4.0	4.0	2.5	4.5	4.0	3.7
Toothsavers Brushing Game	4.9	4.9	4.8	4.8	4.9	5.0	4.5
Toothy: Toothbrush Timer	3.1	4.4	4.2	4.0	3.9	3.4	3.3
Mean score (SD)	4.15 (0.55)	4.58 (0.46)	4.21 (0.60)	3.90 (0.76)	4.30 (0.36)	3.79 (0.71)	3.58 (0.44)
ICC[c,d]	0.85	0.71	0.78	0.98	0.56	0.88	0.84

[a]uMARS: Mobile App Rating Scale user version.

[b]Not available; in 7 out of 17 (41%) apps the "information" quality section could not be assessed because data were not included.

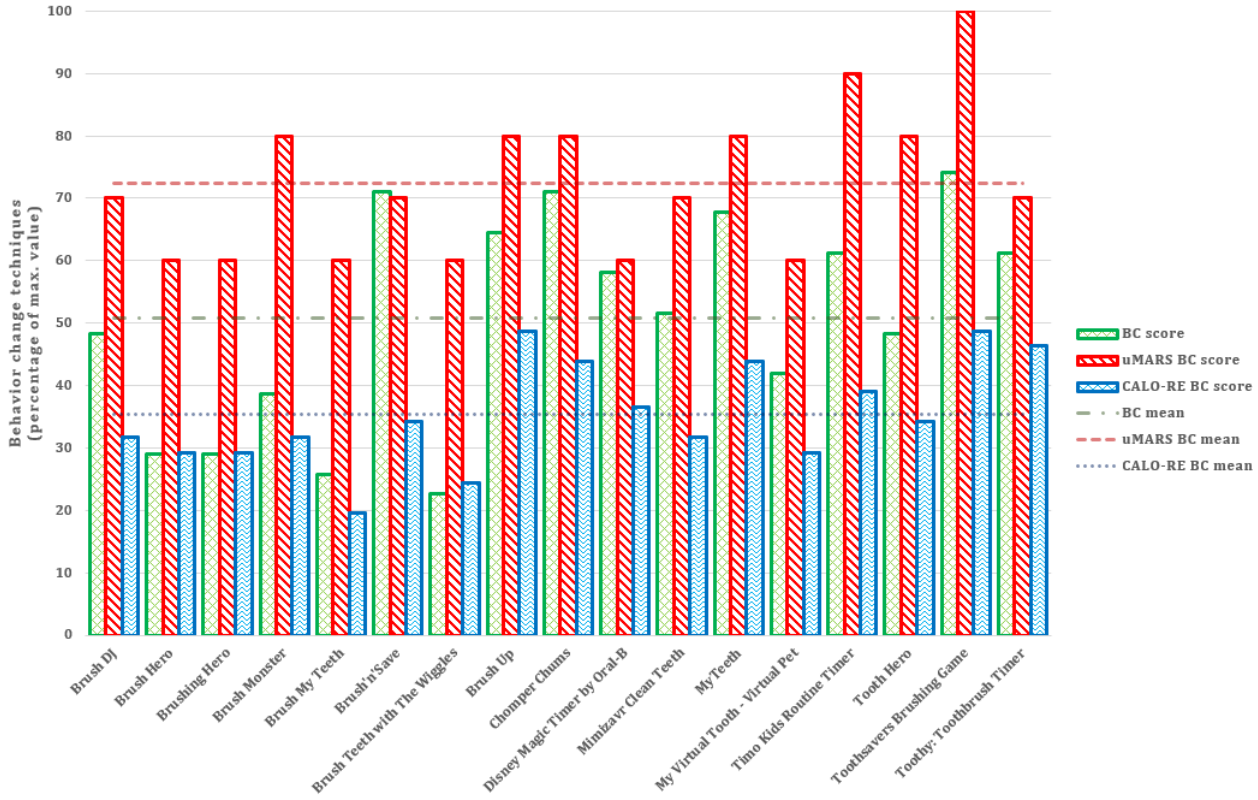
[c]ICC: intraclass correlation coefficient.

[d]A random effects average measures model with absolute agreement was used for calculated ICCs among two sets of ratings.

We also evaluated apps based on "subjective quality," where the mean estimate was good (mean 3.79, SD 0.71) and on "perceived impact," where the mean estimate was also good (mean 3.58, SD 0.44) (Table 2).

Results from three behavior change scores across three instruments indicated that Brush Up [64,65], Toothsavers Brushing Game [79], and Chomper Chums [66,67] included the most strategies targeting behavior change (Figure 7). The most common oral hygiene care behavior change techniques were "prompt intention formation" and "model or demonstrate the behavior."

Figure 7. Number of oral hygiene care behavior change techniques. BC: Behavior Change; CALO-RE: Coventry, Aberdeen, and London-Refined; uMARS: Mobile App Rating Scale user version.



Discussion

Gamification Features

The primary objective of this study was to identify apps that include gamification features focused on motivating children for behavior change to maintain oral self-care and optimal oral hygiene. The study results showed that most of the oral hygiene apps included a relevant number of gamification features and behavior change techniques to perform and maintain oral hygiene in children. Additionally, many apps contained EBD-supported educational content and achieved relatively high scores in app quality ratings. The most commonly identified gamification features were "time pressure" and "audible feedback" in the form of songs that help children to practice oral health care correctly. We compared our gamification results with a study conducted in a pediatric population [81] and found similar results with a focus on "feedback" and "goals" versus "information provision (general)" and "goal setting (behavior)." The authors also concluded that using apps for oral health care can improve motivation for oral self-care in children. Moreover, recent research found that knowledge provision increased motivation in children and that gamification might improve engagement with apps [49].

Apps Content Based on Dentistry Evidence

Our second aim was to determine what should be included in EBD health care apps. A small proportion of apps included EBD and the most frequently included EBD content was brushing time. Oral health care professionals generally recommend at least 2 or 3 minutes for brushing teeth. A study from 2019 showed increased duration of toothbrushing when using the Brush Up app [81].

The brushing technique is important for optimal oral self-care. It is recommended that children use a modified Bass technique because of efficient plaque removal compared with other toothbrushing techniques [82,83]. The scrub brushing technique was the most frequently suggested in the reviewed apps. The scrub brushing technique is the easiest to program into an app and the most commonly used technique in the general population [82,84].

Children should perform oral hygiene activities more than two times a day, and yet, less than half of the apps included proper EBD warnings or reminders for users to brush their teeth at least twice a day. We also found that only a few apps emphasized brushing with a proper toothbrush and toothpaste quantity, as supported by the literature to reduce the incidence of decay [82,85]. The recommended "pea-sized amount," which represents approximately 1000 parts per million (ppm) fluoride, is generally recommended, yet only two apps visualized the correct quantity of toothpaste.

Another important aspect of oral hygiene is the use of a soft or extra-soft toothbrush for safe oral health care [86]. Older brushes lose their plaque removal ability [87,88] and are often contaminated with microbes [89]. Guidelines recommend replacing toothbrushes when they show signs of bristle splaying tear. In addition to oral self-care, less than 20% of apps included the use of dental floss or other interdental brushing approaches. None of the apps promoted the EBD recommendation to first floss and then brush [90,91].

Although the usage of apps to monitor and manage health is increasing, it is difficult to find information about the sources used in the development of the apps [92]. In addition, many app developers focus on the usefulness and ease of use of the app instead of the quality of the content included in the app.

App Quality and Behavior Change Techniques

Our results showed that average users tended to give more positive feedback to apps based on functionality [93] and popularity [94], and there was less focus on usability or quality, which can potentially represent dangerous health information for future users. Therefore, researchers often rely on more complex evaluation tools like the uMARS [95] to obtain a better estimation of app quality. Overall, the Toothsavers Brushing Game [79] had the highest quality score according to the uMARS. Additionally, Toothsavers Brushing Game [79] had the highest scores in all categories of gamification components, EBD oral hygiene-related content, and all three behavior change technique taxonomies, which is consistent with our results. Brush Monster [56,57] and Chomper Chums [66,67] were also visually well designed, included EBD content, and did not have any specific technical problems during their use.

One of the most important behaviors in oral hygiene care change techniques is user interaction with the app [96]. Our results showed that one app included learning of oral self-care with virtual model teeth and enabled the user to perform self-based oral health care interaction in the app. In the apps Brushing Hero [54,55] and Brush Monster [57,58], augmented reality is added to the self-based oral health care interaction, where the user is a role model. Other apps use one or more different teaching approaches for oral self-care in the form of virtual teeth or a virtual tooth, gradually revealing a picture, virtual model teeth, video, and games.

We also must consider the risk of using apps as an intervention for oral hygiene care in children, as the youngest users seem to be most highly affected by and at risk for behavioral addictions [97]. The pattern of smartphone abuse is greatest among young people, and problematic use is linked to self-esteem, impulsivity, self-identity, and self-image. Problematic use is also associated with sleep disturbance, anxiety, stress, and, to a lesser extent, depression [98], as well as impaired psychological well-being, impaired parent and school relationships, and additional behavioral problems [99]. As interventions using apps or games for increasing oral hygiene care quality can be successful, judicious use is advised. Our observation is that using apps for learning, performing, and maintaining oral self-care requires less than 10 minutes every day, and this is much lower than the average time spent on a smartphone.

Observation of Apps and Tips for Future Oral Health Care App Developers

Our study provides app reviewers and developers of oral self-care with structured information on how future apps should be implemented and which EBD content should be included. Apps for supporting oral self-care in oral hygiene should have a combination of gamification features and EBD. Collaboration with oral health care professionals (eg, oral health care organizations and dentists) can improve the content provided in the app and bring EBD self-care information to targeted end users.

We found the following two secondary app payment strategies: “freemium” (eg, in-app purchases) and “free trial” (eg, free to premium version). Three apps had a trial version and required payment over time (Toothy: Toothbrush Timer [80], €1.32/month [US $1.42/month]; Brush Up [64,65], €1.11/month [US $1.20/month] or buy €11.14 [US $12.02]; My Virtual Tooth-Virtual Pet [74,75], €5.49 [US $5.92]). In-app purchases in different forms were offered in five apps, with a minimum price of €0.99 (US $1.07) and maximum price of €11.14 (US $12.02) per item. To increase usage, many mobile game companies provide ‘‘freemium” services, which cost the player nothing for basic usage but need payment with real money for advanced functionality or virtual goods [100]. Liu et al found that the freemium strategy is positively associated with increased sales of paid apps [101]. The most popular in-app purchases were in the form of getting virtual goods (ie, coins), disabling in-app advertisements, getting virtual characters (eg, avatars), and achieving virtual goals (eg, levels). Only three apps in the smartphone app stores provided information that the apps contain advertisements and offer in-app purchases. Affective states, such as playfulness, seem to be one of the most influential factors positively affecting consumers’ intention to pay for mobile game services [102]. Additionally, “stickiness,” that is, the trait of a game that engages users, is found to greatly and positively influence in-app purchases [103].

Positive effective states associated with in-game rewards seem to help the user to play the game better and can enhance both perceived value and customer loyalty [104]. Many smartphone games that provide “freemium“ services are intended to be used by children. Hence, it seems that smartphone game service providers try to use the affective states of children to increase their revenue. This is especially concerning since the preadolescent (and adolescent) brain is still very much developing, especially in terms of reasoning and anticipating what will happen in the future [105]. Only one app (Disney Magic Timer by Oral-B [68,69]) alerted users that real money is needed to make in-app purchases; therefore, parental supervision is recommended. Google Play Store and Apple App Store provide users the opportunity to get a refund of the money paid but within a limited period.

Limitations

As with any review of commercially available tools in app stores, there is time sensitivity of the results because apps are being added and removed from app stores daily. Given that caveat, this review is meant to provide an overview of the state of apps available in this field, with less emphasis on individual apps. Another limitation of this study is that it focused only on features that were freely available in the apps. As such, we may have missed gamification features, EBD contents, and other features in the apps. We also excluded apps that required subscription over time. The last limitation in this study is that we did not make any in-app purchases in paid apps, so we could not present additional results for in-app purchases.

Future Research

Future work should evaluate one of the highest scoring apps (eg, Toothsavers Brushing Game and Chomper Chums) in a randomized controlled trial to evaluate oral hygiene outcomes and motivation for learning, performing, and maintaining oral self-care among children.

Conclusions

In this systematic review, we did not find any app that had all the segments for learning, performing, and maintaining proper oral health care. The Toothsavers Brushing Game app had the highest scores, and if future updates introduce more EBD content, it could be the most appropriate app for learning, performing, and maintaining good oral health care in children. Gamification features with EBD have good potential as new approaches for health care providers to change behavior in the form of learning, performing, and maintaining proper oral hygiene with EBD in the clinical environment.

Acknowledgments

This study was partially supported by the “knowledge through creative pathways 2016-2020” scheme cofunded by the European Union from the European Social Fund and the Republic of Slovenia and the Slovenian Research Agency (grant numbers N2-0101 and P2-0057). In addition, Dr. Masterson Creber is supported by NIH/National Institute of Nursing Research (R00NR016275).

Authors' Contributions

The systematic review presented here was carried out through collaboration among all authors. NF developed a study design and supervised the study. NF, LC, and RMC drafted the manuscript. NF, LG, GS, and PS conducted data collection and analysis. AN interpreted results from the dental hygiene point of view. RMC conducted a comprehensive content review. All authors read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Exemplary gamification features rating criteria for oral hygiene apps.
[DOCX File , 45 KB - mhealth_v8i7e16365_app1.docx]

Multimedia Appendix 2
Oral self-care app segments and evidence-based dentistry.
[DOCX File , 44 KB - mhealth_v8i7e16365_app2.docx]

Multimedia Appendix 3
Exemplary oral health care behavior change techniques according to the Behavior Change score.
[DOCX File , 45 KB - mhealth_v8i7e16365_app3.docx]

Multimedia Appendix 4
Exemplary oral health care behavior change techniques according to the Coventry, Aberdeen, and London-Refined behavior change score.
[DOCX File , 47 KB - mhealth_v8i7e16365_app4.docx]

References

1. Kwan SY, Petersen PE, Pine CM, Borutta A. Health-promoting schools: an opportunity for oral health promotion. Bull World Health Organ 2005 Sep;83(9):677-685 [FREE Full text] [doi: 10.1590/S0042-96862005000900013] [Medline: 16211159]
2. Parmar P, Radha G, Rekha R, Pallavi S, Nagashree S. Promoting oral hygiene and health through school. Int J Oral Health Sci 2016;6(2):70-77. [doi: 10.4103/2231-6027.199989]
3. Azodo CC, Unamatokpa B. Gender difference in oral health perception and practices among Medical House Officers. RusOMJ 2012 Sep;1(2):0208. [doi: 10.15275/rusomj.2012.0208]
4. Petersen P. The World Oral Health Report 2003: continuous improvement of oral health in the 21st century--the approach of the WHO Global Oral Health Programme. Community Dent Oral Epidemiol 2003 Dec;31 Suppl 1:3-23 [FREE Full text] [doi: 10.1046/j..2003.com122.x] [Medline: 15015736]
5. Nunn JH. The burden of oral ill health for children. Arch Dis Child 2006 Mar;91(3):251-253 [FREE Full text] [doi: 10.1136/adc.2005.077016] [Medline: 16492889]

6. Rabiei S, Mohebbi SZ, Patja K, Virtanen JI. Physicians' knowledge of and adherence to improving oral health. BMC Public Health 2012 Oct 09;12:855 [FREE Full text] [doi: 10.1186/1471-2458-12-855] [Medline: 23046660]
7. Bashiru BO, Anthony IN. Oral self-care practices among university students in Port Harcourt, Rivers State. Niger Med J 2014 Nov;55(6):486-489 [FREE Full text] [doi: 10.4103/0300-1652.144703] [Medline: 25538367]
8. Naseem S, Fatima SH, Ghazanfar H, Haq S, Khan NA, Mehmood M, et al. Oral Hygiene Practices and Teeth Cleaning Techniques Among Medical Students. Cureus 2017 Jul 18;9(7):e1487 [FREE Full text] [doi: 10.7759/cureus.1487] [Medline: 28944126]
9. Jabeen C, Umbreen G. Oral Hygiene: Knowledge, Attitude and Practice among School Children, Lahore. J Liaquat Uni Med Health Sci 2017 Sep 30;16(03):176-180. [doi: 10.22442/jlumhs.171630529]
10. Wilson M, Morgan M. A survey of children's toothbrushing habits in Wales: Are parents following the guidelines? Journal of Health Visiting 2016 Feb;4(2):95-99. [doi: 10.12968/johv.2016.4.2.95]
11. Melo P, Fine C, Malone S, Frencken JE, Horn V. The effectiveness of the Brush Day and Night programme in improving children's toothbrushing knowledge and behaviour. Int Dent J 2018 May;68 Suppl 1:7-16. [doi: 10.1111/idj.12410] [Medline: 29660791]
12. Pollick H. The Role of Fluoride in the Prevention of Tooth Decay. Pediatr Clin North Am 2018 Oct;65(5):923-940. [doi: 10.1016/j.pcl.2018.05.014] [Medline: 30213354]
13. Soltani R, Ali EA, Mahaki B, Alipoor M, Sharifirad G. Do maternal oral health-related self-efficacy and knowledge influence oral hygiene behavior of their children? International Journal of Pediatrics 2016;4(7):2035-2042. [doi: 10.22038/ijp.2016.6997]
14. Acharya S, Goyal A, Utreja AK, Mohanty U. Effect of three different motivational techniques on oral hygiene and gingival health of patients undergoing multibracketed orthodontics. Angle Orthod 2011 Sep;81(5):884-888. [doi: 10.2319/112210-680.1] [Medline: 21612316]
15. Meera R, Muthu MS, Phanibabu M, Rathnaprabhu V. First dental visit of a child. J Indian Soc Pedod Prev Dent 2008;26 Suppl 2:S68-S71 [FREE Full text] [Medline: 19075451]
16. Khaleel FL, Ashaari NS, Wook TS, Ismail A. User-enjoyable learning environment based on Gamification elements. 2015 Presented at: 2015 International Conference on Computer, Communications, and Control Technology (I4CT); April 21-23, 2015; Kuching, Malaysia URL: https://www.researchgate.net/publication/304627718_User-enjoyable_learning_environment [doi: 10.1109/I4CT.2015.7219570]
17. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments 2011 Sep 29 [FREE Full text] [doi: 10.1145/2181037.2181040]
18. Subhash S, Cudney EA. Gamified learning in higher education: A systematic review of the literature. Computers in Human Behavior 2018 Oct;87:192-206. [doi: 10.1016/j.chb.2018.05.028]
19. Scott T, Sebastian L, Dirk B. Gamifying information systems-a synthesis of gamification mechanics and dynamics. 2014 Jun 9 Presented at: European Conference on Information Systems (ECIS) 2014; June 9-11, 2014; Tel Aviv, Israel p. 1-17 URL: https://pdfs.semanticscholar.org/a02f/7cc91d26671e591e27b241ed6b17debc2d89.pdf
20. Alaa HQ. Gamification in dentistry: A new era and direction in dental education, awareness and business. 2014 Apr 21 Presented at: 2nd International Conference and Exhibition on Dental & Oral Health; April 21-23, 2014; Crown Plaza Dubai, UAE p. 194-194 URL: https://www.longdom.org/proceedings/gamification-in-dentistry-a-new-era-and-direction-in-dental-education-awareness-and-business-24562.html
21. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The Potential of Mobile Apps for Improving Asthma Self-Management: A Review of Publicly Available and Well-Adopted Asthma Apps. JMIR Mhealth Uhealth 2017 Aug;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
22. Cotton V, Patel MS. Gamification Use and Design in Popular Health and Fitness Mobile Applications. Am J Health Promot 2019 Mar;33(3):448-451 [FREE Full text] [doi: 10.1177/0890117118790394] [Medline: 30049225]
23. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
24. Miller AS, Cafazzo JA, Seto E. A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2016 Jun;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]
25. El-Hilly AA, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, et al. Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study. JMIR Serious Games 2016 Oct 24;4(2):e18 [FREE Full text] [doi: 10.2196/games.5678] [Medline: 27777216]
26. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
27. Google Play. URL: https://play.google.com/store/apps?hl=en_USA [accessed 2019-09-11]
28. App Store. URL: https://www.apple.com/uk/ios/app-store/ [accessed 2019-09-11]
29. Windows Phone Store. URL: https://www.microsoft.com/sl-si/store/apps/windows-phone [accessed 2019-09-11]
30. Amazon Appstore. URL: https://www.amazon.com/mobile-apps/b?ie=UTF8&node=2350149011
31. App World. BlackBerry World URL: https://appworld.blackberry.com/ [accessed 2019-09-11]

32. Galaxy Store. URL: https://www.samsung.com/global/galaxy/apps/galaxy-store/ [accessed 2019-09-11]
33. Stone PW. Popping the (PICO) question in research and evidence-based practice. Appl Nurs Res 2002 Aug;15(3):197-198. [doi: 10.1053/apnr.2002.34181] [Medline: 12173172]
34. Cruz Zapata B, Hernández Niñirola A, Idri A, Fernández-Alemán JL, Toval A. Mobile PHRs compliance with Android and iOS usability guidelines. J Med Syst 2014 Aug 24;38(8):81. [doi: 10.1007/s10916-014-0081-6] [Medline: 24957397]
35. Tiffany B, Blasi P, Catz SL, McClure JB. Mobile Apps for Oral Health Promotion: Content Review and Heuristic Usability Analysis. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e11432 [FREE Full text] [doi: 10.2196/11432] [Medline: 30181114]
36. Pujar P, Subbareddy VV. Evaluation of the tooth brushing skills in children aged 6-12 years. Eur Arch Paediatr Dent 2013 Aug 5;14(4):213-219. [doi: 10.1007/s40368-013-0046-3] [Medline: 23737009]
37. Muller-Bolla M, Courson F. Toothbrushing methods to use in children: a systematic review. Oral Health Prev Dent 2013;11(4):341-347. [doi: 10.3290/j.ohpd.a30602] [Medline: 24046823]
38. Matsui M, Chosa N, Shimoyama Y, Minami K, Kimura S, Kishi M. Effects of tongue cleaning on bacterial flora in tongue coating and dental plaque: a crossover study. BMC Oral Health 2014 Jan 14;14:4 [FREE Full text] [doi: 10.1186/1472-6831-14-4] [Medline: 24423407]
39. Pitts N, Duckworth RM, Marsh P, Mutti B, Parnell C, Zero D. Post-brushing rinsing for the control of dental caries: exploration of the available evidence to establish what advice we should give our patients. Br Dent J 2012 Apr 13;212(7):315-320. [doi: 10.1038/sj.bdj.2012.260] [Medline: 22498529]
40. ADA Division of Communications. For the dental patient. Toothbrush care, cleaning and replacement. J Am Dent Assoc 2006 Mar;137(3):415. [doi: 10.14219/jada.archive.2006.0193] [Medline: 16570477]
41. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
42. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
43. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
44. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
45. 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. Psychology & health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
46. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [Medline: 22833776]
47. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
48. Fijacko N, Stiglic G, Kocbek P, Cilar L, Gosak L. Learning oral hygiene with gamification and smart accessories: a systematic review of mobile phone apps. In: Book of Abstracts International Scientific Conference. 2019 Presented at: International Scientific Conference; June 13, 2019; Maribor, Slovenia. [doi: 10.18690/978-961-286-276-3]
49. Parker K, Bharmal RV, Sharif MO. The availability and characteristics of patient-focused oral hygiene apps. Br Dent J 2019 Apr;226(8):600-604. [doi: 10.1038/s41415-019-0197-7] [Medline: 31028329]
50. Google Play Store. Brush DJ URL: https://play.google.com/store/apps/details?id=uk.co.appware.brushdj&hl=sl [accessed 2020-02-13]
51. Apple App Store. Brush DJ URL: https://apps.apple.com/us/app/brush-dj/id475739913 [accessed 2020-02-13]
52. Google Play Store. Brush Hero URL: https://play.google.com/store/apps/details?id=com.brushheroapp.brushhero [accessed 2020-02-13]
53. Apple App Store. Brush Hero URL: https://apps.apple.com/us/app/brush-hero/id1380844743 [accessed 2020-02-13]
54. Google Play Store. Brushing Hero URL: https://play.google.com/store/apps/details?id=jp.co.litalico.brushinghero [accessed 2020-02-13]
55. Apple App Store. Brushing Hero URL: https://apps.apple.com/us/app/brushing-hero/id1317198419 [accessed 2020-02-13]
56. Google Play Store. Brush Monster - AR Toothbrushing Guide App URL: https://play.google.com/store/apps/details?id=com.kittenplanet.brushmon&hl=sl [accessed 2020-02-13]
57. Apple App Store. Brush Monster - AR Toothbrush URL: https://apps.apple.com/au/app/brush-monster-artoothbrush/id1320759633 [accessed 2020-02-13]
58. Google Play Store. Brush my teeth URL: https://play.google.com/store/apps/details?id=se.skuggan.brush.en&hl=sl [accessed 2020-02-13]
59. Apple App Store. Brush My Teeth URL: https://apps.apple.com/au/app/brush-my-teeth/id974298516 [accessed 2020-02-13]

60. Google Play Store. Brush'n'Save URL: https://play.google.com/store/apps/details?id=com.amanodental.www.HCFM02&hl=sl [accessed 2020-02-13]
61. Apple App Store. Brush'n'Save URL: https://apps.apple.com/us/app/brushnsave/id916125398 [accessed 2020-02-13]
62. Google Play Store. Brush Teeth with The Wiggles URL: https://play.google.com/store/apps/details?id=au.com.australianunity.wigglesbrush&hl=sl [accessed 2020-02-13]
63. Apple App Store. Brush Teeth with The Wiggles URL: https://apps.apple.com/us/app/brush-teeth-with-the-wiggles/id1113805211 [accessed 2020-02-13]
64. Google Play Store. Brush Up URL: https://play.google.com/store/apps/details?id=com.gtw.BrushUp&hl=sl [accessed 2020-02-13]
65. Apple App Store. Brush Up: Toothbrush Trainer URL: https://apps.apple.com/us/app/brush-up-toothbrush-trainer/id967009000 [accessed 2020-02-13]
66. Google Play Store. Chomper Chums URL: https://play.google.com/store/apps/details?id=com.ucci.dtt&hl=sl [accessed 2020-02-13]
67. Apple App Store. Chomper Chums URL: https://apps.apple.com/us/app/chomper-chums/id579421399 [accessed 2020-02-13]
68. Google Play Store. Disney Magic Timer by Oral-B URL: https://play.google.com/store/apps/details?id=com.disneydigitalbooks.DisneyMagicBrushTimer_goo&hl=sl [accessed 2020-02-13]
69. Apple App Store. Disney Magic Timer by Oral-B URL: https://apps.apple.com/us/app/disney-magic-timer-by-oral-b/id747541884 [accessed 2020-02-13]
70. Google Play Store. Mimizavr clean teeth URL: https://play.google.com/store/apps/details?id=ru.mimizavr&hl=sl [accessed 2020-02-13]
71. Apple App Store. Mimizavr Brush Teeth URL: https://apps.apple.com/us/app/mimizavr-brush-teeth/id1242740400 [accessed 2020-02-13]
72. Google Play Store. MyTeeth URL: https://play.google.com/store/apps/details?id=com.thekeptpromise.myteeth&hl=sl [accessed 2020-02-13]
73. Apple App Store. MyTeeth URL: https://apps.apple.com/us/app/myteeth/id715668236 [accessed 2020-02-13]
74. Google Play Store. My Virtual Tooth - Virtual Pet URL: https://play.google.com/store/apps/details?id=com.mytooth.virtualpet&hl=sl [accessed 2020-02-13]
75. Apple App Store. My Virtual Tooth - Virtual Pet URL: https://apps.apple.com/us/app/my-virtual-tooth-virtual-pet/id1129665919 [accessed 2020-02-13]
76. Google Play Store. TIMO Kids Visual Timer - Game for Daily Routine URL: https://play.google.com/store/apps/details?id=com.timo.time&hl=sl [accessed 2020-02-13]
77. Apple App Store. TIMO Kids Visual Timer URL: https://apps.apple.com/us/app/timo-kids-routine-timer/id1417830395 [accessed 2020-02-13]
78. APKPure. Tooth Hero - Brushing Timer Game, Stop Cavities! URL: https://apkpure.com/tooth-hero-brushing-timer-game-stop-cavities/com.newrealmtech.toothhero [accessed 2020-02-13]
79. Apple App Store. Toothsavers Brushing Game URL: https://apps.apple.com/us/app/toothsavers-brushing-game/id734501468 [accessed 2020-02-13]
80. Apple App Store. Toothy: Tooth Brushing Timer URL: https://apps.apple.com/us/app/toothy-tooth-brushing-timer/id1453139747 [accessed 2020-02-13]
81. Jacobson D, Jacobson J, Leong T, Lourenco S, Mancl L, Chi DL. Evaluating Child Toothbrushing Behavior Changes Associated with a Mobile Game App: A Single Arm PrePost Pilot Study. Pediatr Dent 2019 Jul 15;41(4):299-303 [FREE Full text] [Medline: 31439090]
82. Wainwright J, Sheiham A. An analysis of methods of toothbrushing recommended by dental associations, toothpaste and toothbrush companies and in dental texts. Br Dent J 2014 Aug;217(3):E5. [doi: 10.1038/sj.bdj.2014.651] [Medline: 25104719]
83. Janakiram C, Taha F, Joe J. The Efficacy of Plaque Control by Various Toothbrushing Techniques-A Systematic Review and Meta-Analysis. Journal of Clinical & Diagnostic Research 2018;12(11). [doi: 10.7860/JCDR/2018/32186.12204]
84. Baruah K, Thumpala VK, Khetani P, Baruah Q, Tiwari RV, Dixit H. A review on toothbrushes and tooth brushing methods. Int J Pharm Sci Invention 2017 May;6(5):29-38.
85. Valkenburg C, Van der Weijden F, Slot DE. Is plaque regrowth inhibited by dentifrice?: A systematic review and meta-analysis with trial sequential analysis. Int J Dent Hyg 2019 Feb;17(1):27-38. [doi: 10.1111/idh.12364] [Medline: 30169912]
86. Ranzan N, Muniz FW, Rösing CK. Are bristle stiffness and bristle end-shape related to adverse effects on soft tissues during toothbrushing? A systematic review. Int Dent J 2019 Jun;69(3):171-182. [doi: 10.1111/idj.12421] [Medline: 30152076]
87. Van Leeuwen MP, Van der Weijden FA, Slot DE, Rosema MA. Toothbrush wear in relation to toothbrushing effectiveness. Int J Dent Hyg 2019 Feb;17(1):77-84. [doi: 10.1111/idh.12370] [Medline: 30326176]
88. Gundavarapu KC, Ramachandra SS, Dicksit DD. An investigation into toothbrush wear related to months of use among university students. Can J Dent Hyg 2015 Sep 13;49(2):76-80 [FREE Full text]
89. Karibasappa GN, Nagesh L, Sujatha BK. Assessment of microbial contamination of toothbrush head: an in vitro study. Indian J Dent Res 2011;22(1):2-5 [FREE Full text] [doi: 10.4103/0970-9290.79965] [Medline: 21525668]

90. Torkzaban P, Arabi SR, Sabounchi SS, Roshanaei G. The Efficacy of Brushing and Flossing Sequence on Control of Plaque and Gingival Inflammation. Oral Health Prev Dent 2015;13(3):267-273. [doi: 10.3290/j.ohpd.a32678] [Medline: 25197738]
91. Mazhari F, Boskabady M, Moeintaghavi A, Habibi A. The effect of toothbrushing and flossing sequence on interdental plaque reduction and fluoride retention: A randomized controlled clinical trial. J Periodontol 2018;89(7):824-832. [doi: 10.1002/jper.17-0149]
92. Armstrong S. Which app should I use? BMJ 2015 Sep 09;351:h4597. [doi: 10.1136/bmj.h4597] [Medline: 26353800]
93. Iacob C, Veerappa V, Harrison R. What are you complaining about?: a study of online reviews of mobile applications. Station Road, Swindon, United Kingdom: BCS Learning & Development Ltd; 2013 Presented at: The 27th International BCS Human Computer Interaction Conference; September 9-13, 2013; Brunel University, United Kingdom URL: https://dl.acm.org/doi/proceedings/10.5555/2578048 [doi: 10.14236/ewic/hci2013.38]
94. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many Mobile Health Apps Target High-Need, High-Cost Populations, But Gaps Remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
95. LeBeau K, Huey LG, Hart M. Assessing the Quality of Mobile Apps Used by Occupational Therapists: Evaluation Using the User Version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 01;7(5):e13019 [FREE Full text] [doi: 10.2196/13019] [Medline: 31066712]
96. Madan Kumar PD, Mohandoss AA, Walls T, Rooban T, Vernon LT. Using smartphone video "selfies" to monitor change in toothbrushing behavior after a brief intervention: A pilot study. Indian J Dent Res 2016;27(3):268-277 [FREE Full text] [doi: 10.4103/0970-9290.186241] [Medline: 27411655]
97. Alavi SS, Maracy MR, Jannatifard F, Ojaghi R, Rezapour H. The psychometric properties of cellular phone dependency questionnaire in students of Isfahan: A pilot study. J Educ Health Promot 2014;3:71 [FREE Full text] [doi: 10.4103/2277-9531.134822] [Medline: 25077164]
98. De-Sola GJ, Rodríguez DF, Rubio G. Cell-Phone Addiction: A Review. Front Psychiatry 2016;7:175 [FREE Full text] [doi: 10.3389/fpsyt.2016.00175] [Medline: 27822187]
99. Foerster M, Roser K, Schoeni A, Röösli M. Problematic mobile phone use in adolescents: derivation of a short scale MPPUS-10. Int J Public Health 2015 Feb;60(2):277-286. [doi: 10.1007/s00038-015-0660-4] [Medline: 25645102]
100. Staykova KS, Damsgaard J. The race to dominate the mobile payments platform: Entry and expansion strategies. Electronic Commerce Research and Applications 2015 Sep;14(5):319-330. [doi: 10.1016/j.elerap.2015.03.004]
101. Liu CZ, Au YA, Choi HS. Effects of Freemium Strategy in the Mobile App Market: An Empirical Study of Google Play. Journal of Management Information Systems 2015 Mar 09;31(3):326-354. [doi: 10.1080/07421222.2014.995564]
102. Wang W, Li H. Factors influencing mobile services adoption: a brand‐equity perspective. Internet Research 2012 Mar 30;22(2):142-179. [doi: 10.1108/10662241211214548]
103. Hsu CL, Lin JCC. Effect of perceived value and social influences on mobile app stickiness and in-app purchase intention. Technological Forecasting and Social Change 2016 Jul;108:42-53. [doi: 10.1016/j.techfore.2016.04.012]
104. Yi Y, Jeon H. Effects of Loyalty Programs on Value Perception, Program Loyalty, and Brand Loyalty. J Acad Mark Sci 2003;31(3):229-240. [doi: 10.1177/0092070303031003002]
105. Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, et al. Brain development during childhood and adolescence: a longitudinal MRI study. Nat Neurosci 1999 Oct;2(10):861-863. [doi: 10.1038/13158] [Medline: 10491603]

Abbreviations

EBD: evidence-based dentistry
ICC: intraclass correlation coefficient
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
uMARS: Mobile Application Rating Scale user version

Edited by G Eysenbach; submitted 25.09.19; peer-reviewed by N Alshaiban, MA Vicente, C Fernández, L Grepo; comments to author 11.12.19; accepted 23.03.20; published 08.07.20.

Please cite as:
Fijačko N, Gosak L, Cilar L, Novšak A, Creber RM, Skok P, Štiglic G
The Effects of Gamification and Oral Self-Care on Oral Hygiene in Children: Systematic Search in App Stores and Evaluation of Apps
JMIR Mhealth Uhealth 2020;8(7):e16365
URL: https://mhealth.jmir.org/2020/7/e16365
doi:10.2196/16365
PMID:32673235

Original Paper

Mobile Delivery of the Diabetes Prevention Program in People With Prediabetes: Randomized Controlled Trial

Tatiana Toro-Ramos[1], BSc, PhD; Andreas Michaelides[1], PhD; Maria Anton[2], MD; Zulekha Karim[2], DO; Leah Kang-Oh[2], DO; Charalambos Argyrou[2], MSc; Elisavet Loukaidou[2], BA; Marina M Charitou[2], MD; Wilson Sze[2], MD; Joshua D Miller[2], MD

[1]Noom, Inc, New York, NY, United States

[2]Department of Medicine, Division of Endocrinology & Metabolism, Renaissance School of Medicine, Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Andreas Michaelides, PhD
Noom, Inc
229 W 28th St
New York, NY,
United States
Phone: 1 5168087328
Email: andreas@noom.com

Abstract

Background: The Centers for Disease Control and Prevention (CDC) diabetes prevention program (DPP) has formed the foundation for Type 2 Diabetes Mellitus (T2DM) prevention efforts and lifestyle change modifications in multiple care settings. To our knowledge, no randomized controlled trial has verified the efficacy of a fully mobile version of CDC's diabetes prevention program (DPP).

Objective: This study aimed to investigate the long-term weight loss and glycemic efficacy of a mobile-delivered DPP compared with a control group receiving usual medical care.

Methods: Adults with prediabetes (N=202) were recruited from a clinic and randomized to either a mobile-delivered, coach-guided DPP (Noom) or a control group that received regular medical care including a paper-based DPP curriculum and no formal intervention. The intervention group learned how to use the Noom program, how to interact with their coach, and the importance of maintaining motivation. They had access to an interactive coach-to-participant interface and group messaging, daily challenges for behavior change, DPP-based education articles, food logging, and automated feedback. Primary outcomes included changes in weight and hemoglobin A_{1c} (HbA_{1c}) levels at 6 and 12 months, respectively. Exploratory secondary outcomes included program engagement as a predictor of changes in weight and HbA_{1c} levels.

Results: A total of 202 participants were recruited and randomized into the intervention (n=101) or control group (n=99). In the intention-to-treat (ITT) analyses, changes in the participants' weight and BMI were significantly different at 6 months between the intervention and control groups, but there was no difference in HbA_{1c} levels (mean difference 0.004%, SE 0.05; *P*=.94). Weight and BMI were lower in the intervention group by −2.64 kg (SE 0.71; *P*<.001) and −0.99 kg/m2 (SE 0.29; *P*=.001), respectively. These differences persisted at 12 months. However, in the analyses that did not involve ITT, program completers achieved a significant weight loss of 5.6% (SE 0.81; *P*<.001) at 6 months, maintaining 4.7% (SE 0.88; *P*<.001) of their weight loss at 12 months. The control group lost −0.15% at 6 months (SE 0.64; *P*=.85) and gained 0.33% (SE 0.70; *P*=.63) at 12 months. Those randomized to the intervention group who did not start the program had no meaningful weight or HbA_{1c} level change, similar to the control group. At 1 year, the intervention group showed a 0.23% reduction in HbA_{1c} levels; those who completed the intervention showed a 0.28% reduction. Those assigned to the control group had a 0.16% reduction in HbA_{1c} levels.

Conclusions: This novel mobile-delivered DPP achieved significant weight loss reductions for up to 1 year compared with usual care. This type of intervention reduces the risk of overt diabetes without the added barriers of in-person interventions.

Trial Registration: ClinicalTrials.gov NCT03865342; https://clinicaltrials.gov/ct2/show/NCT03865342

(JMIR Mhealth Uhealth 2020;8(7):e17842) doi:10.2196/17842

KEYWORDS

prediabetes; body weight; mHealth; mobile app; mobile phone; randomized controlled trial

Introduction

Background

In the United States, 84.1 million people are living with prediabetes, and 2 million people are diagnosed with diabetes annually [1]. According to the World Health Organization, an estimated 422 million people worldwide had diabetes in 2014, and the prevalence continues to rise [2]. By 2050, the Centers for Disease Control and Prevention (CDC) estimates that 1 in every 3 people globally will have diabetes [3]. It remains a leading cause of death and disability, accounting for over US $327 billion annually in health care costs [4]. Patients with poor glycemic control develop microvascular complications such as blindness and end-stage renal disease as well as macrovascular complications such as heart attack and stroke [5]. The best way to combat type 2 diabetes mellitus (T2DM) is through prevention.

Restoration of normal glucose regulation in persons with prediabetes decreases the risk of developing T2DM and cardiovascular disease [6]. The most effective intervention to date is the CDC's diabetes prevention program (DPP) [7]. Studies have demonstrated a modest amount of weight loss through lifestyle modification, with participants significantly reducing their chances of developing T2DM [7-9]. DPP-based educational curricula have been well studied and validated in diverse patient populations, including in-patient settings [10-13]. As a result, findings from the DPP have formed the foundation for T2DM prevention efforts and lifestyle change modifications in multiple care settings.

Prediabetes is often discovered during routine medical visits by way of hemoglobin A_{1c} (HbA_{1c}) testing in at-risk individuals (based on the American Diabetes Association [ADA] screening criteria) [14]. Face-to-face time with clinicians is often limited, so adequate delivery of DPP-based initiatives is a challenge. Although clinicians recognize diabetes prevention as an urgent public health need that can dramatically affect the well-being of their patients, a lack of funding, collaboration, and other staff support have been reported to be key obstacles for DPP implementation in clinical practice. In addition, patients have expressed low urgency in seeking further health care after a prediabetes diagnosis [15,16]. Furthermore, highly effective, in-person DPPs can have low participation and adherence [17,18]. Mobile interventions require, or are perceived to require, less commitment, thereby overcoming a barrier in treating those at risk where in-person interventions fall short [18].

Evidence-based, scalable interventions for preventive treatment are urgently needed [16]. Exploring novel ways to empower patients to pursue lifestyle changes to prevent or delay the onset of diabetes is critical in addressing the growing diabetes epidemic. Approximately 4 out of 5 adults in the United States own a smartphone [19], and health information has never been more accessible. As such, it would be desirable to have lower cost, less resource-intensive, and scalable programs that are as effective or superior to in-person programs, especially for people who decline to take part in time-intensive face-to-face programs.

Previous Work

Studies using mobile-based platforms have focused largely on weight loss rather than diabetes risk reduction. In searching the literature, one study was found that evaluated the efficacy of a mobile-based platform adapted from the DPP curricula based on weight loss in obese patients. However, the sample size was limited, and the HbA_{1c} level was not an end point [20]. Many virtual DPP programs that utilize the internet and social media currently exist and have shown effectiveness similar to the original in-person DPP. However, fully mobile interventions without in-person components that evaluate long-term results have not been tested by means of a randomized controlled trial (RCT). Therefore, this study offers an opportunity to expand our understanding in this area.

Noom (Noom, Inc) is a mobile-based program that delivers structured curricula and coaches who communicate with users in real time through a web-based dashboard. In an observational study, Noom's DPP program resulted in 85% engagement at the end of the program, in which program completion was defined as having participated in at least one weekly curriculum activity for 9 weeks, per CDC standards and similar to other studies [21]. Upon completion of the core program, participants lost an average 5.6% of body weight [22]. At 65 weeks, the mean weight loss was 6.2% in starters who read one or more lessons per week for ≥4 core weeks, 7.4% in completers who read ≥9 lessons per week on core weeks, and 9.0% in maintenance completers who did any action in postcore weeks (all $P<.001$) [23].

Objective

The purpose of this RCT was to investigate the effectiveness of a novel, fully mobile, coach-enhanced, DPP program compared with a standard care control with paper-based CDC DPP content. The hypothesis was that participants in the intervention group would have a greater reduction in body weight and HbA_{1c} levels at the study end point after the successful completion of the virtual DPP curriculum.

Methods

Recruitment

This parallel RCT took place at Stony Brook Medicine's tertiary care ambulatory clinics from October 2016 to June 2018 in Long Island, New York. Subjects were recruited from general internal medicine, family medicine, and endocrinology practices. The inclusion criteria included patients who were English speaking, were >18 years old, were a referral from the patient's physician, had an HbA_{1c} level of 5.7% to 6.4% within 3 months before study enrollment, and owned a smartphone (Apple or Android). The exclusion criteria included patients who had experienced recent weight loss (by patient report, >5 pounds in the 6 months preceding the enrollment visit), had a previous

diagnosis of type 1 diabetes mellitus or overt T2DM, had serious or persistent mental illnesses, had >72 hours of hospitalization in the past 30 days, who were currently enrolled in a structured weight loss program or within the month preceding study enrollment, were pregnant or nursing, who had given birth within the past 3 months, and had been discouraged by a physician to enroll in a DPP program.

Participants were identified through the electronic medical record based on prediabetes diagnostic criteria (as defined by the ADA [24]). Clinicians at each clinic site were provided with information describing the study and the designated contact for study enrollment. Study personnel contacted individuals for participation, either in person or by telephone call if they had expressed interest in participating in the study. Potential participants were given an information sheet describing the study before enrollment. After indicating an interest in study participation, informed consent was obtained by the study personnel in person or by telephone. All procedures in this study were in accordance with the ethical standards of the World Medical Association Declaration of Helsinki, and all study protocols were approved by Stony Brook University's institutional review board. This trial was retrospectively registered with ClinicalTrials.gov under NCT03865342 on March 9, 2019. The Consolidated Standards of Reporting Trials statement is included in Multimedia Appendix 1.

Randomization

Participants were enrolled and randomized by study coordinators to either the intervention (Noom Coach) or control study arms utilizing a random number generator with a 1:1 allocation ratio, which automatically concealed the previous allocation. The sequence was generated by an external statistician using SAS software, version 9.4, of the SAS system for Windows (SAS Institute Inc). Coordinators set up intervention group participants with the mobile program free of cost and provided control group participants with a printed version of the DPP curriculum [25]. Study coordinators provided Noom with only the first names and email addresses of the randomized participants. No protected health information was provided to Noom by the investigators. Coordinators contacted each participant for the 6- and 12-month visits. Clinicians who gathered follow-up data were unaware of the random allocation, and masking was not broken for the duration of the study.

Procedures

Subjects were weighed at the offices of their physicians at the time of study enrollment, between 5 and 7 months (6-month time point), and between 11 and 13 months (12-month time point) poststudy enrollment. In-clinic weight measurements were recorded in the electronic medical record as part of usual care at the time point of each visit. Patients who attend these clinics are normally scheduled to visit their physician every 6 months; study participants were not asked by study personnel to see their physician any more frequently than as part of usual care. HbA_{1c} testing was completed at baseline, 6, and 12 months. If at any measurement time, the HbA_{1c} level increased above ADA criteria for overt T2DM, participants were counseled and referred to their primary care provider. The baseline HbA_{1c} level was any HbA_{1c} level available in the electronic medical record for the participant in the 3 months before study enrollment. Primary care and endocrinology clinics used the DCA Vantage (Siemens) point-of-care (POC) HbA_{1c} machine (shown to have high levels of accuracy and precision for HbA_{1c} between 5% and 8%) [26,27]. This is a clinical laboratory improvement amendment–waived capillary fingerstick test to allow clinicians to check patients' HbA_{1c} levels in the clinic. POC testing is offered to all patients in these practice sites and was offered to study participants at each of their respective HbA_{1c} testing time points. The HbA_{1c} level from commercial laboratory visits within the required period was used in lieu of POC testing when available. Of all HbA_{1c} measurements for the participants in this study, 99% at baseline, 98% at 6 months, and 95% at 12 months originated from commercial laboratory visits. At baseline and 6 months, the POC values were from the intervention group, and at 12 months, they were equally distributed.

As an incentive for participating, subjects were offered US $10.00 gift cards to Starbucks at the 6- and 12-month follow-up visits at the time of weight and HbA_{1c} measurements.

For the intervention group, within the program, an assigned Noom Coach digitally communicated with participants individually and as a group [23]. During the first week of the study, participants randomized to the intervention learned how to use the Noom program, how to interact with their coach, and the importance of maintaining motivation throughout the program. Participants had mobile access to coach-participant messaging, group messaging, daily challenges for behavior change, the DPP education articles (weekly bite-sized content over 20 weeks for the core portion and up to 52 weeks for the maintenance phase), food logging with color coding, steps and exercise logging, and automated feedback based on food choices (Figure 1). They were asked to log their weight by self-report, meals, and physical activity within the program on a weekly basis. National Diabetes Prevention Program–certified coaches securely monitored participant progress through a web-based dashboard. Participants could communicate as needed to support their individual journeys and could expect to hear from their coach every day.

Coaches working with the participants in the Noom program were trained to meet National Diabetes Prevention Program standards and trained in motivational interviewing techniques. Motivational interviewing is a client-centered therapeutic modality that utilizes positive regard, reflections, and the illumination of client strengths as a route to behavior change [28]. Coaches assisted users in setting specific, measurable, attainable, realistic, and time-based goals on a weekly basis. Coaching functions as a productive addition to a weight loss intervention through the use of accountability, feedback (food logs and choices), problem-solving, and positive reinforcement for the desirable behavior [29]. Motivational interviewing has previously shown promising results in weight loss in women with T2DM [30]. This increase in weight loss for those receiving motivational interviewing was proposed to function through treatment adherence.

Figure 1. Selection of screen pages for mobile health intervention.

Outcomes

Primary outcome measures included a change in weight and HbA_{1c} levels at 6 and 12 months from the start of the program. The CDC's National Diabetes Prevention Program benchmark for weight loss is ≥5% at 6 and 12 months [1]. Secondary outcomes were exploratory and included in-program actions (indicators of program engagement) as predictors of the change in weight and HbA_{1c} levels. Any known occurrence of serious adverse events in study participants, defined as death, serious violent incidents, and formal complaints about the intervention were recorded.

The engagement categories were measured by weekly numbers of logged meals, logged weigh-ins, logged steps, articles read, posts in the group, and messages to the coach. Previous research has suggested that focusing on specific areas of behavioral regulation of food intake [31], physical activity [31], education [32,33], coaching [29], and self-monitoring of weight through regular weigh-ins [34] may promote behavior and lifestyle changes and further weight loss success.

Statistical Analysis

The sample size was determined using the estimated SDs of the change in HbA_{1c} levels from an intervention study of patients with prediabetes (N=129) [35]. Using an SD of 0.36 and α of .05 resulted in a final sample of 224 study participants (112 per group) at 80% power to identify a minimum detectable difference of 0.5% in change in HbA_{1c} levels.

Descriptive statistics were calculated for baseline characteristics, including mean, SD, and 95% CI, to summarize differences from baseline to 6 months and to study the conclusion at 12 months between program intervention and control groups. Linear mixed models tested the null hypothesis that the mean weight and HbA_{1c} levels of the intervention and control groups were equal over time, after adjusting for other covariates. Hypothesis tests are two-sided at the .05 significance level. Multiple linear regression examined in-app actions and engagement variables as predictors of weight loss in the intervention group. Changes in weight and HbA_{1c} levels in the intervention group were analyzed first regardless of program completion. As program engagement is a key factor in attaining clinically meaningful outcomes, we further conducted prespecified analyses based on participants who completed the program as per CDC standards [36]. Maximum likelihood estimates with estimation maximization algorithms were used for missing data. All statistical analyses were performed using SPSS Statistics for Windows (version 21.0; IBM Corp), Minitab (version 17.0; Minitab, LLC), and Mplus (version 8.1; Muthén and Muthén).

Results

Participant and Study Characteristics

Between October 2016 and June 2017, 1513 potential participants were assessed for eligibility, of whom 930 did not meet the inclusion criteria and 381 declined to participate. From this, 202 participants who met the study criteria were recruited and randomized (Figure 2). Participation and retention were high, with 82.2% (166/202) of participants completing all study follow-up visits. At baseline, no differences were observed between groups in weight, BMI, HbA_{1c} level, or demographic characteristics (Table 1). Among the participants, 73.8% (76/103) of the intervention and 69% (68/99) of the control group were female, with a BMI of 31.3 kg/m^2 and 30.9 kg/m^2, respectively. The mean age was 55.7 years for the intervention group and 57.5 years for the control group. No study-related serious adverse events were reported.

Figure 2. Trial profile of recruitment and completion.

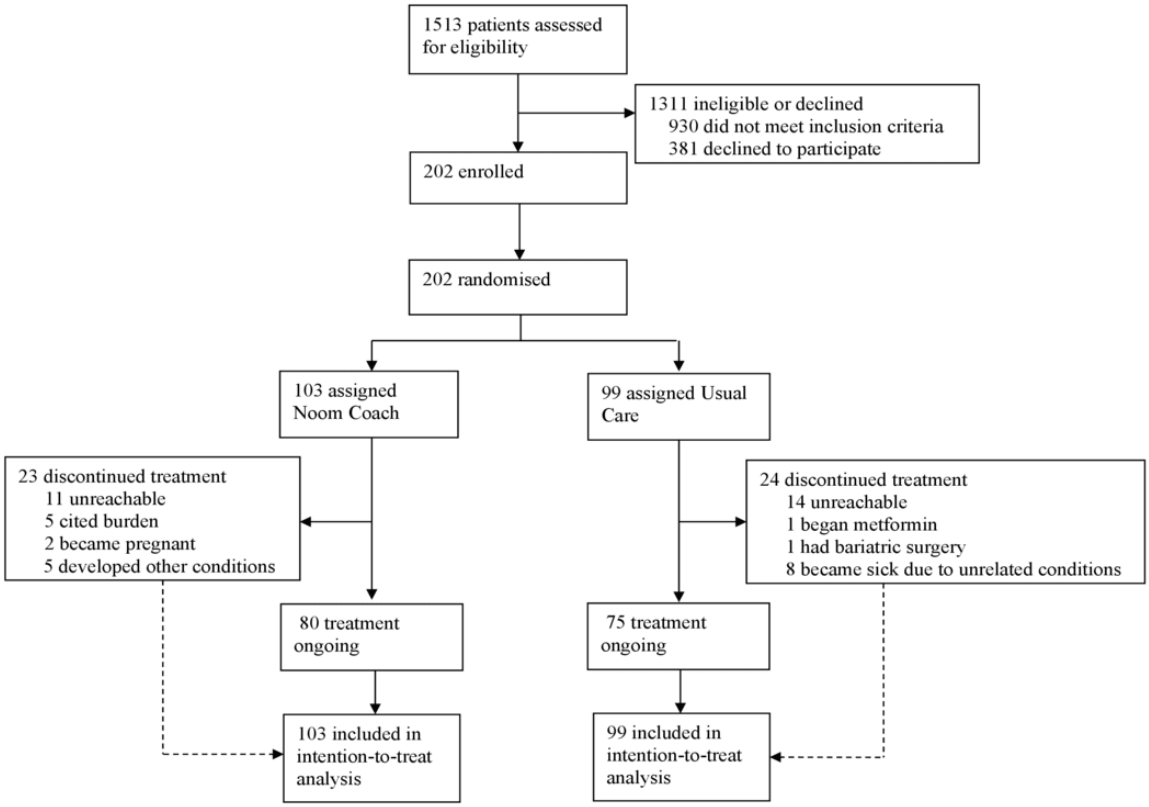


Table 1. Baseline characteristics of the study groups.

Demographic characteristics	Intervention group (n=103)	Control group (n=99)
Women, n (%)	76 (73.8)	68 (69)
Weight (kg), mean (SD)	85.71 (21.47)	85.93 (22.02)
Hemoglobin A_{1c} (%), mean (SD)	5.94 (0.18)	5.93 (0.19)
Height (m), mean (SD)	1.66 (0.08)	1.66 (0.09)
BMI (kg/m^2), mean (SD)	31.25 (6.43)	30.94 (7.23)
Age (years), mean (SD)	55.69 (13.63)	57.54 (12.45)

Among 80 intervention participants who downloaded the program, 27 did not engage meaningfully (completed fewer than 4 in-app actions, for example, read less than one article per week over 4 weeks), 53 started (logged an action and read articles for at least four weeks), and 45 completed the program (logged an action and read articles for at least nine weeks). At 6 months, among those who had weight data, 31% (28/91) of the intervention group (if they started the program or not) and 38% (17/45) of completers lost >5% body weight compared with 14% (11/77) in the control group. At 12 months, among those who had weight data, 27% (25/91) in the intervention group and 38% (17/45) of completers maintained >5% body weight loss compared with 14% (10/72) in the control group.

Similar to the control group, no significant changes in weight or HbA_{1c} levels were seen in those who did not engage meaningfully or start the program. Weight and HbA_{1c} levels by intention-to-treat (ITT) and by completion at 6 and 12 months are shown in Multimedia Appendix 2.

Weight and Hemoglobin A_{1c} Changes at 6 Months

In the ITT analyses, accounting for missing data, changes in weight and BMI were significantly different at 6 months between the intervention and control groups. Weight and BMI were lower in the intervention group at 6 months by −2.64 kg (SE 0.71; $P<.001$) and −0.99 kg/m^2 (SE 0.29; $P<.001$), respectively. No difference was seen in HbA_{1c} levels between groups at 6 months (mean difference 0.004%, SE 0.05; $P=.94$).

In the analysis that did not involve ITT, weight loss was significant over time in the intervention group but not in the control group, and there was a significant interaction of group by time ($P=.04$), indicating that weight loss was dependent on being in the intervention group. Those who completed the program achieved a clinically and statistically significant weight loss of 5.6% at 6 months based on paired *t* tests (Table 2; Figure 3). BMI significantly decreased over time in the intervention completers group versus the control group ($P<.001$; Figure 4). The HbA_{1c} levels significantly decreased over time in both

groups at 6 months, but there was no significant interaction of group by time (P=.35; Figure 5). In the intervention group, those who completed the program did not report a HbA_{1c} level above 6.4% at any time point. In the control group, 2 patients reached HbA_{1c} levels at or above 6.4% at 6 months.

Table 2. Change in body weight and hemoglobin A_{1c} at 6 and 12 months by participant group (intervention, intervention completers, and control group).

Measured values	Intervention group[a]						Intervention completers						Control group[a]			
	6 months, mean (95% CI)	Effect size[b]	*P* value[c]	12 months, mean (95% CI)	Effect size[b]	*P* value[c]	6 months, mean (95% CI)	Effect size[b]	*P* value[c]	12 months, mean (95% CI)	Effect size[b]	*P* value[c]	6 months, mean (95% CI)	*P* value[c]	12 months, mean (95% CI)	*P* value[c]
Weight (kg)	−3.31 (−4.43 to −2.19)	−0.12	<.001	−2.22 (−3.31 to −1.13)	0.01	.02	−4.86 (−6.39 to −3.33)	−0.22	<.001	−3.92 (−5.48 to −2.37)	−0.14	<.001	−0.42 (−1.53 to 0.69)	.45	−0.09 (−1.30 to 1.11)	.88
Weight (%)	−3.69 (−4.89 to −2.48)	−0.63	<.001	−2.54 (−3.74 to −1.33)	−0.40	<.001	−5.59 (−7.22 to −3.95)	−0.99	<.001	−4.66 (−6.42 to −2.90)	−0.72	<.001	−0.15 (−1.42 to 1.11)	.81	0.33 (−1.06 to 1.72)	.63
BMI (kg/m^2)	−1.35 (−1.79 to −0.92)	−0.13	<.001	−0.88 (−1.31 to 0.44)	−0.00	<.001	−1.79 (−2.34 to −1.24)	−0.15	<.001	−1.44 (−2.02 to −0.87)	−0.09	<.001	−0.12 (−0.53 to 0.29)	.56	−0.04 (−0.47 to 0.39)	.86
HbA_{1c}[d] (%)	−0.15 (−0.22 to −0.08)	0.08	<.001	−0.23 (−0.32 to −0.14)	−0.13	<.001	−0.17 (−0.25 to −0.10)	−0.06	<.001	−0.28 (−0.37 to −0.19)	−0.38	<.001	−0.17 (−0.25 to −0.09)	<.001	−0.16 (−0.27 to −0.05)	.01

[a]At 6 months, the intervention group lost 2.64 kg more and had a BMI difference of 0.99 kg/m^2 compared with the control group (both P<.001) in the intention-to-treat (ITT) maximum likelihood estimates missing data analyses. At 12 months, in the ITT analyses, the intervention group lost 1.8 kg more than the control group (P=.01) and had a BMI difference of 0.58 kg/m^2 (P=.01). In the ITT analyses, HbA_{1c} was not different between groups at 6 or 12 months, with a difference of 0.004% (P=.94) and 0.006% (P=.93), respectively.

[b]≤0.2=small effect; >0.2 to <0.8=medium effect; ≥0.8=large effect.

[c]*P* values were obtained from *t* tests using a .05 significance level.

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

Figure 3. Weight change across time points. ITT: intention-to-treat; T0: baseline; T6: 6 months; T12: 12 months. ITT: intention-to-treat; T0: baseline; T6: 6 months; T12: 12 months.

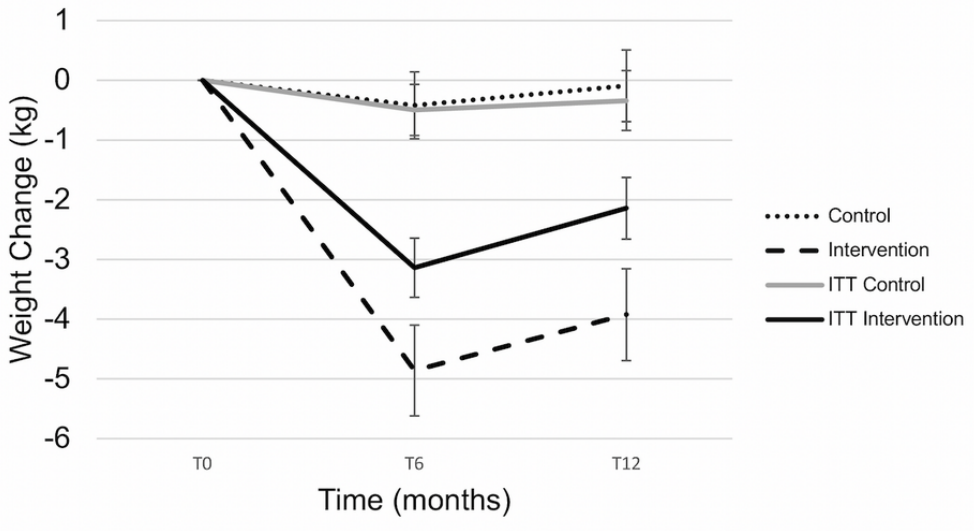


Figure 4. BMI change across time points. ITT: intention-to-treat; T0: baseline; T6: 6 months; T12: 12 months.

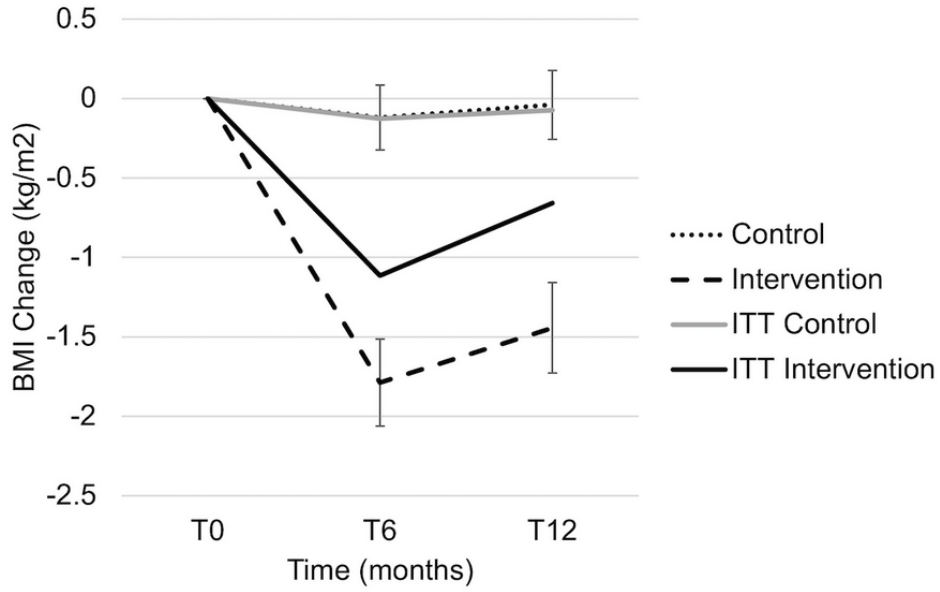
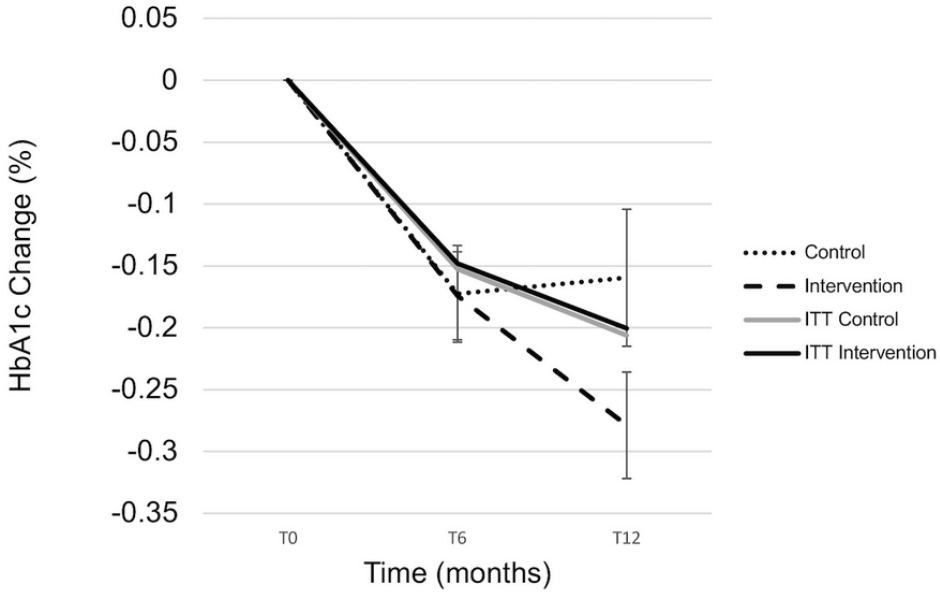


Figure 5. Hemoglobin A_{1c} change across time points. ITT: intention-to-treat; T0: baseline; T6: 6 months; T12: 12 months.



Weight and Hemoglobin A_{1c} Changes at 12 Months

In the ITT analyses, changes in weight and BMI were significantly lower in the intervention group at 12 months by −1.80 kg (SE 0.81; *P*=.01) and −0.58 kg/m^2 (SE 0.24; *P*=.01), respectively. HbA_{1c} levels showed no difference between the groups at 12 months (0.006%; SE 0.07; *P*=.93).

In the analysis that did not involve ITT, an interaction between group and time was found (*P*<.001), indicating that weight loss was dependent on being in the intervention group. Those who completed the program achieved a clinically and statistically significant weight loss of 4.7% at 12 months (Table 2; Figure 3). BMI significantly decreased over time in the intervention completers group versus the control group at 12 months (*P*<.001; Figure 4). At 12 months, the HbA_{1c} levels continued to significantly decrease in the intervention group, which was below the prediabetic values (Figure 5). In the intervention group, those who completed the program did not report HbA_{1c} levels above 6.4%. In the control group, 4 patients reached HbA_{1c} levels at or above 6.4% at 6 months.

Engagement Variables

Completers actively participated in the program (Table 3). Multiple linear regressions controlling for age, sex, baseline BMI (and baseline HbA_{1c} levels for weight change models) observed which program engagement behaviors, including the mean number of weekly weigh-ins, articles read, meals logged, steps, group posts, and messages to coach predicted changes in weight and HbA_{1c} levels at 6 and 12 months. Weight change at 6 and 12 months was also evaluated as a predictor in the models for HbA_{1c} levels change at 6 and 12 months, respectively.

Having a higher baseline BMI and HbA_{1c} status predicted increased weight loss at 6 months (Table 4). Important program engagement predictors of weight loss at 12 months included frequency of weighing-in (β=−0.30; *P*=.01) and logging more steps (β=−0.21; *P*=.08). Weight change at 12 months was predicted by higher meal logging frequency (β=−0.41; *P*=.001). In-app actions did not predict a change in HbA_{1c} levels at 6 and 12 months.

Table 3. Engagement of intervention participants (in-app actions per week as).

In-app activities	Intervention, mean (SD)		Completers, mean (SD)	
	6 months	12 months	6 months	12 months
Logged meals[a]	7.40 (8.03)	5.01 (6.43)	12.58 (6.73)	8.53 (6.39)
Logged weigh-ins[b]	0.42 (0.83)	0.29 (0.72)	0.70 (1.0)	0.48 (0.89)
Logged steps	12,132 (14,131)	9487 (12,494)	19,110 (14,479)	15,152 (13,511)
Articles read	3.92 (5.90)	2.28 (3.80)	6.62 (6.48)	3.86 (4.33)
Group comments[c]	0.14 (0.32)	0.07 (0.17)	0.23 (0.40)	0.11 (0.20)
Messages to coach	1.44 (1.91)	0.94 (1.46)	2.42 (1.98)	1.58 (1.63)

[a]Logged meals refers to the times breakfast, lunch, snack, and dinner were logged per week.

[b]Logged weigh-ins refers to times per week of in-app weight self-reports.

[c]Group comments refers to responses to group posts per week.

Table 4. Multiple linear regression models for in-app actions as predictors of body weight change adjusting for age and gender.

Dependent variable and significant predictors	Beta level	*t* value[a]	*P* value	r^2	Adjusted r^2	*F* value[a]
Weight change at 6 months						
BMI	−0.302	−2.676	.01	0.21	0.24	6.158[b]
HbA_{1c}[c]	−0.305	−2.765	.01	0.21	0.24	6.158
Steps at 6 months	−0.205	−1.815	.08	0.21	0.24	6.158
Weigh-ins at 6 months	−0.296	−2.628	.01	0.21	0.24	6.158
Weight change at 12 months						
Meals at 12 months	−0.405	−3.603	.001	0.16	0.15	12.983[d]

[a]The *df* values for *t* test and *F* test were unavailable.

[b]*P*≤.001.

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

[d]*P*=.001.

Discussion

Principal Findings

This RCT shows that a fully mobile DPP intervention with coaching in adults with prediabetes was effective in significantly reducing weight over 1 year. In the ITT analyses, the intervention group lost 2.64 kg more at 6 months and 1.80 kg more at 12 months compared with the control group, but the HbA_{1c} level was not significantly different between groups over time. Weight loss has been shown to plateau at a 5% to 9% loss 6 months into a program with a slight weight regain, indicating a 4.8% to 8% loss by 12 months with approximately a 3% to 4% loss maintained at 48 months [37]. Our data corroborate these findings in that slight weight regain was seen between 6 and 12 months, with intervention participants not returning to baseline weight.

In the analyses that did not involve ITT, participants in the intervention group who completed the program lost 5.6% and 4.7% body weight at 6 and 12 months, respectively, compared with the control group that had no meaningful weight change. In the analyses that did not involve ITT, HbA_{1c} levels were reduced in both groups at 6 months, but only the intervention group continued to significantly decrease through 12 months. Higher program engagement predicted greater weight loss but not a change in HbA_{1c} levels. To our knowledge, this is the first such RCT demonstrating long-term efficacy for weight loss in a fully mobile DPP intervention.

Comparison With Previous Work

Diabetes prevention interventions have been delivered through multiple means, including in-person, community, the web and mobile, and mixed interventions [35,38-42], reporting weight loss of 3% to 5% or more. In 2009, a web-based DPP intervention showed that completers lost 4.79 kg at 12 months

[42]. To our knowledge, this study is the first to verify the longitudinal efficacy through an RCT of a fully mobile-based DPP encompassing all aspects of a CDC recognized program [43]. It adds to the growing digital DPP intervention literature by demonstrating a clear, ITT reduction in weight at 1 year using a modern state-of-the-art digital mobile health program in a relatively pragmatic setting.

We previously showed significant weight loss in a fully mobile DPP observational study at 24 and 65 weeks [22,23], with program completers achieving higher degrees of weight loss compared with nonstarters and starters. Similar to a web-based DPP intervention with −4.7 kg and −4.0 kg weight loss at 6 and 12 months, respectively, in those who completed the program [44], we found −4.9 kg and −4.0 kg at 6 and 12 months in the intervention completers. Another DPP intervention of older adults found that more participants completed a web-based DPP program compared with an in-person intervention and achieved 5% weight loss over 6 and 12 months [45].

A higher baseline BMI and HbA_{1c} level, logging more steps, and more frequent weigh-ins predicted greater weight loss at 6 months, whereas continuing to log meals at 12 months predicted higher weight loss at 1 year. Previously, weighing-in and logging more meals along with group interaction predicted weight loss at 24 weeks [22], whereas at 65 weeks, meal logging and interacting with a support group continued to be significant weight loss predictors, supporting self-monitoring as a key component of successful weight loss and maintenance [23]. Indeed, social support in the form of in-app group interaction and high self-efficacy evidenced by persistent food logging are crucial factors that predict weight loss success, further supporting the design of personalized intervention strategies through mobile health (mHealth) [46]. Those with higher HbA_{1c} levels had greater reductions in weight, which might imply that borderline HbA_{1c} levels should not necessarily prompt clinicians to prescribe medication before referring them to a DPP, helping reduce the development of overt T2DM without the need for medication. Weight change at 6 or 12 months was not a predictor of HbA_{1c} change, indicating that a larger sample and longer study duration may be needed to establish a stronger association.

This study demonstrates that a control group receiving usual care plus the CDC's DPP written materials did not lead to meaningful weight loss or reduction in HbA_{1c} levels. Previous mobile-based prediabetes and weight loss interventions have shown efficacy, and here we offer further evidence of the feasibility of a fully mobile-based DPP intervention leading to meaningful sustained weight loss for up to 1 year. The efficacy of this approach is clinically significant because it is likely that a larger number of people at risk can participate in such an intervention as smartphone technology removes the barriers of time and accessibility presented by in-person interventions. Mobile interventions can easily be adapted by medical professionals to facilitate patient participation and engagement in DPP programs to reduce T2DM risk and advance patient care while improving their practice's efficiency.

Limitations

Program participation was a limitation of this study. We found that 23 participants declined to download the program or redeem the intervention program's registration code despite remaining enrolled in the study and completing most visits. Another 53 participants did not meaningfully engage with or complete the program. Designing interventions with protocols to re-engage lapsed study participants may improve participation.

The smartphone version of the DPP allows for intensive education that can be disseminated over time due to the various program features for a variety of forms of communication. Studies have shown that patient engagement and diabetes risk awareness are poor [47,48]. Low self-motivation may impair timely treatment, pointing to the added need for intensive education to create true awareness and understanding of the prediabetes condition and potential benefits of a DPP. These factors seem to be significant gaps in clinical care and reporting suboptimal adoption of behaviors for risk reduction in adults with prediabetes [48]. Such findings suggest the need to improve patient education and awareness in the clinical setting to prevent the onset of T2DM and decrease the risk of future complications.

Despite the ability to measure if participants went through the educational content in the program, there was no way to ensure that they read the content fully or engaged with activities or practices that were proposed to be done off of the app. Another limitation of our study is that individuals with prediabetes with a BMI <25 kg/m^2 were included in our analyses, reducing the effect size of weight loss and HbA_{1c} levels. Additionally, although 82% of participants completed all study visits, the study remained underpowered for HbA_{1c} outcomes. A larger sample size would have provided a more robust estimate of the effectiveness of HbA_{1c} levels. Future research is needed to establish wider generalizability and true efficacy for HbA_{1c} reduction resulting from the program intervention.

Despite these limitations, this study is the first mHealth DPP RCT from a program fully recognized by the CDC showing significant long-term weight loss greater than 5% and maintenance up to 1 year. Furthermore, completers in the intervention group had a large enough weight loss that drove significant ITT weight loss. The main outcome measures were reliably obtained in a controlled clinical setting, removing potential measurement errors that can arise through self-reporting. Study participants were predominantly female, which might limit generalizability; however, the sample was recruited from tertiary care outpatient medical practices representative of the general population at risk for T2DM on Long Island. Furthermore, the sample included a broad range of ages ≥18 years and individuals with normal BMI, indicating that intervention findings could apply to the general at-risk population.

Individuals who are more aware of a medical condition or risk are more likely to seek additional medical care and significant lifestyle changes. The perception that body weight poses a health risk to the individual significantly contributes to their efforts to lose weight [49]. People who have been told by physicians that they are at a health risk if they do not change their weight could

have more active program engagement than others, as this advice has been shown to be strongly related to weight loss efforts [49,50]. Future research will focus on testing brief but intensive prediabetes education sessions after diagnosis and before referral to a DPP to aid in a better understanding of the true risks faced with such a diagnosis. A brief preparatory intervention can result in higher uptake of in-person, mobile, web-based, or combination DPPs and effectively prevent T2DM and improve health outcomes in more persons who may have not sought care.

Conclusions

As the first long-term randomized intervention of its kind, the results of this study demonstrate that a novel fully mobile-based smartphone-delivered DPP with human coaching is an effective and powerful tool for attaining clinically and statistically significant weight loss up to 1 year, reducing T2DM risk as well as in-person interventions but without the added barriers. Continuous guidance through coaching, seamless self-monitoring tools, and engaging mobile-delivered DPP content are key to achieving sustained changes in weight and preventing T2DM.

Acknowledgments

The authors would like to thank all participants and physicians at Stony Brook Medicine for their important contribution to this study; without them, this study would not have been possible. For statistics guidance and expertise, the authors are grateful to Dr David González Barreto at the University of Puerto Rico, Mayagüez, and Dr Tom Hildebrandt at the Icahn School of Medicine in New York. The authors also thank Kayla Reynolds, MS, and Dr Martica Heaner, PhD, for their editorial assistance. The authors extend gratitude to the coaches who guided the participants in this study: Peter Athans, MS; Laura Mahoney, MS; and Nancy Sidnam, MS. Stony Brook is a mentor organization for the New York Digital Health Accelerator, a program that pairs providers with emerging technology companies to pilot innovative health care technologies. Through this partnership, Stony Brook and Noom committed to identifying a concrete collaboration opportunity. Noom, Inc owns Noom Coach (the study treatment), helped design the trial, and paid for the participant gift cards, the cartridges for the Siemens DCA-A_{1c} analyzer (for use in primary care and endocrinology clinics), and study personnel. There was no additional funding source for this study.

Conflicts of Interest

TT and AM are employed by Noom, Inc and receive salary and stock options. AM holds a patent pending (3492.004US1) with Noom, Inc. Noom Coach is a mHealth program owned by Noom, Inc. The authors have no additional conflicts of interest, and no competing financial interests exist for the other authors.

Multimedia Appendix 1
Consolidated Standards of Reporting Trials checklist.
[PDF File (Adobe PDF File), 1150 KB - mhealth_v8i7e17842_app1.pdf]

Multimedia Appendix 2
Body weight and hemoglobin A1c at 6 and 12 months by participant group.
[DOCX File , 18 KB - mhealth_v8i7e17842_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1150 KB - mhealth_v8i7e17842_app3.pdf]

References

1. CDC. Centers for Disease Control and Prevention. 2017. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States URL: https://www.cdc.gov/diabetes/data/statistics/statistics-report.html [accessed 2020-05-07]
2. World Health Organization. 2016. Global Report on Diabetes URL: https://www.who.int/diabetes/global-report/en/: [accessed 2019-03-28]
3. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. Popul Health Metr 2010 Oct 22;8:29 [FREE Full text] [doi: 10.1186/1478-7954-8-29] [Medline: 20969750]
4. American Diabetes Association. Economic costs of diabetes in the US In 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
5. Fowler MJ. Microvascular and macrovascular complications of diabetes. Clin Diabetes 2008 Apr 1;26(2):77-82. [doi: 10.2337/diaclin.26.2.77]
6. Perreault L, Temprosa M, Mather KJ, Horton E, Kitabchi A, Larkin M, Diabetes Prevention Program Research Group. Regression from prediabetes to normal glucose regulation is associated with reduction in cardiovascular risk: results from

the diabetes prevention program outcomes study. Diabetes Care 2014 Sep;37(9):2622-2631 [FREE Full text] [doi: 10.2337/dc14-0656] [Medline: 24969574]
7. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
8. Diabetes Prevention Program Research Group. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the diabetes prevention program outcomes study. Lancet Diabetes Endocrinol 2015 Nov;3(11):866-875 [FREE Full text] [doi: 10.1016/S2213-8587(15)00291-0] [Medline: 26377054]
9. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, et al. 10-year follow-up of diabetes incidence and weight loss in the diabetes prevention program outcomes study. Lancet 2009 Nov 14;374(9702):1677-1686 [FREE Full text] [doi: 10.1016/S0140-6736(09)61457-4] [Medline: 19878986]
10. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. J Gen Intern Med 2015 Jan;30(1):107-117 [FREE Full text] [doi: 10.1007/s11606-014-2987-6] [Medline: 25134692]
11. Dunkley AJ, Bodicoat DH, Greaves CJ, Russell C, Yates T, Davies MJ, et al. Diabetes prevention in the real world: effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: a systematic review and meta-analysis. Diabetes Care 2014 Apr;37(4):922-933. [doi: 10.2337/dc13-2195] [Medline: 24652723]
12. McTigue KM, Conroy MB. Use of the internet in the treatment of obesity and prevention of type 2 diabetes in primary care. Proc Nutr Soc 2013 Feb;72(1):98-108. [doi: 10.1017/S0029665112002777] [Medline: 23098133]
13. Ma J, Yank V, Xiao L, Lavori PW, Wilson SR, Rosas LG, et al. Translating the diabetes prevention program lifestyle intervention for weight loss into primary care: a randomized trial. JAMA Intern Med 2013 Jan 28;173(2):113-121 [FREE Full text] [doi: 10.1001/2013.jamainternmed.987] [Medline: 23229846]
14. American Diabetes Association. 3. Prevention or delay of type 2 diabetes. Diabetes Care 2019 Jan;42(Suppl 1):S29-S33. [doi: 10.2337/dc19-S003] [Medline: 30559229]
15. Thomas T, Samuel-Hodge CD, Porterfield DS, Alva ML, Leeman J. Scaling up diabetes prevention programs in north Carolina: perceptions of demand from potential program recipients and providers. Diabetes Educ 2019 Feb;45(1):116-124. [doi: 10.1177/0145721718811564] [Medline: 30413136]
16. Bowen ME, Schmittdiel JA, Kullgren JT, Ackermann RT, O'Brien MJ. Building toward a population-based approach to diabetes screening and prevention for US adults. Curr Diab Rep 2018 Sep 19;18(11):104 [FREE Full text] [doi: 10.1007/s11892-018-1090-5] [Medline: 30229480]
17. Jackson SL, Long Q, Rhee MK, Olson DE, Tomolo AM, Cunningham SA, et al. Weight loss and incidence of diabetes with the veterans health administration MOVE! lifestyle change programme: an observational study. Lancet Diabetes Endocrinol 2015 Mar;3(3):173-180 [FREE Full text] [doi: 10.1016/S2213-8587(14)70267-0] [Medline: 25652129]
18. Griauzde D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A mobile phone-based program to promote healthy behaviors among adults with prediabetes who declined participation in free diabetes prevention programs: mixed-methods pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 9;7(1):e11267. [doi: 10.2196/11267] [Medline: 30626566]
19. Diabetes Pro: American Diabetes Association. 2015. Diabetes Fact Sheet URL: http://professional.diabetes.org/admin/UserFiles/0%20-%20Sean/Documents/Fast_Facts_3-2015.pdf [accessed 2015-08-01]
20. 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]
21. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res 2015 Apr 10;17(4):e92 [FREE Full text] [doi: 10.2196/jmir.4052] [Medline: 25863515]
22. Michaelides A, Raby C, Wood M, Farr K, Toro-Ramos T. Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching. BMJ Open Diabetes Res Care 2016;4(1):e000264 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000264] [Medline: 27651911]
23. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May 3;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
24. American Diabetes Association. Standards of medical care in diabetes-2014. Diabetes Care 2014 Jan;37(Suppl 1):S14-S80. [doi: 10.2337/dc14-S014] [Medline: 24357209]
25. National Institute of Diabetes and Digestive and Kidney. 2006. Your Game Plan to Prevent Type 2 Diabetes URL: https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-type-2-diabetes/game-plan [accessed 2020-05-07]

26. Zercher A, Schulman L, Boone J. Quantitative Measurement of Hemoglobin A_{1c} on the DCA Vantage Point-of-Care Analyzer as a Diagnostic Test for Diabetes: An Internal Validation Study. Elkhart, IN, US: Siemens Healthcare Diagnostics; 2014.
27. Whitley HP, Yong EV, Rasinen C. Selecting an A1c point-of-care instrument. Diabetes Spectr 2015 Aug;28(3):201-208 [FREE Full text] [doi: 10.2337/diaspect.28.3.201] [Medline: 26300614]
28. Rollnick S, Miller WR, Butler CC, Aloia MS. Motivational interviewing in health care: helping patients change behavior. Int J Chron Obstruct Pulmon Dis 2009 Jul 2;5(3):203. [doi: 10.1080/15412550802093108]
29. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am 2011 Dec;34(4):841-859 [FREE Full text] [doi: 10.1016/j.psc.2011.08.006] [Medline: 22098808]
30. West DS, DiLillo V, Bursac Z, Gore SA, Greene PG. Motivational interviewing improves weight loss in women with type 2 diabetes. Diabetes Care 2007 May;30(5):1081-1087. [doi: 10.2337/dc06-1966] [Medline: 17337504]
31. Tronieri JS, Wadden RA. Behavioral assessment of patients with obesity. In: Wadden TA, Bray GA, editors. Handbook of Obesity Treatment. Second Edition. New York, USA: Guilford Press; 2018:253-282.
32. Barnes RD, Ivezaj V, Martino S, Pittman BP, Grilo CM. Back to basics? No weight loss from motivational interviewing compared to nutrition psychoeducation at one-year follow-up. Obesity (Silver Spring) 2017 Dec;25(12):2074-2078 [FREE Full text] [doi: 10.1002/oby.21972] [Medline: 29086484]
33. Barnes RD, White MA, Martino S, Grilo CM. A randomized controlled trial comparing scalable weight loss treatments in primary care. Obesity (Silver Spring) 2014 Dec;22(12):2508-2516 [FREE Full text] [doi: 10.1002/oby.20889] [Medline: 25298016]
34. Butryn ML, Phelan S, Hill JO, Wing RR. Consistent self-monitoring of weight: a key component of successful weight loss maintenance. Obesity (Silver Spring) 2007 Dec;15(12):3091-3096 [FREE Full text] [doi: 10.1038/oby.2007.368] [Medline: 18198319]
35. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
36. Centers for Disease Control and Prevention. 2015. Requirements for CDC Recognition URL: https://www.cdc.gov/diabetes/prevention/lifestyle-program/requirements.html [accessed 2019-03-17]
37. Franz MJ, van Wormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007 Oct;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
38. Sweet CM, Chiguluri V, Gumpina R, Abbott P, Madero EN, Payne M, et al. Outcomes of a digital health program with human coaching for diabetes risk reduction in a medicare population. J Aging Health 2018 Jun;30(5):692-710 [FREE Full text] [doi: 10.1177/0898264316688791] [Medline: 28553807]
39. Shirinzadeh M, Afshin-Pour B, Angeles R, Gaber J, Agarwal G. The effect of community-based programs on diabetes prevention in low- and middle-income countries: a systematic review and meta-analysis. Global Health 2019 Feb 1;15(1):10 [FREE Full text] [doi: 10.1186/s12992-019-0451-4] [Medline: 30709362]
40. Aziz Z, Absetz P, Oldroyd J, Pronk NP, Oldenburg B. A systematic review of real-world diabetes prevention programs: learnings from the last 15 years. Implement Sci 2015 Dec 15;10:172 [FREE Full text] [doi: 10.1186/s13012-015-0354-6] [Medline: 26670418]
41. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via ehealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
42. McTigue KM, Conroy MB, Hess R, Bryce CL, Fiorillo AB, Fischer GS, et al. Using the internet to translate an evidence-based lifestyle intervention into practice. Telemed J E Health 2009 Nov;15(9):851-858. [doi: 10.1089/tmj.2009.0036] [Medline: 19919191]
43. Albright AL, Gregg EW. Preventing type 2 diabetes in communities across the US: the national diabetes prevention program. Am J Prev Med 2013 Apr;44(4 Suppl 4):S346-S351 [FREE Full text] [doi: 10.1016/j.amepre.2012.12.009] [Medline: 23498297]
44. Moin T, Damschroder LJ, AuYoung M, Maciejewski ML, Havens K, Ertl K, et al. Results from a trial of an online diabetes prevention program intervention. Am J Prev Med 2018 Nov;55(5):583-591 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.028] [Medline: 30262149]
45. Lee PG, Damschroder LJ, Holleman R, Moin T, Richardson CR. Older adults and diabetes prevention programs in the veterans health administration. Diabetes Care 2018 Dec;41(12):2644-2647 [FREE Full text] [doi: 10.2337/dc18-1141] [Medline: 30377187]
46. Kim H, Faw M, Michaelides A. Mobile but connected: harnessing the power of self-efficacy and group support for weight loss success through mhealth intervention. J Health Commun 2017 May;22(5):395-402. [doi: 10.1080/10810730.2017.1296510] [Medline: 28339324]

47. Geiss LS, James C, Gregg EW, Albright A, Williamson DF, Cowie CC. Diabetes risk reduction behaviors among US adults with prediabetes. Am J Prev Med 2010 Apr;38(4):403-409. [doi: 10.1016/j.amepre.2009.12.029] [Medline: 20307809]
48. O'Brien MJ, Moran MR, Tang JW, Vargas MC, Talen M, Zimmermann LJ, et al. Patient perceptions about prediabetes and preferences for diabetes prevention. Diabetes Educ 2016 Dec;42(6):667-677 [FREE Full text] [doi: 10.1177/0145721716666678] [Medline: 27621093]
49. Gregory CO, Blanck HM, Gillespie C, Maynard LM, Serdula MK. Perceived health risk of excess body weight among overweight and obese men and women: differences by sex. Prev Med 2008 Jul;47(1):46-52. [doi: 10.1016/j.ypmed.2008.01.008] [Medline: 18289656]
50. Bish CL, Blanck HM, Serdula MK, Marcus M, Kohl HW, Khan LK. Diet and physical activity behaviors among Americans trying to lose weight: 2000 behavioral risk factor surveillance system. Obes Res 2005 Mar;13(3):596-607 [FREE Full text] [doi: 10.1038/oby.2005.64] [Medline: 15833946]

Abbreviations

ADA: American Diabetes Association
CDC: Centers for Disease Control and Prevention
DPP: diabetes prevention program
HbA_{1c}: hemoglobin A_{1c}
ITT: intention-to-treat
mHealth: mobile health
POC: point-of-care
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 17.01.20; peer-reviewed by D Biduski, YH Lin, A De Marchi; comments to author 21.03.20; revised version received 03.04.20; accepted 15.04.20; published 08.07.20.

Please cite as:
Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, Loukaidou E, Charitou MM, Sze W, Miller JD
Mobile Delivery of the Diabetes Prevention Program in People With Prediabetes: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e17842
URL: https://mhealth.jmir.org/2020/7/e17842
doi:10.2196/17842
PMID:32459631

Original Paper

Improving the Quality of Antenatal Care Using Mobile Health in Madagascar: Five-Year Cross-Sectional Study

Anne Caroline Benski[1], MD; Nicole C Schmidt[2], MD; Manuela Viviano[3], MD; Giovanna Stancanelli[4], MD; Adelia Soaroby[5]; Michael R Reich[1,6], PhD

[1]Takemi Program in International Health, Harvard TH Chan School of Public Health, Boston, MA, United States

[2]Katholische Stiftungshochschule München, University of Applied Science, Munich, Germany

[3]Ospedale Regionale di Lugano, Ente Ospedaliero Cantonale, Lugano, Italy

[4]Vita-Salute San Raffaele University, Milano, Italy

[5]Centre Médico-Chirurgical Saint Damien, Ambanja, Madagascar

[6]Department of Global Health & Population, Harvard TH Chan School of Public Health, Boston, MA, United States

Corresponding Author:
Anne Caroline Benski, MD
Takemi Program in International Health
Harvard TH Chan School of Public Health
677 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 787397261
Email: cbenski@hsph.harvard.edu

Abstract

Background: Despite many efforts, maternal mortality remains a major burden in most developing countries. Mobile health (mHealth) has the potential to improve access to obstetric care through apps that help patients and providers.

Objective: This study aimed to use mHealth to provide antenatal care (ANC) to 1446 pregnant women in a rural area in Madagascar and evaluate the quality of ANC provided by an mHealth system designed to change the behaviors of providers and patients.

Methods: We included 1446 women who attended ANC visits in rural Madagascar from 2015 to 2019 using an mHealth system called Pregnancy and Newborn Diagnostic Assessment (PANDA). This cross-sectional study used data from different participants, with information collected over several years, to analyze the outputs related to the quality of ANC over time. Specifically, we examined the timing of the first ANC visit, the relationship between the visit duration and the risk factors among pregnant women, and the number of ANC visits per woman.

Results: Following the implementation of the mHealth system in 2015, we observed that women started to come earlier for their first ANC visit; more women attended their first ANC visit in the second trimester of pregnancy in 2019 than in the previous years (P<.001). In 2019, fewer women attended their first ANC visit in the third trimester (57/277, 20.6%) than in 2015 (147/343, 42.9%). There were statistically significant associations between the ANC visit durations and the risk factors, including age (>35 years; 25.0 min, 95% CI 24.0-25.9), educational level (longer visit for women with lower than primary education and for women who attended university and shorter for women with primary school–level education; 40.7 min, 95% CI 30.2-51.3 and 25.3 min, 95% CI 24.4-26.3 vs 23.3 min, 95% CI 22.9-23.8; P=.001), experience of domestic violence during pregnancy, gravidity, parity, infectious diseases (HIV, malaria, and syphilis), and level of anemia. Statistically significant associations were observed for all quality indicator variables. We observed a statistically significant increase in the number of ANC visits per woman over time from 2015 to 2017; the number of ANC visits per woman then became stable after the third year of implementing the PANDA mHealth system.

Conclusions: This study shows the potential of an mHealth system to improve the quality of ANC, change provider behavior by standardizing ANC visits, and change patient behavior by increasing the willingness to return for subsequent visits and encouraging ANC attendance early in pregnancy. As this is an exploratory study, further studies are necessary to better understand how mHealth can change behavior and identify the conditions required for behavioral changes to persist over time.

(JMIR Mhealth Uhealth 2020;8(7):e18543) doi:10.2196/18543

KEYWORDS

mobile health; maternal health; antenatal care; quality of care; mobile phone

Introduction

Background

Mobile health (mHealth) tools are an innovative technology that can allow patients and their health care providers to effectively access medical data before, during, and after medical appointments. mHealth has the potential to improve the quality of health care through apps that can facilitate communication between patients and health care providers [1]. In the past few decades, most health-related apps have been created for patients rather than for providers [2,3].

Maternal mortality has decreased dramatically in low- and middle-income countries (LMICs), declining by 45% from 2008 to 2013 [4]; however, around the world, as many as 800 women continue to die each day because of complications from pregnancy and delivery [5]. Up to 99% of maternal deaths worldwide occur in LMICs, and researchers have postulated that a lack of access to high-quality maternal and newborn care is a major reason for maternal deaths in these countries [6,7]. Antenatal care (ANC) provides a unique opportunity for screening, diagnosis, and health promotion among pregnant women and their families and communities [8]. Appropriate utilization of ANC services is associated with improved maternal and newborn health as well as a reduction in maternal deaths during pregnancy and childbirth [9-11]. Due to the benefits of ANC, the World Health Organization (WHO) recommends that pregnant women should attend at least four ANC visits to increase opportunities for risk identification, management of pregnancy and/or comorbidities, and health promotion, and since 2016, the WHO recommends eight contacts with health care providers during pregnancy [12]. Delays in the recognition and management of clinical problems that may arise during pregnancy can increase maternal and neonatal mortality [13].

Expanded use of mHealth could help improve health care quality in the context of LMICs [14]. mHealth can offer technical support to assist health care providers in data collection and retrieval, and, even more importantly, it can support the clinical practice of health care providers and improve the quality of care and interactions between patients and health care workers [15]. If mHealth is capable of increasing patients' satisfaction with the clinical approach, this higher satisfaction may translate into increased attendance at clinical visits, which may ultimately reduce adverse obstetric and neonatal outcomes. However, current evidence on this topic is scarce, and little is known about how mHealth can influence the clinical practice of health care providers and the quality of services [16,17].

Objectives

We conducted a study in Madagascar using an mHealth system to record and access women's data during their ANC visits. The main goal of this study was to evaluate the quality of ANC provided with mHealth by measuring the adherence to ANC visits, the timing of the first ANC visit, and the duration of the visits.

Methods

Study Setting and Collaboration

This cross-sectional, observational study was conducted from January 2015 to September 2019 in the Ambanja district in northwestern Madagascar. Data related to ANC visits were collected using an mHealth system called Pregnancy and Newborn Diagnostic Assessment (PANDA). The local health authorities in Ambanja, Madagascar, and the Human Research Ethical Cantonal Board of Geneva, Switzerland (Comité d'éthique de la recherche CER 14-217; project number: CCER PB_2017-00641) approved the study.

We implemented an mHealth system to support providers in conducting ANC visits in Madagascar. Maternal mortality in Madagascar has decreased by over 50% in the last 20 years—from 776 to 353 deaths per 100,000 live births—but maternal mortality remains high and the Millennium Development Goal 5 has not been reached [18]. In the northern part of the country, 82.1% of women receive at least one ANC visit, but only 44% receive four or more ANC visits [19]. In the Ambanja district, only 58% of pregnant women attend four or more ANC visits with any kind of health care provider [20].

The Ambanja district is a rural area with a population of 240,000 inhabitants [21]. Farming is the main economic activity in the region. This study was conducted in the 2 main hospitals in Ambanja city: the Centre Médico Chirurgical Saint Damien (a private nonprofit organization) and Centre Hospitalier de Référence du District de Ambanja (the district public hospital). In the city, ANC is provided in these 2 hospitals and in 18 dispensaries within a 200-km radius of the Ambanja district. In 2015, we provided ANC with the PANDA system in both urban and rural areas; however, in 2016, we changed the strategy and excluded rural dispensaries, providing ANC exclusively in the city of Ambanja to allow for proper follow-up.

A collaboration between the University Hospitals of Geneva and the Centre Médico Chirurgical was established in 2010 for a cervical cancer program and was continued in 2015 with the implementation of the PANDA system. The PANDA team collaborates with the Ministry of Health of Madagascar.

The PANDA System Intervention

The PANDA mHealth system was first implemented in 2015 in a pilot study to assess the system's feasibility and usability in Madagascar [22]. The program was designed to support health care providers who received training on using a smartphone app and on conducting a standardized ANC visit. This mHealth system encourages providers to perform a standardized ANC visit as the program requires the provider to go through all the sections of a standard ANC visit, without skipping any section from the beginning to the end. Standardization is an intended

result of using an mHealth system such as PANDA. PANDA is a mobile app that facilitates the collection of information during ANC and postpartum care visits according to WHO recommendations. The collected data are automatically transmitted to the medical unit's web-based database, which stores a digitized medical record for each patient.

The system comprises the following 3 elements. First, *PANDA point of care* is a kit containing all the instruments needed to take measurements such as height, weight, body temperature, and blood pressure and to screen for syphilis, HIV, malaria, anemia, gestational diabetes, urinary tract infections, and malnutrition. Second, the *PANDA app* is an Android smartphone app made up of 5 sections: sociodemographic information, medical and obstetric history, clinical screening test results, health education and birth plan, and postpartum care. It is an app mostly based on icons and illustrations with very little text. A quarter of each visit is dedicated to educating the patient regarding warning signs during pregnancy, labor, and postpartum as well as birth preparedness. The education section of the visits also focuses on breastfeeding recommendations and overall well-being of women. Alert functions are integrated into the PANDA app to notify the provider of abnormal clinical results, technical problems, or missing patient information, thereby assuring a complete assessment [23]. The providers must go through all 4 modules; otherwise, it is not possible to close the visit. Third, the *PANDA medical unit* is a web-based database that captures the data and results collected during the ANC and postpartum visits and enables the hospital team to open a computerized medical record for each woman to make a diagnosis and define the frequency of follow-up. In addition, after all the data were collected for this study, the PANDA medical unit allowed the team to stratify pregnancies by area, time, and risk conditions. The on-site medical team has access to the PANDA database via username and password; all data are encrypted to ensure privacy. Patients were informed, and they agreed to their data being transmitted to and saved in the medical unit to create their medical records. Patients were asked to provide both paper and electronic consent.

Data were collected during the women's first and subsequent ANC visits. The 3 following types of information were collected: (1) sociodemographic characteristics; (2) medical, surgical, and obstetric history; and (3) results from screening to detect obesity or malnutrition, hypertension or preeclampsia, anemia, HIV, syphilis, malaria, diabetes, infections, and other conditions. The PANDA system collected data on at least 75 items per woman on their past and current medical and obstetric history as well as clinical screening data.

Since 2015, we have trained 13 providers to use the PANDA system. In the PANDA medical unit, it is possible to track the content of the visits, including each provider's activities, such as the number of visits conducted and eventual errors in completing the visits, thus allowing the team to build a learning curve for each provider.

Sample Description

A total of 1446 pregnant women fulfilled the inclusion criteria and were enrolled in the study. All pregnant women, regardless of age or stage of pregnancy, were eligible to participate in the study. The only exclusion criterion was the inability to understand or act as described in a previous publication explaining the acceptability and feasibility of the PANDA mHealth device [22]. As ANC with the PANDA mHealth system was offered to all women receiving routine ANC, the sample is considered to be representative.

Statistical Analysis

We used a convenience sample of 1446 women. We planned to recruit patients from 2015 until the end of September 2020, which resulted in a total of 1446 women.

The data collected with PANDA were digitized as electronic medical records, which were used to analyze maternal morbidity and evaluate the quality of the ANC. Continuous variables are presented as mean (SD) or median, and categorical variables are presented as frequencies and relative percentages. The proportions of patients who tested positive for syphilis, HIV, or malaria are provided with their 95% CIs. Comparisons of categorical variables by year were performed using the chi-square test, and the mean durations of ANC visits were compared by the different categorical variables using the nonparametric Kruskal-Wallis test. We assessed the associations of different patient characteristics, place of residence, year, and visit order (independent variables) with visit duration (dependent variable)—first at the univariate level, using mixed linear regression models with the patient as a random factor to take into account repeated measurements within patients. We then constructed a parsimonious multivariable mixed linear regression model including all the variables that were significantly associated with the visit duration at P=.04. We reported the estimated marginal mean duration and 95% CI for each independent variable from the multivariable model and the associated P values for each category of the variables. Stata Statistical Software, release 16 (StataCorp), was used to describe the study population and to analyze the indicators of the ANC quality.

Results

Demographic Information of the Participants

Most of the 1446 women in the study were recruited in the city of Ambanja. The first table in Multimedia Appendix 1 summarizes the participants' demographic and obstetric history. The mean age of the women was 24.4 years (SD 6.8), and the majority (1185/1441, 82.23%) of the participants were married or living with a partner. Most of the women had primary school–level education or lower, and most (1337/1441, 92.78%) of the participants did not have running water at home. Multiparas accounted for 60% of the participants. In our sample, more than one-third of the pregnancies were unplanned. Data on unplanned pregnancies were only included in the system in 2018 as we noticed during the visits that the women very often said that their pregnancies were not planned. In addition, data on contraceptive use before pregnancy were added in 2018 and 2019; of the women who reported previous contraceptive use in these years, 51.7% (170/329) used long-acting progesterone methods and 32.2% (106/329) used traditional methods. Only 154/985, 15.63% of the participants had been vaccinated for tetanus. A total of 4.66% (67/1438) of women reported

experiencing domestic violence during their pregnancy; importantly, several women only reported this at the end of their ANC visits, after the education session on violence.

Table 1 summarizes the status of syphilis, HIV, and malaria among the pregnant women included in this study. HIV prevalence was 1.3% (95% CI 0.7-1.96%). The prevalence of syphilis infection in our sample was 2.8% (95% CI 2.0-3.8%).

Table 1. Prevalence of malaria, HIV infection, and syphilis among pregnant women in the Ambanja district, Madagascar, from January 13, 2015, to September 20, 2019, at their first antenatal care visit (n=1443).

Variables	Participants, n (%)	95% CI
HIV status		
Positive	18 (1.25)	0.7-1.96
Negative	1290 (89.40)	87.7-90.9
Invalid or not tested	135 (9.36)	7.9-11.0
Syphilis status		
Positive	41 (2.84)	2.0-3.8
Negative	1061 (73.53)	71.2-75.8
Invalid or not tested	341 (23.63)	21.5-25.9
Malaria status		
Positive	23 (1.59)	1.0-2.4
Negative	813 (56.34)	53.7-58.9
Invalid or not tested	607 (42.06)	39.5-44.7

Timing of First Antenatal Care Visit

Table 2 shows that since the implementation of PANDA began in 2015, women tended to attend their first ANC visit earlier in their pregnancy. Comparing the timing of the first ANC visit in terms of pregnancy trimester by year, we observed a significant change from 2015 to 2019. In 2019, fewer women attended their first visit in the third trimester (57/277, 20.6%) than in 2015 (147/343, 42.9%). In addition, more women attended their first ANC visit in the second trimester in 2019 than in the previous years (P<.001).

Table 2. Timing of the first antenatal care visit during pregnancy by year among pregnant women in the Ambanja district, Madagascar.

Pregnancy trimester	Year, n (%)					*P* value
	2015 (n=341)	2016 (n=265)	2017 (n=197)	2018 (n=360)	2019 (n=276)	
First trimester	28 (8.2)	39 (14.7)	28 (14.2)	46 (12.8)	18 (6.5)	<.001
Second trimester	168 (49.0)	148 (55.9)	131 (66.5)	237 (65.8)	201 (72.8)	N/A[a]
Third trimester	147 (42.8)	78 (29.4)	38 (19.3)	77 (21.4)	57 (20.7)	N/A

[a]N/A: not applicable.

Mean Duration of Antenatal Care Visits Across Time and Associated Variables

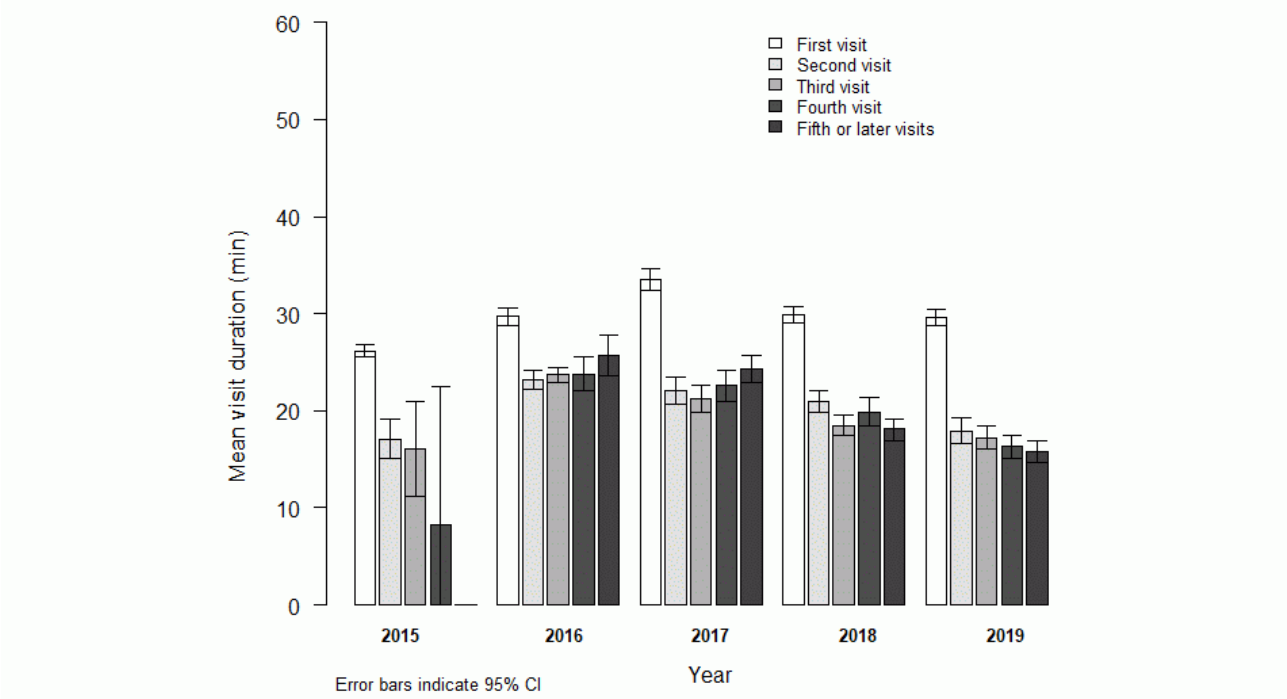
In the univariate analysis, the mean visit duration decreased significantly by approximately 5 min from 2015 to 2019 and varied by approximately 10 min across visit orders (see the second table in Multimedia Appendix 1). There was no statistically significant difference in the mean duration of the visits by the place of residence (rural vs urban). The visit duration was significantly associated with the patient's level of education, with longer visits for patients with less than primary education (mean 40.7 min; 95% CI 30.2-51.3 min) and patients who had completed university education (mean 25.3 min; 95% CI 24.4-26.3 min), compared with patients who had completed primary education (mean 23.3 min; 95% CI 22.9-23.8 min; P=.001) or secondary school (mean 22.8 min; 95% CI 22.0-23.4 min; P=.001). The visit duration also differed significantly by age, with significantly longer visits for women aged at least 35 years (mean 25.0 min; 95% CI 24.0-25.9 min) than for those aged between 16 and 20 years (22.9 min; 95% CI 22.3-23.5 min; P<.001). Patients who reported domestic violence during the visit had longer visit durations than those who did not report domestic violence (P<.001), and those who had running water at home had longer visit durations than those without running water at home. There were also significant differences in the visit duration by patient gravidity (longer duration with higher gravidity), parity (longer duration with higher parity), and trimester of pregnancy (shorter duration in the third [mean 21.7 min; 95% CI 21.3-22.1 min; P<.001] and second [mean 25.5 min; 95% CI 25.1-25.9 min; P<.001] trimesters, compared with the first trimester [mean 29.2 min; 95% CI 27.8-30.5 min]). The providers noted that the visit duration increased with parity and gravidity as they had to go through more detailed information about each past pregnancy and delivery. Active smokers had longer visit durations than did nonsmokers, and women with severe or moderate anemia had longer visit durations than

women with no anemia or mild anemia. Finally, there were significant differences in the visit duration by HIV status (longer duration among patients with positive HIV status compared with those with negative or unknown status), syphilis status (longer duration among patients with unknown status compared with those with positive or negative status), and malaria status (longer duration among patients with positive status compared with those with negative or unknown status). In the multivariable model, the mean visit duration remained independently associated with the year, visit order, level of education, age category, domestic violence, availability of running water at home, and HIV and syphilis status (Multimedia Appendix 1).

Figure 1 displays the mean visit duration by year and visit order. The duration of the first ANC visit remained stable over time. In 2015 and 2019, the mean duration decreased with each subsequent visit. In 2016 and 2017, the fifth visit was longer than the fourth visit, and the fourth visit was longer than the third visit in 2018.

All variables included in the multivariable model were significantly associated with the visit duration. These variables were the year, visit order, education, age, availability of running water at home, HIV status, and syphilis status.

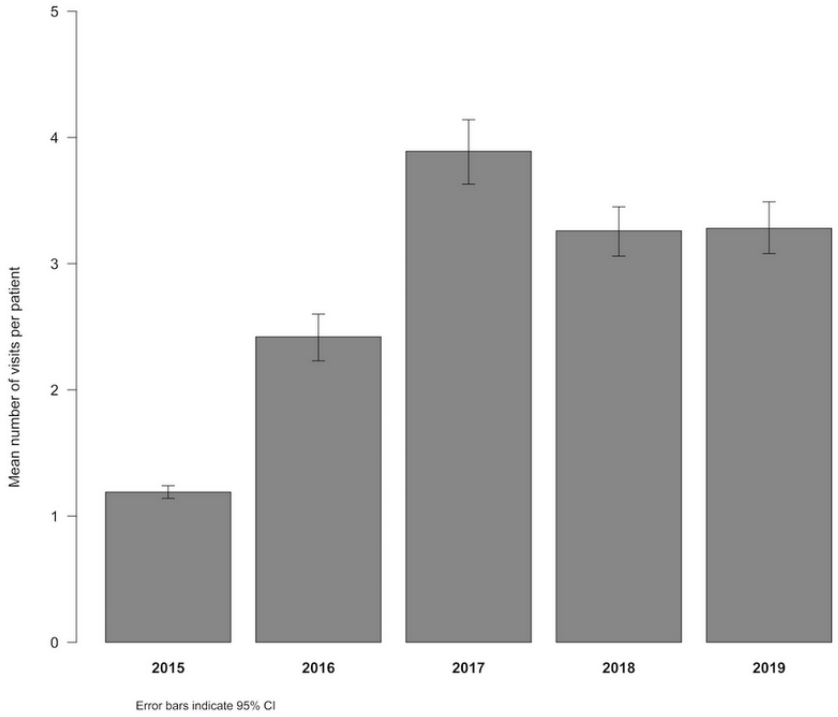
Figure 1. Mean antenatal care visit duration in minutes (95% CI) by year and visit order.



Number of Visits Per Woman by Year and Antenatal Coverage

We observed a statistically significant increase in the number of visits per woman from 2015 to 2017. The number of visits per woman then became stable after the third year of implementing the PANDA system. Figure 2 illustrates the mean number of visits per woman per year.

Figure 2. Mean number of antenatal care visits (95% CI) per patient by year (Kruskal-Wallis test; *P*<.001).



Discussion

Principal Findings and Interpretation

This study has shown the potential of using an mHealth system to encourage providers to follow a standardized ANC visit and also contribute to increase women's adherence and willingness to return for subsequent ANC visits in Ambanja, Madagascar.

According to the United Nations International Children's Emergency Fund data gathered in Madagascar in 2018, most women attended their first ANC visit in the third trimester of pregnancy, with only 45% attending their first visit in the second trimester or earlier [24]. In this study, we found that 73% of the women attended their first visit in the second trimester in 2019. This suggests that a systematically conducted ANC visit, supported and guided by mHealth technology, has the potential to boost women's confidence in health care providers. Similarly, in a randomized controlled trial conducted in 2017 in Ethiopia, Atnafu et al [25] found that, after implementing an mHealth tool to assist health care providers in collecting and organizing patient data in medical records, about 60% of the patients attended their first ANC visit from the fourth to the sixth month of their pregnancies.

There is no recommended benchmark value for the duration of ANC visits as a quality indicator; however, we assumed that a provider would need at least 20 min for the first ANC visit (based on our clinical experience) to cover all the main topics that should routinely be part of a high-quality ANC visit. The duration of the first ANC visit, a quality indicator defined by the WHO, remained stable over the study period, indicating that the use of the mHealth device did not significantly lengthen the visit duration. We also found that the visit duration was positively associated with several patient risk factors. Low education, being over 35 years of age, experiencing domestic violence during pregnancy, having anemia, and having an HIV positive status were all associated with a longer ANC visit duration.

In terms of the completeness of the ANC visits, most pregnant women in our sample were tested for HIV during these visits (92.7%), compared with only 10% of pregnant women in an analysis of national-level data from 2018, which also reported that only 3% were receiving antiretroviral therapy to prevent vertical transmission [26]. Furthermore, HIV prevalence was 1.3% (95% CI 0.7-1.96%) in our sample, compared with 0.5% (95% CI 0.1-0.2%) in the 2018 national-level data. The prevalence of syphilis infection in our sample was 2.8% (95% CI 2.0-3.8%). Previous reports of the seroprevalence of syphilis in Madagascar ranged from 1.6% to 4.5% [27]. In addition, the national-level data indicated that only 68% of pregnant women in Madagascar had their blood pressure measured during ANC visits [24], compared with almost 100% of the women in our sample using the PANDA system.

In this study, we found that the number of ANC visits per pregnant woman tended to increase over the study period. In the study by Atnafu et al, [25] the number of ANC visits varied significantly by region, with 77% of pregnant women attending 1 to 4 visits and only 23% attending more than four ANC visits. It can be assumed that increasing women's willingness to return for subsequent ANC visits may help prevent maternal morbidities that can put the mother and newborn at risk during pregnancy. Rosario et al [28] have shown that birth outcomes are directly related to ANC attendance (stillbirth: unadjusted odds ratio [OR]=0.34, 95% CI 0.16-0.70; abortion: OR=0.07, 95% CI 0.04-0.12) According to the national data, in 2018, only 51% of women in Madagascar attended at least four ANC visits [29].

Strengths of the Study

This study analyzed data from a large population over a 5-year period following the implementation of the PANDA mHealth system to support good quality ANC in a resource-constrained setting (Madagascar). The indicators in our study, such as the timing of the first ANC visit, the visit duration, and the number of visits, strongly suggest that this mHealth system encourages the performance of a standardized ANC visit and thereby facilitates the provision of high-quality ANC services. This study has shown the benefits of the PANDA mHealth system for both providers and patients. The providers received guidance on how to conduct standardized ANC of good quality, and they also received a record of all important patient data that can be used in follow-up visits. The women and their families received ongoing education and encouragement to seek appropriate care throughout the ANC and postpartum visits conducted with the PANDA system. The system improved communication between the health care workers and patients, facilitated continuous education, and encouraged the patients to play a more active role in the decision-making processes related to their health.

Limitations of the Study

Our study has several limitations. The most significant limitation is the lack of a control group to compare with participants using the PANDA mHealth system. Instead, we used national-level surveys as a standard reference—the 2009 Madagascar Demographic and Health Survey and the 2018 Multiple Indicator Cluster Survey. Although our results differed significantly from the findings of these two national-level surveys, the lack of a comparison group and a randomized design limits the conclusiveness of our findings. D-tree International's study on safer deliveries in Tanzania, which also lacked a control group, demonstrated that the examined mHealth program was a success as it reached over 13,000 pregnant women in Zanzibar. The implementation of this mHealth system was described as a success as it was linked to unprecedented rates of both service delivery and postpartum attendance, even though this mHealth system did not provide support for either ANC or postpartum visit content or quality [30]. A second limitation is that we did not measure health outcomes to evaluate the effectiveness of the mHealth intervention; we only began to collect data on postpartum visits and delivery outcomes in the last year of the study. However, the study by Lund et al [31] in Tanzania showed a trend toward improved time and quality after the implementation of a mobile phone intervention for patients in a cluster randomized control study in Tanzania. They found that the mobile phone intervention was associated with not only an increase in ANC attendance, 44% of the women in the intervention group received four or more ANC visits versus 31% in the control group (R, 2.39; 95% CI 1.03-5.55), but also a trend toward improved timing and quality of ANC services, although the difference was not significant. We also did not explore the cost effectiveness or durability of the PANDA system, compared with the existing paper-based checklist. These aspects need to be explored in future studies. Finally, we trained 13 providers to use the PANDA mHealth system, and we did not explore their satisfaction with the system. The providers were not selected randomly; rather, an opportunistic selection was employed.

Several questions about the use and diffusion of the PANDA mHealth system remain open. These questions relate to, for example, dealing with the low quality of service delivery and the scalable and sustainable integration of the PANDA system in the local health system.

Conclusions

This study shows that the PANDA mHealth system has the potential to improve the quality of ANC in a resource-limited setting, modify the behavior of providers by providing standardized ANC visits, and increase patient compliance. Mobile technology should not be considered a stand-alone health intervention for ANC; rather, it is a strategic tool for improving the delivery and quality of maternal health care. Further studies are necessary to better understand the conditions under which behavioral changes occur and persist over time with the use of mHealth systems such as PANDA as well as whether undesired behavioral changes may also arise with the use of mHealth in ANC settings.

Acknowledgments

The authors thank Angèle Gayet-Ageron, MD PhD MPH. for the support on the statistical analysis and Jennifer Barrett, PhD, from Edanz Group for editing the draft of this manuscript.

Authors' Contributions

AB participated in the study design, analysis, and writing. MR, MV, and NS took part in the revision of the paper. GS participated in the final revision of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sociodemographic and clinical characteristics of pregnant women in the Ambanja district, Madagascar, from January 13, 2015, to September 20, 2019, at their first antenatal care visit (Table 1). Sociodemographic and clinical variables that were significantly associated with antenatal care visit duration among pregnant women in the Ambanja district, Madagascar (univariate and multivariable analyses) (Table 4).

[DOCX File , 23 KB - mhealth_v8i7e18543_app1.docx]

References

1. Radin JM, Topol EJ, Steinhubl SR. A digital toolkit for improved maternal health. Lancet 2018 Sep;392(10153):1107. [doi: 10.1016/s0140-6736(18)32294-3]
2. Ilozumba O, Abejirinde IO, Dieleman M, Bardají A, Broerse JE, Van Belle S. Targeting strategies of mHealth interventions for maternal health in low and middle-income countries: a systematic review protocol. BMJ Open 2018 Feb 24;8(2):e019345 [FREE Full text] [doi: 10.1136/bmjopen-2017-019345] [Medline: 29478019]
3. Sigulem D, Ramos M, de Holanda AR. The new medicine: from the paper medical record to the digitized human being. In: Marin H, Massad E, Gutierrez MA, Rodrigues RJ, Sigulem D, editors. Global Health Informatics: How Information Technology Can Change Our Lives in a Globalized World. New York, USA: Academic Press; 2017:152-167.
4. Victora CG, Requejo JH, Barros AJ, Berman P, Bhutta Z, Boerma T, et al. Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. Lancet 2016 May 14;387(10032):2049-2059. [doi: 10.1016/S0140-6736(15)00519-X] [Medline: 26477328]
5. World Health Organization. 2019. Maternal Mortality URL: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality [accessed 2019-11-06]
6. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the sustainable development goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252 [FREE Full text] [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
7. Campbell OM, Graham WJ, Lancet Maternal Survival Series Steering Group. Strategies for reducing maternal mortality: getting on with what works. Lancet 2006 Oct 7;368(9543):1284-1299. [doi: 10.1016/S0140-6736(06)69381-1] [Medline: 17027735]
8. Ogbo FA, Dhami MV, Ude EM, Senanayake P, Osuagwu UL, Awosemo AO, et al. Enablers and barriers to the utilization of antenatal care services in India. Int J Environ Res Public Health 2019 Aug 29;16(17) [FREE Full text] [doi: 10.3390/ijerph16173152] [Medline: 31470550]
9. Downe S, Finlayson K, Tunçalp O, Gülmezoglu AM. What matters to women: a systematic scoping review to identify the processes and outcomes of antenatal care provision that are important to healthy pregnant women. BJOG 2016 Mar;123(4):529-539 [FREE Full text] [doi: 10.1111/1471-0528.13819] [Medline: 26701735]
10. Blencowe H, Cousens S, Jassir F, Say L, Chou D, Mathers C, Lancet Stillbirth Epidemiology Investigator Group. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. Lancet Glob Health 2016 Mar;4(2):e98-108 [FREE Full text] [doi: 10.1016/S2214-109X(15)00275-2] [Medline: 26795602]
11. Dixit P, Khan J, Dwivedi LK, Gupta A. Dimensions of antenatal care service and the alacrity of mothers towards institutional delivery in South and South East Asia. PLoS One 2017;12(7):e0181793 [FREE Full text] [doi: 10.1371/journal.pone.0181793] [Medline: 28742809]
12. World Health Oranization. WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. Geneva, Switzerland: World Health Organization; 2016.
13. Grenier L, Suhowatsky S, Kabue MM, Noguchi LM, Mohan D, Karnad SR, et al. Impact of group antenatal care (G-ANC) versus individual antenatal care (ANC) on quality of care, ANC attendance and facility-based delivery: a pragmatic cluster-randomized controlled trial in Kenya and Nigeria. PLoS One 2019;14(10):e0222177 [FREE Full text] [doi: 10.1371/journal.pone.0222177] [Medline: 31577797]
14. Chatfield A. Maternal Health Task Force. 2014. Reflecting on the 'mHealth for Maternal Health: Bridging the Gaps' Technical Meeting URL: https://www.mhtf.org/2014/04/16/reflecting-on-the-mhealth-for-maternal-health-bridging-the-gaps-technical-meeting/ [accessed 2020-06-09]
15. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
16. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and appropriateness of mhealth interventions for maternal and child health: systematic review. JMIR Mhealth Uhealth 2018 Jan 9;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]
17. Zaidi S, Shaikh SA, Sayani S, Kazi AM, Khoja A, Hussain SS, et al. Operability, acceptability, and usefulness of a mobile app to track routine immunization performance in rural Pakistan: interview study among vaccinators and key informants. JMIR Mhealth Uhealth 2020 Mar 13;8(2):e16081 [FREE Full text] [doi: 10.2196/16081] [Medline: 32053115]
18. United Nations Economic Commission for Africa. 2013. MDG Report 2013 URL: https://www.undp.org/content/dam/undp/library/MDG/english/MDG%20Regional%20Reports/Africa/MDG%20Report2013_ENG_Fin_12June.pdf [accessed 2020-06-09]
19. Institut National de la Statistique Madagascar (INSTAT Madagascar). Institut National de la Statistique et Partenaires Techniques Financiers _Enquête Nationale sur le Suivi des Objectifs du Millénaire pour le Développement à Madagascar URL: https://www.instat.mg/wp-content/uploads/2016/11/INSTAT_Ensomd_Obj4-2012-2013.pdf [accessed 2019-11-10]

20. United Nations International Children's Emergency Fund (UNICEF). Antenatal Care URL: https://data.unicef.org/topic/maternal-health/antenatal-care/ [accessed 2020-11-10]
21. City Population. Ambanja: District in Madagascar URL: https://www.citypopulation.de/php/madagascar-admin.php?adm2id=71719 [accessed 2019-11-10]
22. Benski AC, Stancanelli G, Scaringella S, Herinainasolo JL, Jinoro J, Vassilakos P, et al. Usability and feasibility of a mobile health system to provide comprehensive antenatal care in low-income countries: PANDA mhealth pilot study in Madagascar. J Telemed Telecare 2017 Jul;23(5):536-543. [doi: 10.1177/1357633X16653540] [Medline: 27365320]
23. Salem A, Lacour O, Scaringella S, Herinianasolo J, Benski AC, Stancanelli G, et al. Cross-sectional survey of knowledge of obstetric danger signs among women in rural Madagascar. BMC Pregnancy Childbirth 2018 Feb 5;18(1):46 [FREE Full text] [doi: 10.1186/s12884-018-1664-x] [Medline: 29402226]
24. United Nations International Children's Emergency Fund (UNICEF). 2018. MICS6 Madagascar 2018 URL: https://www.unicef.org/madagascar/media/2451/file/MICS6-Madagascar-2018-Health-maternalnewborn.pdf [accessed 2020-06-09]
25. Atnafu A, Otto K, Herbst CH. The role of mhealth intervention on maternal and child health service delivery: findings from a randomized controlled field trial in rural Ethiopia. Mhealth 2017;3:39 [FREE Full text] [doi: 10.21037/mhealth.2017.08.04] [Medline: 29184891]
26. UNAIDS. Madagascar URL: https://www.unaids.org/en/regionscountries/countries/madagascar [accessed 2019-11-12]
27. Frickmann H, Schwarz N, Girmann M, Hagen R, Poppert S, Crusius S, et al. Serological survey of HIV and syphilis in pregnant women in Madagascar. Trop Med Int Health 2013 Jan;18(1):35-39 [FREE Full text] [doi: 10.1111/tmi.12007] [Medline: 23094758]
28. Rosário EV, Gomes MC, Brito M, Costa D. Determinants of maternal health care and birth outcome in the Dande health and demographic surveillance system area, Angola. PLoS One 2019;14(8):e0221280 [FREE Full text] [doi: 10.1371/journal.pone.0221280] [Medline: 31437180]
29. UNFPA Madagascar,. 2013. Madagascar Millennium Development Goals National Monitoring Survey URL: https://madagascar.unfpa.org/sites/default/files/pub-pdf/OMD_Summary_0.pdf [accessed 2020-06-09]
30. Battle JD, Farrow L, Tibaijuka J, Mitchell M. mHealth for safer deliveries: a mixed methods evaluation of the effect of an integrated mobile health intervention on maternal care utilization. Healthc (Amst) 2015 Dec;3(4):180-184. [doi: 10.1016/j.hjdsi.2015.10.011] [Medline: 26699340]
31. Lund S, Nielsen BB, Hemed M, Boas IM, Said A, Said K, et al. Mobile phones improve antenatal care attendance in Zanzibar: a cluster randomized controlled trial. BMC Pregnancy Childbirth 2014 Jan 17;14:29 [FREE Full text] [doi: 10.1186/1471-2393-14-29] [Medline: 24438517]

Abbreviations

ANC: antenatal care
LMIC: low- and middle-income country
mHealth: mobile health
OR: odds ratio
PANDA: Pregnancy and Newborn Diagnostic Assessment
WHO: World Health Organization

Edited by G Eysenbach; submitted 04.03.20; peer-reviewed by S Farag; comments to author 26.03.20; revised version received 06.04.20; accepted 14.05.20; published 08.07.20.

Please cite as:
Benski AC, Schmidt NC, Viviano M, Stancanelli G, Soaroby A, Reich MR
Improving the Quality of Antenatal Care Using Mobile Health in Madagascar: Five-Year Cross-Sectional Study
JMIR Mhealth Uhealth 2020;8(7):e18543
URL: https://mhealth.jmir.org/2020/7/e18543
doi:10.2196/18543
PMID:32673263

Original Paper

A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study

Carol Ann Maher[1], PhD; Courtney Rose Davis[1], PhD; Rachel Grace Curtis[1], PhD; Camille Elizabeth Short[2], PhD; Karen Joy Murphy[3], PhD

[1]Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

[2]Melbourne Centre for Behaviour Change, School of Psychological Sciences and School of Health Sciences, University of Melbourne, Melbourne, Australia

[3]Alliance for Research in Exercise, Nutrition and Activity, Clinical and Health Sciences, University of South Australia, Adelaide, Australia

Corresponding Author:
Carol Ann Maher, PhD
Alliance for Research in Exercise, Nutrition and Activity
Allied Health and Human Performance
University of South Australia
GPO Box 2471
Adelaide, 5001
Australia
Phone: 61 883022315
Email: carol.maher@unisa.edu.au

Abstract

Background: Poor diet and physical inactivity are leading modifiable causes of death and disease. Advances in artificial intelligence technology present tantalizing opportunities for creating virtual health coaches capable of providing personalized support at scale.

Objective: This proof of concept study aimed to test the feasibility (recruitment and retention) and preliminary efficacy of physical activity and Mediterranean-style dietary intervention (MedLiPal) delivered via artificially intelligent virtual health coach.

Methods: This 12-week single-arm pre-post study took place in Adelaide, Australia, from March to August 2019. Participants were inactive community-dwelling adults aged 45 to 75 years, recruited through news stories, social media posts, and flyers. The program included access to an artificially intelligent chatbot, Paola, who guided participants through a computer-based individualized introductory session, weekly check-ins, and goal setting, and was available 24/7 to answer questions. Participants used a Garmin Vivofit4 tracker to monitor daily steps, a website with educational materials and recipes, and a printed diet and activity log sheet. Primary outcomes included feasibility (based on recruitment and retention) and preliminary efficacy for changing physical activity and diet. Secondary outcomes were body composition (based on height, weight, and waist circumference) and blood pressure.

Results: Over 4 weeks, 99 potential participants registered expressions of interest, with 81 of those screened meeting eligibility criteria. Participants completed a mean of 109.8 (95% CI 1.9-217.7) more minutes of physical activity at week 12 compared with baseline. Mediterranean diet scores increased from a mean of 3.8 out of 14 at baseline, to 9.6 at 12 weeks (mean improvement 5.7 points, 95% CI 4.2-7.3). After 12 weeks, participants lost an average 1.3 kg (95% CI –0.1 to –2.5 kg) and 2.1 cm from their waist circumference (95% CI –3.5 to –0.7 cm). There were no significant changes in blood pressure. Feasibility was excellent in terms of recruitment, retention (90% at 12 weeks), and safety (no adverse events).

Conclusions: An artificially intelligent virtual assistant-led lifestyle-modification intervention was feasible and achieved measurable improvements in physical activity, diet, and body composition at 12 weeks. Future research examining artificially intelligent interventions at scale, and for other health purposes, is warranted.

(JMIR Mhealth Uhealth 2020;8(7):e17558) doi:10.2196/17558

KEYWORDS

virtual assistant; chatbot; Mediterranean diet; physical activity; lifestyle

Introduction

Poor diet and physical inactivity are amongst the leading modifiable causes of death and disease and increase the risk of developing chronic health conditions such as type 2 diabetes, cardiovascular disease, obesity, cancers, depression, and anxiety [1]. In highly developed countries, these unfavorable health behaviors are pervasive; for example, over ninety percent of Australian adults fail the recommended daily intake of vegetables and fruit, and approximately two-thirds do not meet guidelines of 30 min of physical activity most days [2,3].

Despite the overwhelming benefits of healthy eating and physical activity, making and sustaining lifestyle modification is tremendously challenging [4]. Specialized support from health professionals, such as dietitians and exercise physiologists or physiotherapists, can help people achieve lifestyle change. However, health systems' finite budgets typically limit access to such services to specific patient populations, such as those with established chronic disease [5]. Consequently, many people with unfavorable lifestyle patterns, who in time will go on to develop chronic disease, do not have access to health professional support to help them modify their lifestyle and prevent disease.

Technology advances in the fields of artificial intelligence and wearables offer new opportunities to deliver accessible, personalized, and cost-efficient behavior modification programs. In particular, virtual assistants use artificial intelligence to mimic human conversation and can be programmed with scripted conversations, questions and answers, and the ability to provide personalized responses based on input from the user. Basic virtual assistants have limited functionality and restrict users' inputs (eg, by offering questions with multiple-choice response options). For example, weight-loss virtual assistant "Lark" led to significant weight loss, comparable in magnitude to that achieved in in-person lifestyle interventions.[6] However, Lark only allowed users to respond to program-directed questions and did not allow users to ask questions. In contrast, sophisticated virtual assistant technology is capable of recognizing free-written or spoken language (natural language processing), enabling more natural and user-directed communication. Compared to more traditional knowledge transfer methods in digital health (eg, tailored modules), the conversational methods in virtual assistance may be perceived as more personally relevant and humanistic, which is critical given that perceived relevance and the inclusion of social support are associated with increased effect sizes in digital behavior change interventions [7,8].

Preliminary research supports the acceptability and efficacy of natural language virtual assistants in health; however, few studies have been conducted to date. Laranjo and colleagues' 2018 systematic review of studies evaluating natural language processing virtual assistants in healthcare identified just 14 studies, most of which focused on mental health (n=6) or clinician decision making (n=4) [9]. Single studies (n=3) used virtual assistants to provide patients with education and support for asthma, sexual health, and language impairment [9].

Using an identical search strategy, we updated this review in January 2020 and identified a further 11 studies. Similar to the previous review, most recent studies were aimed at mental health self-management (n=5) [10-12]. Single studies (n=3) used virtual assistants to provide patients with education and support for breast cancer [13], genetic counseling [14], and clinician training [15]. Despite the potential for natural language virtual assistants to provide personalized information and support users to engage in positive health behaviors, only three natural language processing virtual assistants focus on lifestyle modification. "J'arrête de fumer" assisted users in quitting smoking by profiling smoking behavior and providing advice and support at times when cravings were likely to occur [16]. A feasibility study published in 2019 employed a virtual assistant "Tess" to support obese adolescents in health-promoting behavior change with success [17]; however, Tess played a support role only and was designed to supplement ongoing in-person hospital-based services. Additionally, "Reflection Companion" prompted participants to reflect on their physical activity patterns, goals, barriers, and enablers of activity but was unable to provide positive affirmation or encouragement, answer participants' questions, or provide tailored advice [18].

In sum, although the use of virtual assistants in healthcare is a rapidly developing field, few virtual assistants use natural language processing to promote lifestyle change and those that do remain basic, performing limited, highly specific processes. To our knowledge, there are currently no published studies examining fully automated, natural language processing artificial intelligence health coaches for lifestyle modification.

This proof-of-concept study aimed to assess the feasibility of an artificial intelligence virtual health coach-led physical activity and diet intervention for community-dwelling middle-aged and older adults (recruitment and retention) and to evaluate the preliminary efficacy of the virtual health coach intervention for improving physical activity, Mediterranean-style diet adherence, and health risk factors.

Methods

The case-controlled study was approved by the University of South Australia Human Research Ethics Committee (ref 201724). All participants provided informed consent before commencing the study.

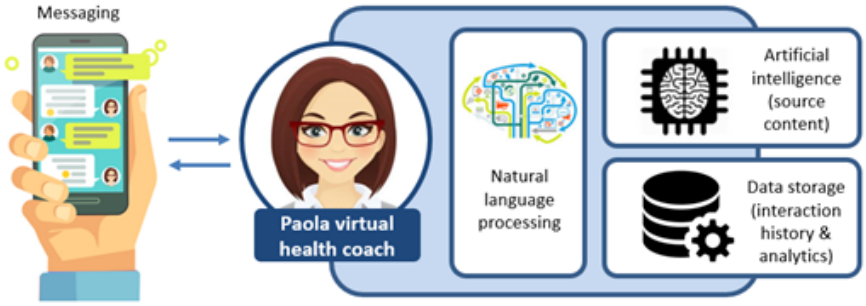
Intervention

The 12-week MedLiPal program assists users in increasing their lifestyle physical activity and adopt a Mediterranean-style diet. It incorporates numerous behavior change techniques to target determinants of lifestyle behaviors, including goal-setting, problem-solving, goal review, self-monitoring with feedback, social support, reattribution, and use of credible sources [19]. The components include Paola, a wearable activity monitor, the MedLiPal website, and a diet and activity log.

'Paola' is a natural language processing virtual health coach (Figure 1), created using IBM Watson Virtual Assistant artificial intelligence software (IBM Corp.). Paola performs three key roles: (1) guides participants through an introductory session during which she teaches them about the principles of physical

activity, goal-setting and self-monitoring, and the Mediterranean diet, including recommended daily and weekly servings (based on Davis protocol) [20]; (2) guides users through 11 weekly check-ins regarding their daily steps and dietary pattern for the past week, and assists users to set personalized step and diet goals for the coming week; and (3) is available 24/7 to answer users' questions. Paola was deployed on the cloud-based instant messaging platform Slack (Slack Technologies). Participants were required to download the Slack application on their device and create a user account to access Paola.

Figure 1. Overview of the technology underpinning Paola, the virtual health coach.



Paola referred to users by their first name and could respond to questions at any time. Frequently asked questions (eg, "How much bread can I eat?") were categorized by the *intent* (eg, "how much" type questions) and by the *entity* (eg, breads and cereals). The natural language processing artificial intelligence software underpinning Paola identified intents and entities and selected an appropriate response from stored dialogue options (eg, "You can have up to 5 servings of breads and cereals per day"). Intents, entities, and dialogue selections were developed through collaboration between the health scientists on the study team and the software developers.

We provided participants with a Garmin Vívofit4 wearable activity monitor, which allows users to monitor their daily physical activity levels (Garmin). During the introductory session, Paola assisted participants to set a personal daily step count goal based on age-based normative values +2000 steps, considering their current activity levels. This daily step goal was revisited and edited at each weekly check-in.

The MedLiPal website housed educational videos, study information, and useful links. During the baseline session, Paola linked participants to the website from Slack, requiring them to log in and view the educational videos. Participants could then visit the website for information and recipes at any time. The diet and activity log was used to record daily steps and dietary servings (Multimedia Appendix 1) [20,21]. Details of the Mediterranean diet recommendations are included as Multimedia Appendix 2.

Participants

Volunteers were recruited during February and March 2019 through mainstream news items, flyers, and social media. Participants were eligible if they were aged 45-75 years, owned an iOS or Android smartphone or tablet with internet access, were not currently meeting the Australian physical activity guidelines, and were not following a Mediterranean dietary pattern. Participants completed an online screening survey, followed by a brief phone interview to confirm eligibility, answer any potential questions participants may have had, and arrange their baseline appointment. Participants' physical activity level was determined based on a single question enquiring whether they did more than 30 minutes of activity on five or more days per week, while Mediterranean diet pattern was determined using the 14-item Mediterranean diet questionnaire, with a score of 7 or less indicating they were not following the Mediterranean diet [22]. Participants were excluded if they were unable to consume a Mediterranean diet due to allergy or other food aversions, had a medical condition precluding them from increasing their physical activity or were pregnant or lactating. All participants read the information sheet and were provided an opportunity to ask questions before providing informed consent and commencing the study.

Procedure

The primary outcomes were total minutes of weekly moderate-to-vigorous physical activity and Mediterranean diet adherence, measured via an online survey at baseline, week 6, and week 12. Total minutes of weekly moderate to vigorous physical activity was assessed using the Active Australia Survey (AAS) [23,24]. Mediterranean diet adherence was measured using a 14-item Australian Mediterranean diet adherence tool, adapted from the Prevención con Dieta Mediterránea (PREDIMED) study [25] to align with the Australian food supply [26]. The Australian Mediterranean diet adherence tool

has been validated relative to the Mediterranean diet score calculated from a 3-day weighed food record r=0.44 [26].

Participants attended in-person assessments at baseline, 6 weeks, and 12 weeks for secondary outcome measures. Secondary outcomes included: body weight (Seca 703), measured while clothed and with shoes removed; height measured without shoes (Seca 206); waist circumference, measured at the midpoint between the pelvic crest and bottom rib, unless a visual narrow was present elsewhere (Lufkin Thinline 2 mm, Apex Tool Group) [27]; systolic and diastolic blood pressure (Omron Healthcare) [26].

Sociodemographic characteristics (sex, age) and medication use were captured in the online survey. Socioeconomic status was measured according to the Socio-Economic Index for Areas disadvantage index, based on postcode [28].

Following the completion of the baseline assessments, participants were provided with a Garmin Vívofit4 activity tracker and access to the virtual health coach and the MedLiPal website. Initial login prompted Paola to begin the introductory session, in which the health coach introduced herself and taught users about the Mediterranean lifestyle and its benefits, the principles of increasing physical activity, goal-setting, and how to use the Garmin Vívofit4 activity tracker, the Mediterranean dietary pattern, and how to self-monitor daily and weekly servings using the MedLiPal weekly log sheet. Paola then invited the participant to converse with her regarding any questions they had about the program. Paola also explained to participants that she was available 24/7 to answer their questions.

Each week, users received an email notification inviting them to complete a weekly check-in with Paola to check their progress, assisting them with setting new weekly goals and answering their questions.

Participants did not receive an honorarium but could keep their Garmin Vívofit4 activity tracker after the study.

Feasibility

Feasibility was judged based on recruitment, retention, and engagement. We sought to recruit 30 participants to the study within 6 weeks, on the rationale that if this could be achieved with no advertising budget, a future definitive trial with a dedicated recruitment budget should be able to recruit approximately 400 participants over a year. We set a retention threshold of 75% at 12 weeks. Behavioral engagement with the MedLiPal program was assessed via virtual health coach usage data (number of weekly check-ins completed). The a priori engagement target was set at 70% (ie, that participants would complete at least 8 of 11 weekly sessions with Paola).

Power and Statistical Analyses

The target sample size was set at n=30, which was considered adequate to answer feasibility research questions. For efficacy outcomes, this sample size, assuming 80% power, an alpha of 0·05, and a single group design with three repeated measures, was able to detect an effect size of d=0.48.

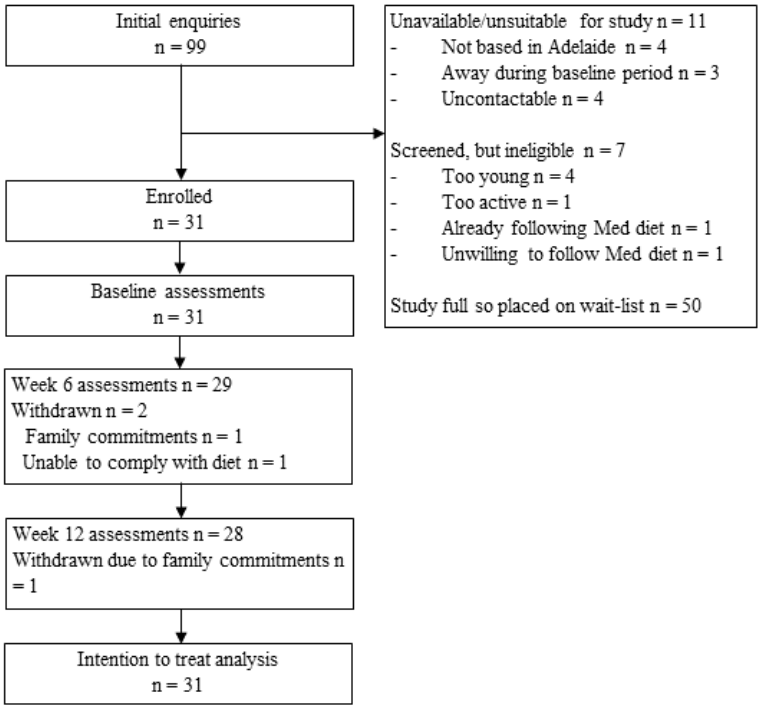
Feasibility data and participants' sociodemographic characteristics were analyzed descriptively, using means and standard deviations, frequencies, and percentages. Efficacy was assessed using an unadjusted repeated measures ANOVA, conducted on an intention-to-treat principle. Given that a small amount of data was missing (16% missing data at week 6, 10% at week 12 for primary outcomes), imputation was achieved using the last observation carried forward. Sensitivity analyses using complete cases were also conducted. All analyses were performed using SPSS (version 26, IBM) with a P value of .05. Pairwise comparisons included Bonferroni adjustment, and exact P values are reported.

Results

Recruitment and Retention

A total of 99 potential participants formally expressed their interest in the study within four weeks of recruitment opening, at which point expressions of interest were closed. The first 38 potential participants completed the eligibility screening interview, of which 31 (82%) were eligible. Participants were sequentially booked for baseline assessments until the study quota of 30 participants was filled, with n=31 ultimately enrolled. Attendance for the 6- and 12-week assessments was 29 and 28, respectively (Figure 2).

Figure 2. Participant flow through the study.



Participant Characteristics

Participants' baseline demographic and clinical characteristics are provided in Table 1. On average, participants were aged 56.2 years (SD 8.0), and two-thirds were women. Most participants were either overweight or obese.

Table 1. Baseline characteristics of participants in the MedLiPal study.

	Men (n=10)	Women (n=21)	All (n=31)
Age (years), mean (SD)	60.6 (8.7)	54.1 (7.0)	56.2 (8.0)
BMI, n (%)			
Underweight	0 (0)	0 (0)	0 (0)
Normal	1 (10)	7 (33)	8 (26)
Overweight	4 (40)	6 (29)	10 (32)
Obese	5 (50)	8 (38)	13 (42)
Weekly MVPA[a] (minutes), mean (SD)	308.7 (172.2)	157.2 (143.0)	206.1 (166.5)
Mediterranean diet adherence score (out of 14), mean (SD)	3.9 (2.1)	3.8 (1.6)	3.8 (1.8)
Waist circumference (cm), mean (SD)	102.4 (8.5)	93.0 (18.0)	96.0 (16.0)
SBP (mmHg)[b], mean (SD)	135.8 (11.1)	120.8 (10.6)	125.7 (12.8)
DBP (mmHg)[c], mean (SD)	85.3 (11.2)	81.6 (5.7)	82.8 (7.9)
Socioeconomic status[d], mean (SD)	997.4 (62.5)	1001.9 (64.3)	1000.5 (62.8)

[a]MVPA: moderate and vigorous physical activity.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]Socio-Economic Index for Areas (SEIFA) national mean = 1000±100.

Behavioral and Health Outcomes

The efficacy results are shown in Table 2. From baseline to week 6, weekly physical activity increased by approximately one hour, and then by a further 50 minutes at 12 weeks. Thus overall, from baseline to 12 weeks, physical activity increased by 109.8 minutes (95% CI 1.9 to 217.7, *P*=.005). Mediterranean diet adherence increased markedly from baseline to week 6 and then was approximately maintained at this level from week 6 to week 12 (mean change baseline to 12 weeks 5.7, 95% CI 4.2 to 7.3, *P*<.001).

On average, participants lost 1.1 kg from baseline to week 6, and then lost a further 0.2 kg to week 12, resulting in an overall average loss of 1.3 kg (95% CI –2·5 to –0·7, *P*=.01). Similarly, waist circumference decreased by 1 cm from baseline to week 6, and then another 1 cm to week 12, leading to an overall loss of –2.1cm (95% CI –3.5 to –0.7, *P*=.003). There was no change in blood pressure (diastolic or systolic) at either time point.

Sensitivity analyses (Multimedia Appendix 3) were conducted using complete case data (n=28). Results were consistent with the intention-to-treat (ITT) analyses, though effect sizes tended to be slightly larger; for example, physical activity increased by 145 minutes (vs 110 minutes in the ITT analyses) and that Mediterranean diet adherence score increased by 6.5 points (vs 5.7 points in the ITT analysis).

Table 2. Outcome measures at baseline, 6 weeks, and 12 weeks.

Outcome measure	Baseline	6 weeks	12 weeks	*F (2, 29), P* for overall model	Difference from baseline to Week 6 (95% CI)[a]	Difference from baseline to Week 12 (95% CI)[a]
	Mean (SD)			Mean (SD)	Mean (SD)	
Weekly total MVPA[b] minutes (min)	206.1 (166.5)	266.8 (207.2)	315.9 (261.7)	6.45, .005	60.8 (–21.7 to 143.3)	109.8 (1.9 to 217.7)
Mediterranean diet adherence score (out of 14)	3.8 (1.8)	9.8 (3.7)	9.6 (3.1)	44.56, <.001	6.0 (4.3 to 7.7)	5.7 (4.2 to 7.3)
Weight (kg)	83.6 (19.0)	82.4 (18.3)	82.3 (18.1)	5.41,.01	–1.1 (–2.0 to –0.3)	–1.3 (–2.5 to –0.1)
Waist circumference (cm)	96.0 (16.0)	95.1 (15.8)	93.9 (15.8)	7.13,.003	–1.0 (–1.8 to –0.1)	–2.1 (–3.5 to –0.7)
SBP[c] (mmHg)	125.7 (12.8)	124.9 (14.5)	125.5 (13.8)	0.11,.90	–0.8 (–4.9 to 3.4)	–0.2 (–5.7 to 5.3)
DBP[d] (mmHg)	82.8 (7.9)	81.8 (9.3)	81.8 (8.8)	0.64, .54	–1.0 (–3.4 to 1.3)	–1.0 (–4.3 to 2.4)

[a]Pairwise comparisons confidence intervals include Bonferroni adjustment for multiple comparisons.

[b]MVPA: moderate and vigorous physical activity.

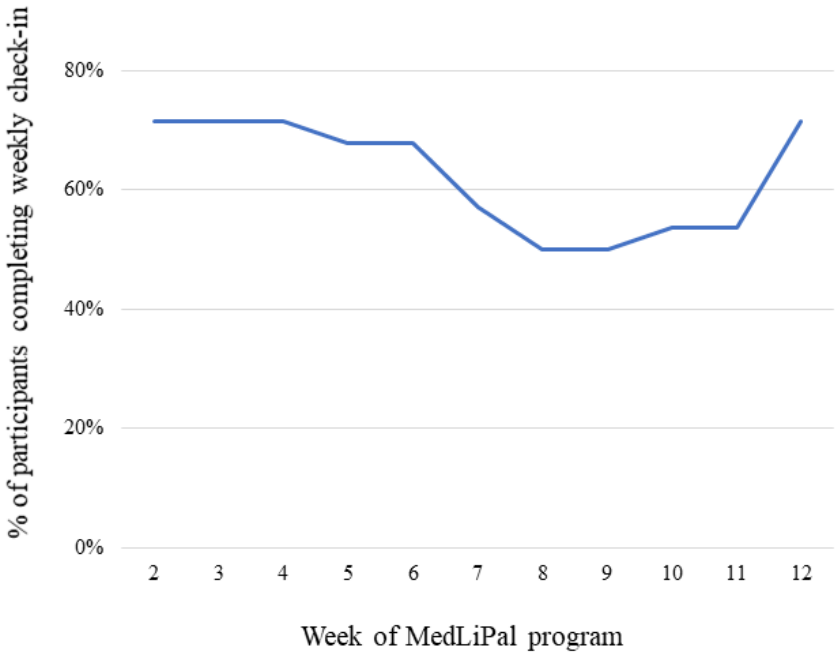
[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

Engagement With the Virtual Health Coach

From Weeks 2-12 of the program, participants were encouraged to complete weekly check-ins with Paola, the virtual health coach. Out of a maximum of 11 possible check-ins, participants completed an average of 6.9 (64%, range 1-11). Engagement with the check-ins across the intervention period is shown in Figure 3. Around 70% of participants completed the check-ins in weeks two, three, four and 12, with engagement gradually sliding to around 50% through weeks eight and nine, before rising again in the final weeks of the program. Participants who completed the first weekly check-in had significantly higher engagement across the intervention period than those who did not complete the first weekly check-in (mean 7.4, SD 3.4 vs mean 4.6, SD 2.6; *P*=.03).

Figure 3. The percentage of participants completing the weekly virtual health coach check-in across the 12-week intervention period.



Other Feasibility Issues

Of the 31 participants meeting all eligibility criteria and enrolled in the study, it became clear during baseline appointments that 3 participants had minimal smartphone skills (eg, did not know how to send a text message or download an app). These participants were reliant on their partners or children to use their phones as part of the study. At the 6-week assessment, several participants reported difficultly consuming the recommended number of food group servings (due to feeling satiated). They were encouraged to ask Paola for advice. Paola's advice to users reporting satiation was to reduce portion sizes across all food groups and to maintain a variety of healthy foods

Adverse Events

No adverse events related to study participation were experienced.

Discussion

Principal Results

This study aimed to assess the feasibility of a highly innovative virtual health coach physical activity and diet intervention for middle-aged and older adults. The study found a high level of interest in the technology-based intervention approach amongst the target demographic and high retention across the 12-week study period. Preliminary evidence suggests that the intervention led to sizable improvements in physical activity, diet, and body composition across 12 weeks.

The intervention approach was well-received by participants, exceeding our a priori expectations for recruitment and retention. Engagement with the virtual health coach (based on completed weekly check-ins) was slightly lower than our a priori threshold of ≥70% (actual completion rate 64%). Interestingly, the most significant drop off with the virtual health coach occurred between the baseline appointment and week two of the program (ie, the first check-in participants were supposed to complete on their own), with around a quarter of participants failing to complete the first weekly check-in. Three participants who enrolled in the study were found to have very low smartphone literacy at the baseline session, which impacted their ability to use the program as intended. In some cases, technical problems with Slack or the software interrupted attempts to check in, which generally were rectified after week 1. It appears that getting participants to complete their first independent check-in successfully may be a strong predictor of ongoing engagement, and a specific target in future program iterations. It may be useful to provide participants with a courtesy follow-up phone call after the first week to ensure they were able to complete the first check-in and proactively troubleshoot issues. Allowances for low smartphone literacy could be addressed in the study protocol, for example, by offering in-person assistance and written step-by-step instructions and troubleshooting information.

Our study attracted mostly women, who were similar to national averages for socioeconomic advantage (national average 1000) [28]. Our sample had poor adherence to Australian healthy eating guidelines (baseline average self-reported vegetable servings was 2.2/day and fruit 1.3/day compared to the recommended 5 and 2 servings per day, respectively) [29]. They also exhibited above-average rates of obesity and overweight compared with population norms [30] and physical activity rates comparable to other Australian studies of older adults measured using the AAS [31-33]. This population appears to be ideally suited to this type of new offering, given that they are not unwell enough to qualify for face-to-face services through the health

system, yet have modifiable risk factors for the future development of chronic diseases such as type 2 diabetes and cardiovascular disease.

Feasibility data collected in the study are invaluable to inform future trials, which could take the form of a traditional randomized controlled trial to determine the efficacy of the current intervention package or utilize a modern research design, such as the micro-randomization trial design, which allows multiple intervention elements to be experimentally evaluated [34]. The latter is useful for optimizing intervention packages for future use. Results suggest there is considerable enthusiasm amongst older adults for this style of intervention and confirm that most older adults have sufficient skills to access a virtual health coach program and use it effectively to achieve sizeable behavior and health change. Future research may also focus on thoroughly understanding the user experience of virtual health coach programs, which would be useful to inform improvements, for example, in terms of the virtual health coach's language style, variety of language, and possibly, sense of humor (which improves the humanity and emotional connection of artificial-intelligence virtual assistants [35]). Given the rapid technological advancements occurring in this field, there are many other opportunities for enhancing this intervention to address poor health behaviors, which may allow targeting of specific health issues, such as weight loss or diabetes self-management.

Comparison With Prior Work

To our knowledge, this is the first study to evaluate the capacity of an advanced virtual assistant technology to deliver a personalized lifestyle intervention amongst older adults. The magnitude of change in physical activity was similar to [36,37] or larger than [38,39] that achieved in previous computer-tailoring interventions. Improvement in Mediterranean diet adherence was much more significant than those achieved by previous technology-based Mediterranean diet interventions [40,41]. The degree of behavior change achieved was approximately in line with that reported in previous studies using intensive one-on-one dietary counseling [26,42]. These exciting findings suggest that artificial intelligence technology may be able to provide practical, supportive dietary counseling at a low cost.

Limitations

This study's primary strength is its ground-breaking intervention approach of using an artificially intelligent virtual health coach to deliver a personalized physical activity and dietary intervention program. Our virtual health coach used natural language processing, which allows users to converse using freely written language, representing a significant advancement over previous computer-tailoring approaches, which typically rely on multiple choice. Other study strengths include its mostly hands-off approach (thus improving the ecological validity of the study compared with a study with a high level of human contact), the use of objective health measures, and the use of intention-to-treat analysis, which ensures that the intervention effects are not over-estimated. Limitations include the pre-post design, which meant the study lacked a randomized control comparator, and follow-up is limited to three months.

Conclusions

Virtual assistant technology offers exciting potential for the delivery of highly personalized, scalable health interventions. They can make evidence-based advice and support available to patients who are not unwell enough to quality for in-person (ie, expensive) services. They also have the potential to be used as adjunct support for those patients with complex and chronic health conditions such as type 2 diabetes or celiac disease, who see a health practitioner, but would benefit from additional support. Furthermore, they present a way of providing a scientifically substantiated dietary pattern and lifestyle programs to rural or remote areas or isolated individuals.

This ground-breaking study confirmed that a virtual health coach diet and physical activity intervention is feasible for older adults and leads to sizable health behavior change and improvements in body composition. Future research examining artificially intelligent interventions at scale, and for other health purposes, is warranted.

Availability of Data and Materials

The datasets used and analyzed for this study are available from the corresponding author on request.

Acknowledgments

CM receives a Career Development Fellowship from the Australian National Health and Medical Research Council (APP1125913). The study was completed with support from the University of South Australia.

The authors acknowledge and thank OutThought for their assistance in developing the virtual health coach platform and ongoing support during data collection and analysis.

Authors' Contributions

CM and KM conceived the study, CES had input into the tailoring of lifestyle advice, CM, KM, CD, and RC designed the study, CD and RC collected the data, CM and CD analyzed the data and CM led the drafting of the manuscript. All authors edited the manuscript, and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physical activity and diet log sheet, for recording daily steps and dietary intake, adapted from Davis et al.
[DOCX File , 17 KB - mhealth_v8i7e17558_app1.docx]

Multimedia Appendix 2
Specific recommendations for the Mediterranean diet, based on MedLey diet.
[DOCX File , 14 KB - mhealth_v8i7e17558_app2.docx]

Multimedia Appendix 3
Outcome measures at baseline, 6 weeks and 12 weeks – complete case analysis (n=28).
[DOCX File , 14 KB - mhealth_v8i7e17558_app3.docx]

References

1. Remington P, Brownson R, Wegner M. Chronic Disease Epidemiology and Control. 3rd ed: American Public Health Association; 2010.
2. 4364.0.55.003 - Australian Health Survey: Updated Results, 2011-2012. Daily Intake of Fruit and Vegetables.: Australian Bureau of Statistics; 2013. URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/C549D4433F6B74D7CA257B8200179569?opendocument [accessed 2019-09-01]
3. 4364.0.55.001 - Australian Health Survey: First Results, 2011-12.: Australian Bureau of Statistics; 2012. URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4364.0.55.001main+features12011-12 [accessed 2019-08-31]
4. Sinclair TJ, Grieve R. Facebook as a source of social connectedness in older adults. Computers in Human Behavior 2017 Jan;66:363-369. [doi: 10.1016/j.chb.2016.10.003]
5. Chronic Disease Management - Provider Information. Australian Government, Department of Health. 2014. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/mbsprimarycare-factsheet-chronicdisease.htm [accessed 2019-08-31]
6. Stein N, Brooks K. A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. JMIR Diabetes 2017 Nov 01;2(2):e28. [doi: 10.2196/diabetes.8590]
7. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the Role of In-Person Components for Online Health Behavior Change Interventions: Can a Digital Person-to-Person Component Suffice? J Med Internet Res 2018 Apr 11;20(4):e144. [doi: 10.2196/jmir.8480]
8. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A Meta-Analysis of Web-Delivered Tailored Health Behavior Change Interventions. Journal of Health Communication 2013 Sep;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727]
9. Laranjo L, Dunn A, Tong H, Kocaballi A, Chen J, Bashir R. Conversational agents in healthcare: a systematic review. JAMIA Open 2018;25(9):58. [doi: 10.1093/jamia/ocy072]
10. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the Chatbot “Vivibot” to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018. [doi: 10.2196/15018]
11. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64. [doi: 10.2196/mental.9782]
12. Suganuma S, Sakamoto D, Shimoyama H. An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. JMIR Ment Health 2018 Jul 31;5(3):e10454. [doi: 10.2196/10454]
13. Chaix B, Bibault J, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot. JMIR Cancer 2019 May 02;5(1):e12856. [doi: 10.2196/12856]
14. Schmidlen T, Schwartz M, DiLoreto K, Kirchner HL, Sturm AC. Patient assessment of chatbots for the scalable delivery of genetic counseling. Jrnl of Gene Coun 2019 Sep 24;28(6):1166-1177. [doi: 10.1002/jgc4.1169]
15. Tanana MJ, Soma CS, Srikumar V, Atkins DC, Imel ZE. Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills. J Med Internet Res 2019 Jul 15;21(7):e12529. [doi: 10.2196/12529]
16. Calvaresi D, Calbimonte J, Debusson F, Najjar A, Schumacher M. Social network chatbots for smoking cessation: agent and multi-agent frameworks. 2019 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence; 14-17 Oct 2019; Thessaloniki, Greece. [doi: 10.1145/3350546.3352532]
17. Stephens, TN, Joerin A, Rauws M. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioural coaching chatbot. TBM 2019;9:440-447. [doi: 10.1093/tbm/ibz043]
18. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection Companion: a conversational system for engaging users in reflection on physical activity. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Jul 05;2(2):1-26. [doi: 10.1145/3214273]

19. 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 Mar 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6]
20. Davis CR, Bryan J, Hodgson JM, Wilson C, Dhillon V, Murphy KJ. A randomised controlled intervention trial evaluating the efficacy of an Australianised Mediterranean diet compared to the habitual Australian diet on cognitive function, psychological wellbeing and cardiovascular health in healthy older adults (MedLey study): protocol paper. BMC Nutr 2015 Nov 19;1(1). [doi: 10.1186/s40795-015-0033-7]
21. Davis C, Bryan J, Hodgson J, Murphy K. Definition of the Mediterranean Diet; A Literature Review. Nutrients 2015 Nov 05;7(11):9139-9153. [doi: 10.3390/nu7115459]
22. Estruch R, Ros E, Salas-Salvadó J, Covas M, Corella D, Arós F, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. N Engl J Med 2018 Jun 21;378(25):e34. [doi: 10.1056/nejmoa1800389]
23. Australian Institute of Health and Welfare (AIHW). The Active Australia Survey: a guide and manual for implementation, analysis and reporting. Report Number: 1740242580. AIHW 2003;Report Number: 1740242580.
24. Brown W, Burton N, Marshall A, Miller Y. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. Aus N Z J Public Health 2008;32(6):41. [doi: 10.1111/j.1753-6405.2008.00305.x]
25. Schröder H, Fitó M, Estruch R, Martínez-González M, Corella D, Salas-Salvadó J. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011;141(6):5. [doi: 10.3945/jn.110.135566]
26. Wade A, Davis C, Dyer K, Hodgson J, Woodman R, Murphy K. A Mediterranean diet supplemented with dairy foods improves markers of cardiovascular risk: results from the MedDairy randomized controlled trial. Am J Clin Nutr 2018;108(6):82. [doi: 10.1093/ajcn/nqy207]
27. Norton K, Olds T. Anthropometrica. Sydney, NSW: UNSW Press; 1996.
28. Australian Bureau of Statistics (ABS). ABS. 2033.0.55.001 - Census of Population and Housing: Socio-economic Indexes for Areas (SEIFA). 2016. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocumOpe [accessed 2019-09-20]
29. National Health and Medical Research Council (NHRMC). Australian Dietary Guidelines Summary. Canberra, NHRMC 2013.
30. Australian Institute of Health and Welfare (AIHW). Data sources for monitoring overweight and obesity in Australia. Canberra, AIHW 2019;Cat. no. PHE 244.
31. Hamdorf P, Starr G, Williams M. A survey of physical-activity levels and functional capacity in older adults in South Australia. J Aging Phys Act 2002;10(3):281-289. [doi: 10.1123/japa.10.3.281]
32. Heesch KC, Hill RL, van Uffelen JG, Brown WJ. Are Active Australia physical activity questions valid for older adults? Journal of Science and Medicine in Sport 2011 May;14(3):233-237. [doi: 10.1016/j.jsams.2010.11.004]
33. Kavanagh AM, Goller JL, King T, Jolley D, Crawford D, Turrell G. Urban area disadvantage and physical activity: a multilevel study in Melbourne, Australia. Journal of Epidemiology & Community Health 2005 Nov 01;59(11):934-940. [doi: 10.1136/jech.2005.035931]
34. 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 Psychology 2015;34(Suppl):1220-1228. [doi: 10.1037/hea0000305]
35. Shah H, Warwick K. Machine humour: examples from Turing test experiments. AI & Soc 2016 Jun 7;32(4):553-561. [doi: 10.1007/s00146-016-0669-0]
36. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The Effectiveness of a Web-Based Computer-Tailored Physical Activity Intervention Using Fitbit Activity Trackers: Randomized Trial. J Med Internet Res 2018 Dec 18;20(12):e11321. [doi: 10.2196/11321]
37. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, de Craen AJ, et al. Effects of a Web-Based Intervention on Physical Activity and Metabolism in Older Adults: Randomized Controlled Trial. J Med Internet Res 2013 Nov 06;15(11):e233. [doi: 10.2196/jmir.2843]
38. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a Web-Based, Computer-Tailored, Pedometer-Based Physical Activity Intervention for Adults: A Cluster Randomized Controlled Trial. J Med Internet Res 2015 Feb 09;17(2):e38. [doi: 10.2196/jmir.3402]
39. Spittaels H, De Bourdeaudhuij I, Brug J, Vandelanotte C. Effectiveness of an online computer-tailored physical activity intervention in a real-life setting. Health Educ Res 2007 Jun 31;22(3):385-396. [doi: 10.1093/her/cyl096] [Medline: 16971674]
40. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MDC, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-Term Effectiveness of a Mobile Phone App for Increasing Physical Activity and Adherence to the Mediterranean Diet in Primary Care: A Randomized Controlled Trial (EVIDENT II Study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]

41. Livingstone KM, Celis-Morales C, Navas-Carretero S, San-Cristobal R, Macready AL, Fallaize R, Food4Me Study. Effect of an Internet-based, personalized nutrition randomized trial on dietary changes associated with the Mediterranean diet: the Food4Me Study. Am J Clin Nutr 2016 Aug 29;104(2):288-297. [doi: 10.3945/ajcn.115.129049] [Medline: 27357094]
42. Davis C, Hodgson J, Bryan J, Garg M, Woodman R, Murphy K. Older Australians Can Achieve High Adherence to the Mediterranean Diet during a 6 Month Randomised Intervention; Results from the Medley Study. Nutrients 2017 May 24;9(6):534 [FREE Full text] [doi: 10.3390/nu9060534] [Medline: 28538676]

Abbreviations

AAS: Active Australia Survey
MVPA: moderate and vigorous physical activity
PREDIMED: Prevención con dieta Mediterránea
SBP: systolic blood pressure
DBP: diastolic blood pressure
MedLiPal: Mediterranean lifestyle and physical activity

Edited by C Dias; submitted 19.12.19; peer-reviewed by R Zhang, M Waller; comments to author 15.02.20; revised version received 09.04.20; accepted 05.05.20; published 10.07.20.

Please cite as:
Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ
A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study
JMIR Mhealth Uhealth 2020;8(7):e17558
URL: https://mhealth.jmir.org/2020/7/e17558
doi:10.2196/17558
PMID:32673246

Original Paper

A Personalized Physical Activity Coaching App for Breast Cancer Survivors: Design Process and Early Prototype Testing

Francisco Monteiro-Guerra[1], MSc; Gabriel Ruiz Signorelli[1,2], MSc; Shreya Tadas[3], MSc; Enrique Dorronzoro Zubiete[4], PhD; Octavio Rivera Romero[4], PhD; Luis Fernandez-Luque[2], PhD; Brian Caulfield[1], PhD

[1]Insight Centre for Data Analytics, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

[2]Salumedia Tecnologias, Seville, Spain

[3]Insight Centre for Data Analytics, School of Computer Science, University College Dublin, Dublin, Ireland

[4]Computer Engineering School, Universidad de Sevilla, Seville, Spain

Corresponding Author:
Luis Fernandez-Luque, PhD
Salumedia Tecnologias
Salumedia Tecnologías S.L.U., Avda. República Argentina
nº 24 Edificio Torre de los Remedios, 5ª planta módulo A
Seville, 41011
Spain
Phone: 34 656930901
Email: luis@salumedia.com

Abstract

Background: Existing evidence supports the many benefits of physical activity (PA) in breast cancer survival. However, few breast cancer survivors adhere to the recommended levels of activity. A PA coaching app that provides personalized feedback, guidance, and motivation to the user might have the potential to engage these individuals in a more active lifestyle, in line with the general recommendations. To develop a successful tool, it is important to involve the end users in the design process and to make theoretically grounded design decisions.

Objective: This study aimed to execute the design process and early prototype evaluation of a personalized PA coaching app for posttreatment breast cancer survivors. In particular, the study explored a design combining behavioral theory and tailored coaching strategies.

Methods: The design process was led by a multidisciplinary team, including technical and health professionals, and involved input from a total of 22 survivors. The process comprised 3 stages. In stage 1, the literature was reviewed and 14 patients were interviewed to understand the needs and considerations of the target population toward PA apps. In stage 2, the global use case for the tool was defined, the features were ideated and refined based on theory, and a digital interactive prototype was created. In stage 3, the prototype went through usability testing with 8 patients and was subjected to quality and behavior change potential evaluations by 2 human-computer interaction experts.

Results: The design process has led to the conceptualization of a personalized coaching app for walking activities that addresses the needs of breast cancer survivors. The main features of the tool include a training plan and schedule, adaptive goal setting, real-time feedback and motivation during walking sessions, activity status through the day, activity history, weekly summary reports, and activity challenges. The system was designed to measure users' cadence during walking, use this measure to infer their training zone, and provide real-time coaching to control the intensity of the walking sessions. The outcomes from user testing and expert evaluation of the digital prototype were very positive, with scores from the system usability scale, mobile app rating scale, and app behavior change scale of 95 out of 100, 4.6 out of 5, and 15 out of 21, respectively.

Conclusions: Implementing a user-centered design approach for the development and early evaluation of an app brings essential considerations to tailor the solution to the user's needs and context. In addition, informing the design on behavioral and tailored coaching theories supports the conceptualization of the PA coaching system. This is critical for optimizing the usability, acceptability, and long-term effectiveness of the tool. After successful early in-laboratory testing, the app will be developed and evaluated in a pilot study in a real-world setting.

(JMIR Mhealth Uhealth 2020;8(7):e17552) doi:10.2196/17552

KEYWORDS

user-centered design; physical activity; coaching; behavior change; mobile app; mobile phone; breast cancer; usability

Introduction

Background

According to the current findings, physical activity (PA) is the most well-established lifestyle factor associated with breast cancer survival [1]. Rapidly accumulating research demonstrates that routine exercise throughout and after treatments offers multiple benefits for breast cancer survivors, including prevention of cancer recurrence, mitigation of treatment side effects (such as lymphedema and fatigue), and improvement of their physical function and quality of life [2-4]. Despite the growing evidence supporting PA in breast cancer survivorship, only a minority of these individuals adhere to the recommended levels of PA [5,6]. Novel and engaging strategies are needed to ensure that participants are adhering to PA that is of high enough intensity and frequency to meet the PA recommendations.

Mobile health (mHealth) has emerged as an important tool for health behavior change interventions [7]. Currently, mobile devices can accurately measure PA at any time and place, creating opportunities to provide real-time tailored support and motivation toward an active lifestyle [8]. In line with this, mobile apps for PA coaching have emerged and have been investigated as a platform to motivate people to be active through recommended goals, feedback on activities performed, and potentially enjoyable experiences [9,10]. An increasing body of evidence indicates that these technology-based interventions may be well received by breast cancer survivors and hold promise for PA promotion initiatives [11-13].

An underlying challenge of these technology-supported interventions is the high attrition rates, with users stopping the use of these systems after a few days or weeks [14-16]. Among the factors associated with user abandonment are that apps are largely targeted at generally healthy individuals and do not address the specific needs of the end users [17]. Studies with breast cancer survivors suggest that the direction of PA systems should meet the detailed requirements of this particular population [13,18], who may be less motivated to engage in PA and who face unique barriers to reaching the recommended level of PA [19,20]. These include physical (eg, fatigue, weight gain, and neuropathy), environmental (eg, lack of knowledge, job and family obligations, and weather), and psychosocial (eg, low confidence and emotional imbalance) limitations [19,20]. Related literature supports the use of a user-centered design (UCD) approach, which focuses on the users and their needs and is considered a prerequisite for useful technology being associated with an increase in the success of these systems [21,22]. Furthermore, it is suggested that breast cancer survivors want a PA app experience targeted not only to their needs on a group level but also tailored to each individual user [13,23,24]. In line with this, personalized or tailored coaching mechanisms can be leveraged to help create experiences that are individualized for each user [25,26]. These are believed to influence the user's attention and contribute to long-term engagement and adherence to these apps [27]. In addition, theory-based behavior change methods have been shown to influence the effectiveness of technology-supported interventions [7] and should be considered in the design of tools that aim to increase PA [8,26].

Currently, a few PA apps reported in the literature target breast cancer survivors [11,24,28,29]; however, these seemed to be in the early stages of development and evaluation and lack proper reporting of design decisions. Only the papers describing the Bounce app presented clear details on the design process, behavioral theory foundations, and features [24,30]. Further research is needed to better understand the design of mobile PA interventions for breast cancer survivors and how to create a more tailored and engaging experience to increase long-term adherence of these users with the coaching systems.

Objectives

This paper aimed to report on the UCD of a personalized PA coaching app for breast cancer survivors that targets the needs of breast cancer survivors at both the group and individual levels. The system is grounded on existing theory, models, and empirical evidence on personalized coaching, behavior change, and linear progression exercise training. The paper describes the 3 design stages of the tool: (1) user and context research, (2) app conceptualization and early prototyping, and (3) prototype testing.

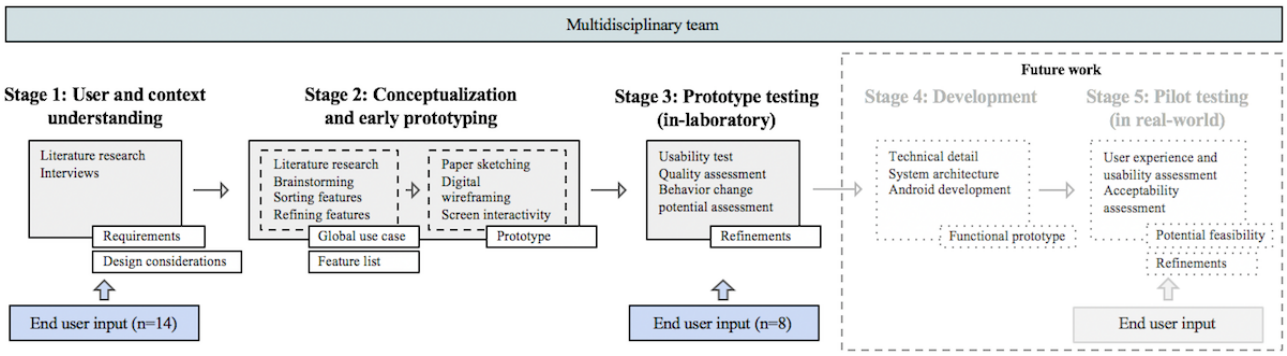
Methods

User-Centered Design Process

A UCD approach was followed. UCD is considered a prerequisite for useful technology and a successful intervention [21,31]. The process involves a multistage, problem-solving process that investigates the needs, desires, and limitations of users to increase the success rate of usability in computerized systems [22].

The proposed design stages were based on Shah's methodological framework [31] and on implementations of such an approach in different studies [32,33]. The design process so far involved 3 stages (Figure 1): stage 1—user and context understanding, stage 2—conceptualization and early prototyping, and stage 3—prototype testing. In stages 4 and 5 (future work), a functional prototype will be developed and then pilot tested for potential feasibility in real-life settings.

Figure 1. Schematic of the app's user-centered design process.



Multidisciplinary Team

A multidisciplinary team guided the design process and included 5 members from different backgrounds and expertise: FG, OR, EZ, and ST have an engineering background (biomedical, computer, electronic, and computer science engineering, respectively) and have experience in the fields of digital health apps, human-computer interaction, participatory health informatics, and user experience; GS has a sports science background, with expertise in the field of PA in oncology. Matilde Mora Fernández also has experience in the field of PA in oncology and provided useful insights to the team at different stages of the design process.

Theoretical Foundation

To guide the design process, we used applicable behavior change theory, with the aim of increasing the long-term effectiveness of the PA system [34]. Appropriate theoretical frameworks (self-determination theory [SDT] [35] and social cognitive theory [SCT] [36,37]) and constructs were identified from related work and empirical evidence to highlight factors, barriers, and determinants that brought important insights into the design of the solution. In addition, to facilitate the design and reporting of theory-based components, the Coventry, Aberdeen, and London-Refined (CALO-RE) taxonomy of behavior change techniques proposed by Michie et al was considered [38]. The CALO-RE taxonomy has been used particularly in PA and healthy eating interventions. It has been reported as a useful tool for researchers to design effective interventions as it provides a more straightforward and simple approach to identifying the factors most likely to create positive effects on PA behavior change [38].

Personalization theory was also explored in the tool design, which may help increase the intended effects of the app communication and, in that way, increase the effectiveness of the behavior change intervention [27]. Op den Akker et al [25] proposed a model for real-time tailoring of PA coaching apps, which defined 7 different methods of personalization. A more recent review by Monteiro-Guerra et al [26] provided a detailed analysis of the existing personalized PA coaching apps and strategies, which provided useful insights for the design of the solution proposed in this work.

Participant Recruitment for User Studies

The inclusion criteria for research participants were (1) to be oncology patients with a history of breast cancer that finished primary curative treatment (surgery, radiotherapy, and chemotherapy), (2) to be aged more than 18 years, (3) to own and use a mobile phone or smartphone, (4) to have the ability to read and speak Spanish, (5) to have no known impairments or comorbidities, and (6) to have no restrictions on PA. The participants were recruited from a specialized oncology clinic by placing a phone call to the eligible individuals identified in the patient database. Recruitment was conducted until saturation of results was reached, which was considered when there was no new information (themes) arising from the qualitative data.

This study was approved by the Research and Ethics Committee of Junta de Andalucia in Spain. Subjects' agreement for participation was obtained through an informed consent process.

Stage 1: User and Context Research

In the first stage, thorough research was performed on the needs and requirements of breast cancer survivors regarding content, expected benefits, features, and personalization and motivational aspects of the proposed tool. For this purpose, we performed (1) a qualitative study with the target users and (2) a review of related literature, from which we identified the consequences for the design of our solution.

Semistructured interviews were conducted with 14 breast cancer survivors. Three of the team members, FG, OR, and EZ, were involved in this study. The interviews involved a combination of open-ended questions on PA adherence and technology interest and a slideshow presentation with examples of PA app features to obtain participants' thoughts and opinions on the featured content. The interviews were audio-recorded, transcribed verbatim, and analyzed using thematic content analysis [39]. For the purposes of this paper, only an overview of the main insights gathered from the aforementioned qualitative study is presented here.

A rapid literature review was performed on February 2019 to complement the findings from the user study on (1) barriers and facilitators for PA; (2) attitudes, needs, and preferences from breast cancer survivors on PA apps; and (3) information on behavioral theory used in PA interventions for breast cancer. A combination of medical and technological keywords was used in the search string: *breast cancer* AND *app** AND *mobile* OR *smartphone* AND *physical activit** OR *exercis** OR *walk**. The searches were performed on 3 web-based databases, PubMED Central, Association for Computing Machinery digital library (ACM), and Scopus, with the use of the selection criteria. Then, a title and abstract review was performed. Other relevant papers

were identified through a snowballing process by screening the reference list of the included ones for related papers [40].

Stage 2: Conceptualization and Early Prototyping

The goal of this phase was to conceptualize the idea of the app, to define a list with potential functionalities, and to create a low-fidelity prototype.

Conceptualization

After the analysis of the literature and clinical guidelines, a high-level concept of the tool was proposed, which is defined as the global use case. The global use case was designed to facilitate communication of the idea of the tool to the design team, without giving too much direction to their thoughts [32]. It was defined by the authors FG, OR, EZ, and GS in a design session, based on the knowledge gathered from the literature, and validated by Matilde Mora Fernández, the collaborator expert in PA and breast cancer survivorship. The validation process relied on a discussion with Matilde Mora Fernández, which revised the global use case and provided suggestions to refine it. The global use case described on a high level the main objectives of the tool, the technological platform, and essential requirements.

The ideation process to elicit features for the mobile app consisted of 2 group sessions with FG, OR, EZ, and GS. The first session was introductory, where the results from stage 1 were discussed and clarified to each of the team members. The second session involved feature elicitation, followed by sorting and selecting features. Each session lasted for 45 min to 1 hour. A brainstorming technique was implemented using the affinity wall method [41]. First, the team members were asked to read through the global use case and the list of user needs and requirements, which was made visible to everyone in 1 slide, and then, for 15 min each member individually wrote their ideas for potential features. Once the ideas were submitted, similar ideas were combined, resulting in a preliminary list of features. The criteria-based evaluation [42], which uses a decision matrix, was used to choose the main ideas, based on the considerations from stage 1. The criteria used for feature ranking were based on the strengths, weaknesses, opportunities, and threats analysis [43]. On the basis of the ranking given and team agreement, the top ideas for features were selected.

These ideated features went through a refinement phase, which involved an iterative process of analyzing design considerations from stage 1, to further specify the app functionalities. In this process, 2 researchers (FG and OR) identified a list of requirements and preferences from breast cancer survivors that were related to each of the elicited features. In addition, considerations were taken from the CALO-RE taxonomy [38], which presents 40 techniques and the psychological constructs each purport to change. We have considered the model of personalization in real-time PA coaching apps [25], which proposes 7 concepts or strategies to adjust the different properties of communication to the users. These design considerations were listed and mapped to each of the main features to help refine the app functionalities and how they needed to be designed. Furthermore, we followed the model on linear progression training and the insights from the experts in exercise and cancer to construct the foundations of the training plan for the app.

Prototyping

On the basis of the ideas generated, the team started the prototyping phase. This phase consisted of the creation of a digital and interactive low-fidelity prototype, in which priority was given neither to content nor visuals [44]. The wireframes were created by FG using the NinjaMock software and were then transferred to the Proto.io software to simulate the interactions between the buttons and screens. The prototype was then reviewed by the rest of the team, which provided suggestions for improvements through the software. The objective of this low-fidelity prototype was to facilitate idea communication and to perform early evaluations, which were conducted in stage 3 of the design process.

Stage 3: Prototype Testing

The third stage of the process sought to explore an early evaluation of the system, involving both user and expert testing. The evaluation was directed toward assessing the usability, quality, and behavior change potential of the concept ideated.

User Evaluation

Usability relates to the extent to which a system can be used by the end users to achieve specified goals with effectiveness, efficiency, and satisfaction. Indicators for these are error rate, task completion time, and a satisfaction rating questionnaire, respectively [45,46]. In this study, a mixed methods approach was followed for usability testing [47,48]. The study was led by FG and OR. The test was conducted in a laboratory setting, which involved the end users interacting with the prototype created in stage 2. A think-aloud procedure was performed. The participants first completed an initial questionnaire covering demographics, technology use and interest, and PA level. Then, after a short explanation, the participants performed 7 predetermined tasks with the interactive prototype. Participants were asked to verbalize their thoughts while completing the tasks. Interactions with the prototype and the participants' comments were recorded using the AZ Screen Recorder app. In addition, the researcher FG observed the participants throughout their task performance, registering the user's comments and suggestions, the number of errors, the task duration, and any indication of needing assistance or confusion. The number of errors was calculated as the number of times there was an erroneous interaction (eg, the task required the user to find the information tab and the user clicked on the profile button). Task duration was calculated from the moment the researcher presented the task until the task was completed, only considering when the task was performed without errors. In addition, participants valued the complexity of each task using the single ease question (SEQ), which is a scale to rate tasks from 1 (very difficult) to 7 (very easy) [49]. All participants completed the Spanish version of the system usability scale (SUS) [50], a 10-item questionnaire used to quickly and accurately assess the usability of a system. Higher scores indicate better usability.

At the end of the session, a short interview was performed to address the users' understanding of particular features and

information provided by the app, the general opinions on the app and its usefulness, and if there was anything missing in the app.

Expert Evaluation

An expert evaluation was performed with the mobile app rating scale (MARS) [51] and the app behavior change scale (ABACUS) [52] to assess the quality of the app and the potential for behavior change, respectively. The MARS was used to examine app elements, such as engagement, functionality, utility, aesthetics, and information. This scale includes 23 items across 5 categories, with each item scored using a series of questions on a 5-point ordinal scale response. An overall functionality score out of 5 was derived using this scale. The ABACUS scale comprises 21 items and was used to examine the potential behavior change of the app in relation to goal setting, action planning, barrier identification, self-monitoring, and feedback. Two technical experts, with previous experience using these tools, performed the evaluation of the concept based on the low-fidelity prototypes and a document detailing the app functionalities.

Analysis

The SUS, MARS, and ABACUS scores were calculated following the standard procedures for each scale. The SUS mean and SD were calculated for the scores across all participants. In addition, the mean and SD of SEQ ratings and task duration, across all participants, were calculated for each particular task and for all tasks. For each evaluator, a MARS score out of 5 was calculated under 3 of the 4 sections of the scale. Some of the aspects covered by the scale, including the section of design aesthetics, were not considered due to the use of a low-fidelity prototype. The mean of these scores produced an overall score for each evaluator, and the mean of the overall scores for each evaluator provided an overall score for the app quality. The 2 evaluators also identified the presence of ABACUS items under each of the 4 sections of the scale. Discrepancies were discussed between the expert evaluators until an agreement was reached. The number of items identified, out of 21 possible items, was added to provide an overall score for the app's behavior change potential.

The audio data, from the think-aloud procedure and the short user interviews, were transcribed, anonymized, and translated. The transcripts were, then, analyzed by FG and OR to identify the salient aspects about usability issues, functionalities liked by the users, and suggestions for improvements.

Results

Participant Characteristics for User Studies

The first user study (interviews in stage 1) was held with 14 participants and the second (usability in stage 3) with 8 participants. In study 1, the participants' ages ranged from 43 to 69 years, with a mean of 52.8 (SD 8.8) years, and in study 2, it ranged from 38 to 63 years, with a mean of 48.4 (SD 8.0) years. The number of years since diagnosis ranged from 2 to 11.5, with a mean of 5.2 (SD 2.9) for participants in study 1, and from 0.5 to 4.5, with a mean of 2.3 (SD 1.6) in study 2. In general, participants were educated and employed, were very interested in technology, had ready access to technological devices, and had shown high usage of a variety of technology functionalities. Participants' access to technology and usage can be found in Multimedia Appendix 1. According to the international physical activity questionnaire-short form [53], most participants had at least a moderate level of PA. However, when looking at the activity type and intensity, 57% (8/14) of participants in study 1 and 50% (4/8) in study 2 did not adhere to the PA guidelines [54]. Participant characteristics are presented in detail in Table 1.

Table 1. Participant characteristics.

Characteristic	User study 1 (n=14), n (%)	User study 2 (n=8), n (%)
Marital status		
Single	4 (29)	N/A[a]
Married	10 (71)	N/A
Divorced	0 (0)	N/A
Education		
Basic school	1 (7)	1 (13)
High school	2 (14)	2 (25)
Higher education	2 (14)	0 (0)
University or college	9 (64)	5 (63)
Current employment status		
Not working	3 (21)	1 (13)
Employed	11 (79)	7 (88)
Receiving pharmacological treatment	10 (71)	8 (100)
Indication for PA[b]	11 (79)	8 (100)
IPAQ-SF[c] level		
High	1 (7)	3 (38)
Moderate	11 (79)	5 (63)
Low	2 (14)	0 (0)
Adheres to PA guidelines (>150 min per week=moderate activity or >75 min per week=vigorous activity)[d]	6 (43)	4 (50)
Interest in technology		
Agree or strongly agree	12 (86)	8 (100)
Neutral	2 (14)	0 (0)
Disagree or strongly disagree	0 (0)	0 (0)
Self-reported skill with technology		
Agree or strongly agree	9 (64)	7 (88)
Neutral	5 (36)	1 (13)
Disagree or strongly disagree	0 (0)	0 (0)
"I like to experiment with new technology"		
Agree or strongly agree	7 (50)	6 (75)
Neutral	5 (36)	2 (25)
Disagree or strongly disagree	2 (14)	0 (0)

[a]N/A: not applicable.

[b]PA: physical activity.

[c]IPAQ-SF: international physical activity questionnaire-short form.

[d]Information inferred from IPAQ-SF answers.

Stage 1: User and Context Research

User Needs, Requirements, and Preferences for Physical Activity Apps

The findings from the qualitative study with 14 breast cancer survivors provided insight on barriers and motivators for PA and opinions on a variety of app-based intervention characteristics. From the literature review, we have included 5 papers to provide a more complete perspective on the barriers and facilitators of PA [19,20,55-57] and 11 papers with at least some information regarding the attitudes, needs, and preferences of breast cancer survivors on PA apps [12,13,18,23,24,29,58-62]. These included papers on user studies with the Bounce app and the Smart After Care app, both designed for breast cancer survivors, which provided relevant

insights to the design. An overview of the barriers and facilitators for PA in breast cancer survivors is provided in Textbox 1.

The list of requirements taken from our interviews and related work is presented in Table 2.

Textbox 1. Overview of barriers and facilitators for physical activity in breast cancer survivors.

- Barriers
  - Lack of time
    - Consequence of job and family responsibilities
  - Lack of confidence
  - Low motivation
  - Physical limitations
    - Fatigue, lymphedema, joint pain, muscular pain, neuropathy, and weight gain
    - Current physical activity (PA) level compared with pretreatment
  - Lack of information/support for PA
    - Lack of information on the type and amount of PA recommended
    - Fear and uncertainty of starting exercising without guidance
  - Misconceptions about PA
    - Fear of potential side effects
  - Emotional imbalance
    - Not feeling good
    - Stress or anxiety
  - Access to facilities
    - Inconvenient timetable or distant location
  - Seasonal weather
- Facilitators
  - Reserve time during the week for PA
  - Knowing and perceiving the benefits
  - Being nudged to be more active
  - Support to overcome insecurities
  - Emotional support
  - Tailored information
  - Prescription of PA by the health care professionals
  - Training plan tailored to their needs and PA level
  - Clear and realistic objectives
  - Quantify activity performed
  - Having an active family
  - Support from family and close friends

Table 2. Overview of breast cancer survivors' requirements and preferences for physical activity apps.

Type	From user study	Literature review—additional insights
General intervention characteristics	• Activity monitoring and feedback • Preference for information on steps, calories, distance, pace, and duration of activity • PA[a] prescription and goal setting • Scheduling tool and activity reminders • Tailored experience • Progress monitoring and visualization • Straightforward representation of activity performed and incremental improvements • Preference for daily and weekly progress feedback • Information on how their progress translates into physiological processes (eg, benefits for weight management) • Simplicity and ease of use • Mixed reactions toward a game-like design (eg, points, rewards, avatars, and competitions) • Consider strategies to manage emotional challenges (eg, encourage connecting with counselor and include relaxation and meditation exercises).	• Preference for walking activities (most appealing and main form of exercise) • Other liked activities included resistance training and yoga • Feedback on time spent in various intensities of activity • Evidence-based content • Attractive design • Friendly graphic displays • Possibility to integrate with wearable activity trackers • Preference for more straightforward representations of numerical data (compared with having gamified themes) • Suggestions for integration of a Newsfeed • Information on benefits and harms/risks of exercise for breast cancer survivors
Personalized experience	• Adaptive activity plan and goals • Progressive but attainable goals • Customizable exercises and exercise schedule • Targeting user characteristics (eg, age, treatment types, and preferences) • Sensitive to PA level and physical limitations/injuries • Individualized progress feedback • Targeting user's situational/external context (eg, weather and location) • Personalized recommendations • Interface simulating a virtual coach	• Incremental levels adjusted to the user experience • Change intensity or amount of exercise program to reflect user's improvement • Ensures correct execution of exercises • Target value-based goals • Suggestions for integration of a symptom tracker • Suggestions for tracking energy level, how they feel, and sleep quality
Positive communication	• Motivation/encouragement • Encouraging prompts during activity • Recognition for achievements • Positive tone • Absence of pressure • Just enough reminders and notifications	• Casual and concise tone • Motivational messages
Social connectedness (varied opinions)	• Involvement of family and close friends • Mixed reactions toward connecting with other peers (eg, social networking, competitions, or ability to see others' progress) • Ability to connect with a professional (eg, a psychooncologist and/or an exercise trainer)	• Role-model narratives • Preference for a more private experience
Trustworthiness	• Transparent data privacy and security • Developed with and validated by clinical/health experts • Include contact information of people involved in the app development	• N/A[b]
Data sharing and portability	• Optional and customizable data sharing • Willingness to share with health care professionals	• Ability to keep an electronic record of their workouts on the app • Ability to download data to a PC[c] • Extensive yet passive data collection

[a]PA: physical activity.

[b]N/A: not applicable.

[c]PC: personal computer.

Behavioral Theory for Physical Activity Interventions in Breast Cancer

From searching the literature for appropriate behavioral foundations for design, there seems to be little consensus on optimal theories and theory integration techniques to change PA behaviors in this population [63-65]. According to some studies, the transtheoretical model of behavior change, SCT, and SDT are the most appropriate models for behavioral interventions for breast cancer survivors [2,18]. In our qualitative study, the findings suggest that the motivational factors and determinants of PA adherence in breast cancer survivors are in line with the constructs of SCT (self-efficacy and expected outcomes) and SDT (competence, autonomy, and social relationships). Together with the CALO-RE taxonomy of behavior change, these insights will facilitate the integration and reporting of behavior change techniques in our solution. The psychological mediators identified in our user study and associated CALO-RE techniques are presented in Multimedia Appendix 2.

Stage 2: Conceptualization and Early Prototyping

Global Use Case

The high-level concept of the tool is defined as a real-time coaching system to support and motivate breast cancer survivors to increase adherence to PA; that is smartphone-based; that simulates the interactions with a PA coach; that ensures that users engage in activities of high enough intensity to meet the PA recommendations; that provides a tailored experience, on a population level and on an individual level; that focuses on walking activities; and that is based on the PA guidelines and recommendations for breast cancer survivors. The ultimate goal of the tool is to help breast cancer survivors reach and maintain the recommended levels of PA.

App Functionalities

One of the main steps of this stage 2 was the ideation process, which resulted in the specification of 7 main app features and their respective refined subfeatures (see Table 3). An extended version of the table, including design considerations from stage 1 and from the CALO-RE taxonomy used for feature refinement, is presented in Multimedia Appendix 3.

Details on the PA program and the personalized coaching system are provided in the following sections. The interactive low-fidelity mock-ups (translated from Spanish to English) are shown in Figures 2 and 3, including the main screens and the guided session simulation, respectively.

Table 3. List of app functionalities.

Main features (ideated)	General description (ideated)	Subfeatures (refined based on considerations from stage 1 and the CALO-RE[a] constructs)
Training plan	Activity program guided by the PA[b] guidelines and recommendations for breast cancer survivors; based on linear progression training; with adaptable levels; visually represented by an activity schedule; includes reminders for activity.	Information about PA program, guidelines and potential benefits for the users; a plan that sets the number of activities per week and its duration and difficulty; a weekly activity schedule; baseline assessment for current PA level; assessment of perceived difficulty; adjustable plan level; push notifications and reminders for activities scheduled; push notifications and reminders to review the plan and reschedule activities.
Adaptive goal setting	Activity objectives adjusted to the user.	Present the user with clear daily objectives in the main screen; set achievable but challenging goals; progress bars; inform of long-term benefits of achieving goals; present automatic adaptation to the user's profile information, progress, user's perceived fatigue and perceived difficulty; notifications of goal adjustments; weekly goal adjustment.
Real-time monitoring, feedback, and motivation during activity sessions	In-session or "workout" coaching; visual and easy to understand; combines real-time monitoring, feedback and motivation.	Predefined walking sessions; guidance to meet the session plan; intuitive interface to provide session information; shows the session progress (time); shows the user's pace in real time through a glanceable visual display; sends cues to control the pace; provides positive reinforcement and recognition; coaching cues are in textual and audio format; shows achievements when the session is concluded with a breakdown of the session: steps taken; calories burned; distance walked and session duration.
Activity status through the day	Feedback on the total activity performed until that point in the day and progress toward the daily goal.	Screen with numeric representations of active time: steps taken, calories burned and distance walked; progress bar showing progress toward the daily goal; option to manually entry activity; encouraging pop-up messages.
Activity history	Tracking past activity; graphic display; simple and intuitive.	History screen; bar chart representation of daily activity in relation to the goal; week-by-week information.
Periodic summary reports	Descriptive summary of the activity performed during the week and the overall progress in the program; tips for improvement; motivation to be active and to follow the program.	Weekly activity reports; presents a breakdown of the activity performed during the week; bar chart representation comparing current week activity with previous weeks; communicates progress in the plan; encourages users to follow the program; provides tips according to the user's physical barriers; informs users of PA benefits.
Challenges[c]	Unexpected activity challenges.	—[d]

[a]CALO-RE: Coventry, Aberdeen, and London—Refined taxonomy.

[b]PA: physical activity.

[c]To be considered in future iterations of the prototype.

[d]Not available.

Figure 2. Screenshots of the main prototype screens. (a) Welcome screen; (b) profile screen; (c) coach (main) screen on session day; (d) coach (main) screen on a step goal day; (e) MyActivity screen, with information on the user's current activity status; and (f) history screen, with information on past activity and access to the weekly summary reports.

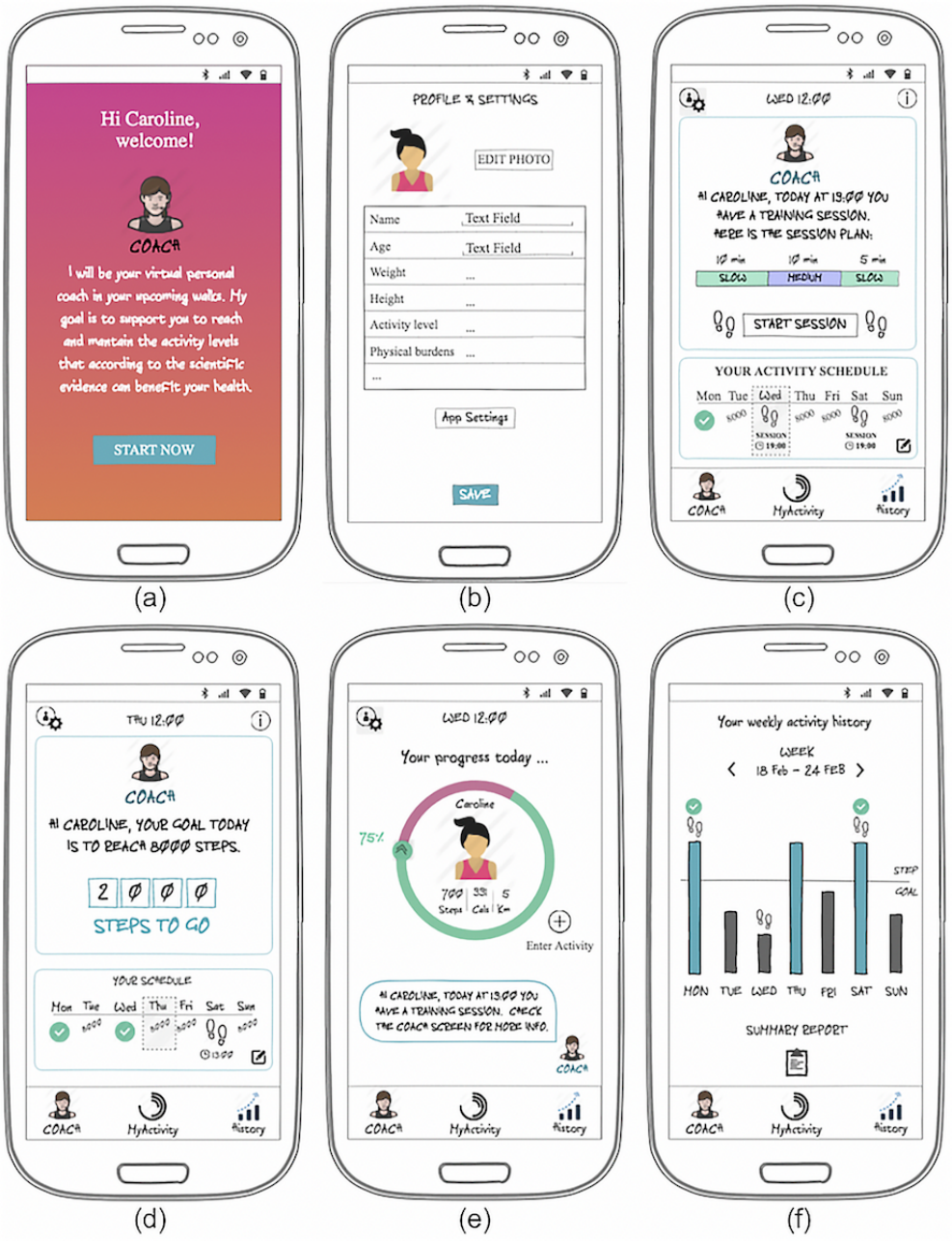
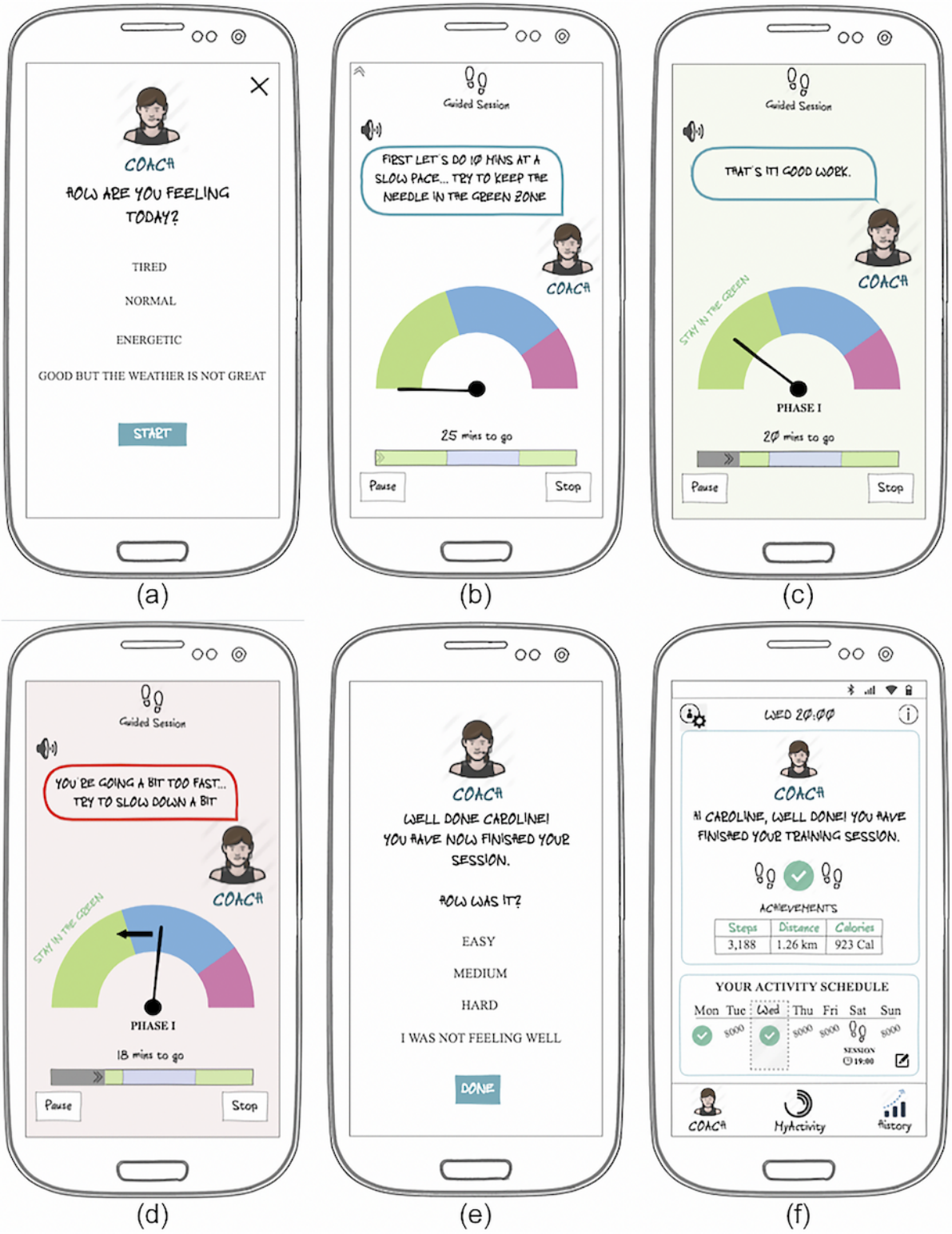


Figure 3. Screenshots of the simulation of a guided session. (a) Presession perceived tiredness, (b) example instruction for session phase duration and pace, (c) example cue to keep up the pace, (d) example cue to slow down the pace to the ideal zone of the current phase, (e) postsession perceived session difficulty, and (f) session achievements in the coach screen.



Physical Activity Program

From the evidence on the topic of PA and breast cancer, it is known that the beneficial effects associated with PA are more pronounced with moderate or vigorous intensity compared with mild intensity physical activities [66]. Breast cancer survivors are recommended to perform at least 120 to 150 min per week of moderate to vigorous activity [54]. This means that a common activity such as walking, which breast cancer survivors have a general preference for, if done at the right intensity and for the right time, can be sufficient to reach the minimum recommended levels of PA. Furthermore, walking is an activity that people can do everywhere and that nowadays can be easily monitored using smartphone technology. Hence, a training program was developed, which targets walking activities and aims to guide

these individuals to reach and maintain the amount and intensity of activity recommended in the guidelines.

The activity program was based on the model of linear progression training to allow for difficulty adjustment according to the user's actual PA level [67]. It will be integrated into a future functional version of the prototype in the form of a training plan. Adjustments will be made session by session, and involve, first, changing the session volume and, in the following session, changing the intensity level. There will be different difficulty levels in the plan: beginner, intermediate, adapted, and advanced. The activity plan will include 3 guided training sessions per week (with controlled intensity), with a minimum duration of 30 min per session. In addition, the user will be encouraged to reach a daily step goal that is adjusted weekly. The objectives in the plan will be gradually more challenging and adjusted regularly. Figure 2 depicts how these training specifications will be represented in the coach tab.

On the basis of the work of Tudor-Locke et al [68], walking cadence or walking pace will be measured to infer the training zone (intensity) of the activity being performed. Hence, based on this and on linear progression training, the training format will be defined as a number of training phases or traits with a certain duration, each on a specific training zone inferred from the walking pace—slow, medium, or fast (see example of session format in the coach tab screenshot, Figure 2).

From the training program specifications, some of the tool features were refined (eg, the coach tab and schedule, in Figure 2, and the real-time guided sessions to control the user's walking pace, in Figure 3).

Personalized Coaching System

The results from stage 1 highlighted the importance of exploring a personalized coaching experience. Therefore, the coaching system in the app will combine a number of strategies proposed in the model of real-time personalization in PA coaching apps. The different characteristics of the app that will be considered for the creation of a personalized experience are described below.

Virtual Coach

A human-like interface between the user and the app will simulate the interactions with a real coach (see virtual coach representations, Figures 2 and 3). It will serve as a sender or source of all the communication provided by the system through a variety of forms: general information, support, tips, activity feedback, motivation, and summary reports. It will provide clear, concise, and positive communication. Such a strategy is expected to increase the feeling of trust and credibility, increase motivation, and increase the feeling of personalized experience and interaction

Targeted Feedback

The app will provide feedback on the activities of the user (number of steps, distance, and calories burned), estimate calories spent by the user based on the user's characteristics, inform the user of their daily personalized goals and progress toward their goal, integrate the user's name in the communication provided by the coach, provide feedback in a weekly summary report considering the user's progress in current and past weeks, and provide tips according to the user's physical burdens.

Adaptive Activity Plan and Goals

Because these individuals vary in PA level and progress and, particularly in this population, there might be physical constraints due to side effects of both cancer and treatment, the app will integrate a rule-based module that adjusts the step goal and training session objectives in line with the model of linear progression training. This will be done considering the user's baseline level, progress, the user's perceived fatigue level (presession), and the user's perceived session difficulty level (postsession).

Real-Time Training Session Guidance

This feature will allow to coach users through the guided sessions, in real time. The coach instructional messages or cues will be designed to be clear, concise, and positive, using short and easy messages, and related to the visual content. It will provide feedback on the user's pace in relation to the ideal pace set for each of the session phases, sending cues to slow down, speed up or to keep pace, inform on time progress through the different phases and the whole session, and give positive reinforcement. Regarding representation, the feedback will be presented in textual, audio, and visual forms. The visual form will use a glanceable display (representing a speedometer) that will provide feedback on the current user's pace. The needle will represent the current user cadence and each color zone at a certain pace (slow—green, moderate—blue, and fast—pink). In addition, a progress bar will be used to indicate the current time, total session time, and progress through the different training phases (including the demarcated color zones). Figure 3 depicts an example real-time guided session simulation.

Stage 3: Prototype Testing

This stage of the process sought to explore an early evaluation of the interactive mock-ups, involving both user and expert testing. The evaluation was directed toward assessing the usability, quality, and behavior change potential of the concept ideated.

User Evaluation

In the task-based session, participants' overall results on the posttask SEQ questionnaire revealed that all tasks were easy or extremely easy to complete, with a mean of 6.6 out of 7 (SD 0.5). Participants' mean duration in completing the tasks was 4.6 (SD 1.2) seconds. The overall mean number of task errors was 0.3 (SD 0.4). Participants suggested that if they had previous practice, they would have easily completed the tasks without errors. Table 4 gives further details on the task-based session results.

The mean SUS score was 95 (SD 6.3) out of 100, which can be considered above average (above a score of 68). The score obtained can be converted to a percentile rank above 95%, which is interpreted as grade A+. This means that the prototype has higher perceived usability than 95% of all products tested with this scale (Multimedia Appendix 4).

From the short usability interviews, overall, the participants were very satisfied with the usability of the app, finding it very intuitive and easy to use. Furthermore, some participants suggested that an app like this would help them to get motivated to do PA, find time to do more activity, serve as a companion, provide support, and make them comply with the PA plan. In addition, participants reported some usability problems and suggested a number of improvements or additional app features. The information synthesized from the short interviews is presented in Textbox 2.

Table 4. Task-based session results.

Tasks	Task description	Participants with errors, n (%)	Errors, mean (SD)	Completion time in seconds, mean (SD)	SEQ[a] score, mean (SD)	Observations
Task 1	Which steps should you take to get to the profile screen?	4 (50)	1.0 (1.1)	4.8 (1.8)	6.5 (0.5)	This was the task with more user errors. Participants suggested the task was easy and that errors were associated with a lack of attention (eg, not taking a look at the whole screen before interacting with the prototype) or due to the lack of a first interaction experience with the prototype.
Task 2	Which steps should you take to find the amount of activity you have done so far today?	2 (25)	0.4 (0.7)	4.1 (3.1)	6.3 (1.2)	A generally easy to complete task. Errors were associated with participants going to the History tab instead of the MyActivity tab. This confusion originated from the ambiguity on how the task was posed, which made both answers correct in some way. In addition, it seemed that the purpose of the MyActivity tab was only clear after having a first look at it.
Task 3	Which steps should you take to reschedule an activity session from Friday to Saturday?	2 (25)	0.5 (0.9)	6.5 (4.3)	5.8 (1.0)	This task had the lowest SEQ[a] score and was the one that took longer to complete. Participants suggested that finding the rescheduling button was easy, but the process of setting the new schedule was not clear.
Task 4	Which steps should you take to start and finish a guided session?	1 (13)	0.1 (0.4)	4.6 (2.7)	6.6 (0.5)	A generally easy task. Error due to lack of attention (eg, not taking a look at the whole screen before interacting with the prototype) or due to the lack of a first interaction experience with the prototype. Participant pressed the MyActivity tab instead of the *Start Session* button. In addition, in the session simulation, some participants thought the green color was for fast pace and pink for slow pace, when in reality it was the opposite.
Task 5	Which steps should you take to find the activity you have done so far this week?	0 (0)	0.0 (0.0)	2.5 (0.8)	7.0 (0.0)	Very easy. No issues.
Task 6	Which steps should you take to find your activity history from last week?	0 (0)	0.0 (0.0)	5.0 (2.3)	7.0 (0.0)	Very easy. No issues.
Task 7	Which steps should you take to find the user manual?	1 (13)	0.1 (0.4)	5.0 (3.3)	6.8 (0.4)	Considered very easy. Error was associated with a lack of attention (eg, not taking a look at the whole screen before interacting with the prototype) or due to the lack of a first interaction experience with the prototype.

[a]SEQ: single ease question.

Textbox 2. Aspects highlighted by participants in the task-based session and interviews.

- Usability issues
  - The session plan phases understood as different session alternatives.
  - Issue with understanding the purpose of MyActivity tab.
  - The scheduling process not clear.
  - The purpose of the session icon in the history tab, to distinguish session days from step goal days, is not clear.
  - Green and pink zones of the speedometers associated with the opposite pace zones (eg, some participants thought green was for fast pace and pink for slow pace).
  - Issue associating the current screen to the corresponding buttons at the bottom.
- Positive perceptions
  - Easy to use and straightforward.
  - Guided session.
  - Prescription of an activity plan.
  - Quantification of activity, knowing calories spent, distance, and steps.
  - Validation by professionals, based on evidence.
  - Access to activity history.
  - May motivate the user to comply with the training.
  - May serve as a companion.
  - May reinforce motivation for physical activity.
- Suggestions for improvements
  - Change text font.
  - Improve graphic design.
  - Include feedback information on time spent walking.
  - Include recommendations and examples of resistance exercises.
  - Include optional short training.
  - Include 'free training', the option to do guided sessions on days without a scheduled session.
  - Include the option to download data to the computer to view annual progress.
  - Induce the user to relate the app and physical activity as another part of the treatment, but one that is fun.
  - Include audio feedback during the guided sessions.
  - Include training or an explanation of how the app works (user manual or guide).
  - Include reminders for activity.
  - Include sleep and weight tracking.

Expert Evaluation

Although not all aspects from the MARS could be addressed, because of the low-fidelity design of the prototype at this stage, the app quality mean score was 4.6 (SD 0.4) out of 5. In particular, for each of the sections on engagement, functionality, and information, the app mean scores were 4.2 (SD 0.6), 4.8 (SD 0.4), and 4.9 (SD 0.1) out of 5, respectively (Multimedia Appendix 5). Regarding the ABACUS scale, the app scored 15 out of 21 possible behavior change techniques (Multimedia Appendix 6). These results suggest good quality and potential for behavior change of the proposed concept.

Discussion

Principal Findings

In this study, we describe a UCD process for the development of a PA coaching app for breast cancer survivors. To our knowledge, this is the first work combining, and reporting in detail, the use of behavioral theory, personalized coaching strategies, and linear progression training for the design of a PA app for breast cancer. The design team gathered the user requirements and insights from experts in exercise and cancer and translated these into the concept of the solution. The concept was refined based on the theoretical foundations, and its viability was confirmed by the technical members of the team. An

interactive low-fidelity prototype of the tool was created and assessed in both user and expert evaluations. The user-centered process provided insight into the needs and preferences of the end users on PA apps, which will increase the likelihood of success of the proposed solution.

This tool was designed to simulate the interactions with an exercise coach. It includes an adaptive walking regimen and a number of personalized coaching features that aim to support and motivate users to adhere to it, with the ultimate goal of helping them to progressively reach and maintain the recommended levels of activity for breast cancer survivors. The main functionalities of the system include a training plan and schedule, adaptive goal setting, real-time feedback and motivation during walking sessions, activity status through the day, activity history, weekly summary reports, and activity challenges. One of the main features of the concept proposed is the ability to provide live coaching during the guided walking sessions, monitoring the user's cadence (through the built-in sensors in the phone), and providing real-time guidance and encouragement to keep the pace within the ideal training zones and to follow the session plan. This particular strategy ensures that users are doing an activity of high enough intensity and duration to meet the PA recommendations.

With regard to personalization, the app aims to (1) provide an automatic and reliable way of providing adaptive training and (2) provide personalized coaching for each user, considering a variety of individualization factors that include the users' personal characteristics (eg, name, age, height, and weight), physical burdens or barriers, baseline PA level, progress, perceived fatigue level, and perceived difficulty. For this purpose, a number of personalization strategies were used, which included a combination of feedback, user targeting, goal setting, and self-learning [26]. Other factors for individualization, including the user's preferences, routine, and the external context (location and weather), will be considered for integration in upcoming iterations of the solution.

The concept covers, at least in some form, 21 out of the 40 behavior change techniques defined in the CALO-RE taxonomy [38]. Some techniques were identified for possible inclusion in future iterations of the concept (eg, C25—involve a written agreement, C27 and C33—prompt self-talk, and C36—encourage stress management), and others were excluded as they were not in line with the user requirements and context (eg, C28—facilitate social comparison and C32—fear arousal). The use of this taxonomy helped in the concept design and in specifying the behavior change components of the tool, which may facilitate future reporting and evaluation of a technology-based PA intervention.

The results from prototype testing with users and experts were promising, with high scores for usability, quality, and behavior change potential. Several considerations can be taken from these evaluations to inform the future refinement of the prototype, which include further exploring engagement strategies, particularly related to entertainment, customization, and interactivity, and to consider other behavior change techniques, such as providing the ability to export data from the app, to suggest restructuring social or physical environment, and to assist with distraction or avoidance. Other system functionalities that might be considered in the future are, for example, to include resistance exercises in the activity program and app, to share experiences with close friends or family members, and to enable data sharing with health care professionals.

Comparison With Previous Work

There is a growing trend in the design of PA coaching systems that are aimed at individuals with chronic conditions. The solution presented here attempts to address the particular needs and requirements of breast cancer survivors for PA apps by taking a UCD approach for the conceptualization of the tool. Few are the systems in the literature designed with such a purpose, with only 4 apps identified that had some component of PA coaching for these individuals [23,24,28,29]. Compared with such apps, our solution targets aerobic exercise, specifically walking, which has been suggested as the preferred activity for these individuals. This allowed us to explore a design that addressed in detail the particular user requirements and preferences associated with such activity type. Although the integration of a component for resistance training might be considered in the future, we believe that it would necessarily bring other considerations for design and different implementation requirements, some of which have already been reported in previous work [24]. Another differentiating factor of this work relies on the attempt to create a personalized app experience for these users, a need that was reported recurrently in related literature on PA apps in breast cancer survivorship. Furthermore, we used the CALO-RE taxonomy, which provided important considerations for the integration of behavior change techniques in the tool design.

Despite some similarities in app functionalities compared with those reported for other populations, the solution proposed in this work differs in some ways and has characteristics that are more particular for breast cancer survivors. For example, studies on a healthy population [69,70] and for people with osteoarthritis [71] have highlighted the importance of social and game-like features (eg, competition). In a study with multiple sclerosis, participants were also interested in gamification and wanted, particularly, a tool focused on fatigue management [72]. Our requirements for breast cancer survivors were not directed toward playful and social experiences or fatigue management, which is similar to the findings of a study on chronic obstructive pulmonary disease patients [73]. Besides, our concept has some characteristics that are more specific to the needs and preferences of our target population, which include a tool for managing time for PA (scheduling feature); a PA program, monitoring and feedback targeted specifically at walking activities; and a personalization system that considers the user's individual characteristics and progress to adjust training and communication provided by a virtual coach.

Related literature reports on key design requirements for a successful PA coaching solution. For example, Consolvo et al [74] demonstrated the importance of giving users proper credit for activities, providing personal awareness of the activity level, supporting social influence, and considering the practical constraints of users' lifestyles. Bielik et al [75] built on those recommendations and highlighted that these systems should

also ensure fair play, provide a variety of motivational tools, provide feedback on activities done, provide short-term and long-term motivation, provide the possibility of integration with existing solutions, and protect users' privacy. Our concept of the tool seems to be in line with most of these recommendations. The design requirement related to the user's privacy is an aspect that also came up in the user research stage and that will need to be carefully considered in the development stage, particularly considering the level of personalization proposed for this system. In addition, as stated by Matthews et al [76], better system credibility support features (eg, trustworthiness, expertise, authority, third-party endorsement, and verifiability) need to be incorporated in PA mobile apps if they aim to achieve the highest level of persuasiveness. Such characteristics were also highlighted in our findings from the initial stage of design. Hence, we believe that pointing out to the users that this app was created in collaboration with other breast cancer survivors and professionals, having content based on the evidence, and simulating the interactions with an activity coach, might increase their trust and, therefore, their interest in the system.

Little is known about how to translate and apply the many high-level theory and design recommendations to these systems. In this paper, we have tried to present in a clear and practical way the UCD of a mobile PA coaching app based on foundations from behavior change and personalization constructs. We expect this to facilitate future research on the design of mHealth solutions, particularly PA coaching systems.

Limitations

The results of the studies with the end users should be considered with caution because of the small sample size. To complement the results of the qualitative study, which included 14 participants, we looked into the literature and compared it with related work. Regarding usability assessment, it is known that tests with 5 participants are able to uncover 85% of usability issues [77,78]. In our study, we had the involvement of 8 participants, and therefore, we think most usability issues have been revealed.

The system evaluation was performed with a low-fidelity prototype, which might have limited the outcomes from the testing stage in some way. For example, interaction of the user with the prototype was limited to the main screen interactions, which did not allow a real simulation of the functionalities provided by the app. In addition, the analysis of some of the testing scales was restricted to the current stage of the prototype. On the other hand, the fact that we have performed such comprehensive early testing provided us with important insights that will feed the next stages of the design and development process.

Future Work

From the concept created in this work, we have started the development process of a functional prototype. The next steps involve more iterations on the concept and further feature refinement to detail the app content, integrate the PA program, integrate an appealing graphic design, and perform further system evaluation. Once the functional prototype is developed, a 2- to 3-week pilot study with breast cancer survivors will be performed to assess potential acceptability, usability, and feasibility.

Conclusions

Centering the design on the end users, breast cancer survivors, and their context revealed valuable requirements and considerations to be taken into account for the design of a tool that aims to address their specific PA needs and motivate these individuals to increase their PA levels. This is essential for increasing the usability and acceptability of the tool. Furthermore, informing the process on the theory and constructs used in tailored PA coaching interventions provided important design insights, which may contribute to the effectiveness, long-term adherence, and acceptability of the system.

At this stage, we have confirmed good usability, quality, and behavior change potential of the prototype in a laboratory setting. A functional prototype will further be tested in pilot and feasibility studies in a real-world environment before it can go through more controlled trials evaluating long-term effectiveness.

This paper details the UCD process for a PA coaching app for breast cancer survivors, which may inform other researchers and developers working in similar mHealth tools.

Acknowledgments

This work was performed in collaboration with Salumedia Tecnologías, the University of Seville, the Oncoavanze Clinic, and the Insight Centre for Data Analytics from the University College Dublin. The study was supported by the European Union's Horizon 2020 research and innovation program: Marie Sklodowska-Curie Actions grant agreement number 722012—Cancer: Activating Technology for Connected Health. For most of the study period FG worked at Salumedia and GS worked at Oncoavanze, until October 2019 and November 2019, respectively. The authors would like to thank Matilde Mora Fernandez, from the University of Seville, for all the useful insights provided during the research. The authors would also like to thank Dr Pedro Valero, medical director and oncologist at Oncoavanze; all the Oncoavanze personnel; and all the patients who participated in the studies for their involvement and interest in this research.

Authors' Contributions

FG was involved in the study design, all data acquisition (interviews and UE), data analysis and interpretation, and all tool design stages and drafted the paper. OR participated in the study design and analysis of the interviews, was involved in most of the tool design stages, and helped to draft the paper. EZ participated in stage 1 of design, expert testing, and interpretation of the data.

GS participated in feature selection and provided important input from an exercise training point of view. ST was involved in the expert testing stage and provided important input from an interhuman interaction and UCD point of view. BC and LL were involved in the study conception and design and helped to draft the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

LL is a shareholder of Salumedia Labs, which is a digital health company that develops and commercializes mHealth solutions for supporting cancer patients. FG worked as a research fellow at Salumedia Labs.

Multimedia Appendix 1
Participants' access to technology and technology usage.
[DOCX File , 3886 KB - mhealth_v8i7e17552_app1.docx]

Multimedia Appendix 2
Table with considerations on psychological mediators of physical activity adherence in breast cancer survivors associated with Coventry, Aberdeen, and London—refined taxonomy constructs.
[DOCX File , 20 KB - mhealth_v8i7e17552_app2.docx]

Multimedia Appendix 3
Extended version of Table 3 with list of app functionalities and related design consequences from stage 1 and from the Coventry, Aberdeen, and London—Refined taxonomy.
[DOCX File , 18 KB - mhealth_v8i7e17552_app3.docx]

Multimedia Appendix 4
Table with system usability scale results.
[DOCX File , 14 KB - mhealth_v8i7e17552_app4.docx]

Multimedia Appendix 5
Table with mobile app rating scale results.
[DOCX File , 15 KB - mhealth_v8i7e17552_app5.docx]

Multimedia Appendix 6
Table with app behavior change scale results.
[DOCX File , 16 KB - mhealth_v8i7e17552_app6.docx]

References

1. Cho J, Jung S, Lee JE, Shim E, Kim NH, Kim Z, et al. A review of breast cancer survivorship issues from survivors' perspectives. J Breast Cancer 2014 Sep;17(3):189-199 [FREE Full text] [doi: 10.4048/jbc.2014.17.3.189] [Medline: 25320616]
2. Bluethmann SM, Bartholomew LK, Murphy CC, Vernon SW. Use of theory in behavior change interventions. Health Educ Behav 2017 Apr;44(2):245-253 [FREE Full text] [doi: 10.1177/1090198116647712] [Medline: 27226430]
3. Duijts SF, Faber MM, Oldenburg HS, van Beurden M, Aaronson NK. Effectiveness of behavioral techniques and physical exercise on psychosocial functioning and health-related quality of life in breast cancer patients and survivors-a meta-analysis. Psychooncology 2011 Feb;20(2):115-126. [doi: 10.1002/pon.1728] [Medline: 20336645]
4. Schmitz KH, Speck RM. Risks and benefits of physical activity among breast cancer survivors who have completed treatment. Womens Health (Lond) 2010 Mar;6(2):221-238. [doi: 10.2217/whe.10.1] [Medline: 20187728]
5. Smith SG, Chagpar AB. Adherence to physical activity guidelines in breast cancer survivors. Am Surg 2010 Sep;76(9):962-965. [Medline: 20836343]
6. Mason C, Alfano CM, Smith AW, Wang C, Neuhouser ML, Duggan C, et al. Long-term physical activity trends in breast cancer survivors. Cancer Epidemiol Biomarkers Prev 2013 Jun;22(6):1153-1161 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0141] [Medline: 23576689]
7. Schueller SM, Muñoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 3;22(6):478-483. [doi: 10.1177/0963721413495872]
8. 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]

9. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: a systematic review. Health Informatics J 2016 Sep;22(3):451-469. [doi: 10.1177/1460458214567004] [Medline: 25649783]
10. O'Reilly GA, Spruijt-Metz D. Current mhealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]
11. Uhm KE, Yoo JS, Chung SH, Lee JD, Lee I, Kim JI, et al. Effects of exercise intervention in breast cancer patients: is mobile health (mhealth) with pedometer more effective than conventional program using brochure? Breast Cancer Res Treat 2017 Feb;161(3):443-452. [doi: 10.1007/s10549-016-4065-8] [Medline: 27933450]
12. Robertson MC, Tsai E, Lyons EJ, Srinivasan S, Swartz MC, Baum ML, et al. Mobile health physical activity intervention preferences in cancer survivors: a qualitative study. JMIR Mhealth Uhealth 2017 Jan 24;5(1):e3 [FREE Full text] [doi: 10.2196/mhealth.6970] [Medline: 28119278]
13. Phillips SM, Courneya KS, Welch WA, Gavin KL, Cottrell A, Nielsen A, et al. Breast cancer survivors' preferences for mhealth physical activity interventions: findings from a mixed methods study. J Cancer Surviv 2019 Apr;13(2):292-305 [FREE Full text] [doi: 10.1007/s11764-019-00751-3] [Medline: 30912011]
14. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 8;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
15. Lazar A, Koehler C, Tanenbaum TJ, Nguyen DH. Why We Use and Abandon Smart Devices. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; September 7-11, 2015; Osaka, Japan. [doi: 10.1002/ijc.25440]
16. Ledger D, McCaffrey D. Medium Corporation. 2014. Inside Wearables: How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2020-04-27]
17. Cabrita M, den Akker H, Tabak M, Hermens HJ, Vollenbroek-Hutten MM. Persuasive technology to support active and healthy ageing: an exploration of past, present, and future. J Biomed Inform 2018 Aug;84:17-30 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.010] [Medline: 29935348]
18. Plank M, Dalal N. Bounce: A Mobile Behavioral Intervention Technology for Breast Cancer Survivors. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7-11, 2016; San Jose, USA. [doi: 10.1145/2851581.2890383]
19. Brunet J, Taran S, Burke S, Sabiston CM. A qualitative exploration of barriers and motivators to physical activity participation in women treated for breast cancer. Disabil Rehabil 2013;35(24):2038-2045. [doi: 10.3109/09638288.2013.802378] [Medline: 23772995]
20. Lavallée JF, Abdin S, Faulkner J, Husted M. Barriers and facilitators to participating in physical activity for adults with breast cancer receiving adjuvant treatment: a qualitative metasynthesis. Psychooncology 2019 Mar;28(3):468-476. [doi: 10.1002/pon.4980] [Medline: 30657225]
21. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Ghaoui C, editor. Bainbridge, W. Encyclopedia of Human Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004. [Medline: 21099019]
22. de Vito DA, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
23. Lee H, Uhm KE, Cheong IY, Yoo JS, Chung SH, Park YH, et al. Patient satisfaction with mobile health (mhealth) application for exercise intervention in breast cancer survivors. J Med Syst 2018 Nov 6;42(12):254. [doi: 10.1007/s10916-018-1096-1] [Medline: 30402781]
24. Marcu G, Misra A, Caro K, Plank M, Leader A, Barsevick A. Bounce: Designing a Physical Activity Intervention for Breast Cancer Survivors. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PrevasiveHealth'18; May 21-24, 2018; New York, USA. [doi: 10.1145/3240925.3240927]
25. op den Akker H, Jones VM, Hermens HJ. Tailoring real-time physical activity coaching systems: a literature survey and model. User Model User-Adap Inter 2014 Jun 25;24(5):351-392 [FREE Full text] [doi: 10.1007/s11257-014-9146-y]
26. Monteiro-Guerra FM, Rivera-Romero O, Fernandez Luque L, Caulfield B. Personalization in real-time physical activity coaching using mobile applications: a scoping review. IEEE J Biomed Health Inform 2019 Nov 19 epub ahead of print. [doi: 10.1109/JBHI.2019.2947243] [Medline: 31751254]
27. Hawkins RP, 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]
28. Economou D, Dwek M, Roberston C, Elliott B, Kounenis T, Azimi T, et al. PhytoCloud: A Gamified Mobile Web Application to Modulate Diet and Physical Activity of Women with Breast Cancer. In: Proceedings of the 30th International Symposium on Computer-Based Medical Systems. 2017 Presented at: CBMS'17; June 22-24, 2017; Thessaloniki, Greece. [doi: 10.1109/cbms.2017.164]
29. Zhang X, Deng Z, Parvinzamir F, Dong F. MyHealthAvatar lifestyle management support for cancer patients. Ecancermedicalscience 2018;12:849 [FREE Full text] [doi: 10.3332/ecancer.2018.849] [Medline: 30079111]

30. Plank M, Dalal N. Bounce: A Mobile Behavioral Intervention Technology for Breast Cancer Survivors. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA'16; May 7-12, 2016; San Jose, USA. [doi: 10.1145/2851581.2890383]
31. Shah SG, Robinson I, Al Shawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
32. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 2;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
33. Timmerman JG, Tönis TM, van Weering MG, Stuiver MM, Wouters MW, van Harten WH, et al. Co-creation of an ICT-supported cancer rehabilitation application for resected lung cancer survivors: design and evaluation. BMC Health Serv Res 2016 Apr 27;16:155 [FREE Full text] [doi: 10.1186/s12913-016-1385-7] [Medline: 27121869]
34. 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]
35. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
36. Bandura A. Self-Efficacy: The Exercise of Control. New York, USA: WH Freeman and Company; 1997.
37. McAuley E, Blissmer B. Self-efficacy determinants and consequences of physical activity. Exerc Sport Sci Rev 2000 Apr;28(2):85-88. [Medline: 10902091]
38. 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]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Wohlin C. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014 Presented at: EASE'14; May 13-14, 2014; London, England. [doi: 10.1145/2601248.2601268]
41. Beyer H, Holtzblatt K. Contextual design. Interactions 1999;6(1):32-42 [FREE Full text] [doi: 10.1145/291224.291229]
42. Wilson C. Brainstorming and Beyond: A User-Centered Design Method. Burlington, Massachusetts: Morgan Kaufmann; 2013.
43. Kumar A, Sharma A. SWOT analysis of requirements engineering for web applications. J Adv Res Sci Eng 2015;4:34-43 [FREE Full text]
44. Pernice K. Nielsen Norman Group. 2016. UX Prototypes: Low Fidelity vs High Fidelity URL: https://www.nngroup.com/articles/ux-prototype-hi-lo-fidelity/ [accessed 2020-04-27]
45. International Organization for Standardization. 1998. ISO 9241-11:1998: Ergonomic Requirements for Office Work With Visual Display Terminals (VDTS)-Part 11: Guidance on Usability URL: https://www.iso.org/standard/16883.html [accessed 2020-04-27]
46. Frøkjær E, Hertzum M, Hornbæk K. Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2000 Presented at: CHI'00; April 1-6, 2000; The Hague, The Netherlands. [doi: 10.1145/332040.332455]
47. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
48. van der Weegen S, Verwey R, Tange HJ, Spreeuwenberg MD, de Witte LP. Usability testing of a monitoring and feedback tool to stimulate physical activity. Patient Prefer Adherence 2014;8:311-322 [FREE Full text] [doi: 10.2147/PPA.S57961] [Medline: 24669188]
49. Sauro J, Dumas JS. Comparison of Three One-Question, Post-Task Usability Questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, Massachusetts. [doi: 10.1145/1518701.1518946]
50. Brooke J. SUS: a 'quick and dirty' Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996:4-7.
51. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
52. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
53. 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]

54. Rock CL, Doyle C, Demark-Wahnefried W, Meyerhardt J, Courneya KS, Schwartz AL, et al. Nutrition and physical activity guidelines for cancer survivors. CA Cancer J Clin 2012;62(4):243-274 [FREE Full text] [doi: 10.3322/caac.21142] [Medline: 22539238]
55. Ventura EE, Ganz PA, Bower JE, Abascal L, Petersen L, Stanton AL, et al. Barriers to physical activity and healthy eating in young breast cancer survivors: modifiable risk factors and associations with body mass index. Breast Cancer Res Treat 2013 Nov;142(2):423-433 [FREE Full text] [doi: 10.1007/s10549-013-2749-x] [Medline: 24177756]
56. Cho D, Park C. Barriers to physical activity and healthy diet among breast cancer survivors: a multilevel perspective. Eur J Cancer Care (Engl) 2018 Jan;27(1):e12772 [FREE Full text] [doi: 10.1111/ecc.12772] [Medline: 28940854]
57. Hefferon K, Murphy H, McLeod J, Mutrie N, Campbell A. Understanding barriers to exercise implementation 5-year post-breast cancer diagnosis: a large-scale qualitative study. Health Educ Res 2013 Oct;28(5):843-856. [doi: 10.1093/her/cyt083] [Medline: 23969632]
58. Phillips SM, Conroy DE, Keadle SK, Pellegrini CA, Lloyd GR, Penedo FJ, et al. Breast cancer survivors' preferences for technology-supported exercise interventions. Support Care Cancer 2017 Oct;25(10):3243-3252 [FREE Full text] [doi: 10.1007/s00520-017-3735-3] [Medline: 28470368]
59. Roberts AL, Potts HW, Koutoukidis DA, Smith L, Fisher A. Breast, prostate, and colorectal cancer survivors' experiences of using publicly available physical activity mobile apps: qualitative study. JMIR Mhealth Uhealth 2019 Jan 4;7(1):e10918 [FREE Full text] [doi: 10.2196/10918] [Medline: 30609982]
60. Nguyen NH, Hadgraft NT, Moore MM, Rosenberg DE, Lynch C, Reeves MM, et al. A qualitative evaluation of breast cancer survivors' acceptance of and preferences for consumer wearable technology activity trackers. Support Care Cancer 2017 Nov;25(11):3375-3384. [doi: 10.1007/s00520-017-3756-y] [Medline: 28540402]
61. Pope Z, Lee JE, Zeng N, Lee HY, Gao Z. Feasibility of smartphone application and social media intervention on breast cancer survivors' health outcomes. Transl Behav Med 2019 Jan 1;9(1):11-22. [doi: 10.1093/tbm/iby002] [Medline: 29471477]
62. Hong Y, Dahlke DV, Ory M, Hochhalter A, Reynolds J, Purcell NP, et al. Designing iCanFit: a mobile-enabled web application to promote physical activity for older cancer survivors. JMIR Res Protoc 2013 Feb 14;2(1):e12 [FREE Full text] [doi: 10.2196/resprot.2440] [Medline: 23612053]
63. Short CE, James EL, Plotnikoff RC. Theory-and evidence-based development and process evaluation of the Move More for Life program: a tailored-print intervention designed to promote physical activity among post-treatment breast cancer survivors. Int J Behav Nutr Phys Act 2013 Nov 5;10:124 [FREE Full text] [doi: 10.1186/1479-5868-10-124] [Medline: 24192320]
64. Basen-Engquist K, Taylor CL, Rosenblum C, Smith MA, Shinn EH, Greisinger A, et al. Randomized pilot test of a lifestyle physical activity intervention for breast cancer survivors. Patient Educ Couns 2006 Dec;64(1-3):225-234. [doi: 10.1016/j.pec.2006.02.006] [Medline: 16843633]
65. Pinto BM, Floyd A. Theories underlying health promotion interventions among cancer survivors. Semin Oncol Nurs 2008 Aug;24(3):153-163. [doi: 10.1016/j.soncn.2008.05.003] [Medline: 18687261]
66. Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. Breast Cancer Res Treat 2013 Feb;137(3):869-882. [doi: 10.1007/s10549-012-2396-7] [Medline: 23274845]
67. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. Tenth Edition. Ontario, Canada: Wolters Kluwer; 2018.
68. Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna JM, Kang M, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. Br J Sports Med 2018 Jun;52(12):776-788 [FREE Full text] [doi: 10.1136/bjsports-2017-097628] [Medline: 29858465]
69. Jee H. Review of researches on smartphone applications for physical activity promotion in healthy adults. J Exerc Rehabil 2017 Feb;13(1):3-11 [FREE Full text] [doi: 10.12965/jer.1732928.464] [Medline: 28349027]
70. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
71. Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an exercise app in the care for people with osteoarthritis: user-driven exploratory study. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e11 [FREE Full text] [doi: 10.2196/mhealth.7734] [Medline: 29326092]
72. Giunti G, Kool J, Romero OR, Zubiete ED. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 9;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
73. Bartlett YK, Webb TL, Hawley MS. Using persuasive technology to increase physical activity in people with chronic obstructive pulmonary disease by encouraging regular walking: a mixed-methods study exploring opinions and preferences. J Med Internet Res 2017 Apr 20;19(4):e124 [FREE Full text] [doi: 10.2196/jmir.6616] [Medline: 28428155]
74. Consolvo S, Everitt K, Smith I, Landay J. Design Requirements for Technologies That Encourage Physical Activity. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006 Presented at: CHI'06; April 22-27, 2006; Montréal, Canada. [doi: 10.1145/1124772.1124840]

75. Bielik P, Tomlein M, Krátky P, Mitrík S, Barla M, Bieliková M. Move2Play: An Innovative Approach to Encouraging People to Be More Physically Active. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. 2012 Presented at: IHI'12; January 28-30, 2012; Miami Florida, USA. [doi: 10.1145/2110363.2110374]
76. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
77. Nielsen J. Nielsen Norman Group. 2000. Why You Only Need to Test With 5 Users URL: http://www.useit.com/alertbox/20000319.html [accessed 2020-04-15]
78. Nielsen J, Landauer TK. A Mathematical Model of the Finding of Usability Problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Presented at: CHI'93; April 24-29, 1993; Amsterdam, The Netherlands. [doi: 10.1145/169059.169166]

Abbreviations

ABACUS: app behavior change scale
CALO-RE: Coventry, Aberdeen, and London—Refined taxonomy
MARS: mobile app rating scale
mHealth: mobile health
PA: physical activity
SCT: social cognitive theory
SDT: self-determination theory
SEQ: single ease question
SUS: system usability scale
UCD: user-centered design

Edited by G Eysenbach; submitted 19.12.19; peer-reviewed by J Hwang, L Grepo, K Blondon, K Riiser; comments to author 06.03.20; revised version received 25.03.20; accepted 29.03.20; published 15.07.20.

Please cite as:
Monteiro-Guerra F, Signorelli GR, Tadas S, Dorronzoro Zubiete E, Rivera Romero O, Fernandez-Luque L, Caulfield B
A Personalized Physical Activity Coaching App for Breast Cancer Survivors: Design Process and Early Prototype Testing
JMIR Mhealth Uhealth 2020;8(7):e17552
URL: https://mhealth.jmir.org/2020/7/e17552
doi:10.2196/17552
PMID:32673271

Original Paper

Formative Evaluation of a Smartphone App for Monitoring Daily Meal Distribution and Food Selection in Adolescents: Acceptability and Usability Study

Billy Langlet[1], PhD; Christos Maramis[2], PhD; Christos Diou[3], PhD; Nikolaos Maglaveras[2], PhD; Petter Fagerberg[1], MSc; Rachel Heimeier[4], PhD; Irini Lekka[2], MSc; Anastasios Delopoulos[3], PhD; Ioannis Ioakimidis[1], PhD

[1]The Innovative Use of Mobile Phones to Promote Physical Activity and Nutrition Across the Lifespan Research Group, Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

[2]Lab of Computing, Medical Informatics and Biomedical Imaging Technologies, Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[3]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

[4]Internationella Engelska Gymnasiet, Stockholm, Sweden

Corresponding Author:
Billy Langlet, PhD
The Innovative Use of Mobile Phones to Promote Physical Activity and Nutrition Across the Lifespan Research Group
Department of Biosciences and Nutrition
Karolinska Institutet
Blickagången 16
Huddinge,
Sweden
Phone: 46 08 524 810 95
Email: billy.langlet@ki.se

Abstract

Background: Obesity interventions face the problem of weight regain after treatment as a result of low compliance. Mobile health (mHealth) technologies could potentially increase compliance and aid both health care providers and patients.

Objective: This study aimed to evaluate the acceptability and usability and define system constraints of an mHealth system used to monitor dietary habits of adolescents in real life, as a first step in the development of a self-monitoring and lifestyle management system against adolescent obesity.

Methods: We recruited 26 students from a high school in Stockholm, Sweden. After a 30-minute information meeting and 5-minute individual instruction on how to use an mHealth system (smartphone with app and two external sensors), participants used it for 2-3 weeks to objectively collect dietary habits. The app and sensors were used by the participants, without supervision, to record as many main meals and snacks as possible in real life. Feasibility was assessed following the "mHealth evidence reporting and assessment checklist," and usability was assessed by questionnaires. Compliance was estimated based on system use, where a registration frequency of 3 main meals (breakfast, lunch, and dinner) per day for the period of the experiment, constituted 100% compliance.

Results: Participants included in the analysis had a mean age of 16.8 years (SD 0.7 years) and BMI of 21.9 kg/m2 (SD 4.1 kg/m2). Due to deviations from study instructions, 2 participants were excluded from the analysis. During the study, 6 participants required additional information on system use. The system received a 'Good' grade (77.1 of 100 points) on the System Usability Scale, with most participants reporting that they were comfortable using the smartphone app. Participants expressed a willingness to use the app mostly at home, but also at school; most of their improvement suggestions concerned design choices for the app. Of all main meals, the registration frequency increased from 70% the first week to 76% the second week. Participants reported that 40% of the registered meals were home-prepared, while 34% of the reported drinks contained sugar. On average, breakfasts took place at 8:30 AM (from 5:00 AM to 2:00 PM), lunches took place at 12:15 PM (from 10:15 AM to 6:15 PM), and dinners took place at 7:30 PM (from 3:00 PM to 11:45 PM). When comparing meal occurrence during weekdays vs weekends, breakfasts and lunches were eaten 3 hours later during weekends, while dinner timing was unaffected.

Conclusions: From an infrastructural and functional perspective, system use was feasible in the current context. The smartphone app appears to have high acceptability and usability in high school students, which are the intended end-users. The system appears

promising as a relatively low-effort method to provide real-life dietary habit measurements associated with overweight and obesity risk.

(JMIR Mhealth Uhealth 2020;8(7):e14778) doi:10.2196/14778

KEYWORDS

mHealth; eHealth; dietary behavior; lifestyle behavioral monitoring; lifestyle interventions; obesity; mobile phone; smartphone; weight management; overweight

Introduction

Obesity prevalence in children and adolescents between the ages of 5 and 19 increased 8-9 times between 1975 and 2016 [1]. Obesity, in turn, increases the risk of physical and mental health complications such as osteoarthritis, type 2 diabetes mellitus, and coronary heart disease [2], along with depression [3]. In addition, a lifetime of obesity increases the risk of disability and loss of income [4].

Comprehensive lifestyle interventions, targeting diet, physical activity, and lifestyle elements in tandem, are recommended by the American Heart Association, American Colleague of Cardiology, and The Obesity Society [5] as a means of accomplishing long-term weight loss and weight maintenance in adults. Indeed, studies have shown that lifestyle interventions can promote moderate weight reduction with low risk of adverse effects [6,7], even in children between the ages of 0 and 18 years, as reported in a recent Cochrane review [8]. However, most lifestyle interventions face the problems of low compliance and high drop-out rates, with interventions often reporting weight regains of almost 50% at the 1-year follow-up mark[9,10]. In primary care, while it seems possible for obesity treatment to be "relatively cost-effective" in the short term [11], long-term weight maintenance requires intensive management in the form of frequent visits and multiple contact hours [12,13]. Others have identified additional means for improving the effectiveness of interventions, including frequent and open communication among patients and health professionals, realistic behavioral goal setting, and continuous patient monitoring [14,15]. In practice though, the inherent time and monetary cost of involved health monitoring often shift the burden to the individuals, requiring them to self-monitor and self-manage their behaviors for longer periods of time [16]. However, this further increases the required effort from the side of the individual, potentially reducing compliance and increasing drop-outs.

To meet this challenge, novel mobile health (mHealth) tools can be developed to assist both patients and health professionals. Indeed, with the accessibility and use of smartphones greatly increasing [17], smartphones are a cheap, accessible option for implementing interventions. This is even more pronounced for interventions targeting children and adolescents since they are early adopters of new technology [18]. Additionally, smartphones (and other wearable electronic devices) create the opportunity for collecting information that may be difficult for users to self-report, such as exact times, locations, and types of behaviors [19], resulting in increased user compliance with smartphone-based interventions [19,20].

Current efforts for the development of large-scale, non-invasive dietary behavioral monitoring systems appropriate for use by children and adolescents in challenging real-life environments, like school, are limited [21]. However, novel methodologies are constantly being tested [22], including our own efforts for collection of large-scale Big Health Data about the dietary and physical activity habits of school children [23]. However, such large-scale deployments require conscious efforts for preliminary testing outside the lab, in the target use environment, to identify potential use constraints and optimize the deployed methodologies.

On a commercial level, smartphone apps claiming to promote weight loss have seen a huge increase recently, but their alleged effectiveness lacks support by large-scale, long-term clinical studies [24]. Meanwhile, the number of mHealth self-management programs against obesity constantly increases, but due to mHealth being in its infancy, evidence for their effectiveness is sparse [25]. Recent meta-analyses have concluded that mHealth interventions performed similarly, if not better, then traditional non-mHealth–based obesity interventions [26-28] but emphasized the heterogeneity in the quality of the existing evidence base. Realizing the lack of consensus on mHealth reports, the World Health Organization has developed a framework for what and how to report on the various components of an mHealth intervention, called the "mHealth evidence reporting and assessment checklist" (mERA), aimed at improving the quality of the existing evidence [25]. Once the risk population has been identified, interventions should identify ways to reach the population in question, performing a formative evaluation, focusing on gathering functional requirements, and developing and testing the technology in a descriptive way [29]. Evaluating the effectiveness of mHealth in the intended population requires the evaluation of at least three components: usage (objective), acceptability (subjective), and feasibility of implementation [25].

In this study, the aim was to evaluate the acceptability, usability, and constraints of a newly developed mHealth system used for monitoring eating behaviors and dietary habits of adolescents in real life, as a first step for the development of a more comprehensive self-monitoring and lifestyle management system. The evaluation of the system was conducted on two levels: subjective self-reports by users and objective compliance estimations of the study protocol (eg, frequency of registering eating events). The system's evaluation followed the mERA checklist for a more transparent report of its feasibility and usability. At its core, the developed system employed a smartphone to collect data from a smartphone app and two additional sensors, and all data were automatically exported to

a centralized data collection platform where the data were analyzed and results collected. The smartphone app allowed participants to self-report eating events throughout the day, while the two additional sensor modules enabled collection of additional objective eating behavior data. The development of these modules, a food scale, and a chewing sensor has been previously described elsewhere [30], and their use was complementary and fully integrated with the system described herein. Finally, this report presents population-level data on dietary habits associated with the development and maintenance of adolescent obesity (ie, frequency of meals consumed in restaurants, frequency of sugary beverage consumption, and analysis of meal timing across days during weeks vs weekends). The detailed analyses of these datasets provide additional evidence of the usefulness of the system, setting the basis for the development of future interventions against obesity in the targeted population.

Methods

Subjects

There were no specific inclusion or exclusion criteria, and all students from 6 classes of the collaborating school were informed of the study. The classes were selected together with the school administration based on their end-of-the-year scheduled obligations and their prior participation in a past relevant study [31]. With 26 system units (a smartphone together with a digital food scale and chewing sensor) available, a first-come-first-serve approach was used for recruitment. Participants 18 years or older provided written consent, while younger participants provided written assent together with written consent by their legal guardians. The forms to be signed were provided to the students after an informational meeting, where students could ask for clarifications regarding the protocol of the study. The researchers returned to the school on two later occasions to collect the signed forms. Ethical approval was provided by the Stockholm Regional Ethics Board (D.nr.: 2015/1824-31), and the presented practices followed the Declaration of Helsinki's guidelines for human research. Participants who completed the study received cinema tickets as compensation for their participation. All the participating students had previously participated in a study including lunch recordings in a school cafeteria, which followed a protocol described elsewhere [31].

Experimental Design

During the study, the subjects participated in a data collection action lasting between 2 and 3 weeks that took place towards the end of the school year. During this period (varying for each student due to their individualized end-of-year obligations), the students were asked to register all their weekday and weekend meals. The basic meal registration was done using a smartphone, and participants could contribute additional eating behavior data by using either the food scale or chewing sensor.

Study Protocol

The protocol of the study was uncontrolled by design, aiming to capture the true dietary habits of the participants, diminishing the required effort on their part. All textual components of the system were in English, in agreement with the teaching language in the selected school. The study began by participants receiving information on system use and study protocol twice: first in a group meeting lasting for 30 minutes and then on an individual level for approximately 5 minutes. During the individual meeting, the students received information on system use, went through all the screens and options of the mobile app, and were handed the smartphone and complementary devices, also signaling the initiation of the data collection. The participants were requested to register all their meals as they happened (not retrospectively) during the study period using the provided smartphone app and were also asked to use the provided food scale to collect additional information for main meals (breakfast, lunch, and dinner) and the chewing sensor to collect information for main meals and snacks. Depending on the students' school schedule and their individual school obligations for the end of the school year, the study duration ranged from 2 to 3 consecutive weeks, after individual arrangements with their supervising teachers. A researcher was available to provide technical and protocol support to the students from Monday to Friday (school days) throughout the study. After the predefined end of the data collection period per student, the participants returned the devices to the researchers and completed the system usability and user experience questionnaires.

Devices and Smartphone App

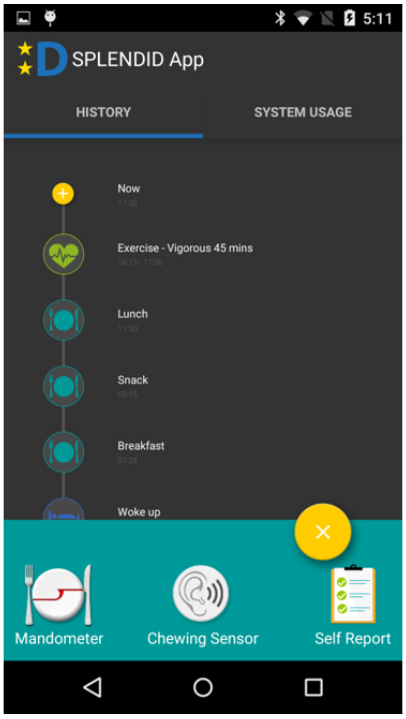
For the duration of the experiment, all participants were provided with an Android smartphone with the required smartphone app installed, as well as a food scale and chewing sensor [30]. We have previously published the methodology and development processes for the integrated sensors (for a system overview, see Figure 1). In summary, the user was presented with 3 options for registering any eating or drinking occurrence: self-register the event using only the mobile app, use the mobile app and food scale to record their plated meals through continuous registration of the food weight remaining on their plate [32], or wear the chewing sensor to record their eating or drinking event automatically through integrated photoplethysmography and acoustic analysis [28]. The provided smartphones did not have SIM cards, dictating the use of Wi-Fi networks (at school or home) to transmit the collected data. All other nonstudy-related smartphone apps and functions (ie, GPS) were turned off. Smartphones were provided to the students by the study because the high prevalence of iOS devices in Sweden [33] would prohibit the participation of otherwise motivated students, and we strived to homogenize the user experience for all the students, providing similar phone models to exclude parameters like varying device performance, custom device interfaces, and differing battery consumption rates.

Registering eating events was supported through a custom developed smartphone app available for the Android operating system. The app allowed registering either a meal or a drink separately (Figure 2) or registering a meal and adding complementary drinks. In all cases, irrespective of the means used, a timestamp was automatically generated and saved by the system. In the case of the chewing sensor, which automatically detected eating and drinking events, the system presented the user with a persistent notification asking for verification of the detection and allowing the user to fill in

additional event information. For every meal, the participants were prompted to answer two additional questions: meal type (breakfast, lunch, dinner, or snack) and where the food was prepared or bought (home, retail store, or restaurant). The maximum number of registerable main meals (breakfast, lunch, and dinner, as characterized by the users) for each participant was 3 per day, with all the remaining eating occasions automatically registered as snacks. Participants could not use multiple methods to register a main meal (ie, self-registration, food scale, and chewing sensor). Once a meal had been registered with one method, it could not be registered again using a different method. However, the users were free to register less than 3 main meals per day, in cases where they skipped or forgot to register some main meals. During drink registration, the participants had the option to include additional information about the type of the consumed drink (sugary drink, coffee/tea, dairy/milk, or water). When the smartphone was connected to Wi-Fi, the registered meals were automatically uploaded online to a study server, allowing the researchers to supervise the progress of the study in real-time using a dedicated web interface. All the communications between the browser and server were encrypted. In conclusion, the system compiled an integrated matrix of meal and drinking event timestamps, irrespective of the registration method. The users were instructed to use the smartphone app and provided weighing scale throughout the day and charge them overnight. Battery concerns dictated the use of the chewing sensor only during the after-school hours (ie, between 5:00 pm and 11:00 pm) for registration of drinks and meals, with the device also being recharged overnight. The participants did not have to charge the provided scales, since a full charge lasted more than 3 weeks.

Figure 1. The system consists of a digital food scale, chewing sensor, smartphone app, and web app.

Figure 2. Main screen of the smartphone app.



System Usability and User Experience Questionnaires

System usability was reported using the System Usability Questionnaire (SUS) [34], and user experience was estimated based on a custom questionnaire developed in discussion with adolescents from the same school in a previous study. In the user experience questionnaire, the smartphone app, digital food scale, and chewing sensor were rated on how comfortable they were to use and if the users perceived that the device use affected their usual behavior, as well as if the technology was potentially usable 'in school', 'at home', or 'outside' on 10-point Likert scales, ranging from 0 for 'Completely disagree' to 9 for 'Completely agree'. As part of the same questionnaire, participants also provided free-text answers on what they would improve for the system and its individual components (smartphone app, digital food scale, and chewing sensor). The questionnaire also contained a free-text question on the price of the system (smartphone app, food scale, chewing sensor, excluding server use), which was compared with the suggested pricing by the developers and manufacturers. The SUS has been validated on multiple occasions [35], while the custom behavior change and comfort questionnaire was previously used in similar studies [31].

Data Analysis

The presented figures were generated using R 3.5.1. Feasibility is presented based on the mERA guidelines, with the checklist available in Multimedia Appendix 1 [25]. Estimation of compliance was made by dividing the number of registered main meals with the maximum number of main meals each participant was expected to register, based on the functionality of the provided system and the provided study instructions (ie, 3 main meals per day: one breakfast, one lunch, and one dinner). The formula for daily participant compliance was compliance = registered main meals/3. Food type frequency and meal timing are presented on a population level, with daily meal distribution presented in 24-hour day cycles starting at 4:00 am (ie, meals reported between 00:00 am and 3:59 am were grouped with meals registered during the previous calendar day) following a previously used practice [36]. All the values presented in the text are mean (SD).

Results

Subjects

Of the 26 adolescents that participated in the study, the datasets provided by 2 participants were excluded from the analysis due to significant deviations from the study protocol, such as device misuse (ie, prohibited use of the phone video camera in school, which is not supported by the current ethical application) and manually unlocking and using additional mobile features (eg, online mobile games). This resulted in 24 of the participants (24/26, 92%) being included in the analyzed study sample (Table 1).

Table 1. Group characteristics.

Characteristics	Total (n=24), mean (SD)	Men (n=7), mean (SD)	Women (n=17), mean (SD)
Age (years)	16.8 (0.7)	17.2 (0.5)	16.6 (0.7)
Weight (kg)	62.2 (14.4)	75.7 (10.1)	56.6 (12.1)
Height (cm)	168.1 (10.3)	181.3 (6.8)	162.7 (5.3)
BMI (kg/m^2)	21.9 (4.1)	23.3 (4.7)	21.3 (3.8)

Feasibility

Analyzing the system feasibility based on the mERA checklist [25], in the current study, phones without SIM cards and with locked hosting features were provided to all participants to ensure user interaction was restricted to the research app and to better manage confidentiality issues, which was a sensitive issue due to data collection by adolescents and especially in the school environment. Conceptually however, since all participants owned a smartphone, the system could easily have been used on the participants' own phones, but that would require that the smartphone app was developed for multiple operating systems and not only Android. The software ran well and displayed properly on Android smartphones with 1.0 GB RAM, 8.0 GB ROM, and at least a 1280x720 HD display. All students had access to Wi-Fi both at home and at school. The battery lives of the smartphone and food scale allowed recordings throughout the day, while battery limitations restricted chewing sensor recordings to 6 hours per day. No protocol adaptations were required during the study as a result of unexpected complications. In total, 6 participants required additional information from researchers on specifics of system use, and 2 food scales had to be replaced during the experiment due to equipment malfunction (excessive battery drainage). All malfunctions were identified and addressed within 1 day. The most time-sensitive and laborious period for the researcher during the study was at the initiation, when information was provided to participants while handing out smartphones and devices. The remainder of the study required limited additional effort, mostly in the form of short interactions with individual students to answer follow-up questions or when data from the web platform indicated equipment malfunction or low compliance. The companies developing the sensors suggested a single purchase cost of €200 for the system, including both sensors and the app. Meanwhile, the price suggested by participants for the system varied greatly, with a mean single purchase cost of €67 (ranging from €14 to €236) or a monthly cost of €6 (ranging from €0 to €19).

System Usability and User Experience

Participants rated the system usability between 47.5 and 97.5, with a mean value of 77.1, which corresponds to a grade B 'Good' rating, on the SUS. No adverse events were reported to researchers during the experiments or to school personnel after

the experiment. The smartphone app and chewing sensor received the highest and lowest ratings for comfort, respectively. For perceived usability, all system components (smartphone app, food scale, and chewing sensor) scored highest at home and lowest outside home (Table 2).

Suggestions for improvements of the smartphone app fell into 4 categories: (1) 5 individuals suggested making it more responsive or faster, (2) 4 wanted a brighter color theme, (3) 3 wanted the function to retrospectively add meals, and (4) 3 requested better app interfaces, without specifying what was lacking.

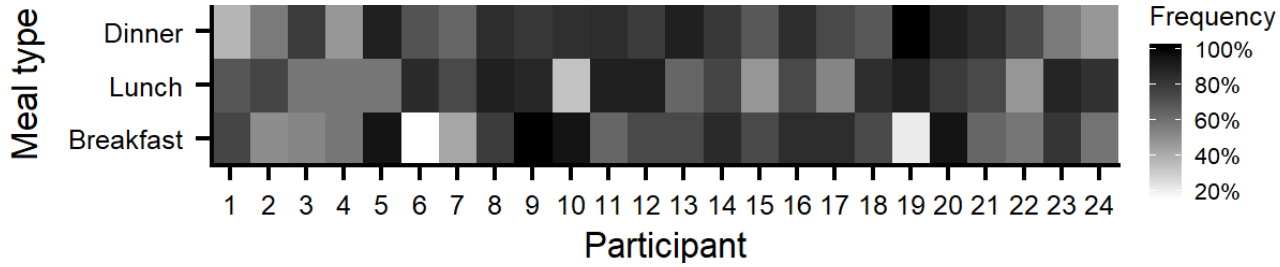
Table 2. Perceived usability and acceptability of the smartphone app and devices, answered using 10-point Likert scales, ranging from 0 (completely disagree) to 9 (completely agree).

Statements	App, mean (SD)	Food scale, mean (SD)	Chewing sensor, mean (SD)
I felt comfortable using the …	7.3 (1.8)	6.4 (2.3)	3.9 (2.8)
Perceived behavioral change from using …	4.0 (2.2)	3.6 (2.6)	3.9 (2.5)
Potential for use of ... in school	6.3 (2.8)	5.4 (3.2)	3.5 (2.7)
Potential for use of … at home	7.1 (2.3)	7.1 (2.5)	4.6 (3.1)
Potential for use of … outside	5.4 (2.5)	3.8 (2.3)	3.1 (2.6)

Registering the Frequency of Main Meals

On average, participants collected data for 18.4 days (SD 1.3 days). On average, each participant registered 63.3 meals (SD 18.0 meals), of which 39.5 meals (SD 6.3 meals) were main meals (breakfast, lunch, and dinner) and 23.8 meals (SD 14.9 meals) were snacks. Of the 39.5 main meals, 12.6 meals (SD 4.0 meals) were breakfasts, 13.2 meals (SD 3.0 meals) were lunches, and 13.7 meals (SD 3.3 meals) were dinners. In 51% (775/1520) of the overall reported meals, participants used a sensor to provide additional data. Of these meals, 87% (674/775) were made using the food scale, and the remaining meals (101/775) were registered with the chewing sensor. The average estimated compliance per participant was 73% (2.2/3 main meals per day) across all study days. To measure changes in compliance over time, only data from the first 2 weeks of system use were compared, during which the mean number of registered meals per participant was 50.2 (SD 13.9 main meals). On the group level, the number of registered meals from the first week was 573, of which 355 were main meals and 218 were snacks. Meanwhile, during the second week, 632 meals were registered, of which 384 were main meals and 248 were snacks. This shows an increased registering frequency from 70% (355/504) to 76% (384/504) of the 3 expected main meals per day from week one to week two. There were large variations between individuals in the type of main meals that were registered in the study (Figure 3).

Figure 3. Frequency of meal registration per participant throughout the study, ranging from 0% to 100% per day for each main meal type.



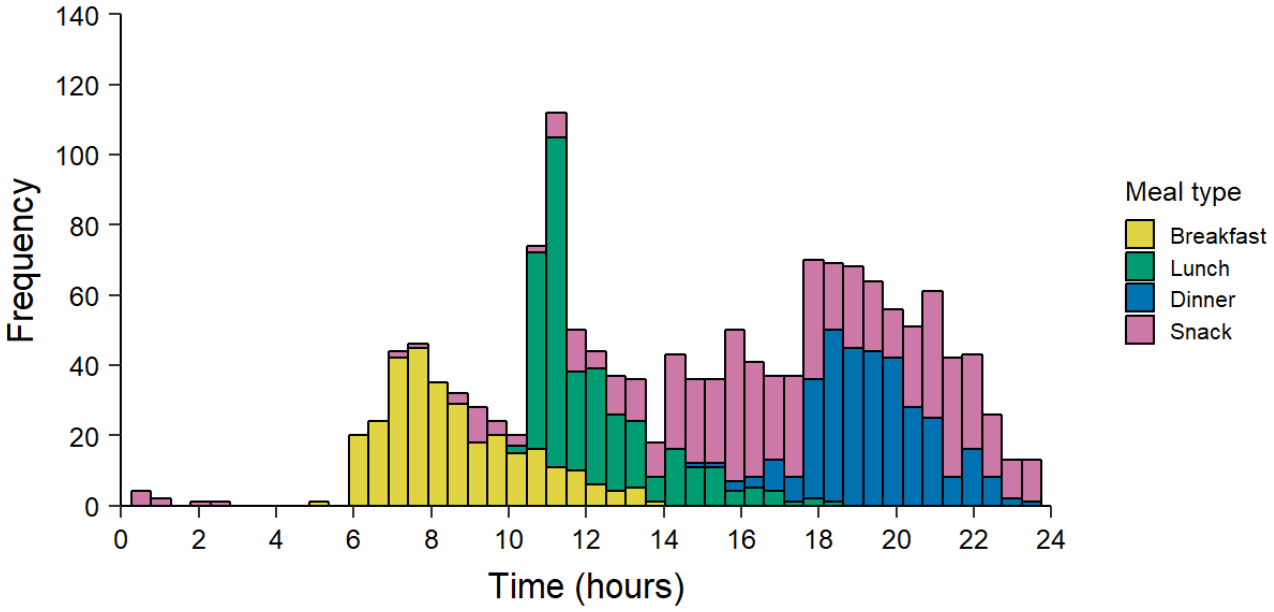
Population Food Selection Frequency

Regarding food preparation, 40% (608/1520) of the registered meals were ranked as home-cooked, 46% (699/1520) as purchased in food stores, and 14% (213/1520) as consumed in restaurants. Regarding drink types, 34% (314/924) of the registered drinks were sugary drinks, 15% (139/924) were dairy-based, 34% (314/924) were water, and 17% (157/924) were tea or coffee.

Population Mealtime Distribution

On average, breakfast took place at around 8:30 am, ranging from 5:00 am to 2:00 pm. On average, lunch took place around 12:15 pm, ranging from 10:15 am to 6:15 pm. Meanwhile, on average, dinners took place around 7:30 pm, ranging from 3:00 pm to 11:45 pm (Figure 4).

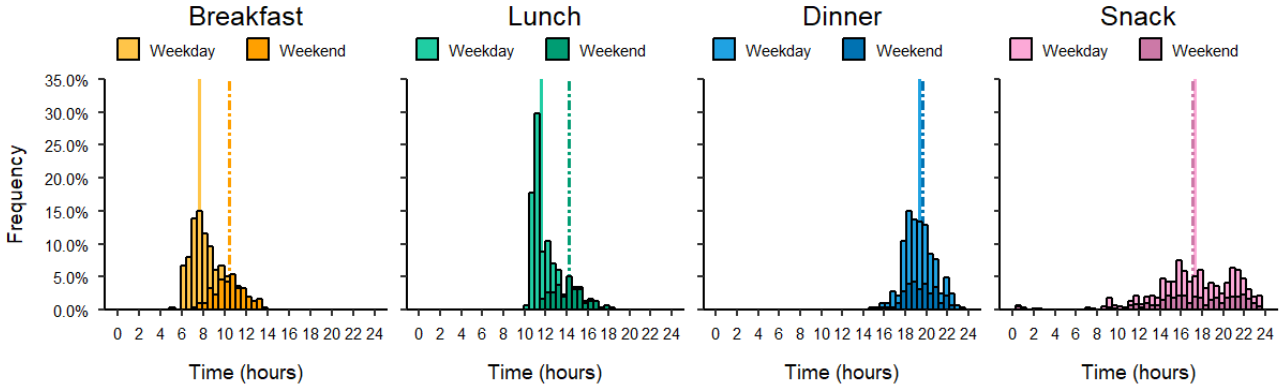
Figure 4. Histogram depicting the meal distribution of breakfast, lunch, dinner, and snacks across the day for all registered meals. Registration was done by 24 participants for a period of 2-3 weeks.



Distribution of Meals Across Weekdays and Weekends

The majority of breakfasts were eaten earlier during weekdays compared to weekends, with a difference of 2 hours 49 minutes between the mean value for weekday and weekend breakfasts. Similarly, lunches were generally eaten earlier during weekdays than during weekends, with a difference of 2 hours 44 minutes. Dinners were eaten slightly earlier during weekdays than during weekends, with a difference of 17 minutes. There was a slight difference between the time (14 minutes) of reported snacks between weekends and weekdays, with snacking overall being spread out throughout the monitored days (Figure 5).

Figure 5. Histogram showing the distribution of all registered meals. Vertical lines in the graphs show the mean values of weekday (solid line) and weekend (dashed line) meal occasions.



Discussion

This study evaluated the perceived acceptability and usability by adolescent participants of a system composed of a smartphone app and two eating behavior sensors, while identifying system constraints for future large-scale deployments. The app was well received by the participants, while the sensors received lower scores of usability and acceptability. From a functional and infrastructural perspective, the system had easy-to-meet requirements, was easy to use, and had high scalability. The usefulness of the system also appeared high, with its ability to collect data on dietary behaviors associated with increased risk of obesity (ie, sugary beverage consumption, restaurant meals, and unstructured meal patterns). From a functional perspective, the technical requirements of the smartphone app were low, including battery consumption, ensuring that the presented app runs on most modern smartphones, with 99% of Swedish 16–25-year-olds using smartphones regularly and a little less than 95% of devices in Sweden fulfilling the 1 GB RAM and 8 GB ROM requirement [33]. In addition, in Sweden, smartphone coverage and accessibility to Wi-Fi and energy outlets are high, making a screening protocol, such as the one explored here, highly feasible from an infrastructural perspective [17], ensuring the scalability of the system. Concerning the system's suggested cost, a discrepancy existed between the system creators and the end users, which is expected due to the targeted age group. For the researcher, little work was required to ensure proper use of the smartphone app and devices, with only one-fourth of participants requiring additional information on specific functions of the system despite the short duration of the initial information meetings. The dedicated research platform enabled real-time

data inspection, which made the identification of malfunctioning devices easy and ensured high data retention throughout the study. Participants found the system easy to use, with the total system receiving a high system usability score. Some participants felt that the smartphone app was a bit slow and unresponsive at times, which could potentially have resulted in lower compliance, but this is not immediately evident in the output dataset. A recent study identified personalization options as a strategy to increase interaction and compliance [37], a potential answer to our user requests (7/24) for improved interfaces. Finally, the possibility of retrospectively registering meals is a feature often seen in comparable apps. Future versions of our system might integrate this functionality to reduce the risk of the user forgetting to report meals as they occur, an issue that might have affected the reported outcomes of this study. However, the addition of retrospective meal reporting has the potential to introduce additional recall bias, resulting in decreased accuracy for the collected measurements [38].

The high frequency of meal registration per individual makes us confident that the mealtime distribution results are representative of the population's actual mealtime distribution. In addition, the increased number of reported meals from week one to week two suggests that a familiarization period, or more training in device use, might increase registering frequency. In a study with a similar sample population (ie, conducted in adolescents in Sweden) that employed a 7-day food record, the number of registered meals was 18% higher than in the current study [39], which might be related to the lack of functionality for retrospective meal reporting in our own system. However, due to the allowance of recall data and differences in main meal definitions (primarily breakfast) and methodology (recollection vs real-time registering) as well as a potentially higher socioeconomic status in the current sample, it is unlikely that these two studies are directly comparable. Future studies should therefore aim to compare the methods employed here with 24-hour/day automatic meal registration methods, such as the ones developed by Sazonov et al [40], Sun et al [41], and Kyritsis et al [42], which have the potential to remove recall and social desirability bias (eg, where individuals only register products that are viewed favorably by others) but may introduce other biases, such as comfort of use. This will be even more relevant once the chewing sensor allows recording for the entire day.

With regards to the sample's frequency of food selection, we found that the number of home-cooked meals was low, with most meals being either store-bought or restaurant-bought, similar to previous reports [43]. Similarly, a large number (one-third) of the reported drinks contained sugar, with the actual portion potentially being even higher since other drink categories might also have contained added sugar (eg, milk/diary). These dietary habits have been associated with increased body weight [44,45]. Meanwhile, the meal timing across days suggests that Swedish adolescents eat their breakfasts and lunches later during weekends than during weekdays. This time shift is partly corroborated by a previous study in young adults, where the first caloric intake occurred later during weekends, while the last energy intake occurred at similar times during weekdays and weekends [36], potentially due to the de facto effect of the school schedule. Overall, due in part to methodological limitations and differences in definitions, there is currently no consensus regarding the effect of meal timing on health [46], with some studies reporting skipping breakfast may increase the risk of obesity and type 2 diabetes [47,48], while others reporting no such effects [49].

The main strengths of the study were a smartphone app designed to allow easy and accurate real-time meal registration, not relying on recall, and a protocol with no restrictions to participants' real-life activities. Another strength, based on the results, was the high usage rate of the system. A strength of the system is its modular design, which enables the additional sensors, ensuring high contextual adaptability. In this study, a digital food scale and chewing sensor were used, but when investigating other diseases, such as diabetes, more appropriate sensors could be added, such as a blood glucose monitor.

One limitation of the study was the low number of available study devices, while a limitation for the generalizability of the results was the low number of male participants, which deviates from the Swedish average [50] but is more in line with the sex distribution of the school in question. Another potential limitation was the de facto assumption of 3 main meals per day for compliance estimation, since meal-skipping studies suggest that not everyone eats breakfast, lunch, and dinner [51]. However, the definition of meal skipping in previous studies varies greatly, causing a large range in the frequency of meal skipping (5% to 83%), which provides no reliable baseline for compliance estimation [51]. A potential alternative for measuring compliance would be to ask the user for meal occurrences in the past 30 minutes at random times during the day and then compare these results with spontaneously self-reported meals [36]. This method was avoided due to concerns that additional interaction with the smartphone app may result in reduced compliance. Another protocol limitation was that our study provided additional study smartphones for the app, but this practice was deemed necessary to include iOS users and homogenize the data collection experience. In addition, our reporting system lacked a dedicated category for reporting "noncaloric soft drinks," which may have resulted in individuals either not reporting those items or reporting it in another category. It should also be noted that the minimalistic reporting approach in our study might introduce additional reporting bias in certain categories. For example, users were able to report milk/dairy-based drinks either in the dedicated category or as "sugary drinks" if they contained sugar. In future iterations of our system, additional user-reporting options can be introduced to resolve this issue (eg, adding the user-selected option to report "added sugar" or "no added sugar" within the milk/diary reports). Finally, one should not ignore the potential observer effect [52], something uniformly affecting the domain of behavioral monitoring, resulting in modification of the observed behaviors due to study participation.

Future studies should aim to repeat the feasibility study by enrolling larger samples and longer data collection periods in order to test the progress of registration compliance over longer periods. In a parallel study in the Netherlands, the chewing sensor received an average wearer comfort score of 3.7 by overweight adults, which is comparable to the score received

in the current study [30]. Future studies may benefit from comparing differences in system perception and user compliance between obese and normal-weight individuals. Additionally, based on existing evidence showing that mHealth is more often used by individuals of higher socioeconomic status, additional studies should also aim to include individuals of lower socioeconomic characteristics [53]. Such deployments have a de facto interest for group-based behavioral comparisons but will also evaluate the scalability potential of the system. For validation purposes, the system should be compared to recall methods (eg, 7-day food record and meal habit questionnaires) as well as other automatic recording methods (eg, eButton and Automatic Ingestion Monitor) [41,54]. In practice, the presented methodologies have been the stepping stone for the extension of the data collection and analysis framework in a follow-up research effort [23]. Specifically, our renewed efforts focus on the collection of lifestyle Big Data from children and adolescents, in an effort to create population-level behavioral profiles (eg, meal frequency, food choice frequency), which will then be used by local and national public health authorities as a helpful tool in their efforts to tackle childhood obesity.

In conclusion, no system constraints related to infrastructure and function were identified for deploying the described smartphone-based system in adolescents of medium to high socioeconomic status in Sweden. Also, the identified population profiles regarding the differences in the timing of meals on weekdays vs weekends and the reported frequency of sugary drinks by adolescents provide valuable preliminary information about the dietary habits of the target population. Additionally, these data point towards the usefulness of comparable mHealth systems in providing health-related behavioral information for such populations. Regarding the system use, high school students are well-versed in the use of smartphones, resulting in high acceptability and usability of the smartphone app, with most suggested improvements being related to design rather than functionality. In line with this, the number of participants that required additional information on device use was low. The high registration frequency of main meals indicates the high usability of the system, which, if coupled with appropriate sensors, can facilitate the collection of reliable food intake data. Overall, the system appears promising as a low-effort method to provide accurate measurements of dietary habits, setting the base for future developments of individual-level and group-level mHealth interventions against adolescent obesity.

Acknowledgments

The work leading to these results received funding from the European Community's ICT Programmes: SPLENDID under Grant Agreement No. 610746 and BigO under Grant Agreement No. 610746. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

Authors' Contributions

Development of the study concept was handled by BL, AD, and II. Project administration was handled by RH, IL, and II, and supervision was handled by RH and II. Methodology and data curation were handled by BL and II. Investigation was performed by BL, PF, and II, and the formal analysis was performed by BL. Visualization of data and writing of the original draft was handled by BL, PF, and II; meanwhile, all authors reviewed and edited the final manuscript. Funding was acquired by AD, NM, and II, and resources were provided by AD, IL, CD, CM, NM, RH, and II.

Conflicts of Interest

During data collection, BL and II received part of their salaries from Mandometer AB (Stockholm, Sweden), owner of the IPR for the Mandometer®. This does not alter the authors' adherence to journal policies.

Multimedia Appendix 1
mHealth evidence reporting and assessment (mERA) checklist.
[PDF File (Adobe PDF File), 115 KB - mhealth_v8i7e14778_app1.pdf]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
2. Kumar S, Kelly AS. Review of Childhood Obesity: From Epidemiology, Etiology, and Comorbidities to Clinical Assessment and Treatment. Mayo Clin Proc 2017 Feb;92(2):251-265. [doi: 10.1016/j.mayocp.2016.09.017] [Medline: 28065514]
3. Brewis AA, Bruening M. Weight Shame, Social Connection, and Depressive Symptoms in Late Adolescence. Int J Environ Res Public Health 2018 May 01;15(5) [FREE Full text] [doi: 10.3390/ijerph15050891] [Medline: 29723962]
4. Neovius K, Rehnberg C, Rasmussen F, Neovius M. Lifetime Productivity Losses Associated with Obesity Status in Early Adulthood. Applied Health Economics and Health Policy 2012 Sep 01;10(5):309-317. [doi: 10.2165/11632770-000000000-00000] [Medline: 22827692]

5. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]
6. Leblanc ES, O'Connor E, Whitlock EP, Patnode CD, Kapka T. Effectiveness of primary care-relevant treatments for obesity in adults: a systematic evidence review for the U.S. Preventive Services Task Force. Ann Intern Med 2011 Oct 04;155(7):434-447. [doi: 10.7326/0003-4819-155-7-201110040-00006] [Medline: 21969342]
7. Beaulac J, Sandre D. Critical review of bariatric surgery, medically supervised diets, and behavioural interventions for weight management in adults. Perspect Public Health 2017 May;137(3):162-172. [doi: 10.1177/1757913916653425] [Medline: 27354536]
8. Waters E, de Silva-Sanigorski A, Hall BJ, Brown T, Campbell KJ, Gao Y, et al. Interventions for preventing obesity in children. Cochrane Database Syst Rev 2011 Dec 07(12):CD001871. [doi: 10.1002/14651858.CD001871.pub3] [Medline: 22161367]
9. Fappa E, Yannakoulia M, Pitsavos C, Skoumas I, Valourdou S, Stefanadis C. Lifestyle intervention in the management of metabolic syndrome: could we improve adherence issues? Nutrition 2008 Mar;24(3):286-291. [doi: 10.1016/j.nut.2007.11.008] [Medline: 18201869]
10. Jeffery RW, Drewnowski A, Epstein LH, Stunkard AJ, Wilson GT, Wing RR, et al. Long-term maintenance of weight loss: current status. Health Psychol 2000 Jan;19(1S):5-16. [doi: 10.1037/0278-6133.19.suppl1.5] [Medline: 10709944]
11. Olsen J, Willaing I, Ladelund S, Jørgensen T, Gundgaard J, Sørensen J. Cost-effectiveness of nutritional counseling for obese patients and patients at risk of ischemic heart disease. Int J Technol Assess Health Care 2005;21(2):194-202. [Medline: 15921059]
12. Walker O, Strong M, Atchinson R, Saunders J, Abbott J. A qualitative study of primary care clinicians' views of treating childhood obesity. BMC Fam Pract 2007 Sep 03;8:50 [FREE Full text] [doi: 10.1186/1471-2296-8-50] [Medline: 17767720]
13. O'Connor EA, Evans CV, Burda BU, Walsh ES, Eder M, Lozano P. Screening for Obesity and Intervention for Weight Management in Children and Adolescents: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2017 Jun 20;317(23):2427-2444. [doi: 10.1001/jama.2017.0332] [Medline: 28632873]
14. Fruh SM. Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. J Am Assoc Nurse Pract 2017 Oct;29(S1):S3-S14 [FREE Full text] [doi: 10.1002/2327-6924.12510] [Medline: 29024553]
15. Bray G, Look M, Ryan D. Treatment of the obese patient in primary care: targeting and meeting goals and expectations. Postgrad Med 2013 Sep;125(5):67-77. [doi: 10.3810/pgm.2013.09.2692] [Medline: 24113665]
16. Chen J, Guedes CM, Cooper BA, Lung AE. Short-Term Efficacy of an Innovative Mobile Phone Technology-Based Intervention for Weight Management for Overweight and Obese Adolescents: Pilot Study. Interact J Med Res 2017 Aug 02;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
17. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies. Switzerland: World Health Organization; 2011:66.
18. Lauricella AR, Cingel DP, Blackwell C, Wartella E, Conway A. The Mobile Generation: Youth and Adolescent Ownership and Use of New Media. Communication Research Reports 2014 Oct 21;31(4):357-364. [doi: 10.1080/08824096.2014.963221]
19. Lee J, Piao M, Byun A, Kim J. A Systematic Review and Meta-Analysis of Intervention for Pediatric Obesity Using Mobile Technology. Stud Health Technol Inform 2016;225:491-494. [Medline: 27332249]
20. Chaplais E, Naughton G, Thivel D, Courteix D, Greene D. Smartphone Interventions for Weight Treatment and Behavioral Change in Pediatric Obesity: A Systematic Review. Telemed J E Health 2015 Oct;21(10):822-830. [doi: 10.1089/tmj.2014.0197] [Medline: 26290954]
21. Vogel C, Zwolinsky S, Griffiths C, Hobbs M, Henderson E, Wilkins E. A Delphi study to build consensus on the definition and use of big data in obesity research. Int J Obes (Lond) 2019 Dec 17;43(12):2573-2586 [FREE Full text] [doi: 10.1038/s41366-018-0313-9] [Medline: 30655580]
22. Everson B, Mackintosh KA, McNarry MA, Todd C, Stratton G. Can Wearable Cameras be Used to Validate School-Aged Children's Lifestyle Behaviours? Children (Basel) 2019 Mar 01;6(2) [FREE Full text] [doi: 10.3390/children6020020] [Medline: 30717207]
23. Diou C, Ioakeimidis I, Charmandari E, Kassari P, Lekka I, Mars M, et al. BigO: Big Data Against Childhood Obesity. : European Society for Paediatric Endocrinology; 2018 Sep 27 Presented at: 57th Annual European Society of Pediatric Endocrinology; 27-29 September 2018; Athens, Greece URL: https://abstracts.eurospe.org/hrp/0089/hrp0089p3-p127
24. Istepanian R, Woodward B. M-health: fundamentals and applications, First edition. Hoboken: Wiley-IEEE Press; Oct 21, 2016.
25. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
26. Bhardwaj NN, Wodajo B, Gochipathala K, Paul DP, Coustasse A. Can mHealth Revolutionize the Way We Manage Adult Obesity? Perspect Health Inf Manag 2017;14(Spring):1a [FREE Full text] [Medline: 28566984]

27. Chen J, Wilkosz ME. Efficacy of technology-based interventions for obesity prevention in adolescents: a systematic review. Adolesc Health Med Ther 2014;5:159-170 [FREE Full text] [doi: 10.2147/AHMT.S39969] [Medline: 25177158]
28. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
29. Higgins DL, O'Reilly K, Tashima N, Crain C, Beeker C, Goldbaum G, et al. Using formative research to lay the foundation for community level HIV prevention efforts: an example from the AIDS Community Demonstration Projects. Public Health Rep 1996;111 Suppl 1:28-35 [FREE Full text] [Medline: 8862154]
30. van den Boer J, van der Lee A, Zhou L, Papapanagiotou V, Diou C, Delopoulos A, et al. The SPLENDID Eating Detection Sensor: Development and Feasibility Study. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e170 [FREE Full text] [doi: 10.2196/mhealth.9781] [Medline: 30181111]
31. Langlet B, Anvret A, Maramis C, Moulos I, Papapanagiotou V, Diou C, et al. Objective measures of eating behaviour in a Swedish high school. Behaviour & Information Technology 2017 May 06;36(10):1005-1013 [FREE Full text] [doi: 10.1080/0144929x.2017.1322146]
32. Esfandiari M, Papapanagiotou V, Diou C, Zandian M, Nolstam J, Södersten P, et al. Control of Eating Behavior Using a Novel Feedback System. J Vis Exp 2018 May 08(135) [FREE Full text] [doi: 10.3791/57432] [Medline: 29806832]
33. Davidsson P, Palm M, Melin Å. Svenskarna och internet 2018. Stockholm, Sweden: Internet Foundation in Sweden; 2018.
34. Jordan P, Thomas B, McClelland I, Weerdmeester B. Usability evaluation in industry. Boca Raton: CRC Press; Jun 11, 1996.
35. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
36. Gill S, Panda S. A Smartphone App Reveals Erratic Diurnal Eating Patterns in Humans that Can Be Modulated for Health Benefits. Cell Metab 2015 Nov 03;22(5):789-798 [FREE Full text] [doi: 10.1016/j.cmet.2015.09.005] [Medline: 26411343]
37. Partridge SR, Redfern J. Strategies to Engage Adolescents in Digital Health Interventions for Obesity Prevention and Management. Healthcare (Basel) 2018 Jun 21;6(3) [FREE Full text] [doi: 10.3390/healthcare6030070] [Medline: 29933550]
38. Dhurandhar NV, Schoeller D, Brown AW, Heymsfield SB, Thomas D, Sørensen TIA, Energy Balance Measurement Working Group. Energy balance measurement: when something is not better than nothing. Int J Obes (Lond) 2015 Jul 13;39(7):1109-1113 [FREE Full text] [doi: 10.1038/ijo.2014.199] [Medline: 25394308]
39. Sjöberg A, Hallberg L, Höglund D, Hulthén L. Meal pattern, food choice, nutrient intake and lifestyle factors in The Göteborg Adolescence Study. Eur J Clin Nutr 2003 Dec;57(12):1569-1578. [doi: 10.1038/sj.ejcn.1601726] [Medline: 14647222]
40. Sazonov ES, Schuckers SAC, Lopez-Meyer P, Makeyev O, Melanson EL, Neuman MR, et al. Toward objective monitoring of ingestive behavior in free-living population. Obesity (Silver Spring) 2009 Oct;17(10):1971-1975 [FREE Full text] [doi: 10.1038/oby.2009.153] [Medline: 19444225]
41. Sun M, Burke LE, Mao Z, Chen Y, Chen H, Bai Y, et al. eButton: A Wearable Computer for Health Monitoring and Personal Assistance. Proc Des Autom Conf 2014;2014:1-6 [FREE Full text] [doi: 10.1145/2593069.2596678] [Medline: 25340176]
42. Kyritsis K, Tatli CL, Diou C, Delopoulos A. Automated analysis of in meal eating behavior using a commercial wristband IMU sensor. Conf Proc IEEE Eng Med Biol Soc 2017 Jul;2017:2843-2846. [doi: 10.1109/EMBC.2017.8037449] [Medline: 29060490]
43. Nago ES, Lachat CK, Dossa RAM, Kolsteren PW. Association of out-of-home eating with anthropometric changes: a systematic review of prospective studies. Crit Rev Food Sci Nutr 2014;54(9):1103-1116. [doi: 10.1080/10408398.2011.627095] [Medline: 24499144]
44. Popkin BM. Sugary beverages represent a threat to global health. Trends Endocrinol Metab 2012 Dec;23(12):591-593. [doi: 10.1016/j.tem.2012.07.003] [Medline: 22867870]
45. Rolls BJ. The Supersizing of America: Portion Size and the Obesity Epidemic. Nutr Today 2003;38(2):42-53. [doi: 10.1097/00017285-200303000-00004] [Medline: 12698053]
46. Leech RM, Worsley A, Timperio A, McNaughton SA. Understanding meal patterns: definitions, methodology and impact on nutrient intake and diet quality. Nutr Res Rev 2015 Jun;28(1):1-21 [FREE Full text] [doi: 10.1017/S0954422414000262] [Medline: 25790334]
47. Bi H, Gan Y, Yang C, Chen Y, Tong X, Lu Z. Breakfast skipping and the risk of type 2 diabetes: a meta-analysis of observational studies. Public Health Nutr 2015 Nov;18(16):3013-3019. [doi: 10.1017/S1368980015000257] [Medline: 25686619]
48. St-Onge M, Ard J, Baskin ML, Chiuve SE, Johnson HM, Kris-Etherton P, American Heart Association Obesity Committee of the Council on Lifestyle Cardiometabolic Health, Council on Cardiovascular Disease in the Young, Council on Clinical Cardiology;, Stroke Council. Meal Timing and Frequency: Implications for Cardiovascular Disease Prevention: A Scientific Statement From the American Heart Association. Circulation 2017 Mar 28;135(9):e96-e121. [doi: 10.1161/CIR.0000000000000476] [Medline: 28137935]

49. Sievert K, Hussain SM, Page MJ, Wang Y, Hughes HJ, Malek M, et al. Effect of breakfast on weight and energy intake: systematic review and meta-analysis of randomised controlled trials. BMJ 2019 Jan 30;364:l42 [FREE Full text] [doi: 10.1136/bmj.l42] [Medline: 30700403]
50. Women and men in Sweden 2018 Facts and figures. Örebro, Sweden: Statistics Sweden; 2018.
51. Pendergast FJ, Livingstone KM, Worsley A, McNaughton SA. Correlates of meal skipping in young adults: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 01;13(1):125 [FREE Full text] [doi: 10.1186/s12966-016-0451-1] [Medline: 27905981]
52. Zhong C, House J. Hawthorne revisited: Organizational implications of the physical work environment. Research in Organizational Behavior 2012 Jan;32:3-22. [doi: 10.1016/j.riob.2012.10.004]
53. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
54. Fontana JM, Farooq M, Sazonov E. Automatic ingestion monitor: a novel wearable device for monitoring of ingestive behavior. IEEE Trans Biomed Eng 2014 Jun;61(6):1772-1779 [FREE Full text] [doi: 10.1109/TBME.2014.2306773] [Medline: 24845288]

Abbreviations

mERA: mHealth evidence reporting and assessment checklist.
mHealth: mobile health.
SUS: System Usability Questionnaire.

Edited by G Eysenbach; submitted 22.05.19; peer-reviewed by J van den Boer, W Jia, J Coleman; comments to author 23.07.19; revised version received 14.09.19; accepted 14.05.20; published 21.07.20.

Please cite as:
Langlet B, Maramis C, Diou C, Maglaveras N, Fagerberg P, Heimeier R, Lekka I, Delopoulos A, Ioakimidis I
Formative Evaluation of a Smartphone App for Monitoring Daily Meal Distribution and Food Selection in Adolescents: Acceptability and Usability Study
JMIR Mhealth Uhealth 2020;8(7):e14778
URL: http://mhealth.jmir.org/2020/7/e14778/
doi:10.2196/14778
PMID:32706684

Original Paper

Usability of Wearable Devices With a Novel Cardiac Force Index for Estimating the Dynamic Cardiac Function: Observational Study

Po-Jen Hsiao[1,2,3,4], MD; Chih-Chien Chiu[4,5], MD; Ke-Hsin Lin[6], MA; Fu-Kang Hu[7], MA; Pei-Jan Tsai[6], MA; Chun-Ting Wu[6], MA; Yuan-Kai Pang[6], MA; Yu Lin[8,9], PhD; Ming-Hao Kuo[7], MA; Kang-Hua Chen[10], PhD; Yi-Syuan Wu[8], MA; Hao-Yi Wu[6,11], MA, RN; Ya-Ting Chang[6], MA; Yu-Tien Chang[6], PhD; Chia-Shiang Cheng[8], MA; Chih-Pin Chuu[12], PhD; Fu-Huang Lin[6], PhD; Chi-Wen Chang[10,13*], PhD; Yuan-Kuei Li[14,15*], MD; Jenq-Shyong Chan[1,2*], MA, MD; Chi-Ming Chu[4,6,7,8,16,17*], PhD

[1]Division of Nephrology, Department of Internal Medicine, Taoyuan Armed Forces General Hospital, Taoyuan, Taiwan

[2]Division of Nephrology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[3]Department of Life Sciences, National Central University, Taoyuan, Taiwan

[4]Big Data Research Center, Fu-Jen Catholic University, New Taipei, Taiwan

[5]Division of Infectious Disease and Tropical Medicine, Department of Internal Medicine, Taoyuan Armed Forces General Hospital, Taoyuan, Taiwan

[6]Division of Biostatistics and Medical Informatics, Department of Epidemiology, School of Public Health, National Defense Medical Center, Taipei, Taiwan

[7]Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan

[8]Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan

[9]Department of Nursing, University of Kang Ning, Tainan, Taiwan

[10]School of Nursing, College of Medicine, Chang Gung University, Taoyuan, Taiwan

[11]Department of Nursing, Tri-Service General Hospital, Taipei, Taiwan

[12]Institute of Cellular and System Medicine, National Health Research Institutes, Miaoli, Taiwan

[13]Division of Pediatric Endocrinology & Genetics, Department of Pediatrics, Chang-Gung Memorial Hospital, Taoyuan, Taiwan

[14]Division of Colorectal Surgery, Department of Surgery, Taoyuan Armed Forces General Hospital, Taoyuan, Taiwan

[15]Department of Biomedical Sciences and Engineering, National Central University, Taoyuan, Taiwan

[16]Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan

[17]Department of Public Health, School of Public Health, China Medical University, Taichung, Taiwan

*these authors contributed equally

Corresponding Author:
Chi-Ming Chu, PhD
Division of Biostatistics and Medical Informatics, Department of Epidemiology
School of Public Health
National Defense Medical Center
Neihu 114
Taipei
Taiwan
Phone: 1 886 2 87923100
Email: chuchiming@web.de

Abstract

Background: Long-distance running can be a form of stress to the heart. Technological improvements combined with the public's gradual turn toward mobile health (mHealth), self-health, and exercise effectiveness have resulted in the widespread use of wearable exercise products. The monitoring of dynamic cardiac function changes during running and running performance should be further studied.

Objective: We investigated the relationship between dynamic cardiac function changes and finish time for 3000-meter runs. Using a wearable device based on a novel cardiac force index (CFI), we explored potential correlations among 3000-meter runners with stronger and weaker cardiac functions during running.

Methods: This study used the American product BioHarness 3.0 (Zephyr Technology Corporation), which can measure basic physiological parameters including heart rate, respiratory rate, temperature, maximum oxygen consumption, and activity. We investigated the correlations among new physiological parameters, including CFI = weight * activity / heart rate, cardiac force ratio (CFR) = CFI of running / CFI of walking, and finish times for 3000-meter runs.

Results: The results showed that waist circumference, smoking, and CFI were the significant factors for qualifying in the 3000-meter run. The prediction model was as follows: ln (3000 meters running performance pass probability / fail results probability) = –2.702 – 0.096 × [waist circumference] – 1.827 × [smoke] + 0.020 × [ACi7]. If smoking and the ACi7 were controlled, contestants with a larger waist circumference tended to fail the qualification based on the formula above. If waist circumference and ACi7 were controlled, smokers tended to fail more often than nonsmokers. Finally, we investigated a new calculation method for monitoring cardiac status during exercise that uses the CFI of walking for the runner as a reference to obtain the ratio between the cardiac force of exercise and that of walking (CFR) to provide a standard for determining if the heart is capable of exercise. A relationship is documented between the CFR and the performance of 3000-meter runs in a healthy 22-year-old person. During the running period, data are obtained while participant slowly runs 3000 meters, and the relationship between the CFR and time is plotted. The runner's CFR varies with changes in activity. Since the runner's acceleration increases, the CFR quickly increases to an explosive peak, indicating the runner's explosive power. At this period, the CFI revealed a 3-fold increase (CFR=3) in a strong heart. After a time lapse, the CFR is approximately 2.5 during an endurance period until finishing the 3000-meter run. Similar correlation is found in a runner with a weak heart, with the CFR at the beginning period being 4 and approximately 2.5 thereafter.

Conclusions: In conclusion, the study results suggested that measuring the real-time CFR changes could be used in a prediction model for 3000-meter running performance.

(JMIR Mhealth Uhealth 2020;8(7):e15331) doi:10.2196/15331

KEYWORDS

cardiac force; running; acceleration; physical activity; heart rate

Introduction

Background

As part of the body's activity, exercise is planned, repeated, and structured to improve or maintain physical health. Exercising regularly and frequently helps to prevent serious illnesses such as heart disease, cardiovascular disease, type 2 diabetes mellitus, and obesity. In recent years, activities such as marathons, triathlons, and road running have become popular sports throughout the world [1-4]. The benefits of endurance exercise for health, such as improving lipid profiles, blood glucose control, blood pressure control, and increased insulin sensitivity, may partly explain the increased participation in marathon races [1-4]. Additionally, marathon runners show higher levels of hardy personality (ie, a group of characteristics related to personal perception of control, commitment, and challenges) than the general population [5]. With regard to gender differences in participation and running performance, Nikolaidis et al [6] noted that the men were faster and older than women, whereas female participation increased disproportionately to that of men, resulting in a decrease in the male-to-female ratio. Most women marathon runners range from 30 to 34 years and most men are 40 to 44 years. Furthermore, the number of master participants increased at a greater rate than that of their younger counterparts [7]. In addition to coordination of the skeletal and muscular systems, running primarily depends on the cardiopulmonary functions of pumping blood oxygen throughout the body. Nevertheless, highly intensive, strenuous physical exercise could have a negative impact on the cardiovascular system [8-10]. Marathon running has become a popular sport, and sudden cardiac arrest is rare but still occurs in approximately 1:100,000 of the general population, with the most common causes being hypertrophic cardiomyopathy or atherosclerotic coronary disease [11]. The incidence of sudden cardiac arrest resulting in death was significantly higher during marathons than during half-marathons and more common among men than in women. Even in young populations, sudden cardiac death may occur during running [11-13]. In addition, the mean age of the nonsurvivors was clearly younger than that of the survivors [11]. The common etiologies of sudden cardiac death in young populations include conduction system abnormalities, focal myocarditis, hypertrophic cardiomyopathy, and arrhythmogenic right ventricular cardiomyopathy, further causing myocardial injury [11-14]. Wearable wireless devices can monitor and offer information about human physiology and increase clinical diagnostic accuracy [15-19]. Combining wearable technology and medicine can further help people to improve their health and quality of life. This technology provides a real-time monitoring system during physical activity, especially exercise [20-24]. To date, these wearable devices can only monitor parameters such as heart rate, respiratory rate, electrocardiography, VO_2 max test, running distance, and running speed [20-26]. They are still unable to determine some indicators precisely, such as the dynamic cardiac function changes in the wearer during running.

Objective

The relationship between cardiac function changes during running and running performance should be further explored. Using a wearable device established in mobile health (mHealth), we explored the potential correlations among 3000-meter runners with stronger and weaker cardiac function during running. The aim of our study was to investigate this issue.

Methods

Study Population

We performed an observational study that was approved by the human trial committee and institutional review board of the Tri-Service General Hospital, Taiwan (TSGH-IRB-1-104-05-147). Military academy students participated in the study from February 2015 to January 2016 in Taiwan. The study's selection conditions were to volunteer to participate in the study and to have the ability to complete a 3000-meter empty-handed run. Education and training personnel assisted participants in completing questionnaire contents and in measuring physiological parameters. Informed written consent was obtained from the participants for publication.

Wearable Device and Cardiac Force Evaluation

This study used the American product called the BioHarness 3.0 (Zephyr Technology Corporation). It is nonpenetrating and contains a 3-axis gyroscope and an accelerometer to distinguish directions in the x, y, and z axes. It has an accelerometer and gyroscope for measuring the angular velocity and a GPS function to provide the speed, distance, and location. It can measure and estimate basic physiological parameters, including heart rate, respiratory rate, acceleration, maximum oxygen consumption, and temperature; thus, it can monitor the body's activity state during running. Furthermore, the device has great reliability and validity and is better than regular heart rate belts that measure only the heart rate [25]. Excellent quality evidence from a recent systematic review study confirmed that the BioHarness 3.0 device could provide reliable and valid measurements of the heart rate across multiple contexts. In addition, it demonstrated good consistency during gold standard comparisons, supporting the validity criterion [25-28].

In exploring related factors pertaining to BioHarness 3.0 parameters and long-distance running, the credibility of the parameters measured by the BioHarness 3.0 have been considered by numerous studies, with one focusing on the measurement of its reliability and validity with 20 healthy male subjects, which consisted of 10 for reliability and 10 for validity [27]. For validity, the oxime part of the BioHarness 3.0 was compared with that of the Finnish T31 coded transmitter (Polar Electro) [27]. Lin et al [28] conducted a survey on 10 super-marathon runners, separating them into two groups and testing the correlations of the running distance with different acceleration levels (3, 6, 8, 9, 10, and 12 km/h). The survey addressed the correlation between the running speed and the 3-axis acceleration gauge with the x, y, and z axes. The results revealed a negative correlation between the y-axis and running at any speed (r=–.89 to –.92, P<.001). These results were similar to this study's results; the activity (r=–.214 to –.317, P<.001) and peak acceleration (r=–.203 to –.226, P=.002) revealed a negative correlation [28].

We further used the BioHarness 3.0 to calculate the cardiac force index (CFI), a new method of detecting the heart's condition instantly and comprehensively without penetration using the following calculation: CFI = weight × activity / heart rate; the CFI of running / CFI of walking ratio is the cardiac force ratio (CFR) [29,30]. Based on these methods, we then further explored the prediction model of running performance during 3000-meter runs.

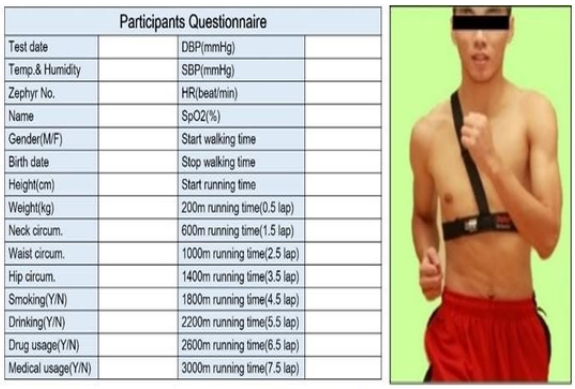
Data Collection

Before starting the 3000-meter empty-handed run, participants were required to squat, stand up, and then jump to check the heart rate belt's 3-axis acceleration gauge function, which determines participant movement status. To ensure the questionnaire data's consistency and completeness, data collection was performed after the participants completed the test and received their physiological measurement data.

Since completing a marathon race requires consistent strength training and an appropriate lifestyle, behaviors such as smoking, physical inactivity, and drinking are considered negative factors [31]. We also checked the baseline demography and associated lifestyle of all the runners in this study. The process of collecting data involved six steps (Figure 1):

- Participants were asked to complete the questionnaire with their name, gender, birth date, and personal health habits including smoking, drinking, drug usage, and medical history. Professional data collecting personnel completed the other blanks after testing the biological parameters and instructing participants in how to wear the BioHarness 3.0 heart rate belt correctly.
- Professional personnel led participants to the test starting position after confirming they were wearing the device correctly. Participants were asked to kneel, stand up, and walk at a normal speed in a half-circle (200 m) to the starting line.
- After participants got to the starting line, they were again asked to kneel and stand up. After ready and start commands from the leader, they began the 3000-meter running test (400 meter lap × 7.5 = 3000 meters) at the speed they see fit.
- After running the 3000 meters, participants were asked to jump once at the finish line and then walk another lap (400 m) at normal walking speed
- After walking the lap, participants were asked to rest until they are not gasping, at which time the professional personnel measured and recorded their biological parameters.
- After these measurements, the professional personnel removed the BioHarness 3.0, and the test was complete.

Figure 1. The process of collecting data followed six steps in this study.



Participants Questionnaire			
Test date		DBP(mmHg)	
Temp.& Humidity		SBP(mmHg)	
Zephyr No.		HR(beat/min)	
Name		SpO2(%)	
Gender(M/F)		Start walking time	
Birth date		Stop walking time	
Height(cm)		Start running time	
Weight(kg)		200m running time(0.5 lap)	
Neck circum.		600m running time(1.5 lap)	
Waist circum.		1000m running time(2.5 lap)	
Hip circum.		1400m running time(3.5 lap)	
Smoking(Y/N)		1800m running time(4.5 lap)	
Drinking(Y/N)		2200m running time(5.5 lap)	
Drug usage(Y/N)		2600m running time(6.5 lap)	
Medical usage(Y/N)		3000m running time(7.5 lap)	

a. Fill in questionnaire and wear Zephyr BioHarness 3.0

b. Walk 200 meters before running

c. Run 3,000 meters

d. Walk 400 meters after running

e. Measure biological parameters

f. Remove Zephyr and complete test

Data Processing and Analysis

Because there are thousands of data points for individuals and manual artifact detection takes a great deal of time, we developed automatically identified data points (automatic filters) for when participants started to walk and run and when they ceased to walk and run. If the heart rate belt is worn improperly, automatic analysis can identify the equipment issues and delete the data. The Bioharness 3.0 can measure all sensor data precisely, such as the dynamic changes of heart rate with movement in the wearer during running (ie, dynamic cardiac function). The BioHarness 3.0 is snapped into an adjustable chest strap belt that contains skin conductive electrodes and captures the heart rate by recording the cardiac electric impulses and saving the output in beats per minute (bpm).

The BioHarness 3.0 parameters are defined as follows:

- Activity per second (ACi): measurement by the sensor of the amount of activity generated every second by the movement of human body parts. Activity level is measured by an accelerometer within the BioHarness 3.0
- Heart rate per second (HRi): number of heartbeats every second while exercise is measured by the sensor; the report from BioHarness 3.0 for the heart rate is in bpm and is provided as an average, at an interval of one second
- Cardiac force index activity (CFIi_AC): activity per second divided by heartbeats per second, where the (physical) activity and heart rate are estimated and measured by the BioHarness 3.0
- Cardiac force index peak acceleration (CFIi_PA): peak acceleration per second divided by heartbeats per second, where the peak acceleration and heart rate are estimated and measured by the BioHarness 3.0

- Cardiac force ratio activity (CFRi_AC): the CFI_AC value of running divided by the CFI_AC value of walking (at a speed of approximately 3 km per hour)
- Cardiac force ratio peak acceleration (CFRi_PA): the CFI_PA value of running divided by the CFI_PA value of walking (at a speed of approximately 3 km per hour)

After saving the analyzable data for variables such as heart rate, peak acceleration, activity per second, and so on in the BioHarness 3.0, the automation we developed calculates the CFIi_AC (the ith second CFI based on AC) and the CFRi_AC (the ith second CFR based on AC) and analyzes each variety of accumulated data according to the minute.

Statistical Analysis

Data Collection

We described the gender, age, height, weight, BMI, neck circumference, waist circumference, hip circumference, hip-waist ratio, alcohol drinking habits, smoking habits, BioHarness 3.0 parameters, and CFI parameters by the number of subjects, percentage, mean, and standard deviation. Statistical analyses were conducted using SPSS Statistics version 20.0 (IBM Corp). All tests were 2-sided, and *P* values <.05 were considered statistically significant.

After the questionnaire collection was completed, the data were filed using Excel 2003 (Microsoft Corp) and coded according to the framework of this study. First, Excel Visual Basic for Application (Microsoft Corp) was used to distinguish among the collected data automatically in terms of the start of walking, end of walking, start of running, and end of running, which allowed for the identification of a cutoff point between walking and running. In addition, the questionnaire included a question on whether the heart rate belt was worn correctly. Therefore, the individuals whose belts were not worn appropriately were excluded after automatic analysis, while the analyzable data were saved. The CFIi_AC, CFIi_PA, CFRi_AC, and CFRi_PA were then calculated. Data analysis was conducted based on the research framework and purpose, which are described below.

Descriptive Statistics

The mean and standard deviation were used to describe continuous variables including age, height, body weight, BMI, neck circumference, waist circumference, hip circumference, waist-hip ratio, blood oxygen before and after the test, systolic pressure before and after the test, diastolic pressure before and after the test, heart rate before and after the test, activity per second, peak acceleration per second, heart beats per second, CFIi_AC, CFIi_PA, CFRi_AC, CFRi_PA, and the completion time of the 3000-meter empty-handed run.

The number and percentage are used to present the distribution of categorical variables including gender, medical history, medication history, smoking habits, drinking habits, and pass/fail results of the 3000-meter empty-handed run.

Inferential Statistics

Chi-square tests were used to determine the correlations between the 5 categorical variables (gender, medical history, medication history, smoking habits, and drinking habits) and the pass/fail results of the 3000-meter empty-handed run.

Independent sample *t* tests are suitable for comparing the means of two groups. The correlations between the 19 continuous variables (including the age, height, body weight, BMI, neck circumference, waist circumference, hip circumference, waist-hip ratio, blood oxygen before and after test, systolic pressure before and after the test, diastolic pressure before and after the test, heart rate before and after the test, activity per second, peak acceleration per second, heart beats per second, CFIi_AC, CFIi_PA, CFRi_AC, and CFRi_PA) and the pass/fail results of the 3000-meter empty-handed run were compared.

Multiple linear regression analysis was used to consider the effects of the two independent variables on dependent variables. For this method, the data type of the dependent variables must be continuous, while the data type of the independent variables can be continuous or categorical. Furthermore, through the establishment of a regression model, the prediction/forecast of dependent variables with independent variables could be achieved. Therefore, this study was intended to analyze 28 independent variables including gender, disease history, medication history, smoking habits, drinking habits, age, height, body weight, BMI, neck circumference, waist circumference, hip circumference, waist-hip ratio, blood oxygen before and after the test, systolic pressure before and after the test, diastolic pressure before and after the test, heart rate before and after the test, activity per second, peak acceleration per second, heart beats per second, CFIi_AC, CFIi_PA, CFRi_AC, and CFRi_PA. The dependent variable was the completion time for the 3000-meter empty-handed run.

In the case that there is only one independent variable, the analysis used is known as a simple logistic regression or univariate logistic regression. In a logistic regression analysis, the independent variables can be continuous or categorical. The purpose of using logistic regression is to establish a practical and reasonable model capable of providing the most concise and fit analysis results. Once established, the model can be used to predict the relationships between the dependent variable and a set of predictor variables. In this study, the logistic regression analysis was employed to establish a predictive model. The independent variables in the analysis included gender, medical history, medication history, smoking habits, drinking habits, age, height, body weight, BMI, neck circumference, waist circumference, hip circumference, waist-hip ratio, blood oxygen before and after the test, systolic pressure before and after the test, diastolic pressure before and after the test, heart rate before and after the test, activity per second, peak acceleration per second, heart beats per second, CFIi_AC, CFIi_PA, CFRi_AC, and CFRi_PA. The dependent variable was the pass/fail result of the 3000-meter empty-handed run.

Generalized estimating equations (GEEs) are primarily used to analyze the dependency of the data samples (including repeated measurement and long-term tracking studies). When the data are longitudinal and they record the state of a subject at different time points, the observation values from the same subject are all considered to have correlation. Since a subject might be tested more than twice in this study, the resulting data were

interdependent. Therefore, the GEE was used in the study, and the covariance matrix was assumed to be AR(1). AR(1) indicated that the next time point was only highly correlated with the previous time point. The dependent variables were continuous variables (the results of the 3000-meter empty-handed run) and binary variables (pass or fail for the 3000-meter empty-handed run).

Results

Demographic Data

A total of 96 voluntary participants completed this study. The following sections show their basic demographic data, predictive physical factors for 3000-meter running performance, models of GEE and logistic regression for their running performance, and the stronger and weaker cardiac forces for the 3000-meter runners. Table 1 provides the descriptive demographic data: the average participant age was 23.25 [SD 3.48] years, average weight 69.40 [SD 9.34] kg, average BMI 22.93 [SD 2.66] kg/m^2, average waist circumference 78.26 [SD 8.20] cm, average hip circumference 95.08 [SD 7.11] cm, and the average waist-to-hip ratio was 0.82 [SD 0.05]. The demographic characteristics were significantly correlated with the 3000-meter running time as follows: age (r=.241, P=.008), body weight (r=.233, P=.047), BMI (r=.284, P=.006), waist circumference (r=.319, P=.005), and waist-to-hip ratio (r=.30, P=.006). These variables correlated positively with the 3000-meter empty-handed running time.

Table 1. Descriptive demographic characteristics of participants (n=96).

Characteristic	Value, n (%)	Mean (SD)	Min-max
Gender, male	96 (100)	—	—
Age in years	96 (100)	23.25 (3.48)	19-38
Height (cm)	96 (100)	173.90 (6.42)	160-188
Weight (kg)	96 (100)	69.40 (9.34)	52-95
BMI (kg/m^2)	96 (100)	22.93 (2.66)	18.18-31.02
Neck circumference (cm)	96 (100)	34.89 (2.79)	25-41
Waist circumference (cm)	96 (100)	78.26 (8.20)	62-102
Hip circumference (cm)	96 (100)	95.08 (7.11)	79-112
Waist-hip ratio	96 (100)	0.82 (0.05)	0.74-0.97
Smoking			
Yes	12 (13)	—	—
No	84 (88)	—	—
Drinking			
Yes	7 (7)	—	—
No	89 (93)	—	—
Medication			
Yes	0 (0)	—	—
No	96 (100)	—	—
Medical history			
Yes	4 (4)	—	—
No	92 (96)	—	—

Prediction of Physical Factors Affecting Completion Times for 3000-Meter Run

To determine the factors that affected the 3000-meter completion times for the 96 valid cases, first we used BioHarness 3.0 parameters in the model, performing a univariate analysis with a simple linear regression to find the appropriate variate to build the 3000-meter running completion time model by performing a stepwise regression analysis on variates with a significant univariate (Table 2). According to the stepwise regression filtering, waist circumference, smoking, ACi1 (physical activity at the 1st minute), and CFRr.AC.w1i6 (CFI_AC.run at the 6th minute divided by CFI_AC.walk at the 1st minute) were significant factors in the completion times. Thus, the prediction model is as follows: 3000-meter running completion time = 1346.499 + 4.152 × [waist circumference] – 47.97 × [CFRr.AC.w1i6] – 5.43 × [ACi1] + 129.60 × [smoking]. From this formula, we know that with every increased centimeter on a participant's waist circumference, the completion time increased by 4.152 seconds. The finishing times for smokers were 129.6 seconds more than for nonsmokers. For each unit increase in the CFRr.AC.w1i6, the running time decreased by 47.97 seconds, and for each unit increase in the ACi1, running time decreased by 5.43 seconds.

Table 2. Physical predictors of 3000-meter running qualification (univariate).

Variable		SE	*P* value	95% CI
Constant	1346.50	207.146	<.001	935.03 to 1757.97
Waist circumference	4.15	1.260	.001	1.64 to 6.66
Smoking	129.60	30.950	<.001	68.13 to 191.07
CFRr.AC.w1i6[a]	–47.97	11.060	<.001	–69.94 to –25.99
ACi=1[b]	–5.43	1.370	<.001	–8.15 to –2.71

[a]CFRr.AC.w1i6: CFI_AC.run at 6th minute divided by CFI_AC.walk at 1st minute.

[b]ACi=1: activity per second accumulated until the 1st minute after the contestant started to run.

We used the BioHarness 3.0 parameters in the model and performed a univariate analysis with binary logistic regression. We then used a forward logistic regression method on the significant variates from the previous univariates to find the predictors affecting the 3000-meter qualification. We established a prediction model by multivariate logistic regression (Table 3). The results showed that the waist circumference, smoking, and ACi7 were significant factors for the 3000-meter run qualification. The prediction model was as follows: ln (3000 meters running performance pass probability / fail results probability) = – 2.702 – 0.096 × [waist circumference] - 1.827 × [smoke] + 0.020 × [ACi7]. If the smoking and ACi7 were controlled, contestants with larger waist circumferences tended to fail the qualification based on the formula above. If the waist circumference and ACi7 were controlled, smokers tended to fail more often than nonsmokers.

Table 3. Physical predictors of 3000-meter running qualification (multivariate logistic regression).

Variate		SE	*P* value	OR[a]	95% CI
Waist circumference	–0.096	0.032	.003	0.909	0.853-0.968
Smoking	–1.827	0.890	.04	0.161	0.028-0.920
ACi=7[b]	0.020	0.007	.004	1.020	1.006-1.034
Constant	–2.702	4.185	.52	0.067	—

[a]OR: odds ratio.

[b]ACi=7: activity per second accumulated until the 7th minute after the contestant started to run.

Comparison of Prediction Model and Actual Test Time of Finishing the 3000-Meter Run

The results showed no significant difference between the estimated time and actual completion time (*P*=.42), indicating no significant difference between the estimated completion time and actual measurement of the number of seconds in this study prediction model (Table 4).

Table 4. Comparison of linear model estimation results and actual measurements of finish times for 3000-meter runs (n=96).

Variate	Mean (SD)	SE	95% CI	*P* value
Linear model estimation	885.42 (79.16)	8.08	869.58-901.26	.42
Actual measurement	885.33 (124.61)	12.72	870.65-900.01	—

Comparison of Pass/Fail Results Predicted by the Logistic Model With Actual Results of 3000-Meter Empty-Handed Run

A McNemar test was employed to compare the pass/fail results estimated using the logistic model with the actual results of the 3000-meter empty-handed run (Table 5). There was a statistically significant correlation between the estimated and actual passing probability (Pearson coefficient *r*=.477, Spearman correlation coefficient=.477, =0.476 [SE 0.090, 95% CI 0.300-0.653], and the area under the receiver operating characteristic curve 0.768, 95% CI 0.687-0.857).

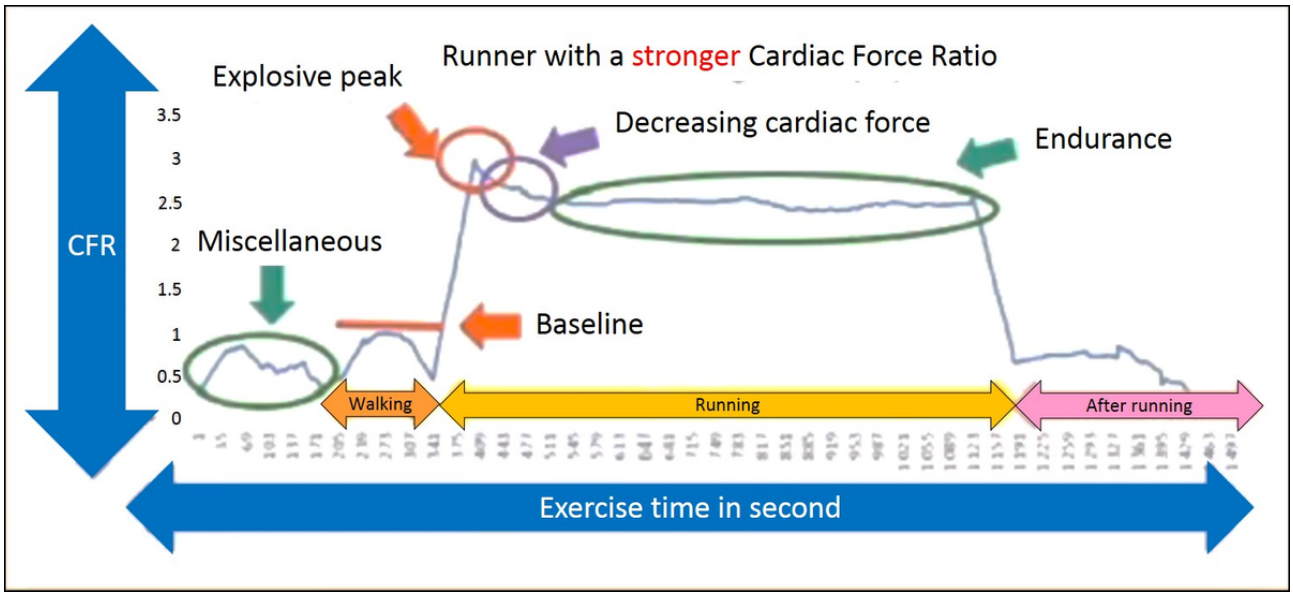
Table 5. Comparison of the pass/fail results of the 3000-meter empty-handed run predicted by the logistic model with actual measured results.

Logical model prediction	Actual measurement		McNemar	Correlation coefficient	
	Fail	Pass	*P* value	Spearman	*P* value
Fail	39 (40.63)	13 (13.54)	.84	.477	<.001
Pass	12 (12.50)	32 (33.33)	—	—	—

Comparison of Stronger and Weaker Cardiac Forces During 3000-Meter Run

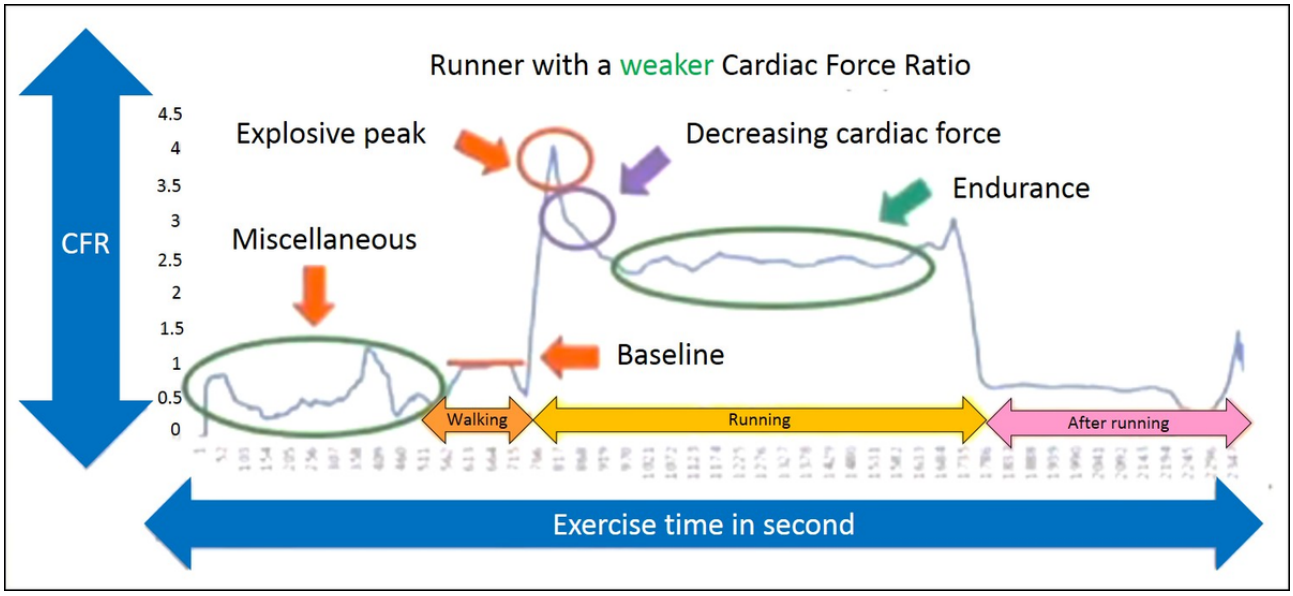
The relationship between the CFR of a healthy 22-year-old person with strong cardiac force and time is illustrated in Figure 2. In this example, a representative runner has a strong heart. The runner uses the method and apparatus for monitoring their cardiac status during exercise according to the protocol, including the walking period, running period, and postrunning period. During the running period, data are obtained while the user slowly runs 3000 meters. The postrunning period comes after the runner has stopped running. Figure 2 uses the maximum CFI of the walking period to serve as a reference to calculate the CFRs of other time periods. The relationship between the CFR and time is plotted. Thus, this finding shows the runner's CFR varying with changes in activity. The signals before the walking period are noise. When the runner changes from the walking period to the running period, the runner converts from a relatively static walking condition to a dynamic running condition. Since the runner's acceleration increases, the CFR quickly increases to an explosive peak, indicating the runner's explosive power. This period shows a CFR of approximately 3. After the explosive peak, the CFR begins to decrease, and after a time lapse, it becomes moderate and gentle, entering an endurance period at approximately 2.5 CFR. After the running period has ended, the postrunning period begins. At this moment, the runner's dynamic condition ends, acceleration decreases very quickly, and the CFR also decreases.

Figure 2. Runner with stronger cardiac force during 3000-meter running.



Another relationship, this time between the CFR of a 44-year-old with weak cardiac force and time is illustrated in Figure 3. The protocol for this runner is the same as that for the first runner (Figure 2), with the walking period, running period, and postrunning period. During the running period, data are obtained while the user slowly runs 3000 meters. The postrunning period occurs after the runner has stopped running. Figure 3 uses the maximum CFI of the walking period as a reference to calculate the CFRs of the other periods. The relationship between the CFR and the time is plotted. Signals from before the walking period are noise. Again, the runner's condition changes from relatively static when walking to dynamic when running, and the CFR reaches an explosive peak with acceleration, at approximately 4. After the explosive peak, the CFR decreases and eventually becomes moderate and gentle, entering an endurance period at approximately 2.5 CFR. After the running period ends, postrunning period begins, the runner's dynamic condition ends, acceleration decreases very quickly, and the CFR also decreases.

Figure 3. Runner with weaker cardiac force during 3000-meter running.



In summary, this study investigates a new calculation method for monitoring cardiac status during exercise that uses the CFI of walking for the runner as a reference to obtain the ratio of between the cardiac force of exercise and that of walking (CFR) to provide a standard for determining if the heart is capable of the exercise. Since the CFR is simply a ratio, it helps eliminate individual differences among the users and errors caused by changing the location of the detection units.

Discussion

Principal Findings

Previous traditional study methods used only parameters such as questionnaires and biochemical values to predict a runner's completion time [27,28]. Our study results revealed that physical characteristics including age, body weight, BMI, waist circumference, and waist-to-hip ratio have impacts on the 3000-meter running qualification. However, this study uses a new method for detecting cardiac status that calculates the instantaneous CFI of the runner by using weight, heart rate, and acceleration or activity level of the user to transform the physiological parameters, which are not quite meaningful individually, into a CFI that is more meaningful. The CFI is provided for detecting the cardiac status of the runner in a dynamic manner. We successfully established a novel prediction model for the running performance based on the real-time cardiac force during running.

In this study, the data obtained from wearable technology were included as variables to explore the factors in the demographic characteristics, physiological parameters, and BioHarness 3.0 parameters that were correlated with the 3000-meter empty-handed run and its pass/fail results to establish a predictive model.

The summary of the results and discussion revealed that certain demographic characteristics in our study were significantly positively correlated with the completion time of the 3000-meter empty-handed run, including age, body weight, BMI, waist circumference, and waist-hip ratio. By contrast, there were no statistically significant correlations between physiological parameters and completion time of the 3000-meter empty-handed run. The BioHarness 3.0 parameters that were significantly negatively correlated with completion time of the 3000-meter empty-handed run included an ACi=1-10, PAi=1, PAi=8, PAi=9, PAi=10, HRi=1-10, CFRr.AC.w1i5, CFRr.AC.w1i6, CFRr.AC.w1i7, CFRr.AC.w1i8, CFRr.AC.w1i9, and CFRr.AC.w1i10.

Among the demographic characteristics examined in our study, the statistically significant variables for the pass/fail results of the 3000-meter empty-handed run included age, waist circumference, hip circumference, waist-hip ratio, and smoking. None of the physiological parameters were statistically significant variables for the pass/fail results of the 3000-meter empty-handed run. Among the BioHarness 3.0 parameters, statistically significant variables for the pass/fail results of the 3000-meter empty-handed run included an ACi=1-10, PAi=1, PAi=6, PAi=7, PAi=8, PAi=9, PAi=10, CFI.AC.WALK=4, and CFI.AC.RUNi=5-10. The HRi was not a statistically significant variable.

In the multivariate linear regression model, waist circumference, ACi=1, and CFRr.AC.w1i6 were able to predict the completion time of the 3000-meter empty-handed run. Moreover, there were no significant differences between the predicted values and the actual measured values. In the logistic regression model, waist circumference, smoking habits, and ACi=7 were able to predict the pass/fail results for the 3000-meter empty-handed run. The correlation between the estimated and actual passing probability was statistically significant.

Finally, a GEE-linear (binary) analysis was conducted on the individuals tested one or more times. The results show that the CFI.PA.WALK=3 was capable of predicting the completion time of the 3000-meter empty-handed run, while the ACi=3 and CFRr.AC.w1i7 were capable of predicting the pass/fail results of the 3000-meter empty-handed run.

Exploration of Correlation Factors Between Demographic Characteristics and 3000-Meter Empty-Handed Run

Age

In 1988, Marti et al [32] explored the correlations between age, body weight, BMI, and lifestyle and 16-km running performance in 4000 joggers. The results of the study showed that age was a predictive factor for completion time of the 16-km run (β=.37, *P*<.001) [32]. In 2009, Leyk et al [31] analyzed 439,278 running times from a results lists of 108 marathon competitions. Their primary findings were there are virtually no relevant running time differences (*P*<.01) in marathon finishers from 20 to 55 years and the majority of middle-aged and elderly athletes have training histories of less than 7 years of running. Lara et al [33] explored the correlations of gender and age with completion time in marathon runners. The results showed that there was a significant positive correlation between the men's age and completion time (Pearson correlation coefficient *r*=.92, *P*<.05). Knechtle et al [34] analyzed the correlation between age and marathon completion time in marathon runners aged 5 to 93 years. The results showed that completion time was increased with age in marathon runners over an age range of 5 to 93 years (Pearson correlation coefficient *r*=.97, R^2=.94, *P*<.001) as well as marathon runners with an age range of 18 to 80 years (Pearson correlation coefficient *r*=.99, R^2=.98, *P*<.001). The results of our study showed that age was significantly positively correlated with completion time of a 3000-meter empty-handed run (seconds) in males aged 19 to 38 years (mean age 23.25 [SD 3.48] years; Pearson correlation coefficient *r*=.241, *P*<.001). The result indicated that the older the males were, the longer it took them to complete the 3000-meter run.

Body Weight and Body Mass Index

The results of our study showed that body weight (69.4 [SD 9.34] kg) was significantly positively correlated with the 3000-meter empty-handed running performance (seconds; Pearson correlation coefficient *r*=.233, *P*<.001). This result indicated that the heavier the body was, the longer it took to complete the 3000-meter run. Previously, Marti et al [32] showed that the BMI value was a predictive factor for completion time (β=0.23, *P*<.001). The results of our study showed that BMI (22.93 [SD 2.66] kg/m^2) was positively correlated with completion time of the 3000-meter empty-handed run (Pearson correlation coefficient *r*=.284, *P*<.001). This finding indicated that the larger the BMI was, the longer the completion time. Our result was consistent with the above literature.

Waist Circumference, Neck Circumference, and Hip Circumference

Recently, many studies have shown that neck and waist circumference are highly correlated and are related to the metabolic status of the body [35-37]. In our study, waist circumference was positively correlated with completion time of the 3,000-meter empty-handed run (Pearson correlation coefficient *r*=.319, *P*<.001). Namely, the larger the waist circumference was, the longer the completion time. Neck circumference was significantly negatively correlated with running performance in the GEE-binary analysis (OR 0.88; *P*=.03): the larger the neck circumference was, the more likely the individual would obtain a failing result. This finding was similar to the waist circumference result. In our study, hip circumference of the individuals who failed the 3000-meter empty-handed run test was larger than that of those who passed the run test (96.44 [SD 7.66] vs 93.55 [SD 6.17], *P*=.046), which was similar to the above results. The waist-hip ratio was also positively correlated with completion time (seconds) of the 3000-meter empty-handed run (Pearson correlation coefficient *r*=.300, *P*<.001), indicating that the larger the waist-hip ratio was, the longer the completion time.

Exploration of Correlation Factors Between Bioharness 3.0 Parameters and 3000-Meter Empty-Handed Run

A study conducted by Johnstone et al [27] showed that the parameters collected by the BioHarness 3.0 have high reliability and validity. The reliability and validity were measured in 20 healthy males (10 subjects for a reliability analysis and 10 for a validity analysis). In terms of the validity of the BioHarness 3.0, blood oxygen measurement was assessed using the T31 (Polar Electro) as the standard while the respiratory rate measurement was assessed using a US-made face mask (Hans Rudolf Inc) as the standard. In addition, subjects performed incremental exercises (walking 4 to 6 kilometers per hour, jogging 8 to 10.5 kilometers per hour, and running 11 kilometers per hour) carrying the portable METAMAX 3B (Cortex Medical; weight 650 g). The heart rate, respiratory rate, and accelerometer were measured. The results showed that heart rate was statistically significantly correlated with the standard value (*r*=.98). Excluding the walking period, a strong correlation was observed between heart rate and the standard at a running speed of 8 to 10.5 kilometers per hour (*r*=.93). As running speed increased, the correlation between heart rate and the standard decreased (11 km/h, *r*=.67) [27]. The results of our study showed that the correlation between the accumulated heart rate per minute and completion time was reduced as running time was prolonged and running speed increased (HRi=1, *r*=–.283; HRi=10, *r*=–.258). Our results were similar to the results of the study described above. The correlation between the amount of activity measured using the BioHarness 3.0 and the standard was statistically significant (*P*<.001, *r*=.91).

In 2014, Lin et al [28] divided 10 ultramarathon runners with distinct performances into two groups and used accelerometers to measure the correlation between acceleration and running distance at different speeds (3, 6, 8, 9, 10, and 12 kilometers per hour). The study examined the correlations between the x, y, and z axes of the triaxial acceleration gauge and running speed. The results showed that the y-axis was negatively correlated with running speed at any speed (*r*=–.89 to –.92, *P*<.001). These results were similar to our results in that activity (*r*=–.214 to –.317, *P*<.001) and peak acceleration (*r*=–.203 to –.226, *P*=.003) were negatively correlated with completion time.

Limitations

There are some limitations to this study. The subjects of this study were students (college-level and graduate-level) and service personnel at a military academy in the northern region

(including voluntary and compulsory service personnel). The subjects were all males, and females were not included in the study. Since the wearable device must be worn tightly against the skin at the lower edge of the chest, most women were unwilling to participate. In addition, the poor availability of the output from the women's physiological/structural relationship data rendered it impossible to include these data for analysis. This study only considered the subjects' smoking, drinking, medication, and medical histories. Potential interference factors (exercise frequency, lifestyle, dietary habit, and environmental factors) were not explored in this study.

Future Works

Our study only included one military college in the northern region as the research subject. We suggest that the sample size in future study be increased to include field troops and national army physical fitness evaluation centers. In addition, wearable technology can be used as one form of physical strength evaluation equipment for the national army to reduce the risk of sport injury and accidents. Completion time was correlated with exercise frequency, lifestyle, dietary habit, environmental factors, and caffeine intake. We suggest that future studies include the above factors as variables. Taking into the consideration the inconvenience among women in wearing the heart rate belt, we also suggest that the belt be worn on various parts of the body and reliability and validity be confirmed. In addition, the staff should confirm that the subjects wear the heart rate band tightly against the skin when collecting the data, which would improve the usability of the data.

In this study, smoking was included as a variable and discussed. The output parameters of the BioHarness 3.0 also include the respiratory rate. Since running is closely related to cardiopulmonary function, we suggest that future studies include the respiratory rate for in-depth exploration. In addition, the output parameters of the BioHarness 3.0 include posture, activity intensities on three different planes (sagittal plane, vertical, Z; cross-sectional plan, mediolateral, X; and longitudinal plane, anteroposterior, Y). We suggest that these parameters be included in subsequent studies for analysis.

We also suggest combining the relevant physiological values embedded in the heart rate belt and the CFI and CFR values discussed in this study with a mobile app, which would allow subjects to understand their own exercise status in real time and provide real-time data. We propose uploading the CFI and CFR obtained in this study to the cloud community for people to compare with themselves and others so the CFI and CFR may serve as incentive parameters for physical activity.

Conclusions

This study used wearable technology and focused on real-time cardiac function changes related to 3000-meter running and qualification results. Based on the cardiac force, we successfully established a reliable prediction model for running performance. In the future, this cardiac force model can be used during running training and may assist in research on the application of the CFI.

Acknowledgments

The authors would like to thank the staff and researchers at the National Defense Medical Center, Tri-Service General Hospital, Taoyuan Armed Forces General Hospital, and Chang Gung University. We are grateful to the Building Interdisciplinary Careers for soldiers, officers, and commanders at the Ministry of National Defense. The funder had no role in the preparation, review, or approval of the manuscript and decision to submit the manuscript for publication. This study was supported by grants from the Research Fund of the Taoyuan Armed Forces General Hospital (AFTYGH-10812 and AFTYGH-D-109035), National Defense Medical Center, and Medical Affairs Bureau, Ministry of National Defense, Taiwan, and Ministry of Science and Technology (104-2623-E-016-004-D and 102-2623-E-016-001-D), Taiwan. Finally, we particularly thank Prof Dr med Shih-Hua Lin, who supports the series studies about cardiac force at the Research Center for Biotechnology and Medicine Policy. The use of cardiac force received the National Innovation Award Healthcare Taiwan 2018.

Authors' Contributions

PJH, CWC, YKL, JSC, and CMC conceived of the study, acquired study funding, led study design and supervised its coordination, and drafted the manuscript for publication. CWC, YKL, JSC, and CMC had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. KHL, FKH, PJT, CTW, YKP, YL, MHK, KHC, YSW, HYW, YTC, YTC, and CSC managed study execution and contributed to drafting of the manuscript. KHL, FKH, PJT, CTW, YKP, YL, KHC, YSW, HYW, CWC, and CMC consulted on data safety, study design and execution, and contributed to drafting of the manuscript. KHL, PJT, CTW, YKP, HYW, and FHL coordinated intervention design and contributed to drafting of the manuscript. KHL, PJT, HYW, and CMC participated in study design, conducted statistical analysis, and contributed to drafting of the manuscript. PJH, CCC, CPC, CWC, YKL, JSC, and CMC contributed to drafting of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Platen P, Schaar B. How to carry out a health-orientated marathon training programme for running and inline skating. Eur J Cardiovasc Prev Rehabil 2003 Aug;10(4):304-312. [doi: 10.1097/00149831-200308000-00015] [Medline: 14555888]
2. Nystoriak MA, Bhatnagar A. Cardiovascular effects and benefits of exercise. Front Cardiovasc Med 2018;5:135 [FREE Full text] [doi: 10.3389/fcvm.2018.00135] [Medline: 30324108]
3. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. N Engl J Med 2001 Sep 13;345(11):790-797. [doi: 10.1056/NEJMoa010492] [Medline: 11556298]
4. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. N Engl J Med 2000 Jul 06;343(1):16-22. [doi: 10.1056/NEJM200007063430103] [Medline: 10882764]
5. Nikolaidis PT, Rosemann T, Knechtle B. A brief review of personality in marathon runners: the role of sex, age and performance level. Sports (Basel) 2018 Sep 18;6(3) [FREE Full text] [doi: 10.3390/sports6030099] [Medline: 30231559]
6. Nikolaidis PT, Rosemann T, Knechtle B. Sex differences in the age of peak marathon race time. Chin J Physiol 2018 Apr 30;61(2):85-91. [doi: 10.4077/CJP.2018.BAG535] [Medline: 29526077]
7. Jokl P, Sethi PM, Cooper AJ. Master's performance in the New York City Marathon 1983-1999. Br J Sports Med 2004 Aug;38(4):408-412 [FREE Full text] [doi: 10.1136/bjsm.2002.003566] [Medline: 15273172]
8. La Gerche A, Heidbuchel H. Can intensive exercise harm the heart? You can get too much of a good thing. Circulation 2014 Sep 16;130(12):992-1002. [doi: 10.1161/CIRCULATIONAHA.114.008141] [Medline: 25223770]
9. Herm J, Töpper A, Wutzler A, Kunze C, Krüll M, Brechtel L, et al. Frequency of exercise-induced ST-T-segment deviations and cardiac arrhythmias in recreational endurance athletes during a marathon race: results of the prospective observational Berlin Beat of Running study. BMJ Open 2017 Aug 03;7(8):e015798 [FREE Full text] [doi: 10.1136/bmjopen-2016-015798] [Medline: 28775185]
10. Spada TC, Silva JMRD, Francisco LS, Marçal LJ, Antonangelo L, Zanetta DMT, et al. High intensity resistance training causes muscle damage and increases biomarkers of acute kidney injury in healthy individuals. PLoS One 2018;13(11):e0205791 [FREE Full text] [doi: 10.1371/journal.pone.0205791] [Medline: 30399190]
11. Kim JH, Malhotra R, Chiampas G, d'Hemecourt P, Troyanos C, Cianca J, Race Associated Cardiac Arrest Event Registry (RACER) Study Group. Cardiac arrest during long-distance running races. N Engl J Med 2012 Jan 12;366(2):130-140. [doi: 10.1056/NEJMoa1106468] [Medline: 22236223]
12. Kiyohara K, Nishiyama C, Kiguchi T, Nishiuchi T, Hayashi Y, Iwami T, et al. Exercise-related out-of-hospital cardiac arrest among the general population in the era of public-access defibrillation: a population-based observation in japan. J Am Heart Assoc 2017 Jun 13;6(6) [FREE Full text] [doi: 10.1161/JAHA.117.005786] [Medline: 28611095]
13. Roberts WO, Roberts DM, Lunos S. Marathon related cardiac arrest risk differences in men and women. Br J Sports Med 2013 Feb;47(3):168-171. [doi: 10.1136/bjsports-2012-091119] [Medline: 22736205]
14. Hart L. Marathon-related cardiac arrest. Clin J Sport Med 2013 Sep;23(5):409-410. [doi: 10.1097/01.jsm.0000433155.97054.c8] [Medline: 23989386]
15. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The accuracy of smart devices for measuring physical activity in daily life: validation study. JMIR Mhealth Uhealth Dec 13 2018;6(12):e10972. [doi: 10.2196/10972] [Medline: 30545810]
16. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental perspectives of a wearable activity tracker for children younger than 13 years: acceptability and usability study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
17. Islam SMS, Cartledge S, Karmakar C, Rawstorn JC, Fraser SF, Chow C, et al. Validation and acceptability of a cuffless wrist-worn wearable blood pressure monitoring device among users and health care professionals: mixed methods study. JMIR Mhealth Uhealth 2019 Sep 14;7(10):e14706 [FREE Full text] [doi: 10.2196/14706] [Medline: 31628788]
18. Graham SA, Jeste DV, Lee EE, Wu T, Tu X, Kim H, et al. Associations between heart rate variability measured with a wrist-worn sensor and older adults' physical function: observational study. JMIR Mhealth Uhealth 2019 Oct 23;7(10):e13757 [FREE Full text] [doi: 10.2196/13757] [Medline: 31647469]
19. Thorpe J, Forchhammer BH, Maier AM. Adapting mobile and wearable technology to provide support and monitoring in rehabilitation for dementia: feasibility case series. JMIR Form Res 2019 Oct 17;3(4):e12346 [FREE Full text] [doi: 10.2196/12346] [Medline: 31625951]
20. Villarejo MV, Zapirain BG, Zorrilla AM. Algorithms based on CWT and classifiers to control cardiac alterations and stress using an ECG and a SCR. Sensors (Basel) 2013 May 10;13(5):6141-6170 [FREE Full text] [doi: 10.3390/s130506141] [Medline: 23666135]
21. Arif M, Kattan A. Physical activities monitoring using wearable acceleration sensors attached to the body. PLoS One 2015;10(7):e0130851 [FREE Full text] [doi: 10.1371/journal.pone.0130851] [Medline: 26203909]
22. Papi E, Murtagh GM, McGregor AH. Wearable technologies in osteoarthritis: a qualitative study of clinicians' preferences. BMJ Open 2016 Jan 25;6(1):e009544 [FREE Full text] [doi: 10.1136/bmjopen-2015-009544] [Medline: 26810998]
23. Ahamed NU, Kobsar D, Benson L, Clermont C, Kohrs R, Osis ST, et al. Using wearable sensors to classify subject-specific running biomechanical gait patterns based on changes in environmental weather conditions. PLoS One 2018;13(9):e0203839 [FREE Full text] [doi: 10.1371/journal.pone.0203839] [Medline: 30226903]

24. Rodríguez-Molina J, Martínez J, Castillejo P, López L. Combining wireless sensor networks and semantic middleware for an Internet of Things-based sportsman/woman monitoring application. Sensors (Basel) 2013 Jan 31;13(2):1787-1835 [FREE Full text] [doi: 10.3390/s130201787] [Medline: 23385405]
25. Nazari G, Bobos P, MacDermid JC, Sinden KE, Richardson J, Tang A. Psychometric properties of the Zephyr bioharness device: a systematic review. BMC Sports Sci Med Rehabil 2018;10:6 [FREE Full text] [doi: 10.1186/s13102-018-0094-4] [Medline: 29484191]
26. Nazari G, MacDermid JC, Sinden KE, Richardson J, Tang A. Reliability of Zephyr Bioharness and Fitbit Charge measures of heart rate and activity at rest, during the modified Canadian aerobic fitness test, and recovery. J Strength Cond Res 2019 Feb;33(2):559-571. [doi: 10.1519/JSC.0000000000001842] [Medline: 30689619]
27. Johnstone JA, Ford PA, Hughes G, Watson T, Mitchell ACS, Garrett AT. Field based reliability and validity of the bioharness™ multivariable monitoring device. J Sports Sci Med 2012;11(4):643-652 [FREE Full text] [Medline: 24150074]
28. Lin S, Sung W, Kuo F, Kuo TBJ, Chen J. Impact of center-of-mass acceleration on the performance of ultramarathon runners. J Hum Kinet 2014 Dec 09;44:41-52 [FREE Full text] [doi: 10.2478/hukin-2014-0109] [Medline: 25713664]
29. Chu, CM. Method for detecting cardiac status, method for monitoring cardiac status during exercise, and apparatus for monitoring cardiac status. United States Patent Office. 2016. URL: https://patents.google.com/patent/US20160058314 [accessed 2020-05-13]
30. Chu, CM. Method for detecting cardiac status, method for monitoring cardiac status during exercise, and apparatus for monitoring cardiac status. Taiwan Patent 2016; No. 408 I546051 2016.
31. Leyk D, Erley O, Gorges W, Ridder D, Rüther T, Wunderlich M, et al. Performance, training and lifestyle parameters of marathon runners aged 20-80 years: results of the PACE-study. Int J Sports Med 2009 May;30(5):360-365. [doi: 10.1055/s-0028-1105935] [Medline: 19277939]
32. Marti B, Abelin T, Minder CE. Relationship of training and life-style to 16-km running time of 4000 joggers. The '84 Berne. Int J Sports Med 1988 Apr;9(2):85-91. [doi: 10.1055/s-2007-1024984] [Medline: 3384526]
33. Lara B, Salinero JJ, Del Coso J. The relationship between age and running time in elite marathoners is U-shaped. Age (Dordr) 2014 Apr;36(2):1003-1008 [FREE Full text] [doi: 10.1007/s11357-013-9614-z] [Medline: 24407890]
34. Knechtle B, Assadi H, Lepers R, Rosemann T, Rüst C. Relationship between age and elite marathon race time in world single age records from 5 to 93 years. BMC Sports Sci Med Rehabil 2014;6:31 [FREE Full text] [doi: 10.1186/2052-1847-6-31] [Medline: 25120915]
35. Liang J, Wang Y, Li H, Liu X, Qiu Q, Qi L. Neck circumference and early stage atherosclerosis: the cardiometabolic risk in Chinese (CRC) study. Cardiovasc Diabetol 2014 Jul 08;13:107 [FREE Full text] [doi: 10.1186/s12933-014-0107-x] [Medline: 25001365]
36. Liu Y, Chang S, Lin W, Hsu J, Chung C, Chang J, et al. Neck circumference as a predictive indicator of CKD for high cardiovascular risk patients. Biomed Res Int 2015;2015:745410 [FREE Full text] [doi: 10.1155/2015/745410] [Medline: 26295050]
37. Hsiao P, Lin H, Chang S, Hsu J, Lin W, Chung C, et al. Albuminuria and neck circumference are determinate factors of successful accurate estimation of glomerular filtration rate in high cardiovascular risk patients. PLoS One 2018;13(2):e0185693 [FREE Full text] [doi: 10.1371/journal.pone.0185693] [Medline: 29394255]

Abbreviations

ACi: activity per second
bpm: beats per minute
CFI: cardiac force index
CFIi_AC: cardiac force index activity
CFIi_PA: cardiac force index peak acceleration
CFR: cardiac force ratio
CFRi_AC: cardiac force ratio activity
CFRi_PA: cardiac force ratio peak acceleration
GEE: generalized estimating equation
HRi: heart rate per second
mHealth: mobile health

Edited by G Eysenbach; submitted 02.07.19; peer-reviewed by L Becker, M Lang, IN Gomez, R Kretschmann; comments to author 01.08.19; revised version received 18.12.19; accepted 22.03.20; published 21.07.20.

Please cite as:
Hsiao PJ, Chiu CC, Lin KH, Hu FK, Tsai PJ, Wu CT, Pang YK, Lin Y, Kuo MH, Chen KH, Wu YS, Wu HY, Chang YT, Chang YT, Cheng CS, Chuu CP, Lin FH, Chang CW, Li YK, Chan JS, Chu CM
Usability of Wearable Devices With a Novel Cardiac Force Index for Estimating the Dynamic Cardiac Function: Observational Study
JMIR Mhealth Uhealth 2020;8(7):e15331
URL: https://mhealth.jmir.org/2020/7/e15331
doi:10.2196/15331
PMID:32706725

Original Paper

Evaluation Criteria for Weight Management Apps: Validation Using a Modified Delphi Process

Noemí Robles[1,2,3*], PhD; Elisa Puigdomènech Puig[1,3,4*], MSc; Corpus Gómez-Calderón[5*], BA; Francesc Saigí-Rubió[6,7*], PhD; Guillem Cuatrecasas Cambra[8*], MD; Alberto Zamora[9,10*], MD, PhD; Montse Moharra[4,11*], BA; Guillermo Paluzié[9*], MD, PhD; Mariona Balfegó[8*], PhD; Carme Carrion[1,2,3,6*], PhD

[1]eHealth Lab Research Group, Universitat Oberta de Catalunya, Barcelona, Spain
[2]eHealth Center, Universitat Oberta de Catalunya, Barcelona, Spain
[3]Red de Investigación en Servicios de Salud en Enfermedades Crónicas, Barcelona, Spain
[4]Agència de Qualitat i Avaluació Sanitàries de Catalunya, Barcelona, Spain
[5]Marina Salud, Alicante, Spain
[6]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
[7]Interdisciplinary Research Group on ICTs, Universitat Oberta de Catalunya, Barcelona, Spain
[8]Clínica Sagrada Família, CPEN SL Servei d'Endocrinologia i Nutrició, Barcelona, Spain
[9]Corporació de Salut del Maresme i la Selva, Hospital de Blanes, Blanes, Spain
[10]Grup de Medicina Traslacional i Ciències de la Decisió, Universitat de Girona, Girona, Spain
[11]Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública, Barcelona, Spain
[*]all authors contributed equally

Corresponding Author:
Elisa Puigdomènech Puig, MSc
Agència de Qualitat i Avaluació Sanitàries de Catalunya
Carrer Roc Boronat 81-95
Barcelona, E08005
Spain
Phone: 34 93 5513476
Email: epuigdomenech@gencat.cat

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2023/1/e47584

Abstract

Background: The use of apps for weight management has increased over recent years; however, there is a lack of evidence regarding the efficacy and safety of these apps. The EVALAPPS project will develop and validate an assessment instrument to specifically assess the safety and efficacy of weight management apps.

Objective: The aim of this study was to reach a consensus among stakeholders on a comprehensive set of criteria to guide development of the EVALAPPS assessment instrument. A modified Delphi process was used in order to verify the robustness of the criteria that had been identified through a literature review and to prioritize a set of the identified criteria.

Methods: Stakeholders (n=31) were invited to participate in a 2-round Delphi process with 114 initial criteria that had been identified from the literature. In round 1, participants rated criteria according to relevance on a scale from 0 (“I suggest this criterion is excluded”) to 5 (“This criterion is extremely relevant”). A criterion was accepted if the median rating was 4 or higher and if the relative intraquartile range was equal to 0.67 or lower. In round 2, participants were asked about criteria that had been discarded in round 1. A prioritization strategy was used to identify crucial criteria according to (1) the importance attributed by participants (criteria with a mean rating of 4.00 or higher), (2) the level of consensus (criteria with a score of 4 or 5 by at least 80% of the participants).

Results: The response rate was 83.9% (26/31) in round 1 and 90.3% (28/31) in round 2. A total of 107 out of 114 criteria (93.9%) were accepted by consensus—105 criteria in round 1 and 2 criteria in round 2. After prioritization, 53 criteria were deemed

crucial. These related mainly to the dimensions of security and privacy (13/53, 24.5%) and usability (9/53, 17.0%), followed by activity data (5/53, 9.4%), clinical effectiveness (5/53, 9.4%), and reliability (5/53, 9.4%).

Conclusions: Results confirmed the robustness of the criteria that were identified, with those relating to security and privacy being deemed most relevant by stakeholders. Additionally, a specific set of criteria based on health indicators (activity data, physical state data, and personal data) was also prioritized.

(JMIR Mhealth Uhealth 2020;8(7):e16899) doi:10.2196/16899

KEYWORDS

mHealth; technology assessment; obesity; overweight; Delphi technique; consensus

Introduction

The prevalence of obesity and being overweight has nearly tripled over the last 30 years and appears likely to continue increasing in the near future [1]. Obesity and being overweight are considered risk factors for type 2 diabetes, cardiovascular diseases, musculoskeletal disorders, and some cancers [2]. As several factors are known to influence being overweight and obesity, prevention and treatment also require a multifactorial approach. One of the main strategies used to reduce this prevalence is the promotion of healthy habits, mainly through diet and exercise plans, which can be reinforced with psychological therapy and behavior change strategies [3].

Over recent years, the health sector has witnessed the development and expansion of health-related mobile apps [4]. It has been estimated that over 325,000 health apps were on the market in 2017 [5]. Such ubiquity can lead to an uncritical, implicit trust in apps—*apptimism* [6]—however, the majority of these apps are rarely downloaded, and their efficacy is not always evident [7]. There has been mixed evidence on whether mHealth apps improve long-term health and well-being [4]. While some apps have demonstrated efficacy in definitive trials, others have performed poorly [8].

A number of studies [9-14] have attempted to identify why health apps are not attaining their goal. Poor quality, a lack of guidance on an app's usefulness, and low levels of support or lack of engagement from health professionals appear to be the most significant factors [9]. The assessment, validation, evaluation, and certification of health apps is a controversial topic among various stakeholders, and some guidelines and frameworks have already been published [10-14]; however, there is a lack of specific instruments to accurately assess health-related apps.

In recent years, there has been a rise in the use of apps intended to prevent weight problems or treat adults who are overweight or obese; these apps facilitate the tracking of physical and dietary patterns, provide recommendations and advice, or include motivational strategies to achieve personalized goals; however, evidence supporting the criteria used to assess the efficacy of these apps is scarce [3,4,15].

The main objective of the EVALAPPS project is to design and validate an instrument to assess the efficacy, safety, and potential effectiveness of health-related apps that are intended to manage weight and prevent obesity. A systematic review [16] to identify efficacy criteria used in previous validation studies has been carried out and a set of criteria were identified. These criteria, together with those defined by recently published frameworks for mHealth assessment, require validation and prioritization in order for the EVALAPPS instrument to be designed.

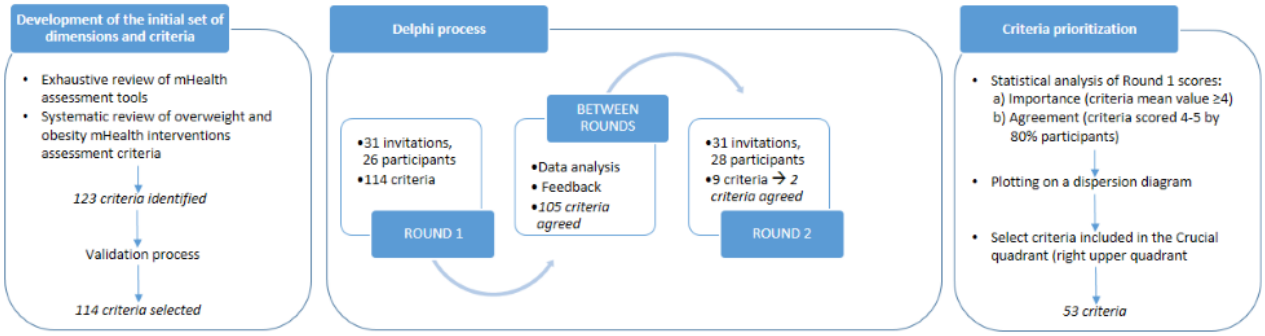
This paper focuses on the process of reaching a consensus among a broad group of stakeholders for a comprehensive set of criteria to be used in the EVALAPPS assessment instrument. First, the robustness of criteria that had been identified through literature review was validated. Second, criteria were prioritized using a quantitative approach based on the importance decided by stakeholders and the level of consensus among the stakeholders.

Methods

Study Design

A modified Delphi process was used with the domains and criteria that had been identified in a systematic review [16] in order to reach a consensus on criteria to be included in the EVALAPPS assessment instrument. The Delphi process is used to achieve expert consensus on a specific theme by voting and providing feedback through several consultation rounds [17]. In this study, the modified Delphi process consisted of an asynchronous online version of the original process. This modification provided the opportunity both to include experts from various geographical locations and to avoid the possibility of the influence of reputation or personality on the discussion. The methodological workflow is shown in Figure 1.

Figure 1. Methodological workflow.



Participant Recruitment

The scientific committee of the EVALAPPS project conducted a snowball sampling procedure to identify potential participants for the modified Delphi process. Snowball sampling is a nonprobability sampling technique in which existing participants recruit from their own contacts [18]. This sampling method is used in qualitative studies intended to reach a consensus and when statistical significance is not required. The initial sample for this study was selected by the research team. A total of 31 participants were selected using their profiles to ensure a representative panel of experts according to areas of expertise (physicians, health professionals, health managers and planners, technical experts such as developers, information and communication technology managers and digital entrepreneurs, and patients) and to ensure gender and geographic diversity. These experts were invited to participate in the modified Delphi process via email. Those who accepted were sent an email with a link to the round 1 survey.

Initial Dimensions and Criteria

Development of the initial set of dimensions and criteria was based on (1) a review of criteria used by several mHealth assessment tools that was conducted through database searches—PubMed, PsycINFO, Scopus, Trip Medical Database, Clinical Trials Register, and Cochrane—and complemented with a manual search (up to May 2018) (Multimedia Appendix 1) and (2) evidence gathered by the EVALAPPS team through a systematic review [16] to identify efficacy, safety, and potential effectiveness criteria used to assess weight, overweight, and obesity management in mHealth interventions.

A total of 123 criteria were identified and classified according to purpose of the app (monitoring, treatment, or guideline), safety and privacy, clinical effectiveness, reliability (quality of contents), usability, functionality (browsing), level of development (interoperability), and 3 health indicators—personal data, physical state data, and activity data (Multimedia Appendix 2).

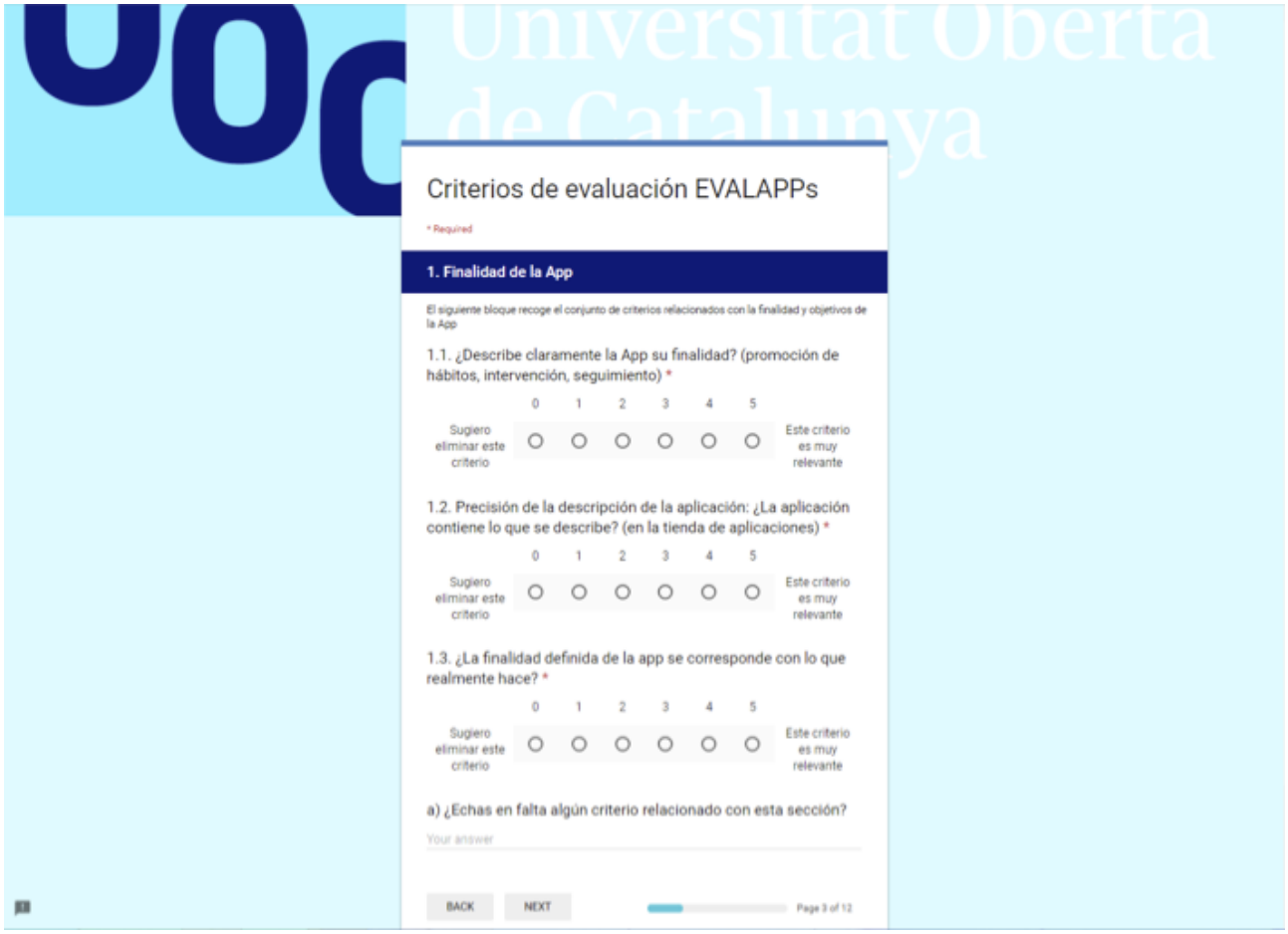
Based on the 123 criteria and 10 dimensions, a pilot survey was designed by a researcher unconnected with the EVALAPPS project. This survey was verified through an iterative (2 rounds) internal validation process by the research team, who reviewed the criteria and proposed corrections and clarifications. After the internal validation process, 114 criteria were selected for the Delphi process—3 in purpose of the app, 24 in safety and privacy, 11 in clinical effectiveness, 8 in reliability, 17 in usability, 9 in functionality, 5 in level of development, 6 in personal data, 15 in physical state data, and 16 in activity data (Multimedia Appendix 3).

Round 1

The first round of the modified Delphi took place from April 14, 2018 to April 21, 2018. The 31 experts that had been selected received an email with information about the project, the objectives of the modified Delphi, and a link to the survey.

The survey was created using Google Forms and contained (1) a questionnaire to gather sociodemographic data about participants (age, gender, professional profile, and degree of expertise in mHealth) and (2) the set of criteria. Criteria were presented clustered by dimensions (Figure 2). Each criterion was polled according to its relevance using a 6-point Likert scale with extremes labeled (0, “I suggest this criterion is excluded”; 5, “This criterion is extremely relevant”). The Google Form included an open-ended blank space at the end of each dimension in which experts could post comments, provide additional information (eg, propose new criteria), and make clarifications. A reminder was sent 2 days before the deadline to those who had not completed the survey.

Figure 2. Screen capture of the online survey for round 1.



Between Rounds

Data obtained in round 1 were collected and analyzed. The participation percentage was calculated and a descriptive analysis of the sample of experts was performed. For each criterion, the mean, standard deviation, median, interquartile range, and relative interquartile range of the ratings were calculated. SPSS software (version 21.0; IBM Corp) was used for statistical analysis. A criterion was accepted if the median sample rating was 4 or higher and the relative interquartile range was 0.67 or lower. Summaries of open-ended answers, including suggestions for changes or additional criteria, were presented to the research team before round 2.

The criteria that did not reach the set agreement level were included in a second round of the modified Delphi process. A new survey was designed with the objective of determining the validity of these criteria. Results obtained in round 1 were shared with experts via email, along with a link to the round 2 survey.

Round 2

Round 2 of the modified Delphi was carried out from May 5, 2018 to May 14, 2018. The survey was created using Google Forms and was distributed to the same experts (n=31) (Figure 3).

The criteria that were discarded in round 1 were presented by dimension. Participants were asked whether they agreed or disagreed with the mean value obtained in round 1 for those criteria. If they disagreed, they were requested to provide a value for the criterion (using the same 6-point Likert scale as in round 1) and their reasoning (open-ended answer). No sociodemographic data were collected in this survey. A gentle reminder was sent 2 days before the deadline to those who had not completed the survey.

For each criterion, the mean, standard deviation, median, interquartile range, relative interquartile range, and also the relative frequency (number of responses that agreed with the round 1 mean/total number of responses) of the ratings were calculated for round 2. A criterion was accepted if the relative frequency of response was equal or greater than 80%. The research team then reviewed all open-ended answers.

Figure 3. Screen capture of the online survey for round 2.

Prioritization

The scores obtained in round 1 were used to prioritize the criteria. Crucial criteria were defined in accordance with Ruiz Olabuénaga et al [19]. Criteria were plotted on a dispersion diagram by (1) the importance and (2) by the level of consensus and were classified as crucial, critical, circumstantial, marginal, or irrelevant according to their position with respect to quadrants defined by importance and level of consensus thresholds, respectively: (1) a mean rating of 4.00 or higher and (2) a level of consensus of at least 80% (ie, at least 80% rated the criteria as either 4 or 5).

Ethics

The EVALAPPS project was approved by the ethics committee of the Universitat Oberta de Catalunya. An informed consent statement including a brief description of the project and the conditions of participation (voluntary participation, confidentiality, and data privacy) was provided for participants to sign prior to participation in each round.

Results

Expert Panel

Of the 31 experts invited to participate, 26 accepted (83.9% response rate) in round 1 and 28 accepted (90.3% response rate) in round 2. Round 1 respondents ranged in age from 31 to 70 years; 66.7% (17/26) were men and 33.3% (9/26) were women. Most participants were clinicians (16/26, 61.5%), followed by researchers (3/26, 11.5%). Most respondents in round 1 (17/26, 65.3%) identified themselves as expert or very expert (4 or 5) with respect to mHealth apps.

Table 1. Expert panel round 1.

Characteristics	Respondents (n=26), n (%)
Gender	
Male	17 (66.7)
Female	9 (33.3)
Age	
<20	0 (0.0)
20-30	0 (0.0)
31-40	7 (26.9)
41-50	8 (30.8)
51-60	7 (26.9)
61-70	4 (15.4)
>70	0 (0.0)
Respondent profile	
Clinicians	16 (61.5)
University or research centers	3 (11.5)
Technology-related position	2 (7.7)
Consultant	2 (7.7)
Other[a]	3 (11.5)
Self-reported knowledge of health apps	
1 (low)	1 (3.8)
2	3 (11.5)
3	5 (19.2)
4	14 (53.8)
5 (very expert)	3 (11.5)

[a]Included 1 expert from an insurance enterprise, 1 person working in a government institution, and 1 retired civil servant.

Round 1

Round 1 voting included 114 potential criteria of which 105 (92.1%) were deemed sufficiently relevant to be included (according to the inclusion thresholds) and 9 (7.9%) whose relevance were considered doubtful. Table 2 shows which criteria did not meet the importance or level of consensus inclusion thresholds during round 1. Respondents (11/26, 42.3%) provided additional comments, observations, and clarifications to the criteria during the round 1. In addition to the open-ended answers, 6 respondents (19%) also suggested additional criteria. The proposed criteria included use of inclusive language, information about risk factors, information about emotional well-being, ability to record diet type (vegan, Mediterranean, etc), functionality to export data in an open format, and distinction between use for kids or adolescents and adults.

Table 2. Criteria whose ratings did not reach the inclusion thresholds in round 1.

Dimension and criteria	Mean (SD)	Median (IQR)	Relative IQR
Security and privacy			
Are cookies added to the device?	2.5 (1.45)	3 (2)	0.67
Clinical effectiveness			
Does the app appear valued with at least a reasonable value in the app store, website, etc)?	3.2 (1.02)	3 (1)	0.33
Does the app avoid the use of logos or other elements that may lead to a conflict of interest?	3.0 (1.46)	3 (2)	0.67
Usability			
Does the app provide information about long-term use?	3.3 (1.08)	3 (1)	0.33
Does the app inform users about possible malfunctions?	3.4 (1.17)	3 (1)	0.33
Does the app include use options for left-handed people?	2.8 (1.43)	3 (2)	0.67
Functionality			
Does the app always need to use an active internet connection?	2.5 (1.33)	3 (2)	0.67
Personal data[a]			
Does the app contain options to record the user's family health history?	3.3 (1.35)	3 (2)	0.67
Physical state data[a]			
In the event that the app contains options to record the height of the user, can it be done progressively over time?	2.8 (1.77)	3 (3)	1.00
Does the app contain options to record the user's resting pulse?	3.3 (1.57)	3.5 (3)	0.86
Activity data[a]			
Does the app contain options to record the user's consumption of other toxins?	3.3 (1.65)	4 (3)	0.75
Does the app contain options for recording the quality of the user's sleep?	3.2 (1.26)	3 (2)	0.67

[a]These dimensions are health indicators.

Round 2

Participants were asked to rerate the 9 criteria whose initial ratings did not reach the inclusion thresholds in round 1 (Table 2). Of the 7 criteria, 2 were included as a result of round 2—"Does the app contain options for recording the quality of the user's sleep?" and "Does the app contain options to register substance abuse?"; the others were rejected with each receiving an opinion from at least one expert about why it should be rejected. A third round of voting was not required.

Priority Criteria

A total of 63 criteria were obtained based on importance (mean rating≥4.00), and a total of 56 criteria were obtained based on level of consensus (at least 80% of ratings≥4). Using either prioritization strategy, these criteria included 14 in the safety and privacy dimension (importance: 14/63, 22.2%; level of consensus: 14/56, 25.0%) and 9 in the usability dimension (importance: 9/63, 14.3.2%; level of consensus: 9/56, 16.0%). When using importance, criteria in activity data (7/63, 11.1%), physical state data (7/63, 11.1%), and reliability (7/63, 11.1%) dimensions were also prioritized, and when using level of consensus, clinical effectiveness (6/56, 10.7%) was prioritized (Table 3).

The top 10 criteria are presented prioritized according to importance (Table 4) and level of consensus (Table 5). Two criteria were found to be high priorities using both methods of prioritization: "Are health recommendations offered by the app based on data collected in accordance with scientific evidence?" (importance: mean 4.77, SD 0.51; level of consensus: 25/26, 96.2%) and "Does the app correctly manage access to personal information through prior approval by the user?" (importance: mean 4.62, SD 0.50; level of consensus: 26/26, 100%). "At the request of the owner, can the supplier delete the app and any related data in the tracking system and documentation of access to the data to avoid any unauthorized access to personal data?" was the only other criteria to have unanimous consensus (26/26, 100%). The dimension with the most criteria prioritized in the top 10 was safety and privacy (importance: 3 criteria; level of consensus: 4 criteria).

When criteria were plotted on a dispersion diagram (Figure 4), there were 53 criteria classified as crucial (Multimedia Appendix 3), mostly from the safety and privacy (13/53, 24.5%) and usability (9/53, 17.0%) dimensions. Dimensions maintained the same proportion of criteria after prioritization (Table 3).

Table 3. Number and percentage of criteria (by dimension) that were prioritized with each strategy.

Dimensions	Criteria, n (%)	Importance, n (%)	Consensus, n (%)	Crucial, n (%)
All	114 (100)	63 (100)	56 (100)	53 (100)
Purpose of the app	3 (2.6)	3 (4.8)	3 (5.4)	3 (5.7)
Safety and privacy	24 (21.1)	14 (22.2)	14 (25.0)	13 (24.5)
Clinical effectiveness	11 (9.6)	5 (7.9)	6 (10.7)	5 (9.4)
Reliability	8 (7.0)	7 (11.1)	5 (8.9)	5 (9.4)
Usability	17 (14.9)	9 (14.3)	9 (16.0)	9 (17.0)
Functionality	9 (7.9)	3 (4.8)	3 (5.4)	3 (5.7)
Level of development	5 (4.4)	4 (6.3)	4 (7.1)	4 (7.5)
Health indicator: personal data	6 (5.3)	4 (6.3)	2 (3.6)	2 (3.8)
Health indicator: physical state data	15 (13.2)	7 (11.1)	4 (7.1)	4 (7.5)
Health indicator: activity data	16 (14.0)	7 (11.1)	6 (10.7)	5 (9.4)

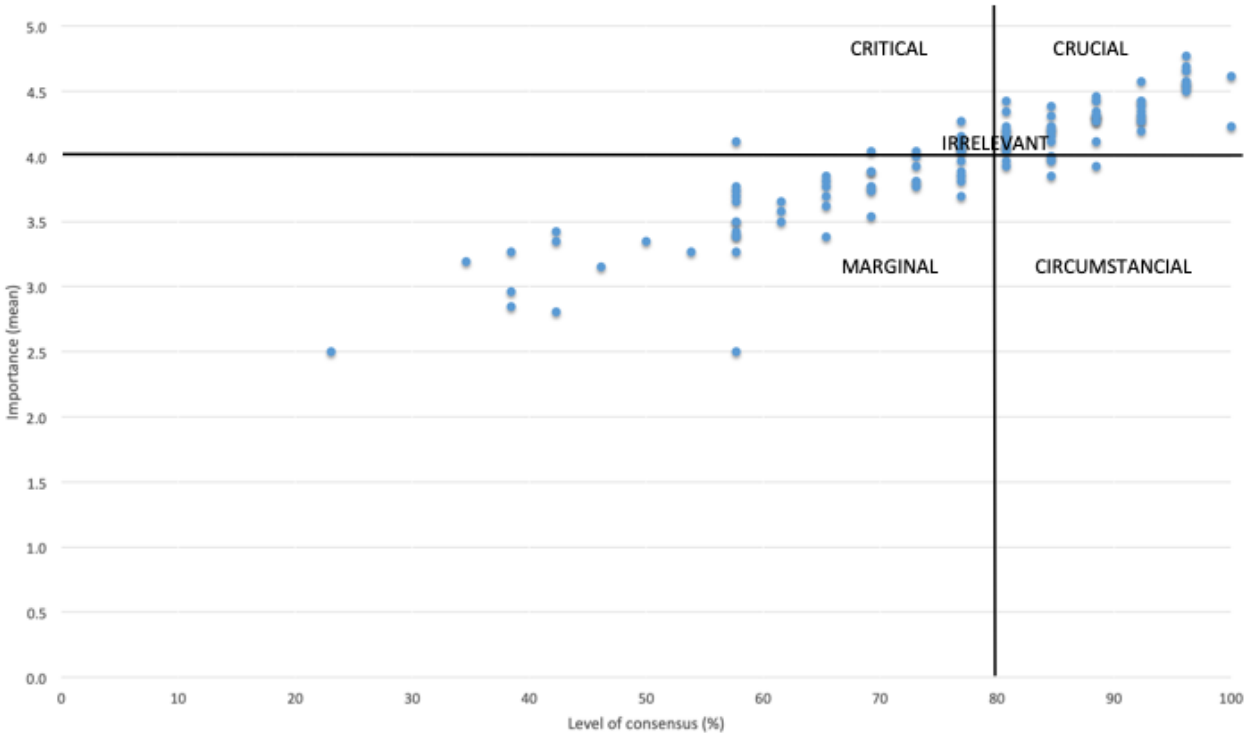
Table 4. The top 10 criteria according to importance.

Criteria	Dimension	Rating, mean (SD)
Are health recommendations offered by the app based on data collected in accordance with scientific evidence?	Clinical effectiveness	4.77 (0.51)
Is the app available on iOS and Android?	Level of development	4.69 (0.68)
Does the defined purpose of the app correspond to what it actually does?	Purpose of the app	4.65 (0.56)
Does the app correctly manage access to personal information through prior approval by the user?	Safety and privacy	4.62 (0.50)
Does the app clearly describe its purpose?	Purpose of the app	4.58 (0.58)
Does the app have a friendly and intuitive interface?	Usability	4.58 (0.58)
Is the app easy to use?	Functionality	4.58 (0.58)
Does the app have a privacy policy?	Safety and privacy	4.54 (0.58)
Is the content of the app correct, well written and relevant to the objective?	Reliability	4.54 (0.58)
Can the user choose not to participate in data transfer?	Safety and privacy	4.50 (1.03)

Table 5. The top 10 criteria according to level of consensus.

Criteria	Dimension	Ratings≥4 (n=26), n (%)
Does the app correctly manage access to personal information through prior approval by the user?	Safety and privacy	(26) 100
At the request of the owner, can the supplier delete the app and any related data in the tracking system and documentation of access to the data to avoid any unauthorized access to personal data?	Safety and privacy	(26) 100
Does the defined purpose of the app correspond to what it actually does?	Purpose of the app	(25) 96.2
Does the app have a privacy policy?	Safety and privacy	(25) 96.2
Can the user choose not to participate in data transfer?	Safety and privacy	(25) 96.2
Are health recommendations offered by the app based on data collected in accordance with scientific evidence?	Clinical effectiveness	(25) 96.2
Is the content of the app correct, well written and relevant to the objective?	Reliability	(25) 96.2
Does the app have a friendly and intuitive interface?	Usability	(25) 96.2
Is the app easy to use?	Functionality	(25) 96.2
Is the app available on iOS and Android?	Level of development	(25) 96.2

Figure 4. Criteria categorization according to their importance and level of consensus.



Discussion

Principal Findings

The EVALAPPS project was established to design and validate an instrument for assessing the efficacy, security, and potential effectiveness of mHealth apps intended to prevent and manage overweight conditions and obesity. The aim of this paper was to achieve consensus among a broad group of Spanish stakeholders on a comprehensive set of criteria, which will be used to guide development of the EVALAPPS assessment instrument. The criteria identified here were not only based on those found in literature, but have also been approved by major stakeholders, such as clinicians and app developers, using a modified Delphi process. The Delphi process was considered the most appropriate method for gathering information from experts in diverse locations in a valid and robust manner, while also avoiding individual influence and dominance. After 2 rounds and in combination with a statistical approach, the Delphi process both confirmed the robustness of the criteria that were identified (107/114, 94% of the criteria reached the consensus level) and generated a list of 53 criteria that were considered crucial for the EVALAPPS evaluation instrument (Multimedia Appendix 3). These criteria were largely from the security and privacy, usability, activity data (a health indicator), clinical effectiveness, and reliability dimensions.

Comparison With Other Studies

There have been several initiatives that have evaluated or certified health apps, in general, and thematic apps, in particular (such as nutrition or mental health–related apps).

In 2016, the World Health Organization mHealth Technical Evidence Review Group published a mobile health evidence reporting and assessment checklist [20] which included 16 items: infrastructure, technology platform, interoperability, intervention delivery, intervention content, usability testing, user feedback, access of individual participants, cost assessment, adoption inputs, limitations for delivery at scale, contextual adaptability, replicability, data security, compliance with national guidelines or regulatory statutes, and fidelity of the intervention. The National Health System of the United Kingdom [21] has also recently published a set of criteria for health app assessment, grouped in the following domains: evidence of effectiveness, regulatory approval, clinical safety, privacy and confidentiality, security, usability and accessibility, interoperability, and technical stability. The Royal College of Physicians in the United Kingdom has published a checklist relating to health app assessment [22] that considers 3 main questions: “Who developed the app and what’s inside it?”; “How well does the app work?”; and “Is there any evidence that the app does actually alleviate the problem?” In 2015, Stoyanov et al [23] published the Mobile App Rating Scale for the evaluation of health apps in general. Mobile App Rating Scale includes 23

criteria grouped in several domains: app quality, engagement, functionality, aesthetics, information, and app subjective quality. The Mobile App Rating Scale also supplies optional items that can be modified to assess knowledge, attitudes, and intention to change, but these are not included in the main Mobile App Rating Scale scoring system.

There are other assessment initiatives focused on specific thematic apps. For instance, in 2017, DiFilippo et al [24] published the Nutrition App Quality Evaluation tool to evaluate nutrition-related apps. The Nutrition App Quality Evaluation includes a set of 25 items grouped in 5 blocks: app purpose, behavior change, knowledge and skill development, app functionality, adding information about the app, and the user. The Nutrition App Quality Evaluation includes 5 additional items to assess app appropriateness for different age groups and 4 additional items to assess apps for specific audiences. For mental health apps, the American Psychiatric Association [25] has proposed an app evaluation model based on 5 steps: gather information, risk/privacy and security, evidence, ease of use, and interoperability.

It is worth noting that the domains and criteria common to most of these initiatives largely relate to information about the app and the target users, content of the intervention, evidence, technological issues, privacy and security, interoperability and usability, and user experience–related issues. These criteria are aligned with domains proposed by international evaluation initiatives such as the European Network on Health Technology Assessment and its core model for the evaluation of health technology [26].

The criteria subjected to the Delphi process were those used in the initiatives and assessment models above, thus the final set of criteria for EVALAPPS aligns with current scholarship in the field. App quality, in terms of ensuring the security and privacy of user data (13/53, 24.5%) and usability (9/53, 17.0%), was considered of higher importance than clinical aspects, both in the literature review and by our professional panel. These results are surprising considering that clinicians formed the majority of the expert panel. Several authors have pointed out that assessing the clinical aspects of health apps is an immense challenge and at times this aspect is not considered or not sufficiently assessed [27]. This could be attributed to several factors, such as lack of a systematic assessment process for evaluating clinical aspects of health apps [27] and the fact that traditional methodologies for demonstrating clinical efficacy are inadequate in the constantly changing and evolving field of mHealth [28], in which technology and research develop independently [29]. Future challenges include the development of new assessment strategies for demonstrating clinical efficacy and achieving a balance in evaluation protocols between the potential effectiveness of health apps and technology-related aspects, such as usability and security.

Finally, this study revealed the relevance of a new dimension—health information collected by the app—which represented 20.7% of the final criteria set. To our knowledge, this is the first model to include criteria specifically for assessing the management of obesity and overweight conditions using mobile tools.

Strengths and Limitations

Criteria were developed through a comprehensive review of evidence and consultation with a diverse expert panel using the Delphi process. A systematic literature review on domains and criteria to be included in the tool contributed to the study's validity, as did the iterative study design and the pilot survey. In fact, the iterative process of this study is one of its main strengths. In addition, open-ended answers were included in each section, thereby enabling experts to provide comments on included criteria or provide new criteria for consideration. Validity was also enhanced by including several types of profiles on the expert panel. This ensured the criteria that were approved were relevant and generalizable.

The Delphi process itself has some limitations. Based on the participation of a small number of experts, the Delphi process can be affected by research questions and panel configuration [30,31]. Respondents from our panel included 26 experts, mainly with clinical profiles. Due to the backgrounds of the experts invited to participate, some criteria may not have been properly addressed, for instance, health professionals may not have expertise on technology privacy and security issues. The majority of experts were from Catalonia, one of the 17 regions of Spain. As the Spanish Health System is organized at the regional level with each autonomous community health system having its own priorities, this panel does not fully guarantee the external validity of results. Finally, this study did not include face-to-face meetings with respondents to discuss ratings, to investigate areas of disagreements, or to gain more in-depth insight.

Future Actions and Conclusions

The findings of this study are important for both professionals and weight management–related technology users. The criteria agreed upon will be included in the EVALAPPS assessment instrument. Future measures include a cocreation session on the appearance and presentation of this instrument. This will be undertaken by the research team and technology experts, who will discuss such aspects as whether the tool should be in a web or app format and what components will be included. The instrument will then be piloted on various weight control apps. The findings of this study contribute to the literature on both mHealth evaluation and weight management apps.

Acknowledgments

All authors contributed equally. The research for this paper was fully funded by the Instituto de Salud Carlos III from the Spanish Ministry of Science, Innovation and Universities, grant number PI16/01764 co-funded by FEDER.

Conflicts of Interest

None declared.

Multimedia Appendix 1
mHealth assessment tools reviewed for the development of the initial set of dimensions and criteria.
[PDF File (Adobe PDF File), 51 KB - mhealth_v8i7e16899_app1.pdf]

Multimedia Appendix 2
Initial set of dimensions and criteria identified - pilot survey format.
[PDF File (Adobe PDF File), 669 KB - mhealth_v8i7e16899_app2.pdf]

Multimedia Appendix 3
List of criteria under consensus.
[PDF File (Adobe PDF File), 92 KB - mhealth_v8i7e16899_app3.pdf]

References

1. WHO - Obesity and overweight Internet. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2019-07-08]
2. Garg SK, Maurer H, Reed K, Selagamsetty R. Diabetes and cancer: two diseases with obesity as a common risk factor. Diabetes Obes Metab 2014 Feb;16(2):97-110 [FREE Full text] [doi: 10.1111/dom.12124] [Medline: 23668396]
3. Castelnuovo G, Manzoni GM, Pietrabissa G, Corti S, Giusti EM, Molinari E, et al. Obesity and outpatient rehabilitation using mobile technologies: the potential mHealth approach. Front Psychol 2014;5:559 [FREE Full text] [doi: 10.3389/fpsyg.2014.00559] [Medline: 24959157]
4. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
5. mHealth app developer economics 2017: current status and trends of the mHealth app market. Research2Guidance. 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-07-08]
6. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 03;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
7. mHealth App Developer Economics 2016. The current status and trends of the mHealth app market. Research2Guidance. 2016. URL: https://research2guidance.com/product/mhealth-app-developer-economics-2016/ [accessed 2019-07-08]
8. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941. [doi: 10.2196/11941] [Medline: 30664463]
9. Main C, Moxham T, Wyatt JC, Kay J, Anderson R, Stein K. Computerised decision support systems in order communication for diagnostic, screening or monitoring test ordering: systematic reviews of the effects and cost-effectiveness of systems. Health Technol Assess 2010 Oct;14(48):1-227 [FREE Full text] [doi: 10.3310/hta14480] [Medline: 21034668]
10. Rapid literature review for National Information Board workstream 1.2. Royal College of Physicians. 2015 Aug. URL: http://harold.thimbleby.net/NICE/NIB%20literature%20review%20final%20report%20v096-harold%20edits.pdf [accessed 2019-07-08]
11. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth assessment: conceptualization of a global framework. JMIR Mhealth Uhealth 2017 May 02;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
12. mHealth MAPS toolkit. mHealth Assessment and Planning for Scale Internet. World Health Organization. 2015. URL: https://www.who.int/life-course/publications/mhealth-toolkit/en/ [accessed 2019-07-08]
13. Høringsnotat Forslag til selvdeklareringsordning helseapplikasjoner Internet. Direktoratet for e-helse. 2016. URL: https://ehelse.no/Documents/E-helsekunnskap/Høringsnotat- Forslag selvdeklareringsordning mobile helseapplikasjoner.pdf [accessed 2019-07-08]
14. Working group on mHealth assessment guidelines Internet. European Commission. 2017. URL: http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3390 [accessed 2019-07-08]
15. Aguilar-Martínez A, Solé-Sedeño JM, Mancebo-Moreno G, Medina FX, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. J Telemed Telecare 2014 Sep;20(6):339-349. [doi: 10.1177/1357633X14537777] [Medline: 24875928]
16. Puigdomenech Puig E, Robles N, Saigí-Rubió F, Zamora A, Moharra M, Paluzie G, et al. Assessment of the efficacy, safety, and effectiveness of weight control and obesity management mobile health interventions: systematic review. JMIR Mhealth Uhealth 2019 Oct 25;7(10):e12612 [FREE Full text] [doi: 10.2196/12612] [Medline: 31654566]

17. de Meyrick J. The Delphi method and health research. Health Education 2003 Feb;103(1):7-16. [doi: 10.1108/09654280310459112]
18. Noy C. Sampling knowledge: the hermeneutics of snowball sampling in qualitative research. International Journal of Social Research Methodology 2008 Oct;11(4):327-344. [doi: 10.1080/13645570701401305]
19. Ruiz Olabuénaga JI. Técnicas de triangulación y control de calidad en la investigación socioeducativa. Bilbao: Fundación Horreum Fundazioa : Mensajero; 2003.
20. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
21. Guidance. Criteria for health app assessment Internet. Public Health England. 2017. URL: https://www.gov.uk/government/publications/health-app-assessment-criteria/criteria-for-health-app-assessment [accessed 2019-07-08]
22. Wyatt JC, Thimbleby H, Rastall P, Hoogewerf J, Wooldridge D, Williams J. What makes a good clinical app? introducing the RCP health informatics unit checklist. Clinical Medicine 2015 Nov 30;15(6):519-521. [doi: 10.7861/clinmedicine.15-6-519] [Medline: 26621937]
23. 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]
24. DiFilippo KN, Huang W, Chapman-Novakofski KM. A new tool for nutrition app quality evaluation (AQEL): development, validation, and reliability testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163. [doi: 10.2196/mhealth.7441] [Medline: 29079554]
25. Mental Helath Apps Internet. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps [accessed 2019-07-08]
26. HTA Core Model Internet. European Network for Health Technology Assessment (EUnetHTA). URL: https://www.eunethta.eu/hta-core-model/ [accessed 2019-07-08]
27. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
28. 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]
29. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
30. Hsu CC, Sandford BA. The Delphi Technique: making sense of consensus. Practical Assessment, Research, and Evaluation 2007;12. [doi: 10.4018/978-1-4666-0074-4.ch011] [Medline: 10633705]
31. Donohoe HM, Needham RD. Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions. International Journal of Tourism Research 2009 Sep;11(5):415-437 [FREE Full text] [doi: 10.1002/jtr.709]

Abbreviations

mHealth: mobile health

Edited by G Eysenbach; submitted 04.11.19; peer-reviewed by E Turk, T van Mierlo; comments to author 30.11.19; revised version received 22.01.20; accepted 13.05.20; published 22.07.20.

Please cite as:
Robles N, Puigdomènech Puig E, Gómez-Calderón C, Saigí-Rubió F, Cuatrecasas Cambra G, Zamora A, Moharra M, Paluzié G, Balfegó M, Carrion C
Evaluation Criteria for Weight Management Apps: Validation Using a Modified Delphi Process
JMIR Mhealth Uhealth 2020;8(7):e16899
URL: http://mhealth.jmir.org/2020/7/e16899/
doi:10.2196/16899
PMID:32706689

Original Paper

A Mobile App to Rapidly Appraise the In-Store Food Environment: Reliability, Utility, and Construct Validity Study

Emma Joy McMahon[1], PhD; Rachael Jaenke[1], BNutrDiet; Julie Brimblecombe[1,2], PhD

[1]Wellbeing and Preventable Chronic Disease Division, Menzies School of Health Research, Charles Darwin University, Darwin, Australia

[2]Department of Nutrition, Dietetics and Food, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

Corresponding Author:
Emma Joy McMahon, PhD
Wellbeing and Preventable Chronic Disease Division
Menzies School of Health Research
Charles Darwin University
Building Red 9
Casuarina campus, Ellengowan Drive
Darwin, 0810
Australia
Phone: 61 731694202
Email: e.j.mcmahon@outlook.com

Abstract

Background: Consumer food environments are increasingly being recognized as influential determinants of food purchasing and subsequent intake and health. We developed a tool to enable efficient, but relatively comprehensive, appraisal of the in-store food environment. The Store Scout mobile app facilitates the evaluation of product (availability and range), placement (visibility, accessibility, proximity to high-traffic areas, and location relative to other products), price (price promotion), and promotion (displays and advertising) across 7 categories of food products, with appraisal given immediately as scores (0-100, where a higher score is more in line with best practice). Primary end users are public health nutritionists and nutritionists employed by store organizations; however, store managers and staff are also potential end users.

Objective: This study aims to evaluate the reliability (interrater reliability and internal consistency), utility (distribution of scores), and construct validity (score by store type) of measurements using the Store Scout mobile app.

Methods: The Store Scout mobile app was used independently by 2 surveyors to evaluate the store environment in 54 stores: 34 metropolitan stores (9 small and 11 large supermarkets, 10 convenience stores, and 4 petrol stations) in Brisbane, Australia, and 20 remote stores (19 small supermarkets and 1 petrol station) in Indigenous Australian communities in Northern Australia. The agreement between surveyors in the overall and category scores was evaluated using intraclass correlation coefficients (ICCs). Interrater reliability of measurement items was assessed using percentage agreement and the Gwet agreement coefficient (AC). Internal consistency was assessed by comparing the responses of items measuring similar aspects of the store environment. We examined the distribution of score values using boxplots and differences by store type using the Kruskal-Wallis test.

Results: The median difference in the overall score between surveyors was 4.4 (range 0.0-11.1), with an ICC of 0.954 (95% CI 0.914-0.975). Most measurement items had very good (n=74/196, 37.8%) or good (n=81/196, 41.3%) interrater reliability using the Gwet AC. A minimal inconsistency of measurement was found. Overall scores ranged from 19.2 to 81.6. There was a significant difference in score by store type (P<.001). Large Brisbane supermarkets scored highest (median 77.4, range 53.2-81.6), whereas small Brisbane supermarkets (median 63.9, range 41.0-71.3) and small remote supermarkets (median 63.8, range 56.5-74.9) scored significantly higher than Brisbane petrol stations (median 33.1, range 19.2-37.8) and convenience stores (median 39.0, range 22.4-63.8).

Conclusions: These findings suggest good reliability and internal consistency of food environment measurements using the Store Scout mobile app. We identified specific aspects that can be improved to further increase the reliability of this tool. We found a good distribution of score values and evidence that scoring could capture differences by store type in line with previous evidence, which gives an indication of construct validity. The Store Scout mobile app shows promise in its capability to measure and track the health-enabling characteristics of store environments.

(JMIR Mhealth Uhealth 2020;8(7):e16971) doi:10.2196/16971

KEYWORDS

mobile apps; reliability and validity; food; diet; environment and public health

Introduction

Background

An unhealthy diet is one of the leading risk factors for noncommunicable diseases, which are now the leading cause of preventable death worldwide [1]. Consumer food retail environments are increasingly being recognized as influential determinants of food purchasing, diet quality, and health outcomes [2-6]. They not only determine the type of foods consumers can access and the convenience and the cost of accessing them but can also influence the appeal of these foods [1]. Health-enabling food environments are those that make healthy diets available, affordable, accessible, and appealing [1].

We developed a mobile app, Store Scout, to support rapid appraisal of the in-store food environment, drawing from a review of existing tools for assessing the healthiness of food environments, evidence for best practice (for healthy food supply or diet), stakeholder consultation, and pilot testing [7]. This tool was initially developed as a paper-based instrument for use in remote stores in Australian Aboriginal and Torres Strait Islander communities [7]. We found the paper version of the tool to be suitable for use outside this setting when pilot testing in nonremote stores in Brisbane (the capital city of Queensland, Australia) and Darwin (the capital city of the Northern Territory, Australia). During pilot testing, we found that retailers in both remote and nonremote samples were frequently interested in feedback on the results of their store. The current mobile app version of this tool enables immediate appraisal of the store environment against best practice (as scores), which may be an effective mechanism for communicating with decision makers to support the implementation and monitoring of health-enabling food environments [7]. Primary end users are public health nutritionists and nutritionists employed by store organizations; however, store managers and staff also identified as potential end users during stakeholder consultation [7].

Objectives

This study aims to evaluate interrater reliability (of measurement items and scores), internal consistency, utility of scoring, and construct validity (scores by type of store) of measurements taken using the Store Scout app. Establishing the interrater reliability of store environment measurements is important to ensure consistency of measurements between different raters [8]. Measurement of internal consistency can also be used to assess the reliability of a measurement; however, it is reported for only 3.5% of studies evaluating store environment measurements using psychometric measures (whereas 16.7% report interrater reliability) [3]. Examining the distribution of scores and variation by store type will be useful in exploring the utility of the scoring system in distinguishing between store environments. Furthermore, we can examine if differences by store type are in line with other evidence as an assessment of construct validity (the extent to which the measure is related to other constructs in an expected way [3]).

Methods

Study Sample

The Store Scout app was used to evaluate the store environment in 2 samples: (1) 20 remote indigenous community stores in the Northern Territory and Northern Queensland, Australia, as part of the Healthy Stores 2020 study (ACTRN12618001588280) [9]; and (2) 34 metropolitan stores in Brisbane, the capital city of Queensland, Australia (including supermarkets, convenience stores, and petrol stations), prioritizing those in the most disadvantaged areas of Brisbane for better comparability with the remote sample. The Australian Bureau of Statistics 2016 Index of Relative Socio-economic Advantage and Disadvantage (IRSAD), an ordinal index score derived from social and economic variables from the 2016 national Australian census (eg, income, employment status, education level, disability and health conditions, and household overcrowding [10]), was used to identify Brisbane suburbs (statistical area level 2) with the highest levels of disadvantage (lowest IRSAD score [11]).

Approximately 19% of Aboriginal and Torres Strait Islander peoples live in remote or very remote areas of Australia in small towns, commonly referred to as communities and/or homelands [12]. These communities vary in size, with most having less than 1000 people. In remote and very remote Aboriginal and Torres Strait Islander communities, most food is sourced from the local community retail store [13]. These stores are small- to medium-sized retail businesses that sell a wide range of food products as well as other household items, and therefore, they can be classified as (small or medium) supermarkets. The Healthy Store 2020 study was conducted in collaboration with the Arnhem Land Progress Aboriginal Corporation (ALPA), one of the largest remote retail store organizations and employers of Aboriginal and Torres Strait Islander people in Australia. This study included 20 stores in remote communities in the Northern Territory and in the Torres Strait Island and Cape York regions of far north Queensland. These communities had a median population of 740 (range 220-2560), and a median 93% of the population identified as Aboriginal or Torres Strait Islander (range 69%-97%) [12]. Further details of this study, which examined the effectiveness of a merchandising strategy to reduce discretionary food purchasing, are available in the study protocol [9].

Brisbane is the capital city of Queensland, Australia, and the third most populous city in Australia. We aimed for a sample of 30 to 50 Brisbane food stores. Food stores were identified in each suburb using web-based maps and business listings, beginning with the most disadvantaged suburb (lowest ISRAD score [11]) and continuing through the suburbs (in order of ascending ISRAD score) until at least 50 stores that appeared to meet the inclusion criteria were identified. The inclusion criteria were Brisbane stores likely to sell foods from all the five food groups as defined by the Australian dietary guidelines (fruit; vegetables; dairy; meat, fish and eggs; and breads and cereals [14]) and a range of both ready-to-eat and

non–ready-to-eat food items (including supermarkets, convenience stores, and some petrol stations and excluding cafes and restaurants that usually sell only ready-to-eat items). Exclusion criteria were stores that were not easily accessible by car (required boat access and >1-hour travel time from central Brisbane), in areas surveyed during paper tool development (South Brisbane areas), and specialty food stores.

Using these methods, 50 stores from 10 suburbs in North, East, and West Brisbane were identified. No stores were in the inner-city suburbs (because of these having lower IRSAD scores; ie, less disadvantage). All 50 stores were visited, although 3 were found to be closed (permanently), 2 did not meet the inclusion criteria (1 store had been converted to a café and 1 store sold spices only), and in 1 store, surveyors reported that the app was not loading. Of the remaining 44 eligible stores, 34 provided verbal consent and 10 did not provide consent (no one available with the appropriate authority to consent or language barrier), giving a 77% participation rate.

Although the 34 Brisbane stores surveyed were in some of the most disadvantaged suburbs of Brisbane, these areas were less disadvantaged relative to the areas the remote stores were located. Of the 20 remote stores, 19 were in the first decile (most disadvantaged) areas of Australia, whereas most Brisbane stores were in the fifth (23/34) and sixth (9/34) deciles (as reported by the Australian Bureau of Statistics deciles of the 2016 ISRAD by statistical area level 2 [11]). The remaining 2 Brisbane stores were in the third and fourth deciles, whereas the remaining remote store was in the third decile [11].

Data Collection

The Store Scout app includes measurement items across 7 sections, one for each of the 7 food and drink categories with questions related to product (availability and product range), placement (visibility, accessibility, proximity to high-traffic areas, and location relative to other products), price (price promotion), and promotion (displays, advertising, or activity) for types of products in each category, including healthier and less healthy options. Definitions related to product type or store environment elements are provided in the tooltips for each measurement item. Healthiness definitions used nutrient criteria (eg, >15 g fiber per 100 g) or descriptive terms (eg, lean meat) based on other nutrition resources or guidelines used in Australia [14-16]. There is an additional section of store manager questions and suggested strategies, but these do not affect scoring and were not collected in this study.

Store Scout scoring is calculated from up to 199 yes or no measurement items, each of which is worth 1 possible point. As 6 scoring measurement items are hidden if the response to a previous measurement item was *no*, there are up to 193 to 199 possible points across the 7 categories (breads and cereals: 20, dairy and eggs: 29, drinks: 28, fruits and vegetables: 30, meals and convenience foods: 24-30, meat and seafood: 21, and snack foods: 41). Practices related to both healthy products and unhealthy (or less healthy) products contribute to scoring. A point is awarded for a *yes* response to practices likely to encourage purchases of healthy products (eg *Fresh fruit at or near checkout*) or discourage purchases of unhealthy or less healthy products (eg *Limited shelf space for unhealthy products*) or for a *no* response for practices likely to encourage purchases of unhealthy (or less healthy) products (eg, *Lollies* [confectionery], *chocolate & chips at or near checkouts*). Scores (range 0-100, a higher score indicates a more *health-enabling* store) are calculated for each category as a percentage of total possible points averaged across the 7 categories to give an overall score.

Data are transferred to a web-based portal once the user connects to the internet. The web-based portal was under development during this study, and there were intermittent issues with data loss during transfer. Where these issues were known ahead of time, screenshots of completed data screens were taken as backup, and data were entered manually; however, some data were missing despite this (Multimedia Appendix 1). Category scores and overall scores were calculated (as described earlier) only for complete data (ie, where there were no missing data).

A total of 68 surveys were conducted in 34 Brisbane food stores in April 2018. In each store, 2 surveyors with nutrition expertise (final year Bachelor of Nutrition and Dietetics students) who had no previous experience with the Store Scout tool before this project used the Store Scout app to simultaneously, but independently, measure the in-store environment. Data collection occurred in 2 stages. Initially, surveyors were given training only on the basics of how to use the app. After 17 stores had been surveyed (stage 1), the surveyors and 2 members of the research team discussed specific questions where agreement was lowest to inform training materials and identify where tooltips could be improved. This discussion mostly related to differences in inclusion or exclusion of products into the specific categories (especially for meals and convenience foods); classification of products as healthy or less healthy; definition of promotional materials; and definitions of key terms such as *high traffic areas*, *convenience meal*, and *limited range*. The 2 surveyors then surveyed the remaining 17 stores (stage 2). Surveyors were asked to count the number of registers as a surrogate for store size.

In the remote store sample, measurement of the store environment using the Store Scout app occurred between September 2018 and May 2019 at the end of each study period: baseline, intervention, and postintervention. Measurements were taken by a single surveyor except for at the end of the intervention period (November 2018 to December 2018), where 2 surveyors independently measured the store environment to enable the assessment of interrater reliability. Surveyors included members of the research team and public health nutritionists, most of whom had no previous experience in using the tool besides a short (approximately 30-60 min) training session.

This project was granted ethics approval by the Top End Human Research Ethics Committee (HREC 2017-2820 and HREC-2018-3048) and the Far North Queensland Human Research Ethics Committee (HREC-18-QCH-23-1211).

Statistical Analysis

We assessed interrater reliability (of measurement items and scores), internal consistency, utility of scoring, and construct validity (scores by store type). The data included depended on

the aim of the analysis and unit of analysis (Multimedia Appendix 2). All analyses were performed using Stata Statistical Software, version 16.0 (StataCorp) [17].

We assessed the interrater reliability of the 196 core (always shown) yes or no measurement items using percentage agreement and the Gwet agreement coefficient (AC) [18]. We also reported kappa statistics for comparability with other tools, although these are less suitable because of their dependence on the homogeneity of marginal distributions [8,19,20]. Coefficients were interpreted using the following cutoffs: slight or poor (<0.2), fair (0.21-0.4), moderate (0.41-0.6), good (0.61-0.8), and very good (0.81-1) [21].

We assessed the interrater reliability of scores (overall and by category) using intraclass correlation coefficients (ICCs) and median and range of the difference between scores (as the difference between scores was not normally distributed). The ICC model and estimates and their 95% CIs were calculated using a one-way random effects model based on an individual rater, as recommended in the guideline by Koo and Li [22] for reliability experiments where subjects are assessed using different sets of raters and the measurement will be taken by a single rater in usual practice (ie, outside of a reliability experiment). ICC values were interpreted using the following cutoffs: poor (<0.5), moderate (0.5-0.75), good (0.75-0.9), and excellent (>0.90) [22].

We did not examine interrater reliability by surveyor group or sample because of the reduced sample size and differential variance in these groups. Instead, we examined the level of agreement between surveyors overall (combined sample) and in each surveyor group or level of training (Brisbane stage 1, Brisbane stage 2, and remote stores) as percentage agreement across all measurement items and median and range of the difference between scores (as the difference between scores was not normally distributed).

Internal consistency was assessed on all surveys collected by comparing responses to related measurement items, classified as higher-order items (eg, “Is there promotional material or activity for healthier breads and cereals?”) and more specific follow-up items (eg, “Does the promotion stand out?”). Surveys where responses to both measurement items were *yes* or *no* were considered internally consistent. Surveys where there was a *no* response to the higher-order measurement item but a *yes* response to the more specific follow-up item were inconsistent. Cases where there was a *yes* response to the higher-order item and a *no* response to the more specific item were excluded (as it was not possible to assess the consistency of these). Internal consistency was calculated as the percentage of surveys where responses were consistent.

We examined the extent to which Store Scout scores resulted in a distribution of values using boxplots (of scores where an average of the 2 values was used if collected by more than one surveyor).

Difference by store type was evaluated using the Kruskal-Wallis equality-of-populations rank test with Dunn pairwise comparison of scores and correction for multiple comparisons (Holm-Bonferroni). It was expected that petrol stations and convenience stores would score lower than supermarkets [23].

Results

Store Characteristics

Table 1 shows the types of stores surveyed in each sample. A cutoff of ≤5 registers was used to classify Brisbane supermarkets as small or large to enable comparison with remote stores, which had 1-4 registers in each store. The remote store that functioned primarily as a petrol station was excluded from the analysis of score by store type.

Table 1. Store characteristics.

Store type	n	Number of registers[a], median (range)	Time taken to survey (min), median (range)
Brisbane stores			
All Brisbane stores	34	3 (1-14)	25 (10-40)
Petrol stations	4	2 (1-3)	15 (10-20)
Convenience stores	10	1 (1-3)	15 (10-30)
Supermarkets			
Both small and large supermarkets	20	6 (2-14)	30 (15-40)
Small supermarkets	9	4 (2-5)	30 (15-30)
Large supermarkets	11	9 (6-14)	30 (20-40)
Remote stores			
Both remote supermarkets and petrol stations	20	2.5 (1-4)	—[b]
Remote supermarkets	19	3 (1-4)	—
Petrol stations	1	2	—

[a]Excludes self-serve registers (only seen in large stores).

[b]Time taken to complete the surveys was not collected in the remote sample.

Interrater Reliability

Most measurement items had >80% agreement between surveyors (n=129/196, 65.8%) and very good (n=74/196, 37.8%) or good (n=81/196, 41.3%) interrater reliability using the Gwet AC (Table 2). There were 3 measurement items that had slight or poor interrater reliability; these combined a judgment of healthiness (healthier or less healthy snack or healthier meals and convenience foods) as well as a subjective judgment of placement or range (*near checkout*, *limited range*, and *same or more space*). Using kappa coefficients, most had moderate (n=72/196, 36.7%) or good or very good (n=72/196, 36.7%) interrater reliability. Of the 18 items with slight or poor interrater reliability assessed by kappa coefficients, most surveyors had >70% (n=11/18) agreement, and some surveyors had >80% (n=5/18) agreement.

Overall, the scores had excellent interrater reliability, as indicated by ICC (0.954; 95% CI 0.914-0.975). The ICC for category scores indicated excellent interrater reliability for fruits and vegetables (0.940; 95% CI 0.898-0.966); good interrater reliability for dairy and eggs (0.881; 95% CI 0.798-0.932), meat and seafood (0.875; 95% CI 0.791-0.926), drinks (0.799; 95% CI 0.671-0.881), snack foods (0.784; 95% CI 0.648-0.872), and breads and cereals (0.727; 95% CI 0.568-0.834); and moderate interrater reliability for meals and convenience foods (0.696; 95% CI 0.517-0.817).

Overall, there was 83.20% (8584/10312) agreement across all measurement items collected by 2 surveyors in the combined sample. The percentage agreement increased from Brisbane stage 1 (n=2634/3278, 80.35%) to stage 2 (n=2764/3176, 87.00%), with the greatest increase within a category being in meal and convenience foods, likely because of clarification of which products fell into this category. Agreement was 83.6% in remote stores, with all categories having >80% agreement (Multimedia Appendix 3).

The median difference in overall scores between surveyors was 4.4 in the combined sample (range 0.0-11.1), with the smallest median difference in the remote sample (Table 3). There were larger differences between surveyors in category scores than overall scores.

Table 2. Store Scout measurement items by interrater reliability category (N=196).

Category	Agreement, n (%)	Gwet agreement coefficient, n (%)	Value, n (%)
Slight or poor (<0.2)	0 (0)	3 (2)	18 (9)
Fair (0.21-0.4)	0 (0)	6 (3)	34 (17)
Moderate (0.41-0.6)	3 (2)	32 (16)	72 (37)
Good (0.61-0.8)	64 (33)	81 (41)	54 (28)
Very good (0.81-1)	129 (66)	74 (38)	18 (9)

Table 3. Score difference between surveyors by category and sample.

Score[a]	All		Brisbane stage 1		Brisbane stage 2		Healthy stores 2020	
	n	Median (range)	n	Median (range)	n	Median (range)	n	Median (range)
Overall	39	4.4 (0.0-11.1)	13	5.5 (0.4-11.1)	15	5.3 (0.7-8.7)	11	2.2 (0.0-6.9)
Breads and cereals	52	5.0 (0.0-45.0)	16	10.0 (0.0-45.0)	16	5.0 (0.0-15.0)	19	5.0 (0.0-30.0)
Dairy and eggs	47	6.9 (0.0-27.6)	15	6.9 (0.0-20.7)	15	3.5 (0.0-27.6)	17	6.9 (0.0-17.3)
Drinks	49	7.1 (0.0-25.0)	17	7.1 (0.0-25.0)	16	9.0 (0.0-25.0)	16	3.6 (0.0-17.9)
Fruits and vegetables	51	6.6 (0.0-20.0)	17	6.6 (0.0-16.7)	16	3.4 (0.0-20.0)	18	6.7 (0.0-20.0)
Meals and convenience	48	12.5 (0.0-41.7)	16	19.1 (0.0-41.7)	16	4.2 (0.0-24.1)	16	8.4 (0.0-33.4)
Meat and seafood	51	9.5 (0.0-38.1)	17	9.5 (0.0-28.6)	15	4.8 (0.0-38.1)	19	9.5 (0.0-23.8)
Snack foods	49	4.9 (0.0-39.0)	16	4.9 (0.0-21.9)	16	6.1 (0.0-39.0)	17	2.5 (0.0-21.9)

[a]Scores were calculated only for complete data (ie, where there were no missing data in the category).

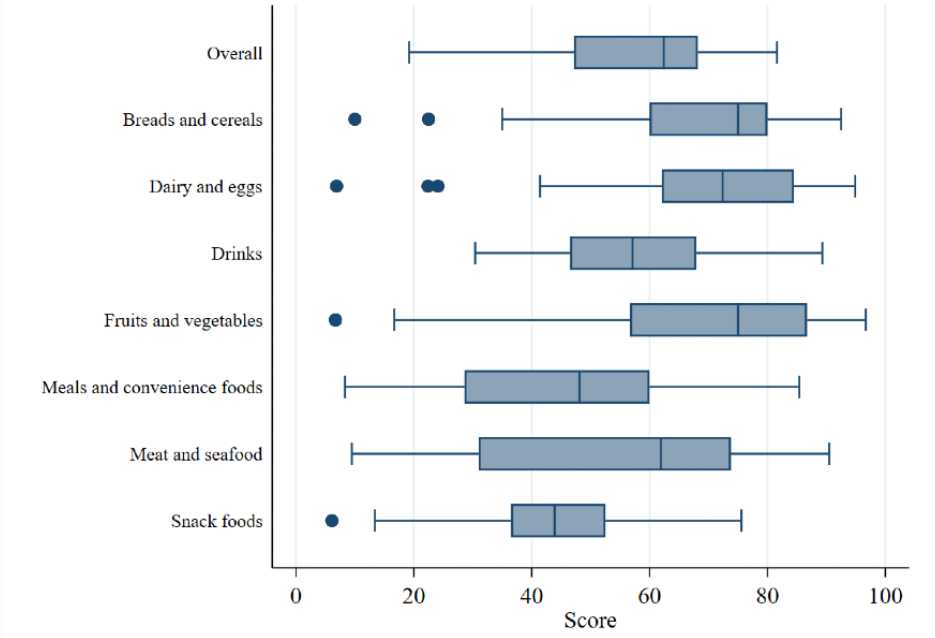
Internal Consistency

There was a small amount of inconsistency with some measurement items. In Brisbane stage 1 (n=34), all measurement items had >89% consistency, and 57% (n=20/35) of measurement items had perfect (100%) internal consistency. In Brisbane stage 2 (n=33 surveys), all measurement items had >95% consistency, and 76% (n=27/35) of measurement items had perfect internal consistency. In the remote sample (n=79 surveys), all measurement items had >98% consistency, and 74% (n=26/35) of measurement items had perfect internal consistency. Multimedia Appendix 4 shows the internal consistency percentage for each measurement item by sample.

Utility of Scoring

Figure 1 shows the distribution of overall and category scores in the 54 stores. Overall scores ranged from 19.2 to 81.6, and there was acceptable variation within each category.

Figure 1. Distribution of overall and category scores for the combined sample of 54 stores.

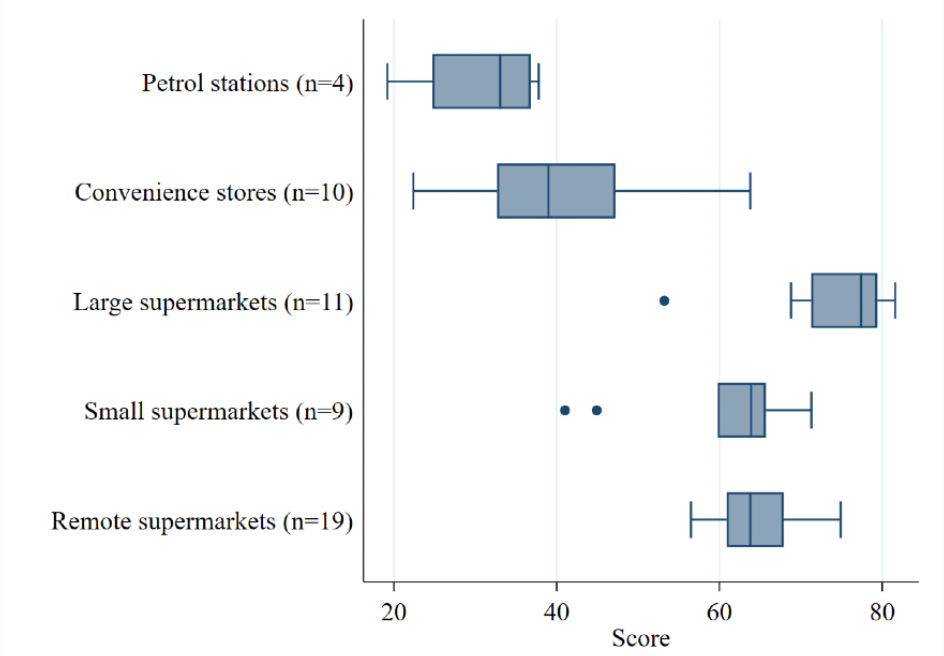


Construct Validity

Figure 2 shows the distribution of scores by store type. There was a significant difference in scores by store type (P<.001 using the Kruskal-Wallis test), with large Brisbane supermarkets scoring significantly higher (median 77.4, range 53.2-81.6) than all other types of stores, and small Brisbane supermarkets (median 63.9, range 41.0-71.3) and remote supermarkets (median 63.8, range 56.5-74.9) scoring significantly higher than Brisbane convenience stores (median 39.0, range 22.4-63.8) and Brisbane petrol stations (median 33.1, range 19.2-37.8). After adjustment for multiple comparisons using the Holm-Bonferroni method, the difference between small Brisbane supermarkets and Brisbane convenience stores or petrol stations was no longer significant (Multimedia Appendix 5).

Multimedia Appendix 6 shows the distribution of points from measurement items related to healthy products (n=156 measurement items) or unhealthy products (n=43 measurement items) by store type. Although large supermarkets had the highest points for measurement items related to healthy products (median 136, range 73-140), they scored lowest (median 14, range 9-33) after petrol stations (median 12, range 11-14) for measurement items related to unhealthy products (least in line with best practice because of more availability, prominent placement, promotion, and/or pricing promotion of unhealthy products). Remote supermarkets scored highest (median 21, range 15-29) for measurement items related to unhealthy products (most in line with best practice because of less availability, prominent placement, promotion, and/or pricing promotion of unhealthy products).

Figure 2. Distribution of overall scores by store type.



Discussion

Principal Findings

We found good to very good interrater reliability of most measurement items, high agreement between scores, and high internal consistency of measurements when the in-store food environment was evaluated using the Store Scout app. We found good variation in scores across the samples, which include a variety of types of stores, including petrol stations, convenience stores, supermarkets, and remote stores, and evidence that the scoring system could capture differences by type of store in line with previous evidence, providing evidence of construct validity.

We found excellent interrater reliability of overall scores and good to excellent interrater reliability for all category scores, except meals and convenience foods scores, which had moderate interrater reliability. Although the median difference between surveyors in the overall score was acceptable (4.4), the median difference for category scores was larger. This is because of the number of measurement items contributing to category scores (20-41 depending on category) versus overall score (193-199), meaning that a difference between surveyors in one measurement item would mean a 2.4%-5% difference in category scores but a 0.3%-0.7% difference in the overall score. This means that overall scores are more reliable than category scores when looking at change (within store) or difference (between stores) with different surveyors.

There were 3 measurement items with slight or poor interrater reliability, as assessed by the Gwet AC. These measurement items required the surveyor to make judgments about placement or product range of healthier or less healthy products. It is likely that the reliability of some measurement items could be improved by more specific guiding definitions; for example, *at or near checkout* may be better defined by a perimeter that would be considered near, and *limited range* with a guideline of how many products or product facings this would be. Assessing if there is greater shelf space dedicated to healthier versus less healthy products is probably one of the more difficult items for surveyors, especially as it often requires assessing the healthfulness of every item in the category, sometimes needing to check the nutritional information, and it can be difficult to estimate space allocation when products are often found in numerous areas of a store. In this case, it may be better to instruct surveyors to focus on the primary area where the product is located. It is likely that the interrater reliability of measurement items will be strengthened by minimizing the subjectivity of measurement items.

We found perfect internal consistency (no identified inconsistency) for most measurement items where this could be assessed. Measurement items where a small degree of inconsistency was observed tended to have less specific wording regarding the product type that was assessed. These mostly had wording such as *healthier option* rather than the specific product type (which is listed in a header above the measurement items) and were examining a specific product type within the higher-order food group category. For example, the *healthier option easy to find* was related to healthier sweet biscuits and cakes, but the surveyor may have answered this more broadly (for any healthier option in the snack foods category). Modifying this to *healthier sweet biscuits or cakes easy to find* may make this clearer.

We found an increase in agreement and internal consistency from Brisbane stage 1 to Brisbane stage 2. In Brisbane stage 1, surveyors had no specific training or preexisting familiarity with the tool and did not have previous experience in retail food environment research but still achieved an 80% agreement rate. After discussing some key definitions (stage 2), this was increased to 87%. These discussions led to improvements in definitions in the app to reduce ambiguity (mostly related to categorization of products in categories and as healthy or unhealthy options), and learning was incorporated into the training of surveyors before data collection in the remote stores.

Surveyors in the remote store sample were similar to those who would use the tool in practice in terms of having experience in retail food environment research and/or practice. Most had either not used the Store Scout tool before or had only done so once or twice, and most received only 30-60 min of training. Despite this, surveyors agreed 83% of the time, and the median difference in overall scores was only 2.2 (range 0-6.9).

Stores servicing smaller populations have reduced opportunities for economies of scale. Remote stores have additional logistical barriers to providing high-quality fresh food because of the long distance the food travels to get to the store [24,25]. High outside temperatures and poor road conditions (or road closures) can contribute to this. Therefore, remote stores have food demand and supply constraints that determine the price, quality, and variety of healthy foods that they can offer. Despite these challenges, we found that the store environment, as assessed by the Store Scout app in the 19 small remote supermarkets, scored similarly to small supermarkets in Brisbane and higher than convenience stores. When only the health-enabling practices related to unhealthy or less healthy products were considered (eg, not stocking or promoting unhealthy products or not placing unhealthy products in high-traffic areas), larger supermarkets and petrol stations scored the least points, whereas the small remote supermarkets scored the most. This is an encouraging finding; there has been considerable work aimed at improving food supply and access in the remote Australian Aboriginal and Torres Strait Islander communities where these stores are located [26-30], which has likely contributed to healthier food environments. ALPA (the organization overseeing the management of these remote stores) has demonstrated long-term commitment to promoting health and nutrition in the communities it serves, as demonstrated by their nutrition policy (first implemented in the early the 1980s), which includes strategies to increase the availability and affordability of nutritious foods [30].

Comparison With Prior Work

We included kappa values to enable comparison with evaluations of other tools and found kappa coefficients that compare reasonably well with other tool evaluations, which generally report coefficients of around 0.7 [3]. Notably, other studies that have used kappa values could not include all measurement items because of the high and/or low prevalence of store traits measured [31,32]. Overall, 2 instances of excellent interrater reliability include that found for the Nutrition Environment Measures Survey in 88 stores (>90% agreement and kappa >0.84 for all items [32]) and the Food Environment Audit for Diverse Neighborhoods in 44 stores (89% of items had kappa >0.60) [31]; however, in these studies, surveyors received considerably more training than this study (2 days and 25 hours of training, respectively) [31,32]. The Store Scout app was designed to be used not only in research but also in practice to rapidly appraise the store environment and provide immediate feedback to retailers and recommendations for improvement. To replicate what was likely to be feasible in practice, we tested interrater reliability where surveyors received no or minimal training in the use of the tool.

We found that the scoring captured differences in the store environment across store types. Supermarkets, particularly large supermarkets, scored highest, indicating the most health-enabling store environments as assessed using the Store Scout app, whereas convenience and petrol stores scored lowest. This is consistent with previous research comparing supermarkets, petrol stations, and convenience stores [23,24,33], indicating construct validity.

There is limited evidence comparing the food environment in remote versus nonremote supermarkets in Australia. Most research has focused on food prices, with food prices consistently higher in remote stores [16], whereas the Store Scout app takes a broader approach to food environment measurement and does not assess price other than price promotions. Cameron et al [34] assessed the shelf space and strategic placement of healthy and discretionary foods in urban, urban fringe, and rural or nonmetropolitan supermarkets in Victoria (from a single major Australian supermarket chain) and found that urban supermarkets had a generally healthier food environment compared with urban fringe and rural or nonmetropolitan stores; however, the urban fringe (most similar to the Brisbane sample in this study) and rural or nonmetropolitan supermarkets had similar healthy food environments overall. This is consistent with the findings in this study, but this is not a direct comparison, as the rural or nonmetropolitan centers in the Cameron study had populations >6000, more than double the largest remote community in the remote sample (median 740, range 220-2560).

Strengths and Limitations

A strength of this study was that measurements were taken as would occur in practice (in physical stores during the day when the store is open) in several different store types and by surveyors who received no or minimal training in the use of the tool. A limitation is that in the remote sample, surveyors were not able to complete surveys simultaneously, which may have reduced interrater reliability, especially for meals and convenience foods as some ready-to-eat items (eg, pies, sandwiches, and salads) may only be available at certain times of the day.

Although we were able to examine internal consistency from up to 146 surveys in 2 very different retail contexts, we could not include all response combinations; however, there were still a median of 136 responses across the 35 measurement items used to assess internal consistency (range 63-146). The issues we experienced with missing data should not be an issue in future, as the app does not allow the surveyor to progress unless all measurement items are completed. The results reported here pertain mostly to surveys completed by surveyors with qualifications related to nutrition and/or public health. We cannot determine if the same degree of reliability and agreement would be achieved if this were not the case. We did not ask surveyors in the remote sample to record the time taken to record surveys; therefore, the median time of 25 min may not be applicable to this setting.

The difference when only practices related to unhealthy (or less healthy) products were considered is important. Approximately one-third of the energy in the Australian diet comes from

discretionary (unhealthy) foods [35], and these foods are disproportionately represented on Australian food store shelves, and more likely to be promoted [36]. Although the Store Scout app assesses the store environment related to healthy and unhealthy (or less healthy) products, there are a greater number of measurement items related to healthy than unhealthy or less healthy products (156 versus 43, respectively), and practices discouraging purchases of unhealthy or less healthy products contribute only approximately 20% to the overall score. We are in the process of testing the criterion validity of the Store Scout mobile app against external measures (such as sales of healthy or unhealthy foods), and as part of this, we will investigate if higher weighting of practices related to unhealthy products leads to better validity of Store Scout scores.

Conclusions

We developed a mobile app to rapidly appraise the in-store food retail environment across a wide range of healthy and unhealthy foods. We tested the Store Scout app in a diverse sample of 54 stores and found good to very good interrater reliability of measurement items and high internal consistency. We identified areas where tooltips and wording of measurement items can be improved to further increase the reliability of this tool. Scores reflecting how health enabling the store environment is are embedded in the app, giving the user a convenient mechanism for communicating results and initiating change with retailers or other stakeholders. We found high agreement between surveyors in overall scores and some category scores. This suggests that Store Scout overall scores are reliable for making comparisons across stores even when the tool is completed by different surveyors; however, category scores were less reliable. We found good distribution of scores and evidence that the scoring system can capture scoring by store type in line with previous evidence, indicating construct validity. The Store Scout mobile app shows promise in its ability to measure and track health-enabling store environments. In future projects, we will further assess the validity of this tool against external measures.

Acknowledgments

The authors would like to acknowledge the participating stores; ALPA; the Healthy Stores 2020 project team including Anthony Gunther (project manager), Emma Chappell (research assistant) and Khia De Silva (ALPA nutritionist); the public health nutritionists who assisted with the Healthy Stores 2020 Store Scout data collection; and Ana Freeman and Fleur Lesslie (Griffith University) for Brisbane data collection. The information and opinions contained in this paper do not necessarily reflect the views or policies of the supporting organizations. EM was supported by a cofunded National Heart and Medical Research Council (NHMRC) and the Australian Heart Foundation Early Career Fellowship (100085). JB was supported by an NHMRC Translating Research into Practice Fellowship (1168333). Healthy Stores 2020 was funded by an NHMRC project grant (1138629). Store Scout tool development was supported by funding from the National Health and Medical Research Council of Australia (ID631947). EM and JB are researchers in an NHMRC-funded Center of Research Excellence in Food Retail Environments for Health (APP1152968). These funding bodies did not have any role in the study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

Authors' Contributions

RJ and JB led the development of the Store Scout mobile app. EM and JB coordinated data collection. EM analyzed data. All authors interpreted the results. EM drafted the manuscript. All authors revised the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Numbers of surveys with complete and missing data.
[DOCX File , 28 KB - mhealth_v8i7e16971_app1.docx]

Multimedia Appendix 2
Data included by analysis type.
[DOCX File , 27 KB - mhealth_v8i7e16971_app2.docx]

Multimedia Appendix 3
Percentage agreement (%) between surveyors of Store Scout measurement items.
[DOCX File , 27 KB - mhealth_v8i7e16971_app3.docx]

Multimedia Appendix 4
Internal consistency of Store Scout measurement items.
[DOCX File , 36 KB - mhealth_v8i7e16971_app4.docx]

Multimedia Appendix 5
Kruskal Wallis equality-of-populations rank test results (*p* values).
[DOCX File , 26 KB - mhealth_v8i7e16971_app5.docx]

Multimedia Appendix 6
Points for measurement items related to healthy products and unhealthy or less healthy products by store type.
[DOCX File , 32 KB - mhealth_v8i7e16971_app6.docx]

References

1. Food and Agriculture Organization. 2016. Influencing Food Environments for Healthy Diets URL: http://www.fao.org/3/a-i6484e.pdf [accessed 2020-05-05]
2. Ni Mhurchu C, Vandevijvere S, Waterlander W, Thornton LE, Kelly B, Cameron AJ, INFORMAS. Monitoring the availability of healthy and unhealthy foods and non-alcoholic beverages in community and consumer retail food environments globally. Obes Rev 2013 Oct;14(Suppl 1):108-119. [doi: 10.1111/obr.12080] [Medline: 24074215]
3. Lytle LA, Sokol RL. Measures of the food environment: a systematic review of the field, 2007-2015. Health Place 2017 Mar;44:18-34. [doi: 10.1016/j.healthplace.2016.12.007] [Medline: 28135633]
4. Adam A, Jensen JD. What is the effectiveness of obesity related interventions at retail grocery stores and supermarkets? -a systematic review. BMC Public Health 2016 Dec 28;16(1):1247 [FREE Full text] [doi: 10.1186/s12889-016-3985-x] [Medline: 28031046]
5. Hartmann-Boyce J, Bianchi F, Piernas C, Payne Riches S, Frie K, Nourse R, et al. Grocery store interventions to change food purchasing behaviors: a systematic review of randomized controlled trials. Am J Clin Nutr 2018 Jun 1;107(6):1004-1016 [FREE Full text] [doi: 10.1093/ajcn/nqy045] [Medline: 29868912]
6. Gittelsohn J, Rowan M, Gadhoke P. Interventions in small food stores to change the food environment, improve diet, and reduce risk of chronic disease. Prev Chronic Dis 2012;9:E59 [FREE Full text] [Medline: 22338599]
7. Jaenke R, van den Boogaard C, Brimblecombe J. Development and Pilot of a Tool to Measure the In-Store Food Environment. In: Dietitians Association of Australia 35th National Conference. 2018 Presented at: DAA'18; May 17-19, 2018; Sydney, Australia URL: https://tinyurl.com/y4d5jn79 [doi: 10.1111/1747-0080.12426]
8. Hernaez R. Reliability and agreement studies: a guide for clinical investigators. Gut 2015 Jul;64(7):1018-1027. [doi: 10.1136/gutjnl-2014-308619] [Medline: 25873640]
9. Brimblecombe J, Ferguson M, McMahon E, Peeters A, Miles E, Wycherley T, et al. Reducing retail merchandising of discretionary food and beverages in remote Indigenous community stores: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Mar 29;8(3):e12646 [FREE Full text] [doi: 10.2196/12646] [Medline: 30924788]
10. Australian Bureau of Statistics, Australian Government. 2018. 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016 URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~IRSAD~20 [accessed 2019-12-17]
11. Australian Bureau of Statistics, Australian Government. 2016. Table 3 Statistical Area Level 2 (SA2) Index of Relative Socio-Economic Advantage and Disadvantage, 2016 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016 [accessed 2019-11-09]
12. Australian Bureau of Statistics, Australian Government. 2018. 3238.0.55.001 - Estimates of Aboriginal and Torres Strait Islander Australians, June 2016 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/3238.0.55.001 [accessed 2019-03-02]
13. Wycherley T, Ferguson M, O'Dea K, McMahon E, Liberato S, Brimblecombe J. Store turnover as a predictor of food and beverage provider turnover and associated dietary intake estimates in very remote indigenous communities. Aust N Z J Public Health 2016 Dec;40(6):569-571. [doi: 10.1111/1753-6405.12571] [Medline: 27624683]
14. The Australian Dietary Guidelines (The Guidelines). 2013. Australian Guidelines Dietary: Eat For Health URL: https://www.eatforhealth.gov.au/guidelines [accessed 2020-05-05]
15. Australian Government Department of Health. 2014. Healthy School Canteens: Guidelines for Healthy Foods and Drinks Supplied In School Canteens URL: https://www1.health.gov.au/internet/main/publishing.nsf/content/5ffb6a30ecee9321ca257bf0001dab17/$file/canteen%20guidelines.pdf [accessed 2019-12-20]
16. NT Health Digital Library: Northern Territory Government. 2016. Northern Territory Market Basket Survey URL: https://digitallibrary.health.nt.gov.au/prodjspui/bitstream/10137/1428/1/2016%20MBS%20Full%20Report.pdf [accessed 2020-05-15]
17. StataCorp. 2019. Stata Statistical Software: Release 16 URL: https://www.stata.com/ [accessed 2020-05-18]
18. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008 May;61(Pt 1):29-48. [doi: 10.1348/000711006X126600] [Medline: 18482474]
19. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [doi: 10.1016/0895-4356(90)90158-l] [Medline: 2348207]
20. Gwet K. Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. Stat Methods Inter-Rater Reliab Assess Series 2002;2(1):9. [doi: 10.1007/springerreference_183942]

21. Gwet K. Handbook of Inter-rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Piedmont, California, United States: Advanced Analytics; 2014.
22. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
23. Black C, Ntani G, Inskip H, Cooper C, Cummins S, Moon G, et al. Measuring the healthfulness of food retail stores: variations by store type and neighbourhood deprivation. Int J Behav Nutr Phys Act 2014 May 23;11:69 [FREE Full text] [doi: 10.1186/1479-5868-11-69] [Medline: 24884529]
24. Whelan J, Millar L, Bell C, Russell C, Grainger F, Allender S, et al. You can't find healthy food in the bush: poor accessibility, availability and adequacy of food in rural Australia. Int J Environ Res Public Health 2018 Oct 21;15(10):E2316 [FREE Full text] [doi: 10.3390/ijerph15102316] [Medline: 30347893]
25. Ferguson M, O'Dea K, Chatfield M, Moodie M, Altman J, Brimblecombe J. The comparative cost of food and beverages at remote indigenous communities, northern territory, Australia. Aust N Z J Public Health 2016 Apr;40(Suppl 1):S21-S26. [doi: 10.1111/1753-6405.12370] [Medline: 25902766]
26. Ferguson M, O'Dea K, Altman J, Moodie M, Brimblecombe J. Health-promoting food pricing policies and decision-making in very remote Aboriginal and Torres Strait Islander community stores in Australia. Int J Environ Res Public Health 2018 Dec 19;15(12):E2908 [FREE Full text] [doi: 10.3390/ijerph15122908] [Medline: 30572601]
27. Gwynn J, Sim K, Searle T, Senior A, Lee A, Brimblecombe J. Effect of nutrition interventions on diet-related and health outcomes of Aboriginal and Torres Strait Islander Australians: a systematic review. BMJ Open 2019 Apr 3;9(4):e025291 [FREE Full text] [doi: 10.1136/bmjopen-2018-025291] [Medline: 30948579]
28. Colles SL, Maypilama E, Brimblecombe J. Food, food choice and nutrition promotion in a remote Australian Aboriginal community. Aust J Prim Health 2014;20(4):365-372. [doi: 10.1071/PY14033] [Medline: 25053144]
29. The Australian Prevention Partnership Centre. 2019. Remote Aboriginal Communities Improve Their Food Security and Diet URL: https://preventioncentre.org.au/news-and-events/prevention-centre-news/research-project-helps-remote-aboriginal-communities-improve-their-food-security-and-diet/ [accessed 2019-07-30]
30. ALPA – Unity Through Enterprise. ALPA Health & Nutrition Policy URL: http://www.alpa.asn.au/alpa-health-nutrition-policy/ [accessed 2020-11-25]
31. Izumi BT, Zenk SN, Schulz AJ, Mentz GB, Sand SL, de Majo RF, et al. Inter-rater reliability of the food environment audit for diverse neighborhoods (FEAD-N). J Urban Health 2012 Jun;89(3):486-499 [FREE Full text] [doi: 10.1007/s11524-011-9657-3] [Medline: 22350513]
32. Glanz K, Sallis JF, Saelens BE, Frank LD. Nutrition environment measures survey in stores (NEMS-S): development and evaluation. Am J Prev Med 2007 Apr;32(4):282-289. [doi: 10.1016/j.amepre.2006.12.019] [Medline: 17383559]
33. Vandevijvere S, Mackay S, D'Souza E, Swinburn B. How Healthy Are New Zealand Food Environments? A Comprehensive Assessment. Auckland, New Zealand: The University of Auckland; 2018.
34. Cameron AJ. The shelf space and strategic placement of healthy and discretionary foods in urban, urban-fringe and rural/non-metropolitan Australian supermarkets. Public Health Nutr 2018 Feb;21(3):593-600. [doi: 10.1017/S1368980017003019] [Medline: 29143716]
35. Australian Bureau of Statistics, Australian Government. 2011. 4364.0.55.007 - Australian Health Survey: Nutrition First Results - Foods and Nutrients, 2011-12 URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4364.0.55.007~2011-12~Main%20Features~Discretionary%20foods~700 [accessed 2020-05-15]
36. Riesenberg D, Backholer K, Zorbas C, Sacks G, Paix A, Marshall J, et al. Price promotions by food category and product healthiness in an Australian supermarket chain, 2017-2018. Am J Public Health 2019 Oct;109(10):1434-1439. [doi: 10.2105/AJPH.2019.305229] [Medline: 31415196]

Abbreviations

AC: agreement coefficient
ALPA: Arnhem Land Progress Aboriginal Corporation
ICC: intraclass correlation coefficient
IRSAD: Index of Relative Socio-economic Advantage and Disadvantage
NHMRC: National Heart and Medical Research Council

Edited by G Eysenbach; submitted 08.11.19; peer-reviewed by L Lytle, G James-Martin; comments to author 16.12.19; revised version received 20.01.20; accepted 24.03.20; published 22.07.20.

Please cite as:
McMahon EJ, Jaenke R, Brimblecombe J
A Mobile App to Rapidly Appraise the In-Store Food Environment: Reliability, Utility, and Construct Validity Study
JMIR Mhealth Uhealth 2020;8(7):e16971
URL: https://mhealth.jmir.org/2020/7/e16971
doi:10.2196/16971
PMID:32706683

Original Paper

Use of a Smartphone App for Weight Loss Versus a Paper-Based Dietary Diary in Overweight Adults: Randomized Controlled Trial

Jeong Sun Ahn[1], MSc; Heejin Lee[1], MSc; Jiae Kim[2], BSc; Haemin Park[2], BSc; Dong Woo Kim[3], PhD; Jung Eun Lee[1], ScD

[1]Department of Food and Nutrition, College of Human Ecology, Seoul National University, Seoul, Republic of Korea

[2]Platform Development Team, Bluecore, Seoul, Republic of Korea

[3]Department of Home Economics, Korea National Open University, Seoul, Republic of Korea

Corresponding Author:
Jung Eun Lee, ScD
Department of Food and Nutrition
College of Human Ecology
Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 2 880 6834
Email: jungelee@snu.ac.kr

Abstract

Background: Mobile health (mHealth) tools may be useful platforms for dietary monitoring and assessment.

Objective: This study aims to evaluate the effectiveness of a mobile dietary self-monitoring app for weight loss versus a paper-based diary among adults with a BMI of 23 kg/m^2 or above.

Methods: A total of 33 men and 17 women aged 18-39 years participated in a 6-week randomized controlled trial. We randomly assigned participants to one of two groups: (1) a smartphone app group (n=25) or (2) a paper-based diary group (n=25). The smartphone app group recorded foods and dietary supplements that they consumed and received immediate dietary feedback using Well-D, a dietary self-monitoring app developed by our team. The paper-based diary group was instructed to record foods or supplements that they consumed using a self-recorded diary. The primary outcomes were weight, BMI, waist circumference, body fat mass, and skeletal muscle mass. We also examined changes in nutrient intake, including energy, carbohydrate, protein, fat, dietary fiber, vitamins, and minerals, using 3-day 24-hour recalls. Differences in changes between the two groups were analyzed using independent t tests or Wilcoxon Mann-Whitney tests. All of the data were analyzed using intent-to-treat analysis.

Results: The mean number of days recorded was 18.5 (SD 14.1) for the app group and 15.5 (SD 10.1) for the paper-based diary group. The differences in changes in weight, BMI, and waist circumference were not significantly different between the app group and paper-based diary group (*P*=.33, .34, and .70, respectively). Similarly, changes in body fat mass or skeletal muscle mass did not differ between the two groups (*P*=.71 and .054, respectively). Although energy intake was reduced in both groups, there was no significant difference in changes in energy intake between the two groups (*P*=.98).

Conclusions: There were no differences in changes in anthropometric measures and nutrient intake between the app group and the paper-based diary group. Both mobile dietary self-monitoring app and paper-based diary may be useful for improving anthropometric measures.

Trial Registration: Clinical Research Information Service KCT0003170; https://cris.nih.go.kr/cris/search/search_result_st01_en.jsp?seq=11642&ltype=&rtype=

(JMIR Mhealth Uhealth 2020;8(7):e14013) doi:10.2196/14013

KEYWORDS

smartphone app; mobile phone; dietary self-monitoring; randomized controlled trial; weight loss

Introduction

Noncommunicable diseases (NCDs) were responsible for 71% of all deaths globally in 2016, and obesity is a risk factor for NCDs like diabetes, coronary heart disease, stroke, and cancer [1]. The World Health Organization (WHO) announced that one of the global NCD targets is to halt the rise of obesity [2]. Despite multifaceted efforts to prevent obesity, the prevalence of obesity has nearly tripled between 1975 and 2016 [3]. In Korea, the prevalence of a BMI 25 kg/m^2 increased in men from 36.6% in 2008 to 44.7% in 2017 [4]. The prevalence of hypercholesterolemia doubled between 2008 and 2018 in both men and women, from 10.6% to 20.9% in men and from 11.8% to 21.4% in women [4]. Increasing obesity is partly due to changes in lifestyle factors, including eating energy-dense foods and foods high in fat and sugars as well as low physical activity [1]. Therefore, dietary intervention is a key strategy to reduce the obesity epidemic.

Dietary modification approaches for obesity management often involve multiple strategies from governments, businesses, communities, individuals, and families. Use of mobile devices, including smartphones, tablets, laptops, wearable devices, and barcode scanners, is thought to increase accessibility at a lower cost by reducing face-to-face, in-person education, clinic visits, and phone calls. Although several studies have shown that mobile interventions led to weight loss [5], a comparison of the effectiveness of mobile health (mHealth) tools for weight loss with conventional methods such as the paper-based diary is needed.

The third global survey on eHealth conducted by the WHO Global Observatory for eHealth defines mHealth as "the use of mobile devices, such as mobile phones, patient monitoring devices, Personal Digital Assistants (PDAs), and wireless devices, for medical and public health practice" [6]. There has been an increase in the number of mHealth solutions because information and communications technology (ICT) has become an integral part of daily life [7]. The WHO announced that 87% (n=109) of countries worldwide have at least one mHealth program in their country [7]. A systematic review of 24 intervention studies examining the potential of mobile apps for health and fitness reported high acceptability of smartphone apps for health behavior change [8].

Mobile technology may be feasible, sustainable, and cost-effective for weight loss. A systematic review of the literature on self-monitoring in weight loss showed that self-monitoring tools (eg, the paper diary, web tools, PDAs, and electronic digital scales) helped individuals lose weight [9]. That systematic review focused on self-monitoring strategies based on works published from 1993 to 2009; most of these studies used a paper diary and 5 used digital technology. A recent qualitative summary of 7 randomized clinical trials suggested that mobile technology interventions facilitated weight loss among individuals who were overweight and obese [10]. Several previous studies investigated the effectiveness of mHealth tools for weight loss or dietary behavior change by comparing them with traditional paper-based methods [11-15] and found that the use of mHealth tools was not superior to paper-based methods.

We have developed a mobile dietary self-monitoring app, Well-D, the features of which have been described elsewhere [16]. In brief, users can log foods or dietary supplements they consume through the Well-D app and receive real-time personalized dietary feedback. Key features of Well-D include sign-up and profile input, log-in, main page, logging meals, food data creation, recipe data creation, logging favorite foods, logging dietary supplements, supplement data creation, display of foods and supplements consumed, and dietary feedback and monitoring.

South Korea had the highest rate of smartphone ownership and internet usage worldwide in 2018, followed by Israel and the Netherlands [17]. Although mHealth has high potential to facilitate health management, health promotion, and disease prevention in South Korea, the effectiveness of morbidity self-management (eg, diabetes [18] and obesity with sleep apnea [19]) through smartphone use has been studied in only a few randomized controlled trials (RCTs). We conducted an RCT in South Korea to examine weight loss using a self-monitoring dietary app. This study aims to evaluate the effectiveness of the dietary self-monitoring app Well-D for weight change by comparing it with paper-based diary use.

Methods

Study Participants

We recruited participants between February 6, 2018, and April 12, 2018, via poster advertisement at Seoul National University and web-based announcements. The inclusion criteria were as follows: (1) 18-40 years of age, (2) BMI ≥23 kg/m^2, (3) willingness for weight loss, and (4) smartphone ownership. We excluded participants if they were pregnant or lactating. This study was approved by the Seoul National University Institutional Review Board (IRB #1710/003-007). The trial was registered with the Clinical Research Information Service (KCT0003170).

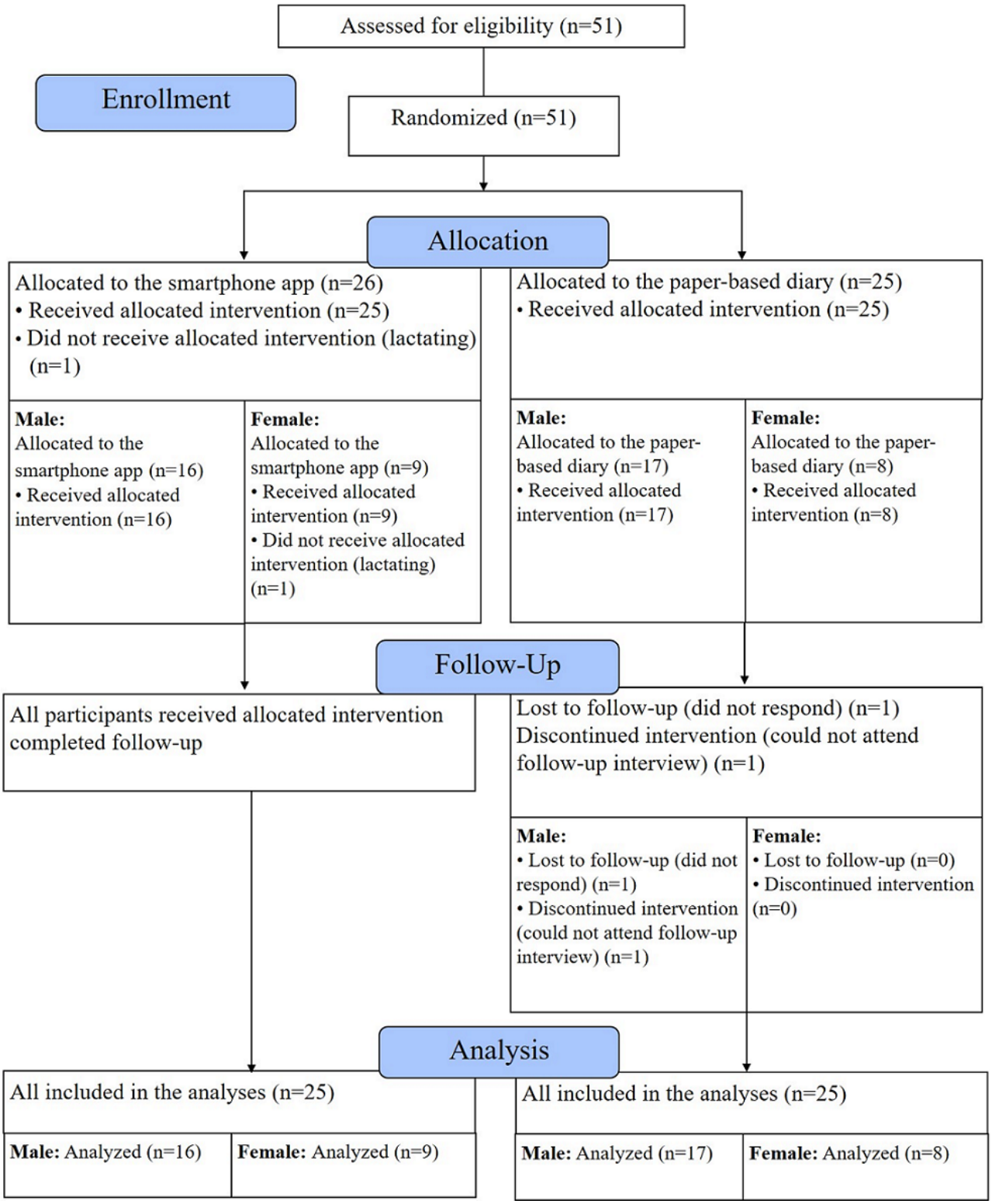
Screening and Randomization

A two-arm, parallel RCT was conducted. Potential participants contacted author JSA via phone to show their willingness to participate in the intervention study. Potential participants were invited to attend a baseline session held at Seoul National University (30-45 minutes). Before starting the baseline session, potential participants reported their age, and their height and weight were measured using a stadiometer to confirm eligibility. All eligible participants returned a written informed consent prior to enrollment. Participants received 20,000 KRW (approximately $17 USD) for attending each of the three visits.

Participants were randomly assigned with a 1:1 allocation to the Well-D app group and the paper-based diary group using a random number table generated by PROC PLAN in SAS version 9.4 (SAS Institute). The allocation sequence was concealed to both JSA and each participant, and the intervention arm allocation was sealed after the participant signed the consent form at the visit site. The participant was informed whether he

or she was assigned to the intervention (smartphone group) or the control group (paper-based diary group). Figure 1 presents the flow diagram of inclusion of participants.

Figure 1. Flow diagram of participants.



Intervention

Both groups were instructed to record foods or supplements that they consumed during the 6-week intervention period. The energy goal was reducing 500 kcal/day from the the estimated energy requirement (EER) in both groups. Participants in the app group received a link to download Well-D, which was developed by our multidisciplinary team (eg, dietitians, nutrition professionals, and software engineers) [16,20]. Well-D was designed for participants to log consumed foods or dietary supplements and learn about adequate nutrient intake according to the Dietary Reference Intakes for Koreans 2015 and major foods contributing to key nutrients, including total energy,

through personalized feedback. A detailed description of Well-D has been reported elsewhere [16].

In the intervention arm, participants were instructed to use Well-D for at least 14 days. Contact was minimal except when technical questions were raised. Since Well-D was implemented as a hybrid app, users who had either iOS or Android smartphones could freely access the app in a network environment. To facilitate app usage, dietitian staff sat with users and helped them register and log into the app. The users also reviewed all the menus in the app with the dietitian staff and were provided a Well-D manual. Users typed in their age, sex, weight, height, and physical activity level during registration. The app provides a database of more than 20,000 foods and recipe items. For foods and dietary supplements that are not available in the database, users could add new food data by typing in the item name and describing the item or adding a photo. The users could also create new recipe data by typing in the ingredients from a food list. Dietitians checked the items that users created and updated the recipe and the food and nutrient database. Based on users' age, sex, BMI, physical activity level, and foods and supplements that they recorded, users received real-time feedback about daily total energy, carbohydrates, protein, total fat, sodium, saturated fat, fiber, sugar, calcium, vitamin C, riboflavin, and food groups on the diabetic exchange list. All of the data on the intake of foods, supplements, and nutrients were collected. We reviewed and downloaded the data from the admin page of the website.

In the paper-based diary group, participants were provided paper-based diaries and pamphlets. The paper-based diary was designed to record the date, time, name and amount of food and ingredients consumed, and the energy intake that participants roughly calculated. Each participant in the paper-based diary was provided a pamphlet that had tips about weight loss strategies and information on the website and URLs available for calculating the energy content of food items. We also provided instructions on how participants could set a proper energy intake goal via the paper-based diary and pamphlet, and participants wrote their energy intake goal for weight loss in the paper-based diary.

Our study design and results were presented according to the CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [21].

Outcome Assessments

Weight, BMI, waist circumference, body fat mass, and skeletal muscle mass were measured as primary outcomes at baseline and after 6 weeks of intervention. Height was measured only at baseline. On the day before measurements, we sent text messages to inform participants to avoid large meals before the visit and to wear light clothing for the measurements. Height was measured twice to the nearest 0.1 cm without shoes using a digital stadiometer (Biospace Korea). Body weight was measured to the nearest 0.1 kg on the Inbody 720 (Biospace Korea) with participants wearing light clothing [22]. BMI was calculated as weight (kg) divided by the squared height (m^2). Waist circumference was measured 1 inch above the umbilicus to the nearest 0.1 cm with a tape measure. For body composition, we assessed body fat mass and skeletal muscle mass using the Inbody 720 [22].

The participants' diets were assessed using scheduled 24-hour recalls (24HRs) for 3 days including 1 day during a weekend. A dietitian conducted the 24HRs using the automated multiple-pass method (AMPM). AMPM uses five steps, including listing foods consumed the previous day, probing for forgotten foods, collecting the time of consumption, collecting descriptions about and amounts of each food, and final questions [23]. On the first day of the 3-day 24HR, participants visited the researcher's office and completed the 24HRs. They completed the other 2 days of 24HRs over the phone. Three-day 24HRs were conducted at baseline and in week 6. We provided a booklet for a sample serving size of foods to all the participants in both groups to help estimate the amount of foods that they consumed. The amount of nutrients were calculated using databases sourced from the Diet Evaluation System by SAS 9.4 and Microsoft Excel 2013 software [24]. If the foods and dietary supplements were not available in the Diet Evaluation System database, we updated the additional food composition databases based on open-source food composition databases from the Ministry of Food and Drug Safety and the National Institute of Agricultural Sciences [25,26]. App and paper-based diary usage was measured by the number of days recorded from baseline to the endpoint visit. We considered it one recording day if users logged at least one food item.

Participants recorded self-reported physical activity using a South Korean version of the Global Physical Activity Questionnaire (GPAQ) [27]. We calculated metabolic equivalent of task (MET)-hours per week using the GPAQ analysis guideline [28]. Participants were advised to maintain their usual physical activity levels during the intervention.

Statistical Analysis

We estimated the sample size to detect statistically significant differences in weight change between the app group and paper-based diary group using a similar previous 6-week trial [11]. By assuming a 0.9 kg difference between the two groups and a standard deviation of 1 kg, the calculated sample size with 80% power was 21 per group. Considering loss to follow-up, we recruited 25 participants per group.

All data were analyzed with intent-to-treat analysis. The differences in changes in anthropometric measures and nutrient intake between the app group and the paper-based diary group were analyzed by independent *t* tests for normally distributed data and Wilcoxon Mann-Whitney tests for skewed data. The differences between pre- and postintervention within each group were compared using paired *t* tests for normally distributed data and Wilcoxon signed-rank tests for skewed data. Data were transformed into normality, if necessary, using Box-Cox power transformations [29]. Carbohydrate, protein, fat, and saturated fat intakes were calculated as a percent of energy. We used a simple linear regression to evaluate the correlation between the number of days spent recording foods and weight change in each group.

There was one missing value for weight and two for other primary outcomes. The missing outcomes were carried forward from the baseline assessments. We also conducted a sensitivity analysis where we used per-protocol analysis by excluding those with missing outcomes.

Results

Table 1 shows baseline characteristics by intervention arms. The mean age of participants was 26.0 years. The mean weight was 77.1 kg and the mean BMI was 26.7 kg/m^2. The mean daily intake of energy was 2166.1 kcal/day. On average, % energy from carbohydrates, protein, and fat were 50.5%, 18.6%, and 30.9%, respectively. There were no significant differences in baseline characteristics between the app group and the paper-based diary group.

After randomization, one lactating participant was excluded (Figure 1). As a result, 33 men and 17 women aged 18-39 years participated in a 6-week RCT, and 32 men and 17 women completed the study. One participant in the paper-based diary group did not respond to final contact whereas another participant responded but could not attend a follow-up interview. This participant provided his weight and the number of days that he used the dietary diary. All participants completed 3-day 24HRs at baseline; 47 participants completed 3-day 24HRs and 1 participant in the app group completed 2 days of 24HRs during the 6-week intervention period.

We found no statistically significant difference in change in body weight between the app group and the paper-based diary group (mean –0.4, SD 1.6 kg vs mean –1.4, SD 2.7 kg;*P*=.33) (Figure 2). Likewise, differences in changes in BMI, waist circumference, body fat mass, and skeletal muscle mass were not statistically significant between the two groups. Mean change and standard deviation between pre- and postintervention are presented in Table 2.

When we compared the preintervention anthropometric measures with the postintervention measures, significant decreases in body weight and BMI were observed in the paper-based diary group (*P*=.02 and .01, respectively), but not in the app group (*P*=.25 and .26, respectively). Waist circumference and body fat mass decreased significantly in both groups. The skeletal muscle mass significantly increased in the app group (*P*=.048). Overall, the results were similar in the sensitivity analysis where a per-protocol analysis was conducted by excluding those with missing anthropometric measures (Multimedia Appendix 1).

Table 1. Baseline characteristics of participants by intervention arms.

Characteristics	Total	App group	Paper-based diary group	*P* value[a]
Sex, n (%)				.77
Male	33 (66)	16 (64)	17 (68)	
Female	17 (34)	9 (36)	8 (32)	
Age (year), mean (SD)	26.0 (4.8)	26.5 (5.3)	25.6 (4.3)	.50
Weight (kg), mean (SD)	77.1 (11.5)	77.9 (12.9)	76.3 (10.2)	.62
BMI (kg/m^2), mean (SD)	26.7 (2.7)	27.1 (3.0)	26.4 (2.5)	.54
Waist circumference (cm), mean (SD)	91.7 (9.3)	93.1 (9.6)	90.3 (8.9)	.25
Body fat mass (kg), mean (SD)	23.3 (6.3)	24.2 (5.6)	22.3 (6.9)	.29
Skeletal muscle mass (kg), mean (SD)	30.3 (6.1)	30.2 (6.5)	30.4 (5.7)	.91
Total physical activity (MET[b]-hours/week), mean (SD)	26.4 (25.6)	25.3 (23.4)	27.6 (28.1)	.87
Energy (kcal/day), mean (SD)	2166.1 (546.5)	2270.3 (522.1)	2061.9 (560.8)	.18
Carbohydrate (% energy/day)	50.5 (6.7)	48.8 (7.8)	52.1 (5.1)	.11
Protein (% energy/day)	18.6 (4.0)	19.3 (4.2)	17.8 (3.6)	.34
Fat (% energy/day)	30.9 (5.7)	31.7 (6.7)	30.0 (4.4)	.34
Total dietary fiber (g/day), mean (SD)	17.0 (6.1)	16.7 (5.9)	17.4 (6.4)	.69
Calcium (mg/day), mean (SD)	530.8 (237.9)	512.8 (185.3)	548.8 (283.8)	.74
Sodium (mg/day), mean (SD)	4021.7 (1157.2)	4110.6 (1124.0)	3932.7 (1205.8)	>.99

[a]Chi-square tests were used for categorical variables, and independent *t* tests or Wilcoxon Mann-Whitney tests were used for continuous variables.

[b]MET: metabolic equivalent task.

Figure 2. Differences in change between the app group and the paper-based diary group in terms of (A) body weight, (B) BMI, (C) waist circumference, (D) body fat mass, and (E) skeletal muscle mass. The asterisk denotes *P*<.05 (significant difference) for comparisons of anthropometric measures pre- to postintervention.

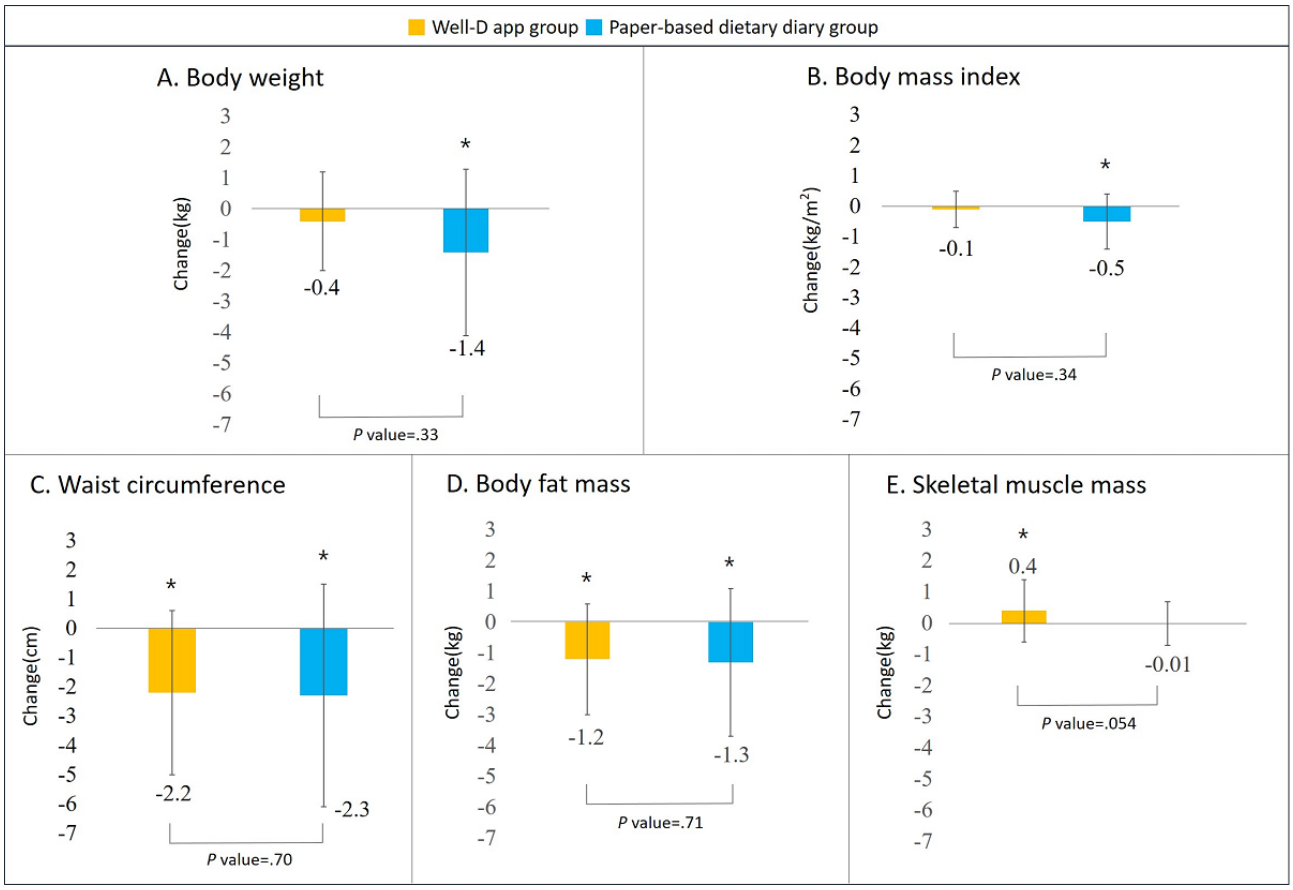


Table 2. Differences in anthropometric measures between the app group and the paper-based diary group (intent-to-treat analysis).

Measure	App group (n=25), mean (SD)				Paper-based diary group (n=25), mean (SD)				*P* value[a]
	Baseline	6 weeks	Change[b]	*P* value[c]	Baseline	6 weeks	Change[b]	*P* value[c]	
Weight (kg)	78.0 (12.9)	77.6 (13.0)	–0.4 (1.6)	.25	76.3 (10.2)	75.0 (9.3)	–1.4 (2.7)	.02	.33
BMI (kg/m^2)	27.1 (3.0)	26.9 (3.0)	–0.1 (0.6)	.26	26.4 (2.5)	25.9 (2.2)	–0.5 (0.9)	.01	.34
Waist circumference (cm)	93.1 (9.6)	90.9 (9.2)	–2.2 (2.8)	<.001	90.3 (9.0)	88.1 (7.1)	–2.3 (3.8)	.004	.70
Body fat mass (kg)	24.2 (5.6)	23.0 (6.1)	–1.2 (1.8)	.004	22.3 (6.9)	21.0 (5.9)	–1.3 (2.4)	.01	.71
Skeletal muscle mass (kg)	30.2 (6.5)	30.6 (6.6)	0.4 (1.0)	.048	30.4 (5.7)	30.4 (5.8)	–0.01 (0.7)	.48	.054

[a]Independent *t* tests or Wilcoxon Mann-Whitney tests were used to assess the differences in percent changes in anthropometric measures between the app group and the paper-based diary group.

[b]Changes were calculated as postintervention anthropometric measures minus preintervention anthropometric measures.

[c]Paired *t* tests or Wilcoxon signed-rank tests were used to assess the differences in percent changes in anthropometric measures pre- to postintervention.

Differences in changes in nutrient intake between the app group and the paper-based diary group were not statistically significant (Table 3). When we compared nutrient intake assessed from preintervention 24HRs with postintervention 24HRs, we observed significant reductions in cholesterol, calcium, phosphorus, and potassium only in the paper-based diary group (all *P*s<.05). When we compared nutrient intake levels from the Well-D app with those from the 24-hour recalls (24HR) among 25 participants in the intervention arm, we found moderate-to-high correlations (from 0.43 for thiamine to 0.71 for iron) (Table 4).

Over the 6-week intervention period, differences in the number of days recorded was not significant between the app group and the paper-based diary group (mean 18.5, SD 14.1 vs mean 15.5, SD 10.1, respectively; *P*=.67). We examined whether the degree of weight loss was associated with the number of days spent using the app or the paper-based diary (Figure 3). The change in body weight from pre- to postintervention tended to increase

according to the number of days participants recorded an entry. However, the beta coefficient was not statistically significant.

When we counted the number of participants who recorded food items in each week, we found higher proportions of recording in week 1 and week 2 in both groups than in later weeks and a higher proportion of recording, in general, in the app group than in the paper-based diary group. In all, 51.43% (12.86 on average per day for 7 days) of participants recorded food items in the app group and 38.10% (8 on average per day) of participants recorded food items in the paper-based diary group in week 1, but the proportions decreased to 36.00% (9 on average per day) in week 5 in the app group and 32.65% (6.86 on average per day) in the paper-based dairy group (Table 5).

Table 3. Differences in changes in nutrient intake assessed from 24-hour recalls between the app group and the paper-based diary group (intent-to-treat analysis).

Characteristic	App group (n=25), mean (SD)			Paper-based diary group (n=25), mean (SD)			*P* value[a]
	Baseline	6 weeks	*P* value[b]	Baseline	6 weeks	*P* value[b]	
Energy (kcal/day)	2269.7 (522.8)	1983.5 (365.3)	.04	2061.9 (560.8	1780.6 (571.0)	.06	.98
Carbohydrate (% energy/day)	48.8 (7.8)	48.9 (8.5)	.95	52.2 (5.1)	49.8 (8.3)	.20	.34
Protein (% energy/day)	19.3 (4.2)	19.7 (4.4)	.68	17.8 (3.6)	17.4 (3.5)	.63	.60
Fat (% energy/day)	31.9 (7.0)	31.4 (7.4)	.70	30.0 (4.4)	32.8 (7.6)	.13	.18
Saturated fat (% energy/day)	12.1 (6.2)	12.2 (6.5)	.83	10.4 (2.3)	10.1 (2.4)	.50	.76
Total dietary fiber (g/day)	16.7 (5.9)	15.3 (5.1)	.42	17.4 (6.4)	15.0 (5.7)	.10	.64
Cholesterol (mg/day)	371.9 (122.7)	363.8 (160.3)	.70	363.0 (146.2)	289.0 (100.7)	.04	.35
Calcium (mg/day)	512.8 (185.3)	525.2 (312.4)	.57	548.8 (283.8)	438.9 (241.9)	.01	.23
Phosphorus (mg/day)	1076.5 (300.2)	1041.9 (297.1)	.63	1067.6 (281.1)	878.1 (268.6)	.01	.14
Iron (mg/day)	16.8 (12.3)	14.5 (7.9)	.43	19.7 (32.3)	17.9 (32.4)	.05	.30
Sodium (mg/day)	4110.6 (1124)	3833.4 (1176.3)	.23	3932.7 (1205.8)	3531.9 (1596.0)	.21	.77
Potassium (mg/day)	2254.4 (562.3)	2318.1 (743.5)	.77	2325.9 (658.3)	2015.7 (587.8)	.01	.07
Vitamin A (μg RE[c]/day)	742.0 (937.1)	785.2 (1087.6)	.83	588.4 (716.9)	574.1 (580.0)	.84	.55
Thiamine (mg/day)	2.7 (5.1)	2.9 (5.9)	.22	6.3 (21.5)	1.6 (1.8)	.05	.42
Riboflavin (mg/day)	2.9 (5.3)	2.9 (5.8)	.78	5.9 (21.1)	1.5 (1.3)	.09	.35
Niacin (mg/day)	25.7 (28.7)	24.3 (30.8)	.76	25.2 (28.7)	18 (10.2)	.26	.54
Vitamin C (mg/day)	276.4 (338.0)	201.7 (314.2)	.07	127.3 (120.0)	154.7 (281.8)	.42	.46

[a]Independent *t* tests and Wilcoxon Mann-Whitney tests were used to assess group differences.

[b]Paired *t* tests or Wilcoxon signed-rank tests were used to assess differences in percent changes in nutrient intake pre- to postintervention.

[c]RE: retinol equivalents.

Table 4. A comparison of nutrient intake levels from the Well-D app with those from the 24-hour recalls (n=25).

Nutrient[a]	Correlation coefficient	*P* value[b]
Energy (kcal/day)	0.68	<.01
Carbohydrate (% energy/day)	0.55	<.01
Protein (% energy/day)	0.60	<.01
Fat (% energy/day)	0.57	<.01
Saturated fat (% energy/day)	0.66	<.001
Total dietary fiber (g/day)	0.51	.01
Cholesterol (mg/day)	0.47	.02
Calcium (mg/day)	0.49	.01
Phosphorus (mg/day)	0.47	.02
Iron (mg/day)	0.71	<.001
Sodium (mg/day)	0.56	<.01
Potassium (mg/day)	0.56	<.01
Vitamin A (μg RE/day)	0.54	.01
Thiamine (mg/day)	0.43	.03
Riboflavin (mg/day)	0.50	.01
Niacin (mg/day)	0.70	<.001
Vitamin C (mg/day)	0.66	<.001

[a]The residual method was used to adjust for nutrient intake.

[b]Either Pearson correlation or Spearman correlation was used.

Figure 3. Weight loss by the number of days of dietary recording in the app group (left) and the paper-based diary group (right).

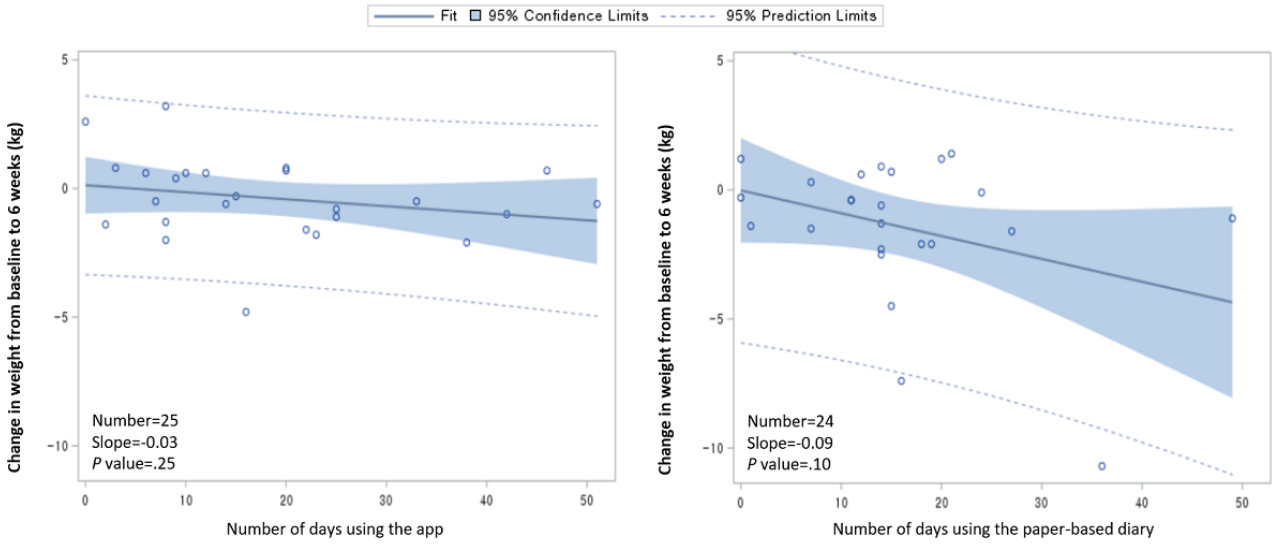


Table 5. Mean number (percentage) of days participants spent recording food items for each week in the app group and in the paper-based diary group.

Week	App group (n=25), mean (%)	Paper-diary group (n=21)[a], mean (%)
Week 1	12.86 (51.43)	8.00 (38.10)
Week 2	11.43 (45.71)	8.14 (38.78)
Week 3	9.57 (38.29)	8.57 (40.82)
Week 4	9.00 (36.00)	6.86 (32.65)
Week 5	9.00 (36.00)	6.86 (32.65)
Week 6	10.00 (40.00)	4.29 (20.41)

[a]Out of 25, one was lost to follow-up. Three did not provide information on dates recorded, but provided the number of days recorded.

Discussion

Principal Findings

We conducted an RCT on weight loss in young adults with a BMI ≥23 kg/m^2 to compare the effectiveness of a mobile dietary self-monitoring app versus a paper-based diary. In our study, 66% of participants were men, which reflects the sex ratio of overweight individuals in Korea [4]. The mean age was 26 years since we recruited participants in college. The average total energy intake was 2166.1 kcal/day with an average of 50.5% of energy derived from carbohydrates, 30.9% energy from fat, and 18.6% from protein. The average sodium intake was 4021.7 mg/day.

In summary, we did not find significant differences in changes in body weight, BMI, waist circumference, body fat mass, or skeletal muscle mass between the app group and the paper-based diary group. Additionally, the changes in nutrient intake were not different between the two groups. However, when we compared anthropometric changes from pre- to postintervention, we found reductions in weight and BMI in the paper-based group and reductions in waist circumference and body fat mass in both groups. Skeletal muscle mass increased slightly in the app group. We also found decreases in total energy intake in both groups and as well as decreases in intake of cholesterol, calcium, phosphorus, and potassium in the paper-based diary group, but not in the app group. When we compared nutrient intake from the Well-D app with 24HRs, we found modest-to-high correlations, suggesting potential use of Well-D for dietary assessment.

Comparison With Previous Work

Previous studies found that the effectiveness of mHealth technologies was similar to the paper-based diary [11-13]. A 24-month randomized trial with 210 overweight or obese adults compared the effectiveness of a PDA for self-monitoring diet, a PDA with daily feedback, and a paper diary. There were no differences in changes in percentage weight over time among the three groups, but a significant weight change over time was found only for the PDA with feedback group [12]. The mean percentages of weight change from baseline to 24 months were –1.94% for the paper diary (not significant), –1.38% for the PDA (not significant), and –2.32% for the PDA + feedback (significant). The study suggested that daily feedback enhanced adherence to self-monitoring and the effectiveness of weight loss. Another randomized trial compared weight loss between the Lose it mobile app and a traditional paper-and-pencil method among 47 overweight or obese adults during an 8-week intervention. Weight was reduced over the course of the study in both groups, but there was no difference in weight change between groups. This study showed that the number of days spent recording diet was significantly higher in the app group than the paper-and-pencil group [13]. The 6-month trial of the My Meal Mate app showed that, among 128 overweight participants, weight change over time was significantly greater in the smartphone group than in the website group, but there was no difference in the smartphone group compared to the diary group [11]. In the intent-to-treat analysis of that study, mean weight changes from baseline to 6 months were –4.6 kg in the smartphone app group, –2.9 kg in the diary group, and –1.3 kg in the website group. The smartphone-based behavioral obesity treatment (SMART) randomized clinical trial of 276 adults with overweight or obesity over 18 months recently published findings of similar weight loss across the three groups: group-based treatment with meetings (GROUP), SMART-based treatment, and a control condition with paper diaries (CONTROL) [30]. In that study, although retention was higher in both GROUP and SMART groups than the CONTROL group, weight changes did not differ at the 6-month, 12-month, or 18-month assessment. Consistent with a few previous studies, we found that changes in anthropometric measures were not different between the app group and paper-based diary group.

There have been only a few studies that have supported the effectiveness of an app to improve users' diet compared to a paper-based method [14,15]. A 4-week pilot intervention study with the MyFitnessPal app among 30 healthy adults showed that the app users who received feedback on sodium intake significantly decreased their urinary sodium compared to the paper-based journal group [15]. In another two-phase crossover intervention study, 34 adolescents recorded food intake using a paper diary and the FoodWiz2 app during each 4-week intervention period [14]. The app phase showed a significant reduction in the consumption of chocolate snacks and fizzy drinks in comparison to the paper diary phase [14]. In our study, when we compared nutrient intake between the app group and the paper-based diary group, we did not find significant differences. However, we found decreases in total energy intake in both app and paper-based diary groups.

Several studies conducted in South Korea reported potential weight change when using mHealth tools. A longitudinal study with a median follow-up of 275 days investigated the

effectiveness of the Noom Coach app on weight reduction among 35,921 South Korean adults with a BMI ≥23 kg/m^2, who recorded dietary data two or more times a month for 6 consecutive months [31]. That study found that 22.7% of app users reduced their weight by more than 10% compared to the baseline weight. Similarly, a weight loss intervention study of 104 Korean adults aged 20-60 years with a BMI ≥23 kg/m^2 showed that the use of the Noom Coach app with daily dietary coaching resulted in a significant weight change of –7.5% after a 15-week intervention period compared to baseline [32].

Additionally, a randomized trial that compared weight loss according to the frequency of usage of the My Meal Mate app found that participants in the highest frequency-of-use category lost an average of 6.4 kg more than those in the lowest frequency-of-use category during the 6-month intervention period [33]. A cohort study showed that more app-based recordings were associated with greater weight loss [31]. In the Cell Phone Intervention For You (CITY) trial, increasing app use per day was associated with increasing weight loss [34]. In our study, we observed a tendency for greater weight loss as the number of recordings increased, albeit without statistical significance.

Strengths and Limitations

Our study was a randomized, parallel trial with a high follow-up rate. Because we developed the app, we had full access to the data. We also assessed participants' dietary information using two 3-day 24HRs, and therefore we were able to compare nutrient intake between the two groups. However, our study had several limitations. Because the population was composed of young adults, the results may not be generalizable to children or older people. Since we included participants with a BMI ≥23 kg/m^2, the magnitude of weight change during the 6-week intervention period may not be large enough to see differences. Furthermore, the sample size was small and the study period was relatively short. Well-D did not have an exercise tracking function, and we were not able to track participants' exercise levels. However, we asked participants to maintain their usual physical activity levels. When we examined their usual exercise levels at baseline and at postintervention, we found no significant differences.

Conclusions

We conducted an RCT to evaluate the effectiveness of a mobile dietary self-monitoring app for weight loss versus a paper-based diary. We found that participants reduced their energy intake, waist circumference, and body fat mass in both groups. However, we found no difference in changes between the app group and paper-based diary group. Our study suggests that both the smartphone app and the paper-based diary method may exhibit similar effectiveness for short-term body fat loss. Our findings may contribute to understanding and implementing mHealth interventions in health care services. Further prospective or intervention studies with larger sample sizes and long-term follow-up are warranted to explore the effectiveness of mHealth tools for the management of common chronic diseases in Korea.

Acknowledgments

This research was supported by the Ministry of Science and ICT, Korea, under the Information Technology Research Center support program (IITP-2019-2014-1-00720) supervised by the Institute for Information & Communications Technology Promotion.

Authors' Contributions

JSA and JEL conceptualized the study design and drafted the manuscript. JSA, HL, and JEL conducted data collection and statistical analysis. All authors contributed to data acquisition and design of the Well-D app. JK and HP devised the app. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

JK and HP were employed by Bluecore Co, Ltd. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Differences in anthropometric measures between the app group and the paper-based diary group (per-protocol analysis).
[DOCX File , 28 KB - mhealth_v8i7e14013_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH (V1.6.1).
[PDF File (Adobe PDF File), 1507 KB - mhealth_v8i7e14013_app2.pdf]

References

1. World Health Organization. Noncommunicable diseases country profiles 2018. Geneva, Switzerland: World Health Organization; 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/274512/9789241514620-eng.pdf?sequence=1&isAllowed=y [accessed 2020-03-04]

2. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. Geneva, Switzerland: World Health Organization URL: https://apps.who.int/iris/bitstream/handle/10665/94384/9789241506236_eng.pdf?sequence=1 [accessed 2020-03-04]
3. NCD Risk Factor Collaboration. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
4. Korea Centers for Disease Control and Prevention. Korea Health Statistics 2018: Korea National Health and Nutrition Examination Survey (KNHANES VII-3). Cheongju: Minister of Korea Centers for Disease Control and Prevention; 2020. URL: https://knhanes.cdc.go.kr/knhanes/sub04/sub04_03.do?classType=7 [accessed 2020-03-04]
5. Lyzwinski LN. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Pers Med 2014;4(3):311-385 [FREE Full text] [doi: 10.3390/jpm4030311] [Medline: 25563356]
6. World Health Organization. Third Global Survey on eHealth 2015. The use of eHealth in support of universal health coverage. Geneva, Switzerland: World Health Organization; 2015. URL: https://www.who.int/goe/survey/goe_2015_survey_en.pdf?ua=1 [accessed 2020-03-04]
7. World Health Organization. Global diffusion of eHealth: making universal health coverage achievable. Report of the third global survey on eHealth. Geneva, Switzerland: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1 [accessed 2020-03-05]
8. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
9. 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]
10. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley WH. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]
11. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
12. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
13. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014 Oct;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
14. Jimoh F, Lund EK, Harvey LJ, Frost C, Lay WJ, Roe MA, et al. Comparing Diet and Exercise Monitoring Using Smartphone App and Paper Diary: A Two-Phase Intervention Study. JMIR Mhealth Uhealth 2018 Jan 15;6(1):e17 [FREE Full text] [doi: 10.2196/mhealth.7702] [Medline: 29335239]
15. Ipjian ML, Johnston CS. Smartphone technology facilitates dietary change in healthy adults. Nutrition 2017 Jan;33:343-347. [doi: 10.1016/j.nut.2016.08.003] [Medline: 27742102]
16. Ahn JS, Kim DW, Kim J, Park H, Lee JE. Development of a Smartphone Application for Dietary Self-Monitoring. Front Nutr 2019;6:149 [FREE Full text] [doi: 10.3389/fnut.2019.00149] [Medline: 31608283]
17. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 Feb. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-03-09]
18. Noh J, Cho Y, Nam H, Kim J, Kim D, Yoo H, et al. Web-based comprehensive information system for self-management of diabetes mellitus. Diabetes Technol Ther 2010 May;12(5):333-337. [doi: 10.1089/dia.2009.0122] [Medline: 20388042]
19. Cho S, Wee JH, Yoo S, Heo E, Ryu B, Kim Y, et al. Effect of Lifestyle Modification Using a Smartphone Application on Obesity With Obstructive Sleep Apnea: A Short-term, Randomized Controlled Study. Clin Exp Otorhinolaryngol 2018 Sep;11(3):192-198 [FREE Full text] [doi: 10.21053/ceo.2017.01284] [Medline: 29374961]
20. Well-D app. URL: http://welld.blue-core.com/mweb/logout.do [accessed 2020-07-19]
21. 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]
22. McLester CN, Nickerson BS, Kliszczewicz BM, McLester JR. Reliability and Agreement of Various InBody Body Composition Analyzers as Compared to Dual-Energy X-Ray Absorptiometry in Healthy Men and Women. J Clin Densitom 2018 Nov 03. [doi: 10.1016/j.jocd.2018.10.008] [Medline: 30472111]
23. Conway JM, Ingwersen LA, Vinyard BT, Moshfegh AJ. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. Am J Clin Nutr 2003 May;77(5):1171-1178 [FREE Full text] [doi: 10.1093/ajcn/77.5.1171] [Medline: 12716668]

24. Jung H, Lee S, Kim D, Noh H, Song S, Kang M. Development and feasibility of a web-based program 'Diet Evaluation System (DES)' in urban and community nutrition survey in Korea. Korean J Health Promot 2013;13(3):107-115 [FREE Full text]
25. Ministry of Food and Drug Safety. Food and Nutrient Data System. URL: http://www.foodsafetykorea.go.kr/fcdb/ [accessed 2020-05-18]
26. National Institute of Agricultural Sciences of Korea. Standard food composition table 9th revision. Rural Development Administration of Korea. URL: http://koreanfood.rda.go.kr/kfi/fct/fctFoodSrch/list [accessed 2018-05-08]
27. 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]
28. World Health Organization. Global physical activity questionnaire (GPAQ). Geneva, Switzerland: World Health Organization URL: https://www.who.int/ncds/surveillance/steps/GPAQ/en/ [accessed 2018-05-07]
29. Box GEP, Cox DR. An Analysis of Transformations. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;26(2):211-243. [doi: 10.1111/j.2517-6161.1964.tb00553.x]
30. Thomas JG, Bond DS, Raynor HA, Papandonatos GD, Wing RR. Comparison of Smartphone-Based Behavioral Obesity Treatment With Gold Standard Group Treatment and Control: A Randomized Trial. Obesity (Silver Spring) 2019 Apr;27(4):572-580 [FREE Full text] [doi: 10.1002/oby.22410] [Medline: 30779333]
31. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
32. Toro-Ramos T, Lee D, Kim Y, Michaelides A, Oh TJ, Kim KM, et al. Effectiveness of a Smartphone Application for the Management of Metabolic Syndrome Components Focusing on Weight Loss: A Preliminary Study. Metab Syndr Relat Disord 2017 Dec;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
33. Carter MC, Burley VJ, Cade JE. Weight Loss Associated With Different Patterns of Self-Monitoring Using the Mobile Phone App My Meal Mate. JMIR Mhealth Uhealth 2017 Feb 02;5(2):e8 [FREE Full text] [doi: 10.2196/mhealth.4520] [Medline: 28153814]
34. Lin P, Grambow S, Intille S, Gallis JA, Lazenka T, Bosworth H, et al. The Association Between Engagement and Weight Loss Through Personal Coaching and Cell Phone Interventions in Young Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Oct 18;6(10):e10471 [FREE Full text] [doi: 10.2196/10471] [Medline: 30341051]

Abbreviations

24HR: 24-hour recall
AMPM: automated multiple-pass method
CITY: Cell Phone Intervention For You
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GPAQ: Global Physical Activity Questionnaire
ICT: information and communications technology
MET: metabolic equivalent of task
mHealth: mobile health
NCD: noncommunicable disease
PDA: Personal Digital Assistant
RCT: randomized controlled trial
SMART: smartphone-based behavioral obesity treatment
WHO: World Health Organization

Edited by G Eysenbach; submitted 15.03.19; peer-reviewed by R Berry, R Ciptaningtyas, V Salvador; comments to author 09.12.19; revised version received 26.03.20; accepted 03.06.20; published 31.07.20.

Please cite as:
Ahn JS, Lee H, Kim J, Park H, Kim DW, Lee JE
Use of a Smartphone App for Weight Loss Versus a Paper-Based Dietary Diary in Overweight Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e14013
URL: http://mhealth.jmir.org/2020/7/e14013/
doi:10.2196/14013
PMID:32735225

Original Paper

Effectiveness of a Mobile App Intervention for Anxiety and Depression Symptoms in University Students: Randomized Controlled Trial

Tayla McCloud[1], MSc; Rebecca Jones[1], MSc; Gemma Lewis[1], PhD; Vaughan Bell[2], DClinPsy; Elias Tsakanikos[3], PhD

[1]Division of Psychiatry, University College London, London, United Kingdom

[2]Division of Psychology & Language Sciences, University College London, London, United Kingdom

[3]Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Tayla McCloud, MSc
Division of Psychiatry
University College London
149 Tottenham Court Road
London, W1T 7NF
United Kingdom
Phone: 44 2031087765
Email: t.mccloud@ucl.ac.uk

Abstract

Background: Depression and anxiety symptoms are common among university students, but many do not receive treatment. This is often because of lack of availability, reluctance to seek help, and not meeting the diagnostic criteria required to access services. Internet-based interventions, including smartphone apps, can overcome these issues. However, a large number of apps are available, each with little evidence of their effectiveness.

Objective: This study aims to evaluate for the first time the effectiveness of a self-guided mobile app, Feel Stress Free, for the treatment of depression and anxiety symptoms in students.

Methods: A web-based randomized controlled trial compared a cognitive behavioral therapy (CBT)–based mobile app Feel Stress Free with a wait-list control. University students self-identified as experiencing symptoms of anxiety or depression and were randomized to 6 weeks of intervention (n=84) or control (n=84), unblinded. The app is self-guided and incorporates behavioral relaxation activities, mood tracking and thought challenging, and minigames. Participants completed the Hospital Anxiety and Depression Scale online at baseline and every fortnight.

Results: At week 6, the primary end point, there was evidence that the Feel Stress Free app reduced depression symptoms (mean difference −1.56; 95% CI −2.67 to −0.44; *P*=.006) but only very weak evidence that it reduced anxiety symptoms (mean difference −1.36; 95% CI −2.93 to 0.21; *P*=.09). At week 4, there was evidence to support the effectiveness of the intervention for anxiety symptoms (mean difference −1.94; 95% CI −3.11 to −0.77; *P*=.001) and, though weaker, depression symptoms (mean difference −1.08; 95% CI −2.12 to −0.04; *P*=.04). At week 6, 83% (34/41) of participants indicated that they were using the app weekly or more frequently.

Conclusions: The Feel Stress Free app is a promising mobile intervention for treating symptoms of anxiety and depression in students and overcomes many of the barriers to traditional CBT. Further research is needed to establish its effectiveness at and beyond 6 weeks.

Trial Registration: ClinicalTrials.gov NCT03032952; https://clinicaltrials.gov/ct2/show/NCT03032952

(JMIR Mhealth Uhealth 2020;8(7):e15418) doi:10.2196/15418

KEYWORDS

anxiety; depression; cognitive behavioral therapy; eHealth; online intervention; mobile apps; randomized controlled trial; mobile phone

Introduction

Depression and anxiety are common and disabling disorders and often co-occur [1]. Around 25% of people with depression or anxiety experience symptoms before the age of 20 years [2], and rates are high among university students relative to other sections of the population [3]. A systematic review of international studies estimated the prevalence of depression to be 30.6% among university students [4]. The prevalence of depression is also increasing: a recent report indicated that from 2006 and 2007 to 2015 and 2016, the number of higher education students in the United Kingdom who disclosed mental health disorders to their institution rose five-fold, and university deaths by suicide increased by 79% [3]. Early intervention could prevent adverse outcomes often associated with anxiety and depression, such as substance misuse, educational underachievement, and suicide [5,6]. However, for university students, there can be numerous barriers to help-seeking, including lack of time, privacy concerns, financial constraints, a lack of perceived need for formal help, and stigma [7]. Innovative approaches are therefore needed to address the high burden of mental health problems among university students.

The most established psychological treatment for depression and anxiety, cognitive behavioral therapy (CBT), often has long waiting lists in the United Kingdom, and services are not evenly spread throughout the country [8]. Students who would benefit from timely CBT may also not meet the criteria required to access services, particularly if their symptoms appear to be mild (though still distressing). University mental health services are struggling with increasing demand [3], and a collective of executive heads representing 136 universities in the United Kingdom (Universities UK) has called for higher education leaders to prioritize student mental health care as imperative [9]. In the absence of professional help, self-help approaches have been shown to be a somewhat effective alternative that are highly valued by young people, particularly when in a digital form [10,11].

There is evidence that computerized forms of self-directed CBT (cCBT) can be as effective as traditional CBT in the treatment of depression and anxiety [10,12-14], and cCBT is now recommended in the United Kingdom for treating subclinical to moderate depression [15]. Mobile CBT apps used primarily on smartphones represent an opportunity to distribute cCBT to the 2.6 billion active smartphone users worldwide [16], at a low cost to the provider and the user. Among those aged between 18 and 24 years in the United Kingdom, it was recently estimated that 93% owned or had ready access to a smartphone [17], and on average over 4 hours per day is spent using them [18]. This eliminates the need for therapist input, formal help-seeking, or a clinical diagnosis, and the lack of waiting time has been shown to be particularly attractive among a UK student sample [19].

Promising findings have been reported regarding the effectiveness of app-based interventions for depression and anxiety in student populations. In 2014, a systematic review by Davies et al [20] suggested that internet- and computer-based interventions could be beneficial in improving depression and anxiety, particularly as an adjunct to university support services. Indeed, internet-based self-help is often recommended by university counseling services struggling to cope with high demand [21]. More recently, a 2019 systematic review [22] on digital health interventions for improving depression and anxiety among students found that mobile-based interventions such as apps appear to be as promising as computer-, web-, and virtual reality–based interventions. However, the authors found a comparative scarcity of research in this area: only 8 out of 71 included studies tested interventions delivered via mobile phones. The issue remains that many mental health apps are available, but there is little or no evidence of the effectiveness of the vast majority of these apps [23-26].

The aim of this study (ClinicalTrials.gov registration NCT03032952) was to examine for the first time the effectiveness of a particular CBT-based mobile app, *Feel Stress Free*, as a treatment for symptoms of anxiety and depression among university students.

Methods

Design

This was a 6-week, web-based, parallel group, unblinded randomized controlled trial, with a wait-list control. Participants were individually randomized in a 1:1 ratio.

Participants and Setting

Eligible participants were aged 18 years or over; scored 8 or above on one or both subscales of the Hospital Anxiety and Depression Scale (HADS), indicating at least a possible case of depression and/or anxiety [27]; were currently a student at 1 of the 4 partnered universities; had access to an Apple or Android phone or tablet or a computer with Firefox, Safari, or Chrome installed; and were computer and internet literate.

A total of 4 universities that partnered with Thrive Therapeutic Software Limited agreed to take part: University College London (UCL), School of Oriental and African Studies University of London, University of Buckingham, and University of Roehampton. Students were recruited between March and June 2016 through their university student union or student welfare services via email, poster and social media advertisements, and university welfare staff recommendations. The recruited participants were directed to the Thrive website, where they could enroll by entering their university email address. Participant IDs were then provided via an email to this address, with a link to the web-based information sheet and consent form. Participants were always contacted via their university email to prevent multiple sign-ups. Participants were not compensated in any way for their participation. All data were collected online.

Signing the consent form was the only time during the trial when the participants had to give their names; they were contacted via email and identified using participant IDs only from this point onward. Participants answered a series of questions confirming that they had read the information about the trial and understood what would be asked of them. If a participant answered *No* to any of the questions, a representative of the trial emailed them to clarify any queries and ensured that

their informed consent was given before continuing. Next, participants were sent the web-based baseline questionnaires via email, which contained demographic questions and the HADS, which was also used to determine eligibility.

Sample Size

Sample size was estimated using both subscales of the self-rated HADS as coprimary outcome measures, with the aim of detecting an intervention effect of half an SD. As the estimated SD differs for each subscale, we chose the most conservative value and calculated that at least 64 participants were needed in each arm, with a 2-tailed significance level of .05 and 80% power. Owing to the high dropout rates usually observed in web-based trials [28], it was decided that up to 300 participants would be randomized.

Randomization and Allocation

Participants were individually randomized in batches of 30 each time this number of students had been screened and confirmed as eligible. Random numbers were generated by a statistician (RJ) to allocate participants within each batch to the 2 study arms in a 1:1 ratio, using prespecified code written in Stata (StataCorp; version 14) [29]. The list of participant IDs and group allocations was then returned to the researcher (TM). At the end of the recruitment period, a final batch of 18 participants was randomized.

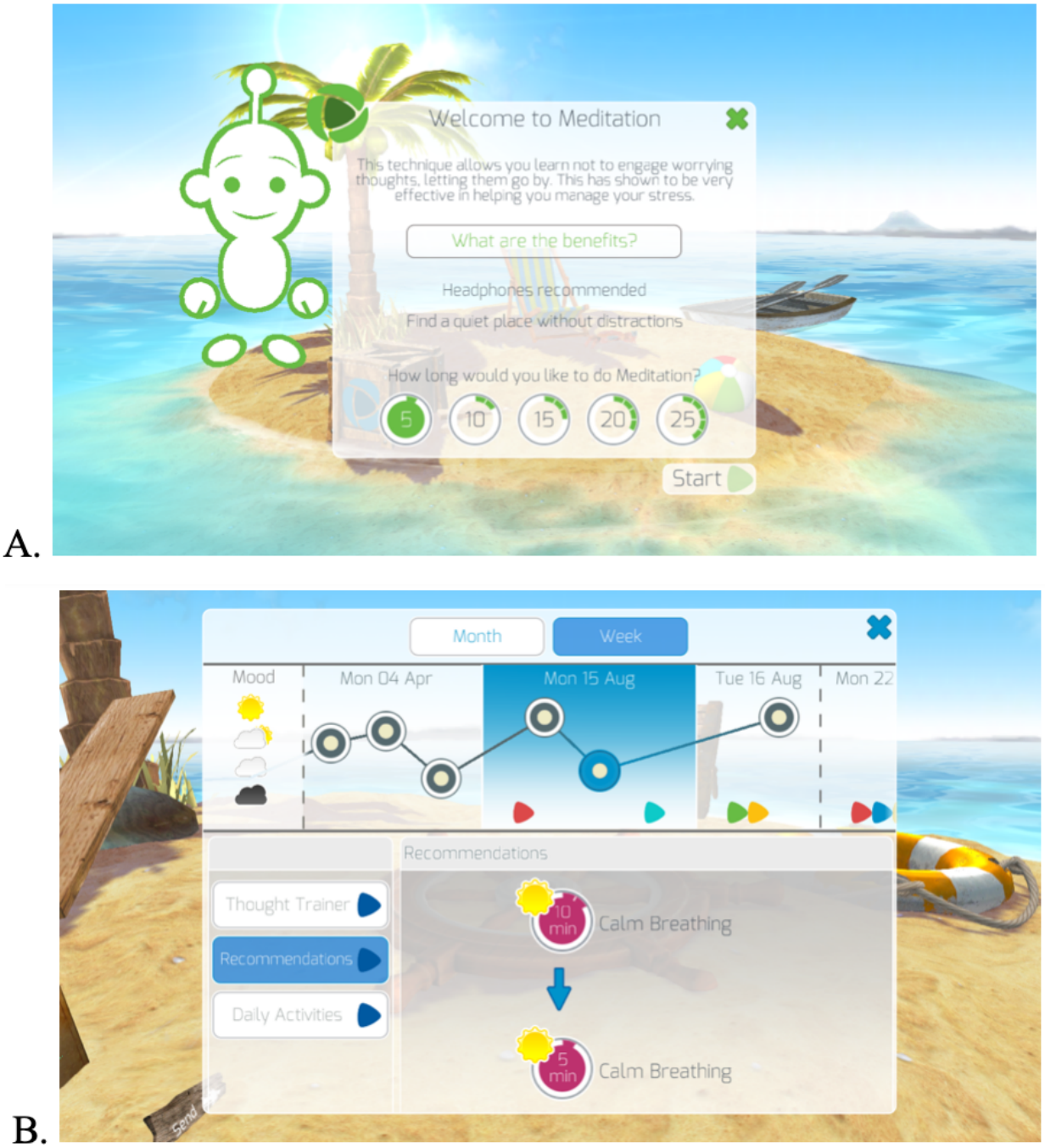
The researcher (TM) then emailed a link to download the *Feel Stress Free* app to those allocated to the intervention group. As participants were required to sign up for an account to use the app, it was possible to monitor the participants who had downloaded the app, and reminders were sent to those who had not yet done so. Those allocated to the wait-list group were sent an email informing them that they would receive access to the app at the end of the trial. Day 1 of the trial was defined as the date of randomization for each batch of participants. Owing to the nature of a wait-list control group, we could not blind the researcher or participants to group allocation. However, apart from the allocation email, all participants received exactly the same emails regardless of group, and trial staff had no other way of influencing the participants.

Intervention

The *Feel Stress Free* app (version 1.5; Figure 1), developed by Thrive Therapeutic Software Limited, uses CBT-based activities to help users manage symptoms of depression and anxiety. The app comprises 4 behavioral relaxation activities—calm breathing, mindfulness-style meditation, deep muscle relaxation, and self-hypnosis; one cognitive activity, incorporating both mood tracking and thought challenging; a relaxing minigame; and a feature for positive messages in a bottle. *Feel Stress Free* is self-guided (fully automated, with no additional human involvement), and individuals are led around the app by a friendly robot character that makes activity recommendations. Each activity has several options for duration and a short audiovisual guide explaining its use and benefits. Participants were instructed to use the app at least once per week, spending at least 10 min on one or more of the main activities, throughout the trial. There were no prompts or reminders to use the app. Further details of the app and its basis can be found in Multimedia Appendix 1. A very similar version of this app (version 1.3) was also tested as a treatment for agoraphobia in 2017 [30]. More information can be found at https://thrive.uk.com/ [31].

Feel Stress Free can be used on any Apple or Android smartphone or tablet or any computer with Firefox, Safari, or Chrome installed. Feel Stress Free is available on the web, although it is primarily a mobile app. Users must be connected to the internet. Participants randomized to receive the intervention were able to download and access the app and all its features free of charge. Those in the wait-list group were able to do so at the end of the study. The app was offered to participants exactly as it is publicly available. No app usage data were available to study researchers, as per Thrive's privacy policy. Participants in both groups were not limited in the additional care they could receive throughout the trial but were asked at baseline whether they were receiving any concurrent treatment (medication or psychological interventions) for anxiety or depression.

Figure 1. Screenshots of the Feel Stress Free app on a smartphone, showing the landing page for the meditation activity (A) and the mood tracker with an activity recommendation (B).



Measures

The HADS [27] is an established measure of the severity of anxiety and depression symptoms and has been validated for online use in a student population [32]. There are 7 items scored from 0 to 3 on each subscale, giving a possible range of 0 to 21 for each; a score between 8 and 10 inclusive indicates a possible case, and a score of 11 or above indicates a probable case [27]. Specificities and sensitivities are usually reported to be 80% or higher in UK-based research [33]. A previous study did not reveal any differences between online and pen-and-paper versions of the scale [32]. The internal consistency (Cronbach alpha) of the HADS in this study was .68 and .77 at baseline for the anxiety and depression subscales, respectively, and .81 and .86, respectively, at week 6.

The joint primary outcomes were depression and anxiety symptom severity at week 6, as measured by the 2 subscales of the HADS (the HADS-Anxiety Subscale, HADS-A, for anxiety symptoms and the HADS-Depression Subscale, HADS-D, for depression symptoms). Secondary outcomes were HADS-A and HADS-D scores as repeated measures at baseline (screening), week 2, and week 4. The link to complete the web-based questionnaires was sent to all participants via email on the first day of each of the relevant weeks, followed by prompts throughout the week.

Participants were asked how often they had been using the app in the past fortnight as a measure of treatment adherence (*not at all*, *fortnightly*, *weekly*, *a few times*, *several times a week*, *once a day*, *more than once a day*, or *I am not in the app group*). A response indicating weekly or more frequent usage was considered to indicate adherence. Participants also had the opportunity to indicate whether they had experienced any adverse events in the past 2 weeks. If a participant indicated that they had experienced an adverse event, details of the event and its severity (on a scale of 1-3, with 1 being mild and 3 being severe) were requested within the questionnaire.

Statistical Analyses

All main analyses were intention-to-treat (ITT) analyses. The primary outcomes (HADS-D and HADS-A scores at 6 weeks) were analyzed using linear mixed models (a separate model for each outcome), with the scores from each time point treated as a repeated measures outcome. Models were adjusted for age, gender, and concurrent treatment, as these variables were expected to be strongly associated with the outcome. An interaction between each covariate and time permitted the effect of the covariates to differ at each time point. We first report the effect on the primary outcomes (HADS-D and HADS-A scores at week 6) and then on the secondary outcomes. A random effect of participant with an unstructured residual covariance matrix allowed for correlations between these repeated measures on individuals over time. Fixed effects of treatment (intervention vs wait-list group), time (baseline and 2, 4, and 6 weeks follow-up), and the interaction between treatment and time were specified. The estimated baseline score was constrained to be identical in the 2 study arms, equivalent to adjusting for baseline and permitting the relationship between the baseline and follow-up scores to differ at each time point. Standardized effect sizes (ESs) were produced by performing similar analyses with outcome variables standardized by the mean and SD of the whole sample at baseline. A per-protocol analysis was also undertaken to examine the effectiveness of the intervention under ideal conditions. This analysis compared all participants in the wait-list group with only those participants in the intervention group who had indicated treatment adherence at all 3 time points, using statistical models similar to those of the main trial analysis. All analyses were conducted using Stata version 14 (StataCorp) [29].

Ethical Approval

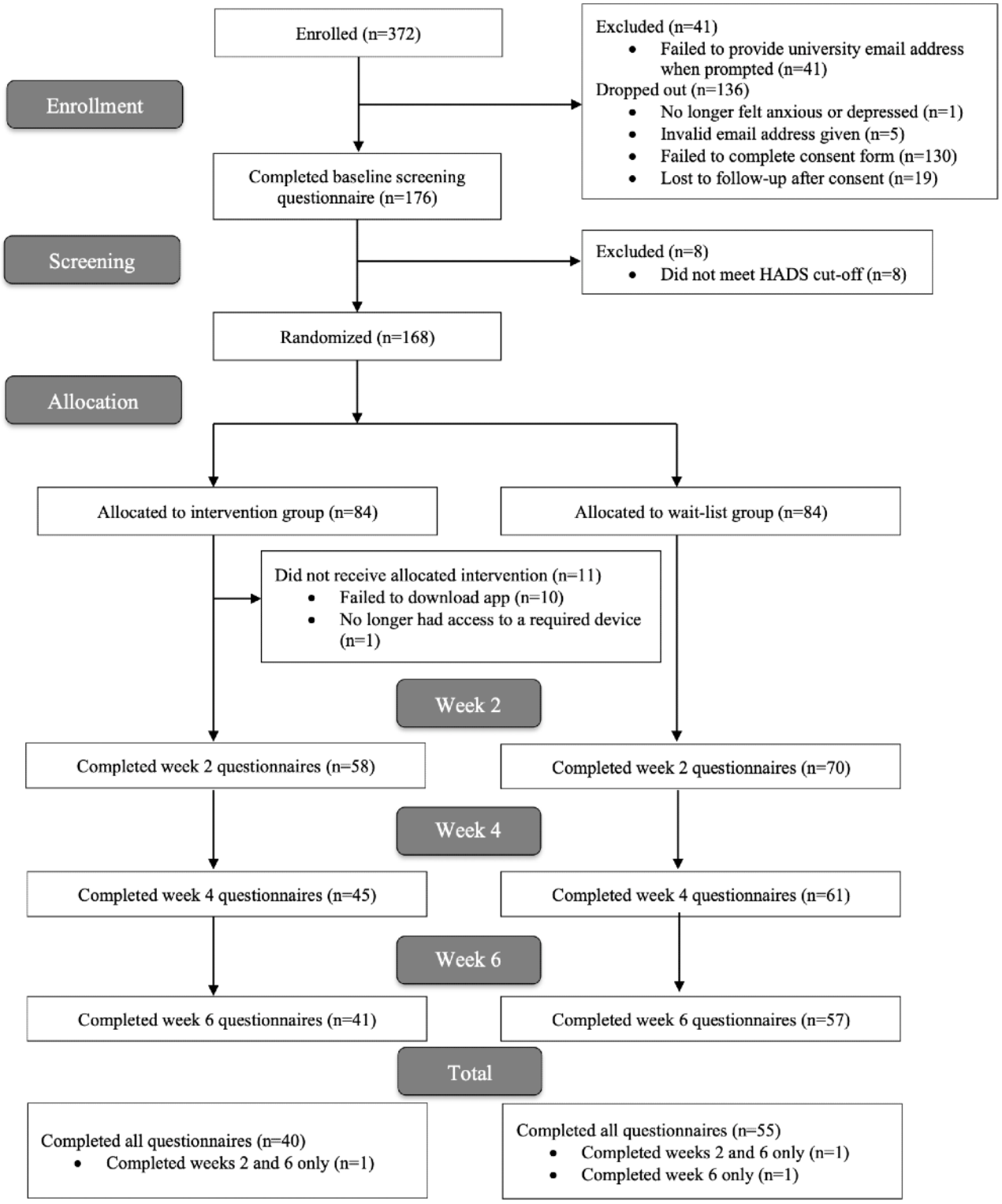
This research was approved by the UCL Ethics Committee (reference number 8227/001). All participants gave informed consent via a web-based form before participating, and adverse events were monitored at each time point. Data were anonymized, stored, and protected according to the UK Data Protection Act (2018) and the General Data Protection Regulation guidelines (2018). The trial was registered and reported in accordance with the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH) checklist [34]. Protocol changes can be found in Multimedia Appendix 2.

Results

Participants

During the recruitment period, 372 students enrolled in the trial; 195 out of 372 (52.4%) completed the consent form. The baseline questionnaires were sent to these consenting participants, and 176 of the 195 participants (90.3%) completed them, 8 of whom were ineligible. The remaining 168 participants were randomized—84 to each study arm. Figure 2 shows the flow of participants through the trial.

Figure 2. Flow of participants through the trial.



In the intervention group, one participant was unable to download the app owing to an unexpected lack of access to a device. A further 10 participants did not download the app. All participants were sent questionnaires at all 3 follow-up time points, regardless of whether they had completed the previous questionnaires. All participants were included in the main analyses in the groups to which they were randomized.

Baseline Characteristics

The mean age of all 168 participants was 24.3 years (SD 6.71; range 18-54 years), 82.7% (139/168) of the participants were female, and 61.9% (104/168) were undergraduate students. At baseline, the mean score on HADS-A was 13.7 (SD 3.33) and on HADS-D was 8.31 (SD 3.96). At baseline, 45% (75/168) of the participants reported that they would have an exam or a dissertation deadline during the trial period, and 25.0% (42/168) reported that they were receiving another treatment concurrently.

These characteristics were well balanced across groups, apart from concurrent treatment; more participants in the intervention group were receiving other forms of treatment during the study than in the wait-list control group (30% and 20%, respectively). Baseline demographic and clinical information for all randomized participants are presented in Table 1.

Table 1. Baseline demographic and clinical characteristics by study arm.

Characteristics	Intervention group (n=84)	Wait-list group (n=84)
Age (years), mean (SD)	25.1 (7.68)	23.5 (5.53)
Gender, n (%)		
Male	13 (15)	10 (12)
Female	69 (82)	74 (88)
Prefer not to say	2 (2)	0 (0)
University, n (%)		
University College London	37 (44)	37 (44)
University of Roehampton	34 (40)	35 (42)
School of Oriental and African Studies, University of London	8 (10)	7 (8)
University of Buckingham	5 (6)	5 (6)
Graduate status, n (%)		
Undergraduate	51 (61)	53 (63)
Postgraduate	33 (39)	31 (37)
Assessment during trial, n (%)	38 (45)	37 (44)
Concurrent treatment, n (%)	25 (30)	17 (20)
Participants with HADS[a] score >8[b], n (%)		
HADS-Anxiety Subscale (HADS-A)	83 (99)	82 (98)
HADS-Depression Subscale (HADS-D)	46 (55)	45 (54)
Comorbid depression and anxiety	45 (54)	43 (51)

[a]HADS: Hospital Anxiety and Depression Scale.

[b]A Hospital Anxiety and Depression Scale (HADS) subscale score of more than 8 indicates possible depression or anxiety. Participants who met this criterion on both subscales of the HADS are additionally indicated as having comorbid depression and anxiety, respectively.

Main Analyses

The mean scores for each HADS subscale by study arm and time point are shown in Table 2. We found evidence that the *Feel Stress Free* app reduced depression at 6 weeks follow-up (adjusted mean difference [MD] −1.56; 95% CI −2.67 to −0.44; *P*=.006; standardized ES=0.39); but only very weak evidence was found of a reduction in anxiety (adjusted MD −1.36; 95% CI −2.93 to 0.21; *P*=.09). At week 4, there was evidence of the effectiveness of the app in reducing symptoms of both depression (adjusted MD −1.08; 95% CI −2.12 to −0.04; *P*=.04; ES=0.27) and anxiety (adjusted MD −1.94; 95% CI −3.11 to −0.77; *P*=.001; ES=0.58). There was weak evidence for an effect of the intervention on anxiety symptoms at week 2, but no evidence of a treatment effect for depression symptoms. These results were consistent with the unadjusted model, with the exception of week 2, where there was weak evidence of the app's effectiveness in reducing both anxiety and depression symptoms in the unadjusted model. The multilevel model results are presented in Table 3. Figure 3 shows the HADS anxiety and depression subscale scores by study arm, estimated by model 2. A sensitivity analysis adjusting additionally for university can be found in Multimedia Appendix 3 (first table).

Table 2. Hospital Anxiety and Depression Scale anxiety and depression scores by study arm at baseline and at 2, 4, and 6 weeks of follow-up.

Scale and time point	Intervention group		Wait-list group	
	n	Mean (SD)	n	Mean (SD)
HADS[a]-Anxiety Subscale				
Week 6	41	10.8 (4.25)	57	11.8 (4.60)
Week 4	45	10.1 (3.85)	61	12.1 (4.19)
Week 2	58	11.2 (3.75)	70	12.6 (3.88)
Baseline	84	13.4 (3.25)	84	13.9 (3.41)
HADS-Depression Subscale				
Week 6	41	5.8 (3.72)	57	6.6 (4.07)
Week 4	45	5.9 (3.63)	61	6.6 (3.61)
Week 2	58	6.3 (3.20)	70	6.9 (3.84)
Baseline	84	8.3 (3.73)	84	8.3 (4.20)

[a]HADS: Hospital Anxiety and Depression Scale.

Table 3. Estimated effect of Feel Stress Free intervention on Hospital Anxiety and Depression Scale anxiety and depression scores at 2, 4, and 6 weeks of follow-up.

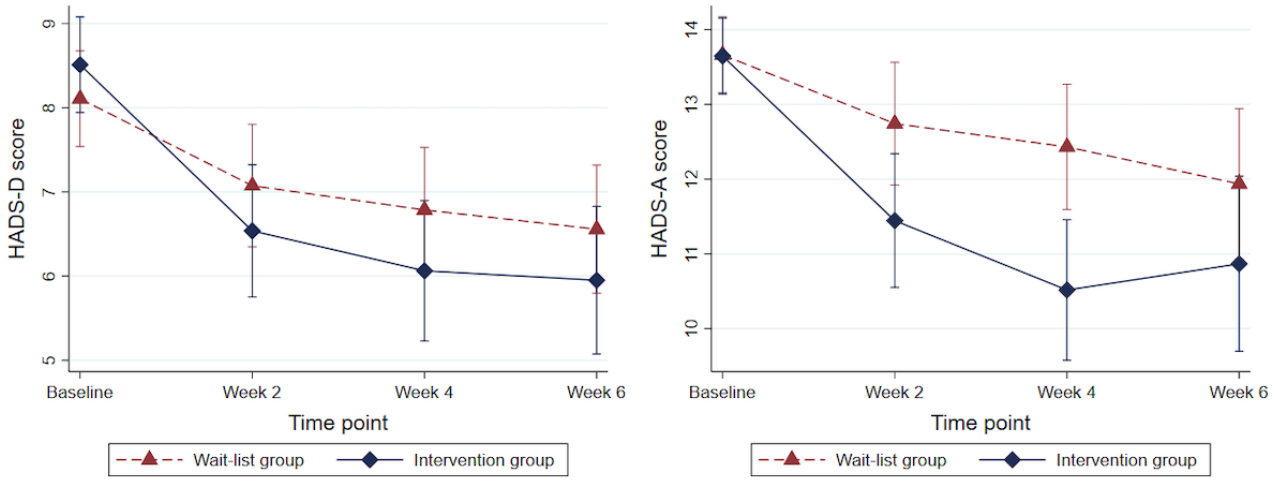
Timepoint	Model 1[a]			Model 2[a]		
	Estimate (95% CI)	*P* value	Effect size[b]	Estimate (95% CI)	*P* value	Effect size[b]
HADS[c]-Anxiety Subscale						
Primary						
Week 6	−1.08 (−2.62 to 0.47)	.17	0.32	−1.36 (−2.93 to 0.21)	.09	0.41
Secondary						
Week 4	−1.94 (−3.06 to −0.82)	.001	0.58	−1.94 (−3.11 to −0.77)	.001	0.58
Week 2	−1.27 (−2.39 to −0.15)	.03	0.38	−1.10 (−2.28 to 0.07)	.07	0.33
HADS-Depression Subscale						
Primary						
Week 6	−1.26 (−2.37 to −0.16)	.03	0.32	−1.56 (−2.67 to −0.44)	.006	0.39
Secondary						
Week 4	−1.20 (−2.21 to −0.19)	.02	0.30	−1.08 (−2.12 to −0.04)	.04	0.27
Week 2	−0.98 (−1.91 to −0.06)	.04	0.25	−0.67 (−1.62 to 0.27)	.16	0.17

[a]Estimates are from linear mixed models, with scores from each time point treated as a repeated measures outcome. In both models, the baseline score was constrained to be identical in the 2 study arms, equivalent to adjusting for baseline. Model 2 adjusted additionally for age, gender, and presence of concurrent treatment.

[b]Effect size standardized by mean and SD of sample at baseline.

[c]HADS: Hospital Anxiety and Depression Scale.

Figure 3. Hospital Anxiety and Depression Scale anxiety and depression scores by study arm, estimated by the adjusted multilevel model at each time point. Error bars represent 95% CIs. HADS-A: Hospital Anxiety and Depression Scale-Anxiety Subscale; HADS-D: Hospital Anxiety and Depression Scale-Depression Subscale.



Attrition and Adherence

Overall, 58.3% (98/168) of the randomized participants completed the week 6 questionnaires. A larger proportion of participants in the wait-list group provided outcome data at week 6 (57/84, 68%) than in the intervention group (41/84, 49%; χ^2_1=6.3; *P*=.01). We used univariable logistic regression models to identify baseline variables associated with having incomplete outcome data (missing questionnaire data at any time point). Postgraduates were less likely to have missing outcome data than undergraduates (odds ratio 0.44; 95% CI 0.23 to 0.84; *P*=.01), but there were no differences in terms of any other baseline characteristics. Therefore, we conducted a sensitivity analysis adjusting for graduate status to investigate the possible effect of missing outcome data. The treatment effect was similar. More details can be found in Multimedia Appendix 3 (second table).

Adherence to the treatment was defined as using the app weekly or more frequently. By this definition, 98% of the participants reported adherence at week 2, 89% at week 4, and 83% at week 6 (Table 4). Of the participants who completed all the follow-up questionnaires, 80% (32/40) indicated adherence at every time point.

Table 4. Usage data for the intervention group at each time point. Adherence was defined as using the app weekly or more frequently, measured via self-report.

Usage	Week 2 (n=56), n (%)	Week 4 (n=45), n (%)	Week 6 (n=41), n (%)
Less than weekly	1 (2)	5 (11)	7 (17)
Weekly	9 (16)	10 (22)	6 (15)
A few times a week	28 (48)	19 (42)	20 (49)
Several times a week	17 (29)	8 (18)	5 (12)
Daily or more	3 (5)	3 (7)	3 (7)

Per-Protocol Analysis

We performed a prespecified per-protocol analysis including only those in the intervention group who indicated treatment adherence at every time point (32/84, 38%) and all participants in the wait-list control group (n=84). Although no longer an unbiased sample, the estimated adjusted treatment effects at the primary end point (week 6) were similar to those from the ITT analysis for both anxiety symptoms (adjusted MD −1.37; 95% CI −3.17 to 0.42; *P*=.13) and depression symptoms (adjusted MD −1.88; 95% CI −3.12 to −0.64; *P*=.003). At week 4, there was also evidence to support the effectiveness of the app at reducing both anxiety (adjusted MD −2.11; 95% CI −3.50 to −0.73, *P*=.003) and depression symptoms (adjusted MD −1.13; 95% CI −2.35 to 0.08; *P*=.07), again in line with the ITT analysis.

Safety

Two participants in the app group reported adverse events associated with the intervention. One participant reported at week 4 that they were “feeling reliant on it” and “recognising deeper thoughts and emotions” and rated the severity of these unpleasant effects as mild. The second participant reported that they experienced feelings of stress and anxiety when they had technical difficulties with the app, which they reported as being of moderate severity at week 2 and of mild severity at week 6. There were no other reports of any adverse events experienced as a result of the intervention.

Discussion

Principal Findings

We found preliminary evidence that the *Feel Stress Free* app reduced depression and anxiety symptoms after 6 weeks. At secondary time points, we observed greater reductions in depression and anxiety symptoms in the intervention group compared with the control group after 4 weeks. Treatment adherence and usage of the app were encouraging (83% reported using the app weekly or more frequently at week 6), and very few adverse events were reported. These results provide preliminary support for the effectiveness of this mobile CBT-based app in treating depression and anxiety symptoms in a student population.

The standardized ESs for statistically significant comparisons were 0.39 (week 6) and 0.27 (week 4) for depression, which are considered small [35], and 0.58 (week 4) for anxiety, which are considered medium [35]. These are in line with other recent studies that compare CBT apps with wait-list or inactive controls; meta-analyses by Firth et al reported standardized mean difference ESs of 0.45 for anxiety [36] and 0.56 for depression [37]. Although these meta-analyses reported smaller ESs for studies comparing similar apps with active controls (0.19 and 0.22 for anxiety and depression, respectively), for students on a waiting list or for whom therapist-supported CBT or face-to-face CBT is not an option, a readily available intervention that produces even a relatively small effect after 4 weeks would be of value.

Although a minimum clinically important difference (MCID) has not been established for the HADS in the general population, the best estimate we have is from trials of those with chronic pulmonary obstructive disorder. Although MCIDs are a complex issue likely to be best represented by percentage rather than absolute score changes [38], in these trials, a change in score of 1.5 points is generally accepted as the MCID [39,40]. Our adjusted mean difference estimates slightly exceeded this value at week 6 for depression but was below it at week 4. For anxiety, the adjusted mean difference estimate was below this value (and not statistically significant) at week 6 but exceeded this value at week 4. On the basis of 95% CIs for these estimates, we cannot rule out the potential for a clinically significant effect of this intervention, but we also cannot rule out the values in the lower range, which would be small and not clinically important. Overall, these findings should be treated with caution regarding clinical importance.

Strengths

This study avoided many methodological issues common throughout the internet-based CBT literature. In particular, adherence to the app-based treatment was clearly defined and measured and not just assumed from the completion of outcome measures. This allowed us to make observations regarding app usage separate from study participation—for instance, postgraduates were less likely to have missing data than undergraduates, but not more likely to indicate treatment adherence, as may have been assumed otherwise. We can, therefore, report the important positive outcome that 83% of those in the intervention group continued to use the app regularly at week 6.

Offering the app to users as it would be in a real-world setting and without supervision allowed the study to have a higher level of external validity than most trials. In addition, the ITT approach to analysis provides a more valid estimate of the effectiveness of the intervention, by reflecting the protocol deviations and noncompliance common in clinical practice. Participants self-administered all outcome measures online and had no personalized contact with trial personnel, which means that observer bias can be discounted.

This study contributes new data to the body of literature surrounding cCBT and more specifically to the emerging field of mobile- and app-based CBT. Our findings suggest that mobile CBT without therapist support or a structured session-based approach may improve symptoms of anxiety and depression in a sample with a range of symptom severities, including subclinical and severe. This study also provides preliminary evidence for the effectiveness of the attributes of this particular app, which can be directly compared with other apps of this kind to consider the ideal characteristics of cCBT and internet-based CBT to optimize effectiveness.

Limitations

This study has several limitations. One issue was attrition, as only 58% of the participants completed the week 6 questionnaires. However, this rate is comparable with what is seen at similar time points in other trials of mobile- and web-based CBT where therapist contact is minimal or absent [28,41,42]. Furthermore, although the sample size at week 6 was relatively small, the multilevel modeling analysis technique used allows all participants to contribute to analyses, even those with missing follow-up data, which increases the statistical power. Nevertheless, future research should focus on trying to ascertain the reasons behind dropouts, particularly in relation to app design. In this study, it is unclear whether those in the intervention group who stopped completing questionnaires were doing so because of the questionnaires or because they did not want or need to use the app any more. Fewer participants dropped out in the wait-list group (Table 2), which may be because they remained interested in the trial as a way to access the app. Overall, the attrition observed appears to be realistic for a web-based trial comparing a CBT-based app with a wait-list control group [42]. Future web-based trials could try to incorporate some elements of interaction with participants to improve retention, for example, an in-person consent procedure and demonstration of the app and regular reminders to use the app.

It must be considered that our estimate of adherence was based on self-reported data subject to social desirability bias, and the true rate of adherence may be lower. However, if we did overestimate adherence, the fact that low levels of usage produced a benefit would provide support for the effectiveness of the app even without frequent use, in line with findings by Firth et al [36]. Nevertheless, more detailed data on participants' usage of the app, potentially including qualitative feedback on its perceived usefulness and user experience, would be beneficial to the field. Future studies should incorporate direct monitoring

of the participants' app usage where possible to ascertain whether there is a dose-response relationship and whether usage of certain activities within the app may be associated with better outcomes and/or preferred by participants [43].

A further limitation of this study is the use of a wait-list control group. As it was not possible to blind participants, the groups differed in their expectation of improvement, and a placebo effect in the intervention group could have inflated ESs or made a type I error more likely. In particular, a so-called *digital placebo* has been suggested by some researchers, in that the observed benefits could arise from the increased use of the electronic device itself [44]. Nevertheless, this intervention requires such little investment on the part of distributors and patients alike that this would not discredit its use entirely, particularly for those who are waiting for another treatment. In this way, our comparison between a mobile CBT-based app and a control group lacking in expectation of improvement can be considered to reflect the real-world situation of those at whom this intervention is aimed and should be considered in this context of increased external validity. In addition, the wait-list group also showed improvement during the trial, thus reducing the differences between groups. This is common in trials and likely because of spontaneous improvement; fluctuations in symptoms; and regression to the mean, all of which accompany the passage of time [45]. However, we cannot rule out the possibility that the control group sought out and used other CBT-based mental health apps or other treatments during the course of the trial, thus diminishing the observed effect of the *Feel Stress Free* app.

Our sample was 85% female, which could limit the generalizability of our results. Although we had more female students taking part than expected, some gender imbalance is common in trials and likely to partly reflect gender differences in prevalence and help-seeking behaviors. Further research could investigate whether female students are more interested in mobile app interventions than male students, and apps could be then tailored accordingly. The results of this trial may also not be generalizable beyond university students, but arguably, these are the individuals who are most likely to benefit from the accessibility of this type of treatment [46]. The sample was a mix of those with subclinical and clinical symptoms to reflect this population. The promising results indicate the possible effectiveness of the app in potentially treating and preventing clinical levels of symptoms.

Future Directions

Further research is needed to confirm the effectiveness of this mobile intervention at and beyond 6 weeks, particularly for anxiety. Our results are in line with the oft-reported finding that the majority of the improvement associated with a CBT app (and traditional CBT [47]) is seen in the short term [48,49], but a longer follow-up period will be needed to ascertain whether our week 6 results represent a fluctuation typically seen around the 6-week time point. To rule out the digital placebo effect as an explanation for our results, this app should also be compared with an active, smartphone-based control. As more evidence is gathered for individual CBT apps, comparisons between different apps are warranted, followed by a consideration of the characteristics that the most effective apps have in common—a recent meta-analysis attempted this but was not able to draw robust conclusions [37].

This app could be offered to university students, particularly where the demand for therapies exceeds the provision. The app could be offered to those who are on the waiting list for traditional CBT; it may be particularly useful when a rapid improvement is imperative, such as before exams. Trying self-directed forms of CBT may also make individuals more likely to seek traditional CBT [50] and can be an alternative for those who are unable to attend therapy. Students are unlikely to have a problem accessing a device on which the app can be used, and benefits were observed regardless of age and gender. Internet-based interventions such as *Feel Stress Free* would also fit well within a stepped care model, such as that currently in place in the United Kingdom; it could be offered during the period of guided self-help [15,,51], followed by traditional CBT if this is not beneficial. Future research could also explore the use of internet-based interventions as an adjunct to traditional CBT, for use between sessions.

Conclusions

We found preliminary evidence to support the use of *Feel Stress Free*, a CBT-based mobile app, as a short-term intervention for students experiencing symptoms of depression and anxiety. Further research is needed to establish the potential benefit of the app, in particular by comparing it with an active control in a larger sample less vulnerable to dropouts.

Acknowledgments

TM is funded by the Economic and Social Research Council and the Medical Research Council.

Conflicts of Interest

Dr Andres Fonseca is a consultant psychiatrist and Co-Founder of Thrive Therapeutic Software Ltd, the company that developed the app. This trial was designed in collaboration with Dr Fonseca, and Thrive provided the app to participants for free, assisted with technical difficulties and contacting universities, and covered recruitment costs. Preparation of this manuscript was completed by the authors independently of Thrive, and Thrive did not have any involvement in review or approval of the final manuscript. The authors report no other conflicts of interest.

Multimedia Appendix 1

Feel Stress Free app activities, rationale and screenshots.
[DOCX File , 1412 KB - mhealth_v8i7e15418_app1.docx]

Multimedia Appendix 2
Protocol changes.
[DOCX File , 12 KB - mhealth_v8i7e15418_app2.docx]

Multimedia Appendix 3
Sensitivity analyses.
[DOCX File , 18 KB - mhealth_v8i7e15418_app3.docx]

Multimedia Appendix 4
CONSORT EHEALTH form V1.6.
[PDF File (Adobe PDF File), 25307 KB - mhealth_v8i7e15418_app4.pdf]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the national comorbidity survey replication (NCS-R). J Am Med Assoc 2003 Jun 18;289(23):3095-3105. [doi: 10.1001/jama.289.23.3095] [Medline: 12813115]
2. Kessler RC, Angermeyer M, Anthony JC, de Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the world health organization's world mental health survey initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
3. Thorley C. Not By Degrees: Improving Student Mental Health in the UK's Universities. IPPR. 2017. URL: http://www.ippr.org/research/publications/not-by-degrees [accessed 2020-06-16]
4. Ibrahim AK, Kelly SJ, Adams CE, Glazebrook C. A systematic review of studies of depression prevalence in university students. J Psychiatr Res 2013 Mar;47(3):391-400. [doi: 10.1016/j.jpsychires.2012.11.015] [Medline: 23260171]
5. Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA. Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. J Abnorm Psychol 1993 Feb;102(1):133-144. [doi: 10.1037//0021-843x.102.1.133] [Medline: 8436689]
6. Woodward LJ, Fergusson DM. Life course outcomes of young people with anxiety disorders in adolescence. J Am Acad Child Adolesc Psychiatry 2001 Sep;40(9):1086-1093. [doi: 10.1097/00004583-200109000-00018] [Medline: 11556633]
7. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]
8. Cavanagh K. Geographic inequity in the availability of cognitive behavioural therapy in England and Wales: a 10-year update. Behav Cogn Psychother 2014 Jul;42(4):497-501. [doi: 10.1017/S1352465813000568] [Medline: 23829940]
9. #stepchange: Mental Health in Higher Education. Universities UK. URL: http://www.universitiesuk.ac.uk/stepchange [accessed 2019-02-26]
10. Kaltenthaler E, Brazier J, de Nigris E, Tumur I, Ferriter M, Beverley C, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess 2006 Sep;10(33):iii, xi-ixiv, 1 [FREE Full text] [doi: 10.3310/hta10330] [Medline: 16959169]
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
13. 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. BMJ Open 2014 Jan 9;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
14. 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]
15. Depression in Adults: Recognition and Management (CG90). The National Institute for Health and Care Excellence (NICE). 2009. URL: https://www.nice.org.uk/guidance/cg90/chapter/1-Guidance [accessed 2019-02-26]
16. Mobile Internet: Smartphones for Everyone? Nakono — UK Consumer Product Reviews & Market Research. 2016. URL: https://www.nakono.com/report/mobile-internet-smartphones-for-everyone/ [accessed 2016-09-05]
17. Global Mobile Consumer Survey: UK Cut. Deloitte. 2019. URL: https://www2.deloitte.com/uk/en/pages/technology-media-and-telecommunications/articles/mobile-consumer.html [accessed 2020-01-06]

18. UK Digital Market Overview Report Q3 2019. UK Online Measurement (UKOM). London, UK; 2019. URL: https://ukom.uk.net/digital-market-overview/167-q3-2019-uk-digital-market-overview-report.php [accessed 2020-01-06]
19. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
20. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
21. Kitzrow M. The mental health needs of today's college students: challenges and recommendations. J Stud Aff Res Pract 2003;41(1):167-181. [doi: 10.2202/1949-6605.1310]
22. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
23. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
24. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 7;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
25. Stoll R, Pina A, Gary K, Amresh A. Usability of a smartphone application to support the prevention and early intervention of anxiety in youth. Cogn Behav Pract 2017 Nov;24(4):393-404 [FREE Full text] [doi: 10.1016/j.cbpra.2016.11.002] [Medline: 29056845]
26. Evidence Standards Framework for Digital Health Technologies. The National Institute for Health and Care Excellence (NICE). 2018. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2019-02-26]
27. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
28. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
29. StataCorp M, Banks J. tata Statistical Software: Release 14. Texas: College Station; 2015. URL: https://www.stata.com/
30. Christoforou M, Fonseca JA, Tsakanikos E. Two novel cognitive behavioral therapy-based mobile apps for agoraphobia: randomized controlled trial. J Med Internet Res 2017 Nov 24;19(11):e398 [FREE Full text] [doi: 10.2196/jmir.7747] [Medline: 29175809]
31. Herrmann C. International experiences with the hospital anxiety and depression scale--a review of validation data and clinical results. J Psychosom Res 1997 Jan;42(1):17-41. [doi: 10.1016/s0022-3999(96)00216-4] [Medline: 9055211]
32. Whitehead L. Methodological issues in Internet-mediated research: a randomized comparison of internet versus mailed questionnaires. J Med Internet Res 2011 Dec 4;13(4):e109 [FREE Full text] [doi: 10.2196/jmir.1593] [Medline: 22155721]
33. Herrmann C. International experiences with the hospital anxiety and depression scale--a review of validation data and clinical results. J Psychosom Res 1997 Jan;42(1):17-41. [doi: 10.1016/s0022-3999(96)00216-4] [Medline: 9055211]
34. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
35. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Hillsdale, NJ: Erlbaum; 1988.
36. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
37. 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]
38. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, et al. Minimal clinically important difference on the beck depression inventory--ii according to the patient's perspective. Psychol Med 2015 Nov;45(15):3269-3279 [FREE Full text] [doi: 10.1017/S0033291715001270] [Medline: 26165748]
39. Puhan MA, Frey M, Büchi S, Schünemann HJ. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. Health Qual Life Outcomes 2008 Jul 2;6:46 [FREE Full text] [doi: 10.1186/1477-7525-6-46] [Medline: 18597689]
40. Bhandari NJ, Jain T, Marolda C, ZuWallack RL. Comprehensive pulmonary rehabilitation results in clinically meaningful improvements in anxiety and depression in patients with chronic obstructive pulmonary disease. J Cardiopulm Rehabil Prev 2013;33(2):123-127. [doi: 10.1097/HCR.0b013e31828254d4] [Medline: 23399845]
41. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the internet (ODIN): a randomized controlled trial of an internet depression skills intervention program. J Med Internet Res 2002 Dec;4(3):E14 [FREE Full text] [doi: 10.2196/jmir.4.3.e14] [Medline: 12554545]

42. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
43. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
44. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
45. Rutherford BR, Mori S, Sneed JR, Pimontel MA, Roose SP. Contribution of spontaneous improvement to placebo response in depression: a meta-analytic review. J Psychiatr Res 2012 Jun;46(6):697-702 [FREE Full text] [doi: 10.1016/j.jpsychires.2012.02.008] [Medline: 22410207]
46. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]
47. Ilardi S, Craighead W. The role of nonspecific factors in cognitive-behavior therapy for depression. Clin Psychol Sci Pract 1994;1(2):138-155. [doi: 10.1111/j.1468-2850.1994.tb00016.x]
48. Warmerdam L, van Straten A, Jongsma J, Twisk J, Cuijpers P. Online cognitive behavioral therapy and problem-solving therapy for depressive symptoms: exploring mechanisms of change. J Behav Ther Exp Psychiatry 2010 Mar;41(1):64-70. [doi: 10.1016/j.jbtep.2009.10.003] [Medline: 19913781]
49. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs computer. BMC Psychiatry 2013 Feb 7;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
50. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an internet-based cognitive behaviour therapy program. Aust N Z J Psychiatry 2006 Jan;40(1):59-62. [doi: 10.1080/j.1440-1614.2006.01743.x] [Medline: 16403040]
51. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]

Abbreviations

CBT: cognitive behavioral therapy
cCBT: computerized forms of self-directed cognitive behavioral therapy
ES: effect size
HADS: Hospital Anxiety and Depression Scale
HADS-A: Hospital Anxiety and Depression Scale-Anxiety Subscale
HADS-D: Hospital Anxiety and Depression Scale-Depression Subscale
ITT: intention-to-treat
MCID: minimum clinically important difference
UCL: University College London

Edited by G Eysenbach; submitted 12.07.19; peer-reviewed by R Morriss, K Kaipainen, A Daros, P Yellowlees; comments to author 21.10.19; revised version received 14.02.20; accepted 26.04.20; published 31.07.20.

Please cite as:
McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E
Effectiveness of a Mobile App Intervention for Anxiety and Depression Symptoms in University Students: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e15418
URL: http://mhealth.jmir.org/2020/7/e15418/
doi:10.2196/15418
PMID:32735221

Original Paper

Effects of an mHealth Brisk Walking Intervention on Increasing Physical Activity in Older People With Cognitive Frailty: Pilot Randomized Controlled Trial

Rick YC Kwan[1*], PhD; Deborah Lee[1*], MSc; Paul H Lee[2*], PhD; Mimi Tse[1*], PhD; Daphne SK Cheung[1*], PhD; Ladda Thiamwong[3*], PhD; Kup-Sze Choi[2*], PhD

[1]Centre for Gerontological Nursing, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, Hong Kong (China)

[2]School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (China)

[3]College of Nursing, University of Central Florida, Orlando, FL, United States

[*]all authors contributed equally

Corresponding Author:
Rick YC Kwan, PhD
Centre for Gerontological Nursing, School of Nursing
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Hong Kong (China)
Phone: 852 27666546 ext 6546
Email: rick.kwan@polyu.edu.hk

Abstract

Background: Cognitive frailty is the coexistence of physical frailty and cognitive impairment and is an at-risk state for many adverse health outcomes. Moderate-to-vigorous physical activity (MVPA) is protective against the progression of cognitive frailty. Physical inactivity is common in older people, and brisk walking is a feasible form of physical activity that can enhance their MVPA. Mobile health (mHealth) employing persuasive technology has been successful in increasing the levels of physical activity in older people. However, its feasibility and effects on older people with cognitive frailty are unclear.

Objective: We aimed to identify the issues related to the feasibility of an mHealth intervention and the trial (ie, recruitment, retention, participation, and compliance) and to examine the effects of the intervention on cognitive function, physical frailty, walking time, and MVPA.

Methods: An open-label, parallel design, randomized controlled trial (RCT) was employed. The eligibility criteria for the participants were age ≥60 years, having cognitive frailty, and having physical inactivity. In the intervention group, participants received both conventional behavior change intervention and mHealth (ie, smartphone-assisted program using Samsung Health and WhatsApp) interventions. In the control group, participants received conventional behavior change intervention only. The outcomes included cognitive function, frailty, walking time, and MVPA. Permuted block randomization in 1:1 ratio was used. The feasibility issue was described in terms of participant recruitment, retention, participation, and compliance. Wilcoxon signed-rank test was used to test the within-group effects in both groups separately.

Results: We recruited 99 participants; 33 eligible participants were randomized into either the intervention group (n=16) or the control (n=17) group. The median age was 71.0 years (IQR 9.0) and the majority of them were females (28/33, 85%). The recruitment rate was 33% (33/99), the participant retention rate was 91% (30/33), and the attendance rate of all the face-to-face sessions was 100% (33/33). The majority of the smartphone messages were read by the participants within 30 minutes (91/216, 42.1%). ActiGraph (58/66 days, 88%) and smartphone (54/56 days, 97%) wearing compliances were good. After the interventions, cognitive function improvement was significant in both the intervention (*P*=.003) and the control (*P*=.009) groups. The increase in frailty reduction (*P*=.005), walking time (*P*=.03), step count (*P*=.02), brisk walking time (*P*=.009), peak cadence (*P*=.003), and MVPA time (*P*=.02) were significant only in the intervention group.

Conclusions: Our mHealth intervention is feasible for implementation in older people with cognitive impairment and is effective at enhancing compliance with the brisk walking training program delivered by the conventional behavior change interventions. We provide preliminary evidence that this mHealth intervention can increase MVPA time to an extent sufficient to yield clinical

benefits (ie, reduction in cognitive frailty). A full-powered and assessor-blinded RCT should be employed in the future to warrant these effects.

Trial Registration: HKU Clinical Trials Registry HKUCTR-2283; http://www.hkuctr.com/Study/Show/31df4708944944bd99e730d839db4756

(JMIR Mhealth Uhealth 2020;8(7):e16596) doi:10.2196/16596

KEYWORDS

cognitive frailty; brisk walking; smartphone; moderate-to-vigorous physical activity; older people

Introduction

Cognitive frailty is a heterogeneous clinical syndrome characterized by the coexistence of both physical frailty and cognitive impairment without being severe enough to fulfil the criteria for dementia [1]. Physical frailty is a phenotype of slowness, muscle weakness, less physical activity, exhaustion, and weight loss [2]. Cognitive frailty is common in community-dwelling older people, with a prevalence of 2.4%-8.9% [3,4]. It is an at-risk state for many adverse health outcomes, including dementia, dependency, and mortality [5-7].

Physical activity is protective against the progression of cognitive frailty because it optimizes the neurobiological conditions that cause cognitive frailty (eg, glucose metabolism, sarcopenia, insulin resistance) [8-10]. Physical inactivity is defined by the World Health Organization (WHO) as the performance of less than 150 minutes of moderate-to-vigorous physical activity (MVPA) per week [11]. The WHO has also advised that engaging in MVPA for 150 minutes per week, with each session lasting not less than 10 minutes, will lead to health benefits [11]. Physical inactivity is one of the key phenotypical characteristics of frailty [2]. There is evidence that the intensity of physical activity plays an important role in yielding favorable health outcomes in older people. A study showed that replacing sitting time with MVPA was associated with a significant decrease in frailty but replacing sitting time with light physical activity was not [12]. Despite the beneficial effects of MVPA, physical inactivity is still common in older people, with a prevalence of over 60% in the United States and over 40% in China [13,14]. A systematic review showed that the prevalence of physical inactivity increases with age in older people [15].

Walking is the most common and inexpensive form of physical activity for older people and makes up approximately 80% of the total amount of physical activity that they engage in during their leisure time [16]. An ActiGraph accelerometer validation study showed that walking at a speed of >3.2 km/h or >100 steps/min fulfilled the criteria of MVPA, which is 3 metabolic equivalents in older people [17,18]. Therefore, brisk walking is a feasible form of physical activity for older people to enhance their MVPA.

Behavioral change interventions employing various behavioral change techniques have been successful at changing the behavior of different populations, from being physically inactive to becoming physically active [19]. In the process of ageing, many factors hinder older people from leading a physically active life, including poor health, a lack of company, a lack of interest, a lack of skills, and a lack of opportunities [20,21]. These factors can diminish the effects of a behavioral change intervention aimed at increasing their levels of physical activity. A systematic review showed that the effect size of a behavioral change intervention in older people was small (d=0.14) because many self-regulation intervention techniques that are effective for younger adults may not be effective for older adults [22]. There is a need to maximize the effect for this vulnerable group to ensure that older people with cognitive frailty are physically active enough to promote their health.

Mobile health (mHealth, also known as eHealth) refers to the health services delivered or enhanced through mobile/electronic-related technology [23]. Persuasive technology is a branch of mHealth in which the aim is to use digital technology to guide users to change their attitudes and behavior by enhancing the effects of behavioral change techniques [24]. mHealth interventions show promise in encouraging older people to increase their levels of physical activity as reported in systematic reviews [25-28]. Because there is a lack of properly designed trials in this area, none of these systematic reviews drew conclusions on whether the mHealth is more effective or can enhance the effect of the conventional intervention to promote physical activity in older people.

Older people with cognitive frailty are more vulnerable than robust older people to engage in physical activity because cognitive impairment and physical frailty reduce their intrinsic motivation through mechanisms such as neurological and muscular damage [29,30]. They have much lower baseline motivation than the robust older people. Baseline motivation is known to be associated with the effect of behavioral changes [31]. Moreover, compared to the robust people, older people with cognitive frailty have a lower baseline physical activity level [2]. The effect of a behavioral intervention promoting physical activity is known to be weaker in people with lower baseline physical activity [32]. Thus, the promising effect of eHealth interventions in robust older people may be hindered in translating to older people with cognitive frailty, let alone their effects on the clinical outcomes of cognitive frailty.

In the previous trials, most of the mHealth interventions that were included in the reviews used only websites, DVDs, and texting instead of face-to-face physical activity training and health education [25-28]. Recently, studies have shown that the newly developed technologies, including wearable devices and social media, showed promising effects in promoting physical activity in many populations (eg, young people, cancer survivors) [33-35]. However, there is a lack of trials examining whether the new mHealth technologies are more effective than the conventional behavior change programs (eg, face-to-face

behavioral counselling, education). We hypothesize that the mHealth intervention employing the recently developed technologies is feasible and can enhance the behavioral change effect to promote physical activity in older people with cognitive frailty and eventually lead to favorable clinical outcomes (ie, amelioration of cognitive frailty). The aims of this study were to identify issues relating to the feasibility of the interventions and the trial (ie, recruitment, retention, participation, and compliance) and to examine the preliminary effects of the intervention by testing the following hypothesis: (1) the mHealth intervention significantly increases cognitive function, reduces physical frailty, increases walking behaviors, and increases MVPA and (2) the conventional behavior change intervention does not significantly increase cognitive function, reduce physical frailty, increase walking behaviors, or increase MVPA.

Methods

Trial Design

An open-label, parallel design (1:1 ratio), randomized controlled trial (RCT) was employed. This pilot trial has been registered with the Hong Kong University Clinical Trials Registry (HKUCTR-2283). This section reports the methods employed in the trial by following the CONSORT-EHEALTH checklist (Multimedia Appendix 1) [36].

Settings

This study was conducted in community settings. The participants were recruited from 2 elderly community centers in Hong Kong during the period of May 2018 to June 2019. Elderly community centers regularly provide recreational and social activities for community-dwelling people above the age of 60 years.

Participants

The inclusion criteria for the participants were age ≥60 years and having cognitive frailty, which was operationally defined as the coexistence of mild cognitive impairment (MCI) and physical frailty (including both frailty and prefrailty) [6]. MCI was confirmed by the following criteria put forward by the National Institute on Aging-Alzheimer's Association [37]: (1) self-reported or informant-reported cognitive complaints, (2) objective cognitive impairment, as defined by a Clinical Dementia Rating of 0.5 and a Montreal Cognitive Assessment (MoCA) score of <25 [38,39], (3) preservation of one's independence, as defined by the Lawton's Instrumental Activity of Daily Living score of >14 [40], and (4) no diagnosed dementia, as observed in the medical record, as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria for Major Neurocognitive Disorders [41]. Physical frailty from being prefrail to frail was defined by a Fried Frailty Index (FFI) score of 1-5 [2]. Physical inactivity was operationally defined by MVPA<150 min in the last 7 days. It was measured by structured interviews guided by a list of MVPAs with reference to the Physical Activity Scale for the Elderly [42]. The exclusion criteria for the participants were having impaired mobility because older people cannot briskly walk outdoors, as defined by the modified Functional Ambulatory Classification score of <7 [43], depressive symptomatology because they have poor motivation as defined by a Geriatric Depression Scale score of ≥8 [44], or probable dementia (ie, MoCA<20 or clinical dementia rating ≥1) [45] because the underdiagnosis of dementia is common, implying that some cases of dementia do not appear on medical records [46].

Intervention Group and Control Group Interventions

As shown in Table 1, in the intervention group, both the conventional behavior change intervention and the mHealth intervention will be implemented. In the control group, only the conventional behavior change intervention will be implemented.

Table 1. Intervention groups and intervention procedures.

Groups and intervention procedures	Conventional behavior change intervention	mHealth intervention
Intervention group		
Reducing the difficulty of changing the target behavior by setting short-term goals	Face-to-face meetings	WhatsApp
Personalization of goals	Face-to-face meetings	WhatsApp + Samsung Health
Messages of praise	Face-to-face meetings	WhatsApp + Samsung Health
e-reminders	N/A[a]	WhatsApp + Samsung Health
Use of validity tested-devices	N/A	Samsung Galaxy smartphone
Integration of self-tracking	N/A	Samsung Health
e-coaching	N/A	WhatsApp + Samsung Health
Control group		
Brief activity counselling	Face-to-face meetings	N/A
Telephone follow-up	Telephone	N/A
Health education	Face-to-face meetings	N/A
Exercise training (brisk walking)	Face-to-face meetings	N/A

[a]Not applicable.

Conventional Behavior Change Intervention

Conventional behavior change interventions have been shown to have a significant effect on prompting physically inactive older people to increase their levels of physical activity by employing various measures [22]. A large-scale study (N=878) showed that a behavioral change intervention that included brief activity counselling using motivational interviewing, regular telephone and face-to-face support, health education, and exercise training significantly increased the physical activity of the older people [47]. As shown in Table 1, these 4 components were adopted in this study to formulate the conventional behavior change intervention. To materialize these components, the procedures in a large-scale study were adopted for the *motivational interviewing* and *regular telephone support* components [47]. For the *health education* component, the educational resources for cognitive frailty from both the National Health Service and the Alzheimer's Association were utilized because they provide internationally reliable health education information that is comprehensible to laypeople [48,49]. The *exercise training* (ie, training in brisk walking) component adopted the contents of the brisk walking training program devised by the Hong Kong Leisure and Cultural Services Department [50].

mHealth Intervention

A systematic review of 32 publications indicated that mHealth interventions are effective when they include (1) reduction in the difficulty of the target behaviors to be changed, (2) personalization of goals, (3) messages of praise, (4) e-reminders, (5) use of validity-tested devices, (6) integration of self-tracking, and (7) e-coaching [51]. As shown in Table 1, these 7 components were adopted in this study to formulate the mHealth intervention.

To materialize these components, we adopted a simple target behavior (ie, brisk walking) that reduces the difficulty of the *target behavior to be changed*. The Samsung Galaxy smartphone J2 with 2 apps (ie, Samsung Health and WhatsApp) was used. This smartphone was chosen because it has many accurate sensors (eg, a triaxial accelerometer) and is a *validity-tested device* that can be used to accurately measure step counts and walking velocity when placed in a pant pocket or a backpack during free walking and has a mean absolute percentage error of less than 3 [52]. Samsung Health is a physical activity autotracking app. It autonomously and continually monitors the walking behaviors (eg, steps, walking speed, walking time, physical activity intensity) of the users. It also coaches users to set individualized goals, logs all the physical activity data, provides immediate onsite rewards and performance reviews, suggests tailored walking, provides real-time feedback, and allows comparisons to be made with peer participants. Samsung Health is used because it allows *e-reminders* to be sent to the users to remind them to perform their walking exercise and provides *self-tracking* of the walking behaviors and the amounts of physical activity with immediate feedback at the scene. WhatsApp is a communication app that allows users to send text messages and voice messages, to make voice calls and video calls, and to share images, documents, user locations, and other media. WhatsApp was used because it allows *e-coaching, personalization of goal settings, and messages of praise* to be remotely provided by the research assistants.

Implementation Procedures

A trained interventionist provided the interventions. One interventionist was trained to deliver the interventions to all the participants because we wanted to minimize the inter-interventionist variance during the stage of the pilot trial. The interventionist was a baccalaureate graduate with a major in psychology who had received theoretical training related to behavioral change. Before the implementation of the interventions, the interventionist completed training in brisk

walking provided by a nursing academic specializing in gerontology.

As shown in Table 2, the total intervention period lasted for 12 weeks. In both the groups, health education was launched in week 1. Exercise training (ie, training in brisk walking) was conducted in weeks 1 and 2. One training session was conducted in an elderly center and another in a real practice environment close to the participants (ie, a public park). Face-to-face meetings were conducted 3 times in weeks 4, 8, and 12. In the control group, in addition to the above, 2 telephone follow-up meetings were conducted in weeks 6 and 10. In the intervention group, the follow-up sessions were conducted using a smartphone with WhatsApp and Samsung Health, instead of with a telephone, from weeks 3 to 12, immediately after the participants received training on how to use the smartphone. Messages (eg, praise messages, e-reminders, personalized goals, coaching) were sent to the participants at least once a week.

Table 2. Intervention implementation schedule.

Groups and interventions	Weeks											
	1	2	3	4	5	6	7	8	9	10	11	12
Intervention group												
mHealth[a]	N/A[b]	N/A	SF[c]	SF	SF	SF	SF	SF	SF	SF	SF	SF
CBC[d]	ET1[e] and HE[f]	ET2	ST[g]	F2F[h]	N/A	N/A	N/A	F2F	N/A	N/A	N/A	F2F
Control group												
CBC	ET1 and HE	ET2	N/A	F2F	N/A	TF[i]	N/A	F2F	N/A	TF	N/A	F2F

[a]mHealth: mobile health.

[b]Not applicable.

[c]SF: smartphone follow-up (WhatsApp + Samsung Health)

[d]CBC: conventional behavior change.

[e]ET: exercise training.

[f]HE: health education.

[g]ST: smartphone training.

[h]F2F: face-to-face meeting.

[i]TF: telephone follow-up.

Tailored dosages were employed. Tailoring refers to the adjustment of intervention implementation strategies to address varying levels of the barrier to changing particular behaviors. Tailoring is needed because these barriers affect the effectiveness of the behavioral change [53]. The varying barrier in this population is mainly the level of the physical fitness at baseline. The training target was adjusted to the baseline physical fitness. The latest guideline recommends that older people should be encouraged to increase their level of activity by small amounts, rather than focusing on immediately reaching the recommended level [54].

As shown in Table 3 and Table 4, the intervention was tailored to 2 aspects: (1) setting personalized goals and (2) contacting participants. The personalized physical activity goals were set according to 4 principles: (1) practice availability, (2) baseline fitness, (3) previous performance, and (4) personal wish. WhatsApp messages were sent to the participants in response to 3 triggers: (1) weekly routine messages, (2) when there was no brisk walking for more than 2 days, and (3) when the weekly goal was achieved earlier than expected. The details of the tailoring procedures are shown in Multimedia Appendix 2.

Therefore, the weekly training goals (eg, relating to walking speed, walking time, number of steps) were tailored according to the participants' individual level of physical fitness at baseline and to their progress. The training goal at the first level was to increase the number of MVPA minutes, while the goal at the second level was to increase the number of MVPA minutes to above 150 min/week at a rate agreed upon by both the participants and the research assistants.

Table 3. Description of the tailored strategies as per personalized goals in the mHealth intervention.

Principles	Actions
Practice availability	Set number of time slots and set available periods to practice brisk walking
Baseline fitness	Prefrail: start with a goal of 5-10 sessions/week*, Frail: start with a goal of 3-5 sessions/week*
Previous performance	Achievers: add 2-5 minutes/session and add 1 session/week; Nonachievers: goal remains as the best performance the participant has ever achieved in the previous weeks
Personal wish	Compromised with the participants according to their wishes (eg, confidence, motivation)

*A session is preset as a 10-minute brisk walking session.

Table 4. Description of the tailored strategies by contacting the participants in the mHealth intervention.

Triggers	Contents
Weekly routine messages	A set of messages on the health benefits of brisk walking and a summary of the participants' performance in the previous week
When there is no brisk walking for more than 2 days	e-reminders and e-coaching
When the weekly goal is achieved before the end of the week	Messages of praise

Outcomes

Trained research assistants collected the data at the baseline, which is the week before the randomization and commencement of the intervention (T0) and at 1 week after the completion of the intervention (T1). The demographic and clinical characteristics were measured at baseline. The outcome variables were collected at both T0 and T1. The 4 outcomes were as follows:

1. Cognitive function was measured using the MoCA [38]. MoCA contains 30 dichotomous items. The correct answer on 1 item will be accorded a score of 1 point. Total scores range from 0 to 30, with a higher score indicating better cognitive function. The test has been found to have good validity in detecting MCI (sensitivity 0.90, specificity 1.00) [38].
2. Frailty was measured using the FFI [2], which quantifies the phenotypes of frailty according to 5 components (ie, weight loss, exhaustion, low physical activity, slow gait, and weakness) by using physical performance tests and questionnaires following Fried's guideline. FFI scores range from 0 to 5, with 1 point assigned for the presence of 1 component. A higher FFI score indicates a higher frailty level. Those with 0, 1-2, or 3-5 point(s) are classified respectively as robust, prefrail, or frail, respectively.
3. Walking was measured using a wrist-worn ActiGraph GT3X+, which is a valid step counter because it can estimate 97.5% of the observed steps in a free-living environment [55]. Walking was quantified according to the total walking time, brisk walking time (>100 steps/min), step counts, and 1-minute peak cadence. The participants were instructed to wear the ActiGraph during the data collection time intervals (ie, T0 and T1) for 24 hours a day and for 7 days. They were allowed to remove the ActiGraph only on special occasions (eg, bathing) that were expected to amount to less than 1 hour per day. Although Samsung Health is valid for counting steps, its data could not be extracted into minute-by-minute units for a precise data analysis. An ActiGraph was therefore used to measure the amount of walking that the participants engage in.
4. Physical activity was also measured using a wrist-worn ActiGraph GT3X+ because it has a good criterion validity for differentiating MVPA from non-MVPA compared to indirect calorimetry (cutoff 6,367, sensitivity 0.70, specificity 0.87, area under the curve 0.83; $P<.001$) in older people when walking on a treadmill at different speeds [56]. An MVPA minute is defined as a minute in which the ActiGraph has recorded physical movements (ie, vector magnitude) of above 6367 counts/min [56]. Only 10 minutes of continuous MVPA minutes were counted as valid MVPA minutes because only sessions of more than 10 minutes of continuous MVPA are considered beneficial, as recommended by the WHO [11]. Physical activity is quantified by valid MVPA minutes measured in 7 consecutive days because a 7-day interval is adequate for understanding the pattern of the routine physical activity engaged in by an individual and is widely used as a standard in studies on physical activity [55,57]. Only MVPA minutes measured on valid days count as valid MVPA minutes (ie, ActiGraph wearing time >10 hours/day) for a valid period (ie, valid days >3 days) [58,59].

Sample Size

There are no previous studies reporting the effects of an eHealth intervention promoting physical activity specifically in older people with cognitive frailty. In a systematic review, eHealth interventions were found to increase MVPA for 8.6-16.0 min/day in the general older population [28]. A study showed that the older people with MCI generally performed MVPA for 24.1 (SD 18.7) min/day [60]. Considering that the effect on older people with MCI is lower because of their lower motivation [29], we adopted a conservative approach to estimate a between-group difference of 8.6 MVPA min/day with a baseline MVPA time of 24.1 (SD 18.7) min/day (ie, d=0.46).

This is a pilot trial, which was not aimed to test the effect by a full-powered study. Instead, this pilot trial aimed to estimate the preliminary effect in a small sample to provide a reference of effect for the main trials to estimate the sample size. Following Cock's methods to estimate the sample size in pilot trials, 34 participants are needed to produce a one-sided 90% confidence limit of the effect size for the main trial (ie, d=0.46) [61]. Assuming the attrition rate of 20% in this pilot trial, we aimed to recruit 34-40 persons in total.

Randomization

Permuted block randomization with block sizes of 8 people in a ratio of 1:1 was used. A random allocation sequence list was generated using a web-based app, that is, Random.org [62]. An independent research assistant who did not participate in any other parts of the research generated and maintained the random allocation sequence list. This independent research assistant assigned group labels to the participants according to the sequence of their entry, referring to the random allocation sequence list to ensure that the other members of the research team did not foresee the group allocation.

Statistical Methods

To describe the issue of feasibility (ie, recruitment, retention, participation, and compliance), frequencies and percentages were used. Nonparametric tests were conducted to examine the effects of the intervention. To test the hypothesis on the effects, a Wilcoxon signed-rank test was used to compare the outcome variables before and after the interventions to test the within-group effect in both the groups separately. The level of significance was .05. Missing values were replaced by the last observed values. An intention-to-treat analysis was conducted to interpret the hypotheses [63].

Ethics

The participants gave their signed informed consent to participate in the study before the study was conducted. Ethical approval was obtained from the Human Subject Ethics subcommittee, The Hong Kong Polytechnic University (Reference Number: HSEARS20180406003).

Results

Demographic Data

As shown in Figure 1, we screened 99 participants referred to us by the elderly community centers. Sixty-six people were excluded because they did not fulfil the criteria for eligibility or did not give their consent to participate. Thirty-three participants were randomized into either the intervention (n=16) or the control (n=17) group. One participant in the intervention group was lost to follow-up because of hospitalization, and 2 participants in the control group were lost to follow-up because they could not be reached.

As shown in Table 5, the median age of the participants was 71.0 (IQR 9.0) years. The majority of the participants were females (28/33, 85%), had completed secondary school or above (17/33, 52%), and had 1-2 chronic illnesses (17/33, 52%). At the baseline, the median MoCA was 21.0 (IQR 6.5) and the median FFI was 2.0 (IQR 1.5). The median walking time was 149.8 min/day (IQR 77.6). The median step count was 12,256.1 steps/day (IQR 4540.2). The median brisk walking time was 2.6 min/day (IQR 4.8). The 1-min peak cadence was 118.0 steps/min (IQR 20.5). The median MVPA time was 23.0 min/week (IQR 85.0) and 9.0 min/valid day (IQR 22.0). There was no significant difference in all of these variables between the groups at baseline.

Figure 1. CONSORT flow diagram.

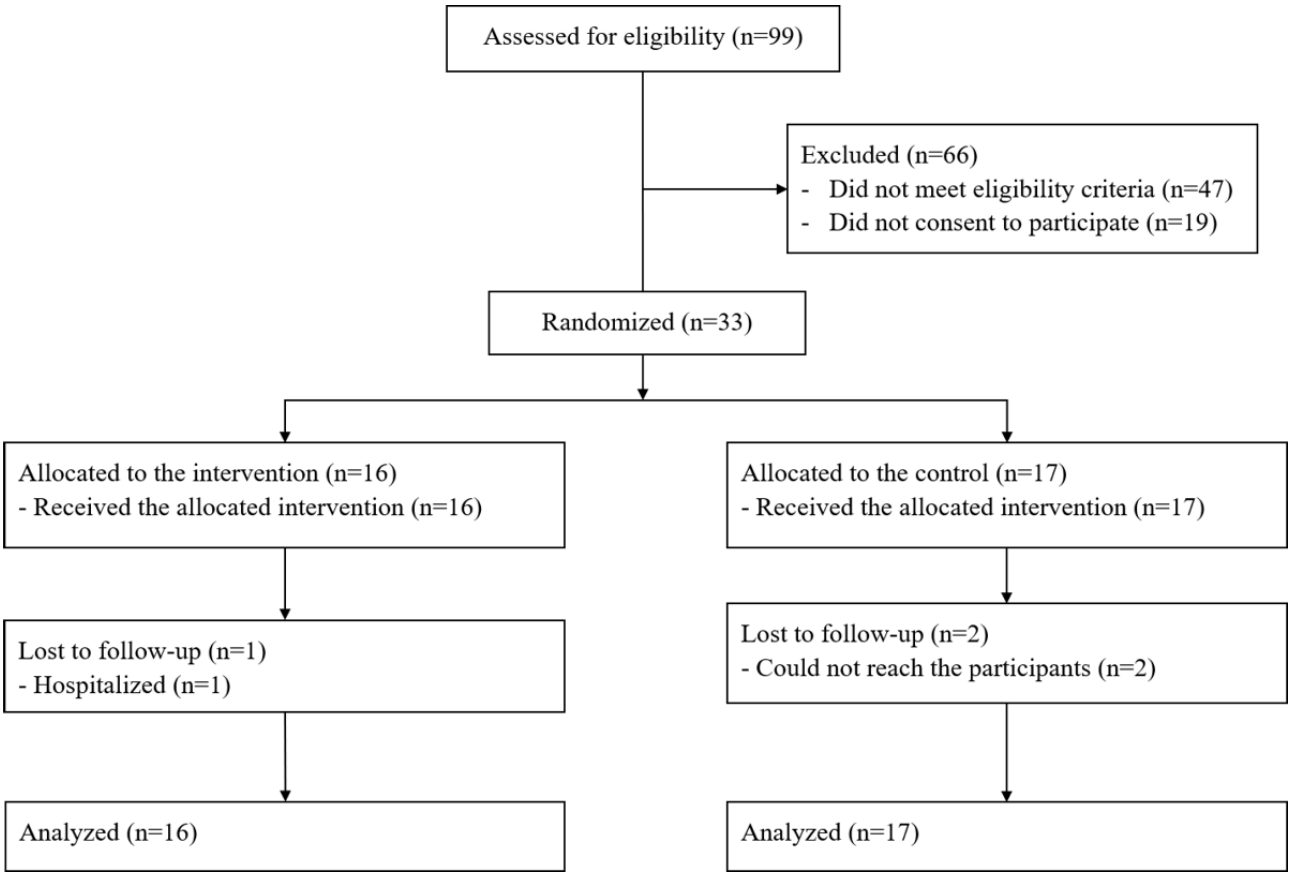


Table 5. Demographic data and outcomes at baseline.

Demographic data	Total population (N=33)	Intervention group (n=16)	Control group (n=17)	*P* value
Gender, n (%)				.58
Male	5 (15)	3 (19)	2 (12)	
Female	28 (85)	16 (81)	15 (88)	
Age in years, median (IQR)	71.0 (9)	70.5 (7)	71.0 (14)	.36
Level of education, n (%)				.79
Secondary or above	17 (52)	9 (56)	8 (47)	
Primary	13 (39)	5 (31)	8 (47)	
No formal education	3 (9)	2 (12)	1 (6)	
Chronic illnesses, n (%)				.90
0	8 (24)	4 (25)	4 (23)	
1-2	17 (52)	8 (50)	9 (53)	
3 or above	8 (24)	4 (25)	4 (23)	
Cognitive function (MoCA[a]), median (IQR)	21.0 (6.0)	21.5 (7.5)	20.0 (5)	.66
Frailty (FFI[b]), median (IQR)	2.0 (1.5)	2.0 (0.8)	2.0 (2.0)	.58
Physical activity, median (IQR)				
Physical activity (PASE[c])	67.6 (40.3)	62.8 (49.0)	67.6 (31.4)	.18
Hand-grip strength (kg)	17.3 (2.3)	17.1 (2.3)	17.3 (5.0)	.89
Walking speed (6MWT[d])	7.6 (1.7)	7.2 (1.8)	7.7 (2.1)	.21
Walking, median (IQR)				
Walking time (min/day)	149.8 (77.6)	158.5 (169.9)	143.4 (64.5)	.14
Step count (step/day)	12,256.1 (4540.2)	13,057.0 (9644.4)	11,620.7 (2863.5)	.08
Brisk walking time (min/day)	2.6 (4.8)	3.0 (4.1)	1.0 (4.6)	.11
1-min peak cadence (step/min)	118.0 (20.5)	122.0 (14.5)	117.0 (24.5)	.34
MVPA [e] , median (IQR)				
MVPA time (min/week)	23.0 (85.0)	31.5 (226.3)	12.0 (38.5)	.18
MVPA time (min/valid day)	9.0 (22.0)	8.5 (44.5)	9.0 (14.0)	.38

[a]MoCA: Montreal Cognitive Assessment.

[b]FFI: Fried frailty index.

[c]PASE: Physical Activity Scale for the Elderly.

[d]6MWT: six-minute walk test.

[e]MVPA: moderate-to-vigorous physical activity.

Objective 1: Feasibility Issues

As shown in Figure 1, the recruitment rate was 33% (33/99). The participant retention rate was 91% (30/33) with 3 participants lost to follow-up. The intervention participation was good. The attendance rate at all the face-to-face training sessions (ie, health education, training in brisk walking, smartphone training, and face-to-face sharing sessions) was 100% (33/33).

As shown in Table 6, in the intervention group (n=16), the mean number of WhatsApp messages received by each participant was 13.5 (SD 6.2) and the mean number of WhatsApp messages sent by each participant was 9.7 (SD 4.5). Most of the messages were read by the participants withing 30 minutes (91/216, 42.1%). The smartphone wearing compliance was good. Almost all of the participants wore the smartphone, as defined by a recording by Samsung Health of at least 1 hour of walking, every day throughout the intervention period with the mean number of wearing days per participant of 54 (SD 1.2) (ie, 97% of the expected wearing days). The reasons for not wearing the smartphone included difficulty in learning how to use it, as stated by 1 participant, who midway through the trial refused to continue wearing the smartphone to implement the intervention, while another participant failed to take the smartphone along on a 2-day trip abroad.

As shown in Table 6, in both groups (N=33), the ActiGraph wearing compliance was very good. The mean number of valid

wearing days was 6.9 (SD 0.3) days/week in the intervention group and 6.1 (SD 2.1) days/week in the control group at T0 and 6.4 (SD 0.5) days/week in the intervention and 5.2 (SD 2.7) days/week in the control group at T1. The mean valid wearing time was 875.8 (SD 129.6) min/day in the intervention group and 812.5 (SD 231.9) min/day in the control group at T0 and 901.3 (SD 128.8) min/day in the intervention group and 868.8 (SD 88.8) min/day in the control group at T1. All ActiGraphs were returned with valid data at T0 and the majority of the ActiGraphs in both the groups were returned with valid ActiGraph data (ie, adequate valid wearing minutes and days for analysis) at T1.

Table 6. Smartphone and ActiGraph use and compliance.

Use of smartphones/ActiGraph	Intervention group, (n=16)	Control group, (n=17)
WhatsApp messages in the whole intervention period (10 weeks), mean (SD)		
Messages received by each participant	13.5 (6.2)	N/A[a]
Messages sent by each participant	9.7 (4.5)	N/A
WhatsApp message reading time (total messages=216), n (%)		
<30 min	91 (42.1)	N/A
30 min-1 h	35 (16.2)	N/A
1-2 h	17 (7.9)	N/A
2-24 h	36 (16.7)	N/A
Unread	37 (17.1)	N/A
Smartphone compliance, mean (SD)		
Valid wearing days per participant (range 0-56 days)	54.1 (1.2)	N/A
ActiGraph compliance at T0[b] , mean (SD)		
Valid wearing day per participant (range 0-7 days)	6.9 (0.3)	6.1 (2.1)
Valid wearing minute per participant per day (range 0-1440 min)	875.8 (129.6)	812.5 (231.9)
ActiGraph compliance at T1[c] , mean (SD)		
Valid wearing day per participant (range 0-7 days)	6.4 (0.5)	5.2 (2.7)
Valid wearing minute per participant per day (range 0-1440 min)	901.3 (128.8)	868.8 (88.8)
Returns with valid ActiGraph data at T0, n (%)		
Valid return	16 (100)	17 (100)
Invalid return	0	0
Unworn	0	0
Returns with valid ActiGraph data at T1, n (%)		
Valid return	13 (81)	12 (71)
Invalid return	2 (13)	3 (18)
Unworn	1 (6)	2 (12)

[a]Not applicable.

[b]T0: beginning of the intervention.

[c]T1: one week after the intervention.

Objective 2: Effects

As shown in Table 7, after the interventions, cognitive function significantly improved in both the intervention (median difference 2.5, *P*=.003) and control (median difference 1.0, *P*=.009) groups. There was a significant reduction in frailty after the intervention in the intervention group (FFI median difference –1.0, *P*=.005). Walking time (median difference 57.9 min/day, *P*=.03), step count (median difference 3778.9, *P*=.02), brisk walking time (median difference 3.1 min/day, *P*=.009), and peak cadence (median difference 7.0 steps/min, *P*=.003) increased significantly after the intervention in the intervention group only but not in the control group. MVPA time (median difference 86 min/week, *P*=.04; median difference 18.5 min/valid day, *P*=.02) increased significantly after the intervention in the intervention group only. The effect sizes of all the outcomes (ie, Cohen *d* between T0 and T1) in the intervention group were higher than those in the control group.

Table 7. Within-group effects of the interventions before and after the interventions in each group.

Outcome variables	Intervention group (n=16)				Control group (n=17)			
	T0[a]	T1[b]	*P* value[c]	ES[d]	T0	T1	*P* value	ES
Cognitive function (MoCA[e]), median (IQR)	21.5 (7.5)	24.0 (6.5)	.003	0.70	20.0 (5.5)	21.0 (6.0)	.009	0.35
Frailty (FFI[f]), median (IQR)	2.0 (1.0)	1.0 (0.8)	.007	–1.41	2.0 (2.0)	1.0 (2.0)	.06	–0.29
Physical activity, median (IQR)								
Physical activity (PASE[g]),	62.8 (49.0)	99.4 (46.6)	.002	1.13	67.6 (31.4)	77.4 (35.4)	.43	0.21
Hand-grip strength (kg)	17.0 (2.2)	18.7 (5.7)	.009	0.66	17.3 (5.0)	16.8 (5.5)	.41	0.12
Walking speed (6MWT[h])	7.2 (1.8)	5.5 (1.2)	.001	–1.32	7.7 (2.1)	7.3 (1.7)	.23	–0.33
Walking, median (IQR)								
Walking time (min/day)	158.4 (169.9)	216.3 (106.9)	.03	0.23	143.4 (64.5)	141.3 (126.6)	.91	–0.05
Step count (step/day)	13,057.0 (9644.4)	16,835.9 (7019.3)	.02	0.39	11,620.7 (2863.5)	10,935.3 (7174.9)	.80	0.03
Brisk walking time (min/day)	3.0 (4.1)	6.1 (6.7)	.009	0.58	1.0 (4.6)	1.0 (8.0)	.26	0.34
1-min peak cadence (step/min)	122.0 (14.5)	129.0 (27.3)	.003	0.72	117.0 (24.5)	118.0 (14.5)	.28	0.21
MVPA								
MVPA[i] time (min/week)	31.5 (226.3)	117.5 (294.5)	.04	0.35	12.0 (38.5)	14.0 (28.0)	.07	–0.43
MVPA time (min/valid day)	8.5 (44.5)	27.0 (37.8)	.048	0.49	9.0 (14.0)	9.0 (20.0)	.08	0.27

[a]T0: beginning of the intervention.

[b]T1: one week after the intervention.

[c]Results based on Wilcoxon signed-rank test.

[d]ES: effect size (Cohen *d* between T0 and T1).

[e]MoCA: Montreal Cognitive Assessment.

[f]FFI: Fried frailty index.

[g]PASE: Physical Activity Scale for the Elderly.

[h]6MWT: six-minute walk test.

[i]MVPA: moderate-to-vigorous physical activity.

Discussion

To summarize the results, this is the first study to provide direct evidence to show that mHealth interventions are not only effective at changing the behavior of young people or healthy older people but also at increasing the walking and MVPA time of older people with cognitive frailty. The extent of the increase through this mHealth intervention was enough to lead to a significant reduction in frailty in the intervention group but not in the control group. The extent of the increase in the MVPA time led to an increase in the cognitive function in both the groups, but the effect size was larger in the intervention group. Our findings confirm that the implementation of mHealth interventions by using smartphones and ActiGraphs is highly feasible, with good recruitment, compliance, and retention.

The participants in our study in Hong Kong walked (12,256.1 steps/day) much more than those in a study conducted in the United States (5660.8 steps/day) [64]. The reason for this is that the pedestrian infrastructure in Hong Kong is good, and people in Hong Kong spend a lot of time walking from one place to another, as reported in a previous study (569 min/week) [65]. This supports the idea that using brisk walking as the target exercise in pedestrian-friendly societies is highly feasible. Although their walking time is relative long, the brisk walking and MVPA times of older people with cognitive frailty are still suboptimal.

Recently, it has been proposed that cognitive frailty can potentially be reversed [5,66,67]. However, there is a lack of direct evidence to show that interventions are effective at reversing cognitive frailty. This study provides direct evidence that cognitive frailty is reversible by increasing MVPA. Further studies should examine whether this reversal of cognitive frailty can be sustained by engagement in physical activities and whether it will eventually lead to a reduced risk of dementia and onset of dependence.

Longitudinal observational studies have shown that there is a significant association between high levels of physical activity and reduced risks of cognitive decline in older people [68]. A systematic review showed that there are only a limited number of trials supporting the claim that physical activity improves the cognitive function in older people with MCI [69]. However, in these studies, physical vulnerability (ie, frailty) was not taken

into account when selecting participants with MCI. There is evidence that the majority of people with MCI have a coexisting condition of frailty (ie, cognitive frailty), which is associated with adverse health outcomes (eg, malnutrition, depression) to a significantly larger extent than solely having MCI [70]. This shows that people with MCI without frailty may not be the group to target to benefit from physical activity nor is their condition necessarily associated with adverse health outcomes. This study provides new knowledge that physical activity training for people with MCI plus concurrent frailty (ie, cognitive frailty) can improve their cognitive functions. Further studies should examine whether increased physical activity can lead to a reduction in adverse health outcomes (eg, malnutrition, depression).

Recent systematic reviews have shown that eHealth interventions are effective at encouraging older people (age >50 years) to engage in physical activity [25]. In these studies, there is no evidence to show that eHealth is also effective among vulnerable older people (ie, older people with cognitive frailty). Fatigue (also known as exhaustion) is also a phenotypic component of frailty [2]. Fatigue contributes to reduced physical activity in older people, making it more difficult to implement an intervention designed to enhance their engagement in physical activity [71]. As mentioned earlier, the most frequently cited barriers to participation in physical activity include a lack of company, a lack of interest, a lack of opportunities to engage in sports, and a lack of transportation [20]. An attempt was made in this study to develop an intervention by solving these problems through integrating eHealth technology with training in brisk walking in daily-living venues. For example, we began with individualized starting doses and targets to reduce the difficulties of participation for people with frailty (ie, tailored dosages), formed groups on an e-platform to enhance social support (ie, WhatsApp), used a physical activity–tracking app to enhance interest (ie, Samsung Health), provided an opportunity-free form of sports (ie, brisk walking), and offered a transportation-free form of exercise (ie, walking in public parks close to the participants' homes). This study provides proof that the eHealth intervention resolves the abovementioned barriers to engaging in physical activity faced by vulnerable older people with cognitive frailty and creates a more favorable environment for them. Therefore, these reasons show the feasibility of implementing this intervention. This intervention is also effective at increasing the MVPA of older people, to an extent that is sufficient to lead to clinical improvements in their cognitive function and physical frailty. A longer period of engagement in increased physical activity can even lead to neurological changes in older people with MCI (eg, increase hippocampal volume) [72]. Further studies should be conducted to examine the long-term effect of eHealth interventions on MVPA as well as on beneficial neurological changes.

Apart from resolving the barriers faced by the participants, the eHealth intervention in our study provided more e-contacts to the participants. A systematic review showed that remote feedback (eg, e-contact) is at least equally effective as supervised training and is more effective than usual care or no treatment, and more contacts are associated with better compliance with physical activity in older people [73]. This intervention employed mHealth technologies to strengthen multiple behavioral change techniques such as e-contacts and social support. mHealth interventions probably allow more frequent, higher quality, and more cost-effective contacts compared with conventional behavior change interventions (eg, face-to-face or telephone-based interventions). With the advancement of smart technology, nonhuman conversational agents powered by artificial intelligence (eg, Chatbot) can also play the role of e-coaching and connect with e-contacts to promote people's health [74]. Further studies should examine the components (eg, social support, contact) that are more effective in yielding clinical benefits and those components should be strengthened to achieve these benefits. Further, mHealth interventions in the future should be upgraded with new technologies.

Older people, particularly those with chronic conditions, find it difficult to meet MVPA exercise targets (ie, >150 min/week) [54]. This study showed that even if the participants could not achieve the median MVPA performance of 150 min/week during the intervention, after the intervention, beneficial clinical effects on cognitive frailty were also observed. This study recommends implementing an eHealth intervention targeted at increasing the MVPA time of older people with cognitive frailty to an extent that they can tolerate, even for those who are unable to attain an MVPA level of >150 min/week.

There are several limitations to this study. First, a relatively small sample of participants was recruited, which may limit the generalizability. The effect size with a better confidence should be confirmed by a large-scale RCT. Second, the data collectors were not blinded; therefore, the results may be affected by observation bias. Third, the participants in this study had a high level of step count (median 12,256.1 steps/day); therefore, this study has limited generalizability to people who do not walk often. Further, participants in the intervention had slightly more active walking behaviors because of random variation in a small sample, particularly an almost significantly higher level of step count (P=.08). The participants in the intervention group may be more motivated to start with the intervention. Fourth, significant improvement in the cognitive function was also observed in the control group without significant improvement in the physical activity. This may probably be caused by the practice effects of the MoCA [75]. The actual cognition-enhancing effect of mHealth intervention should be interpreted with caution. Fifth, in the mHealth group, the participants unavoidably received more e-contacts than the control group. The number of contacts may possibly be a confounding factor. These factors should be controlled in the main trials. Finally, this study did not measure the engagement (eg, contacts or social interactions) in both the groups. This study could not conclude whether the possibly increased engagement is caused by the mHealth measures.

In conclusion, it is feasible to implement this mHealth intervention in older people with cognitive impairment. The mHealth intervention is effective at enhancing compliance with the brisk walking training program delivered by conventional behavior change interventions. This study provides preliminary evidence that this mHealth intervention can increase MVPA time to an extent that is sufficient to yield clinical benefits (ie,

a reduction in cognitive frailty). Future studies should employ a full-powered and assessor-blinded RCT to warrant these effects.

Acknowledgments

This project is funded by the School of Nursing, The Hong Kong Polytechnic University (BE08). We would like to thank our collaborators Mr. Gary Sham and Ms. Cathy Cheung from the Central and Western District Elderly Community Centre, St. James' Settlement, and Mr. Law from Jane Shu Tsao Neighborhood Elderly Centre, Hong Kong Mutual Encouragement Association Limited, for supporting this project by providing venues and for coordinating participant recruitment.

Authors' Contributions

RYCK, DSKC, and MT wrote the Introduction, Methods, and Discussion sections of this manuscript. DL conducted the data collection and prepared all the tables and figures. PHL wrote the statistical analysis, conducted all the statistical analysis, and wrote the Results section. LT and KSC wrote, revised, and commented on the Discussion section of this manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH (V 1.6.1): Submission/Publication Form.
[PDF File (Adobe PDF File), 1776 KB - mhealth_v8i7e16596_app1.pdf]

Multimedia Appendix 2
Intervention tailoring strategies on setting personalized goals and contacting subjects.
[DOCX File , 16 KB - mhealth_v8i7e16596_app2.docx]

References

1. Kelaiditi E, Cesari M, Canevelli M, van Kan GA, Ousset PJ, Gillette-Guyonnet S, IANA/IAGG. Cognitive frailty: rational and definition from an (I.A.N.A./I.A.G.G.) international consensus group. J Nutr Health Aging 2013 Sep;17(9):726-734. [doi: 10.1007/s12603-013-0367-2] [Medline: 24154642]
2. Fried, 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. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
3. Feng L, Zin Nyunt MS, Gao Q, Feng L, Yap KB, Ng TP. Cognitive Frailty and Adverse Health Outcomes: Findings From the Singapore Longitudinal Ageing Studies (SLAS). J Am Med Dir Assoc 2017 Mar 01;18(3):252-258 [FREE Full text] [doi: 10.1016/j.jamda.2016.09.015] [Medline: 27838339]
4. Roppolo M, Mulasso A, Rabaglietti E. Cognitive Frailty in Italian Community-Dwelling Older Adults: Prevalence Rate and Its Association with Disability. J Nutr Health Aging 2017;21(6):631-636. [doi: 10.1007/s12603-016-0828-5]
5. Panza F, Lozupone M, Solfrizzi V, Stallone R, Bellomo A, Greco A, et al. Cognitive frailty: a potential target for secondary prevention of dementia. Expert Opin Drug Metab Toxicol 2017 Oct;13(10):1023-1027. [doi: 10.1080/17425255.2017.1372424] [Medline: 28849681]
6. Ruan Q, Yu Z, Chen M, Bao Z, Li J, He W. Cognitive frailty, a novel target for the prevention of elderly dependency. Ageing Res Rev 2015 Mar;20:1-10. [doi: 10.1016/j.arr.2014.12.004] [Medline: 25555677]
7. Panza F, Lozupone M, Solfrizzi V, Sardone R, Dibello V, Di Lena L, et al. Different Cognitive Frailty Models and Health- and Cognitive-related Outcomes in Older Age: From Epidemiology to Prevention. J Alzheimers Dis 2018;62(3):993-1012 [FREE Full text] [doi: 10.3233/JAD-170963] [Medline: 29562543]
8. Cesari M, Penninx BWJH, Pahor M, Lauretani F, Corsi AM, Rhys Williams G, et al. Inflammatory markers and physical performance in older persons: the InCHIANTI study. J Gerontol A Biol Sci Med Sci 2004 Mar;59(3):242-248. [doi: 10.1093/gerona/59.3.m242] [Medline: 15031308]
9. Dougherty RJ, Schultz SA, Kirby TK, Boots EA, Oh JM, Edwards D, et al. Moderate Physical Activity is Associated with Cerebral Glucose Metabolism in Adults at Risk for Alzheimer's Disease. JAD 2017 Jun 23;58(4):1089-1097. [doi: 10.3233/jad-161067]
10. Steffl M, Bohannon RW, Sontakova L, Tufano JJ, Shiells K, Holmerova I. Relationship between sarcopenia and physical activity in older people: a systematic review and meta-analysis. Clin Interv Aging 2017;12:835-845 [FREE Full text] [doi: 10.2147/CIA.S132940] [Medline: 28553092]
11. World Health Organization. Global recommendations on physical activity for health. Switzerland: WHO; 2010.

12. Mañas A, Del Pozo-Cruz B, Guadalupe-Grau A, Marín-Puyalto J, Alfaro-Acha A, Rodríguez-Mañas L, et al. Reallocating Accelerometer-Assessed Sedentary Time to Light or Moderate- to Vigorous-Intensity Physical Activity Reduces Frailty Levels in Older Adults: An Isotemporal Substitution Approach in the TSHA Study. J Am Med Dir Assoc 2018 Feb;19(2):185.e1-185.e6. [doi: 10.1016/j.jamda.2017.11.003] [Medline: 29269096]
13. Li X, Liu Q, Wang K, Li R, Hou Y, Li Z, et al. Prevalence and correlates of physical inactivity among elderly Chinese. In: Osteoarthritis and Cartilage.: Elsevier; 2017 Apr Presented at: OARSI World Congress on Osteoarthritis; 27-30 April 2017; Las Vegas, NV, USA p. S203. [doi: 10.1016/j.joca.2017.02.349]
14. Bauman A, Petersen C, Blond K, Rangul V, Hardy L. The descriptive epidemiology of sedentary behavior. In: Leitzmann M, Jochem C, Schmid D, editors. Sedentary Behaviour Epidemiology. Cham: Springer; 2018:73-106.
15. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. BMC Public Health 2013 May 06;13:449 [FREE Full text] [doi: 10.1186/1471-2458-13-449] [Medline: 23648225]
16. Monteiro, Conde WL, Matsudo SM, Matsudo VR, Bonseñor IM, Lotufo PA. A descriptive epidemiology of leisure-time physical activity in Brazil, 1996-1997. Rev Panam Salud Publica 2003 Oct;14(4):246-254. [doi: 10.1590/s1020-49892003000900005] [Medline: 14662075]
17. Barnett A, van den Hoek D, Barnett D, Cerin E. Measuring moderate-intensity walking in older adults using the ActiGraph accelerometer. BMC Geriatr 2016 Dec 08;16(1):211 [FREE Full text] [doi: 10.1186/s12877-016-0380-5] [Medline: 27931188]
18. Brown, Harhay MO, Harhay MN. Walking cadence and mortality among community-dwelling older adults. J Gen Intern Med 2014 Sep;29(9):1263-1269 [FREE Full text] [doi: 10.1007/s11606-014-2926-6] [Medline: 24934147]
19. Samdal, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
20. Moschny A, Platen P, Klaassen-Mielke R, Trampisch U, Hinrichs T. Barriers to physical activity in older adults in Germany: a cross-sectional study. Int J Behav Nutr Phys Act 2011 Nov 02;8:121 [FREE Full text] [doi: 10.1186/1479-5868-8-121] [Medline: 22047024]
21. Baert V, Gorus E, Mets T, Geerts C, Bautmans I. Motivators and barriers for physical activity in the oldest old: a systematic review. Ageing Res Rev 2011 Sep;10(4):464-474. [doi: 10.1016/j.arr.2011.04.001] [Medline: 21570493]
22. French DP, Olander EK, Chisholm A, Mc Sharry J. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
23. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
24. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive Technology in Mobile Applications Promoting Physical Activity: a Systematic Review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
25. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile Health Interventions for Physical Activity, Sedentary Behavior, and Sleep in Adults Aged 50 Years and Older: A Systematic Literature Review. J Aging Phys Act 2019 Aug 01;27(4):565-593. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
26. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: A systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
27. Joseph, Royse KE, Benitez TJ. A Systematic Review of Electronic and Mobile Health (e- and mHealth) Physical Activity Interventions for African American and Hispanic Women. J Phys Act Health 2019 Mar 01;16(3):230-239. [doi: 10.1123/jpah.2018-0103] [Medline: 30782040]
28. Kwan, Salihu D, Lee PH, Tse M, Cheung DSK, Roopsawang I, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020;17:7 [FREE Full text] [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]
29. Forstmeier S, Maercker A, Maier W, van den Bussche H, Riedel-Heller S, Kaduszkiewicz H, AgeCoDe Study Group. Motivational reserve: motivation-related occupational abilities and risk of mild cognitive impairment and Alzheimer disease. Psychol Aging 2012 Jun;27(2):353-363. [doi: 10.1037/a0025117] [Medline: 21875213]
30. Evans W, Paolisso G, Abbatecola AM, Corsonello A, Bustacchini S, Strollo F, et al. Frailty and muscle metabolism dysregulation in the elderly. Biogerontology 2010 Oct;11(5):527-536. [doi: 10.1007/s10522-010-9297-0] [Medline: 20683658]
31. Coa K, Patrick H. Baseline Motivation Type as a Predictor of Dropout in a Healthy Eating Text Messaging Program. JMIR Mhealth Uhealth 2016 Sep 29;4(3):e114 [FREE Full text] [doi: 10.2196/mhealth.5992] [Medline: 27688034]
32. Motl, McAuley E, Dlugonski D. Reactivity in baseline accelerometer data from a physical activity behavioral intervention. Health Psychol 2012 Mar;31(2):172-175. [doi: 10.1037/a0025965] [Medline: 22023436]
33. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]

34. Viner, Gireesh A, Stiglic N, Hudson L, Goddings A, Ward J, et al. Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data. Lancet Child Adolesc Health 2019 Oct;3(10):685-696. [doi: 10.1016/S2352-4642(19)30186-5] [Medline: 31420213]
35. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile Apps for Health Behavior Change in Physical Activity, Diet, Drug and Alcohol Use, and Mental Health: Systematic Review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
36. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
37. Albert, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]
38. Nasreddine, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
39. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
40. Ho, Fong TCT, Hon T, Chan WC, Kwan JSK, Chiu PKC, et al. Psychometric validation of Fuld Object Memory Evaluation in older adults with cognitive impairments. Aging Ment Health 2019 Jun;23(6):711-717. [doi: 10.1080/13607863.2018.1442414] [Medline: 29517270]
41. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Washington, DC: American Psychiatric Association; 2013.
42. Ngai, Cheung RTH, Lam PL, Chiu JKW, Fung EYH. Validation and reliability of the Physical Activity Scale for the Elderly in Chinese population. J Rehabil Med 2012 May;44(5):462-465 [FREE Full text] [doi: 10.2340/16501977-0953] [Medline: 22549657]
43. Rosanna Chau MW, Chan SP, Wong YW, Lau MYP. Reliability and validity of the Modified Functional Ambulation Classification in patients with hip fracture. Hong Kong Physiotherapy Journal 2013 Jun;31(1):41-44. [doi: 10.1016/j.hkpj.2013.01.041]
44. Chi I, Yip PSF, Chiu HFK, Chou KL, Chan KS, Kwan CW, et al. Prevalence of depression and its correlates in Hong Kong's Chinese older adults. Am J Geriatr Psychiatry 2005 May;13(5):409-416. [doi: 10.1176/appi.ajgp.13.5.409] [Medline: 15879590]
45. Tan J, Li N, Gao J, Wang L, Zhao Y, Yu B, et al. Optimal cutoff scores for dementia and mild cognitive impairment of the Montreal Cognitive Assessment among elderly and oldest-old Chinese population. J Alzheimers Dis 2015;43(4):1403-1412. [doi: 10.3233/JAD-141278] [Medline: 25147113]
46. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. Aging Ment Health 2011 Nov;15(8):978-984. [doi: 10.1080/13607863.2011.596805] [Medline: 21777080]
47. Kerse N, Elley CR, Robinson E, Arroll B. Is physical activity counseling effective for older people? A cluster randomized, controlled trial in primary care. J Am Geriatr Soc 2005 Nov;53(11):1951-1956. [doi: 10.1111/j.1532-5415.2005.00466.x] [Medline: 16274377]
48. National Health Service, England. Frailty resources.: NHS URL: https://www.england.nhs.uk/ourwork/clinical-policy/older-people/frailty/frailty-resources/ [accessed 2019-10-01]
49. Alzheimer's Association. Mild Cognitive Impairment (MCI).: Alzheimer's Association URL: https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment [accessed 2019-10-01]
50. Leisure and Cultural Services Department. Be Smart and Healthy, Let's Take a Brisk Walk Daily. Hong Kong: Sport for All: Leisure and Cultural Services Department URL: https://www.lcsd.gov.hk/en/sportforall/fitness.html [accessed 2019-10-01]
51. Lentferink, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key Components in eHealth Interventions Combining Self-Tracking and Persuasive eCoaching to Promote a Healthier Lifestyle: A Scoping Review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
52. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018 Jul;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
53. Powell, Beidas RS, Lewis CC, Aarons GA, McMillen JC, Proctor EK, et al. Methods to Improve the Selection and Tailoring of Implementation Strategies. J Behav Health Serv Res 2017 Apr;44(2):177-194 [FREE Full text] [doi: 10.1007/s11414-015-9475-6] [Medline: 26289563]

54. Sparling, Howard BJ, Dunstan DW, Owen N. Recommendations for physical activity in older adults. BMJ 2015 Jan 21;350:h100. [doi: 10.1136/bmj.h100] [Medline: 25608694]
55. Lee, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
56. Kwan R, Lee D, Lee P, Liu J. Measurement of physical activity in older people by wrist-worn actigraph: A validation study. 2019 Presented at: Leadership Conference on Ageing; 12-13 Nov 2019; Austin URL: https://www.nhcgne.org/ADCaregivingConference
57. Sasaki, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
58. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act 2011 Jun 16;8:62 [FREE Full text] [doi: 10.1186/1479-5868-8-62] [Medline: 21679426]
59. Tudor-Locke C, Camhi S, Troiano R. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. Prev Chronic Dis 2012;9:E113 [FREE Full text] [doi: 10.5888/pcd9.110332] [Medline: 22698174]
60. Doi T, Makizako H, Shimada H, Tsutsumimoto K, Hotta R, Nakakubo S, et al. Objectively measured physical activity, brain atrophy, and white matter lesions in older adults with mild cognitive impairment. Exp Gerontol 2015 Feb;62:1-6. [doi: 10.1016/j.exger.2014.12.011] [Medline: 25528600]
61. Cocks K, Torgerson DJ. Sample size calculations for pilot randomized trials: a confidence interval approach. J Clin Epidemiol 2013 Feb;66(2):197-201. [doi: 10.1016/j.jclinepi.2012.09.002] [Medline: 23195919]
62. Random.org. URL: https://www.random.org/ [accessed 2019-02-02]
63. Gupta S. Intention-to-treat concept: A review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
64. Davis MG, Fox KR, Hillsdon M, Sharp DJ, Coulson JC, Thompson JL. Objectively Measured Physical Activity in a Diverse Sample of Older Urban UK Adults. Medicine & Science in Sports & Exercise 2011;43(4):647-654. [doi: 10.1249/mss.0b013e3181f36196]
65. Cerin E, Sit CH, Barnett A, Johnston JM, Cheung M, Chan W. Ageing in an ultra-dense metropolis: perceived neighbourhood characteristics and utilitarian walking in Hong Kong elders. Public Health Nutr 2012 Aug 21;17(1):225-232. [doi: 10.1017/s1368980012003862]
66. Solfrizzi V, Scafato E, Lozupone M, Seripa D, Giannini M, Sardone R, Italian Longitudinal Study on Aging Working Group. Additive Role of a Potentially Reversible Cognitive Frailty Model and Inflammatory State on the Risk of Disability: The Italian Longitudinal Study on Aging. Am J Geriatr Psychiatry 2017 Nov;25(11):1236-1248. [doi: 10.1016/j.jagp.2017.05.018] [Medline: 28689645]
67. Solfrizzi V, Scafato E, Seripa D, Lozupone M, Imbimbo BP, D'Amato A, Italian Longitudinal Study on Aging Working Group. Reversible Cognitive Frailty, Dementia, and All-Cause Mortality. The Italian Longitudinal Study on Aging. J Am Med Dir Assoc 2017 Jan;18(1):89.e1-89.e8. [doi: 10.1016/j.jamda.2016.10.012] [Medline: 28012505]
68. Blondell SJ, Hammersley-Mather R, Veerman JL. Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. BMC Public Health 2014 May 27;14:510 [FREE Full text] [doi: 10.1186/1471-2458-14-510] [Medline: 24885250]
69. Gates N, Fiatarone Singh MA, Sachdev PS, Valenzuela M. The effect of exercise training on cognitive function in older adults with mild cognitive impairment: a meta-analysis of randomized controlled trials. Am J Geriatr Psychiatry 2013 Nov;21(11):1086-1097. [doi: 10.1016/j.jagp.2013.02.018] [Medline: 23831175]
70. Kwan RYC, Leung AYM, Yee A, Lau LT, Xu XY, Dai DLK. Cognitive Frailty and Its Association with Nutrition and Depression in Community-Dwelling Older People. J Nutr Health Aging 2019;23(10):943-948. [doi: 10.1007/s12603-019-1258-y] [Medline: 31781723]
71. Egerton T, Chastin SFM, Stensvold D, Helbostad JL. Fatigue May Contribute to Reduced Physical Activity Among Older People: An Observational Study. J Gerontol A Biol Sci Med Sci 2016 May;71(5):670-676. [doi: 10.1093/gerona/glv150] [Medline: 26347508]
72. ten Brinke LF, Bolandzadeh N, Nagamatsu LS, Hsu CL, Davis JC, Miran-Khan K, et al. Aerobic exercise increases hippocampal volume in older women with probable mild cognitive impairment: a 6-month randomised controlled trial. Br J Sports Med 2015 Feb;49(4):248-254 [FREE Full text] [doi: 10.1136/bjsports-2013-093184] [Medline: 24711660]
73. Geraedts H, Zijlstra A, Bulstra SK, Stevens M, Zijlstra W. Effects of remote feedback in home-based physical activity interventions for older adults: a systematic review. Patient Educ Couns 2013 Apr;91(1):14-24. [doi: 10.1016/j.pec.2012.10.018] [Medline: 23194823]
74. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]

75. Cooley, Heaps JM, Bolzenius JD, Salminen LE, Baker LM, Scott SE, et al. Longitudinal Change in Performance on the Montreal Cognitive Assessment in Older Adults. Clin Neuropsychol 2015;29(6):824-835 [FREE Full text] [doi: 10.1080/13854046.2015.1087596] [Medline: 26373627]

Abbreviations

FFI: Fried frailty index
MCI: mild cognitive impairment
mHealth: mobile health
MoCA: Montreal Cognitive Assessment
MVPA: moderate-to-vigorous physical activity
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 08.10.19; peer-reviewed by D Yoon, A Aswani, J Chan, GY Liao; comments to author 24.04.20; revised version received 12.05.20; accepted 14.06.20; published 31.07.20.

Please cite as:
Kwan RYC, Lee D, Lee PH, Tse M, Cheung DSK, Thiamwong L, Choi KS
Effects of an mHealth Brisk Walking Intervention on Increasing Physical Activity in Older People With Cognitive Frailty: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e16596
URL: https://mhealth.jmir.org/2020/7/e16596
doi:10.2196/16596
PMID:32735218

Original Paper

A One-Step, Streamlined Children's Vision Screening Solution Based on Smartphone Imaging for Resource-Limited Areas: Design and Preliminary Field Evaluation

Shuoxin Ma[1,2], BSc; Yongqing Guan[3], MD; Yazhen Yuan[3], MD; Yuan Tai[2], BSc; Tan Wang[2], MSc

[1]College of Software Engineering, Southeast University, Nanjing, China

[2]TerryDr Infomation Technology, Nanjing, China

[3]The Fourth Hospital of Hebei Medical University, ShiJiaZhuang, China

Corresponding Author:
Shuoxin Ma, BSc
TerryDr Infomation Technology
Room A3-701
#180 RuanJianDaDao, YuHuaTai District
Nanjing, 210000
China
Phone: 86 13813998278
Email: shuoxinma@163.com

Abstract

Background: Young children's vision screening, as part of a preventative health care service, produces great value for developing regions. Besides yielding a high return on investment from forestalling surgeries using a low-cost intervention at a young age, it improves school performance and thus boosts future labor force quality. Leveraging low-skilled health care workers with smartphones and automated diagnosis to offer such programs can be a scalable model in resource-limited areas.

Objective: This study aimed to develop and evaluate an effective, efficient, and comprehensive vision screening solution for school children in resource-limited areas. First, such an exam would need to cover the major risk factors of amblyopia and myopia, 2 major sources of vision impairment effectively preventable at a young age. Second, the solution must be integrated with digital patient record-keeping for long-term monitoring and popular statistical analysis. Last, it should utilize low-skilled technicians and only low-cost tools that are available in a typical school in developing regions, without compromising quality or efficiency.

Methods: A workflow for the screening program was designed and a smartphone app was developed to implement it. In the standardized screening procedure, a young child went through the smartphone-based photoscreening in a dark room. The child held a smartphone in front of their forehead, displaying pre-entered personal information as a quick response code that duplexed as a reference of scale. In one 10-second procedure, the child's personal information and interpupillary distance, relative visual axis alignment, and refractive error ranges were measured and analyzed automatically using image processing and artificial intelligence algorithms. The child's risk for strabismus, myopia, and anisometropia was then derived and consultation given.

Results: A preliminary evaluation of the solution was conducted alongside yearly physical exams in Luoyang, Henan, People's Republic of China. It covered 20 students with suspected strabismus and 80 randomly selected students, aged evenly between 8 and 10. Each examinee took about 1 minute, and a streamlined workflow allowed 3 exams to run in parallel. The 1-shot and 2-shot measurement success rates were 87% and 100%, respectively. The sensitivity and specificity of strabismus detection were 0.80 and 0.98, respectively. The sensitivity and specificity of myopia detection were 0.83 and 1.00, respectively. The sensitivity and specificity of anisometropia detection were 0.80 and 1.00, respectively.

Conclusions: The proposed vision screening program is effective, efficient, and scalable. Compared with previously published studies on utilizing a smartphone for an automated Hirschberg test and photorefraction screening, this comprehensive solution is optimized for practicality and robustness, and is thus better ready-to-deploy. Our evaluation validated the achievement of the program's design specifications.

(JMIR Mhealth Uhealth 2020;8(7):e18226) doi:10.2196/18226

KEYWORDS

vision screening; resource-limited application; photorefraction; strabismus; myopia; anisometropia; mHealth; screening

Introduction

High myopia and amblyopia are 2 major causes of vision impairment, with roots and traits in childhood [1,2]. Compared to other vision disorders, such as glaucoma and macular degradation, the primary risk factors for these 2 disorders are more diagnosable [3]: refractive error can be tracked for both myopia and amblyopia [4]. Additionally for amblyopia, the presence of strabismus can be watched for [5]. Furthermore, the responsiveness of these disorders to simple and cost-effective interventions at early stages [6-8] makes screening the young more valuable. Yet, since the early-stage impacts are subjective, children may not discuss them with their parents and miss the chance to receive cost-effective treatment. This is especially worrisome in developing areas, where the parental generation has a low rate of myopia, and thus poor understanding of its risks, and a high fertility rate thinly divides parental care [9,10]. In fact, the rate of myopia is quickly rising globally, and is rampant not only in eastern Asia but also across the developed world and among all ethnic groups [11]. At earlier stages, mild myopia disrupts daily life and impairs children's learning capabilities. At later stages, severe myopia has a high possibility of resulting in retinal detachment, a blinding and recurring complication [12], among others [13]. Amblyopia generally does not lead to full blindness but does not respond as well to treatment in adulthood.

In 2019, the Chinese central government announced guidelines to cover 185 million kindergarten to grade 12 students with yearly vision screening, focusing on myopia. The recommended procedure includes an eye chart and automatic refractor exams. In practice, however, a number of problems limited the implementation scale. First, the shortage and uneven distribution of systematically trained optometrists, the qualified examiners, will hardly improve in a few years [14]. Second, physical access to remote areas is difficult and time-consuming, especially with high-precision optical measurement equipment that may not endure the transportation. Last, with benefits taking a decade or longer to materialize, investments of US $10 to $20 per student per year by local governments on the program are prohibitive. As a result, while pilot programs are conducted in richer cities, other areas, especially remote ones, are not following.

Luoyang city in Henan, one of the most populated but less developed provinces, leads experimenting with efficient and scalable vision screening programs. While internet-enabled systems and semiautomated workflows have achieved the expected results in the city centers and suburban towns, costs to cover rural areas are well above budget due to transporting technicians and equipment to reach sparse populations. Although some lightweight alternatives have been proposed and evaluated in academia, they suffer from one or more shortcomings: dependency on a specific device, lack of robustness, and lack of practical workflow integration [15-17]. For example, GoCheck Kids [18-20], one recognized commercial product, requires specific iPhone models.

In this study, a new photoscreening solution for resource-limited areas was developed and evaluated. With a smartphone-based automated Hirschberg test and photorefraction, risk levels of strabismus, myopia, and anisometropia can be detected on the spot. Powered by deep learning and image-processing algorithms, the measurement and analysis procedures are fully automated and robust. By replacing dedicated devices with widely available smartphones and conducting an exam that resembles normal photo-taking, no equipment or trained technician needs to be transported. Following a streamlined workflow, about 200 students can be screened and logged with a compact team of 1 supervising teacher and 1 consulting optometrist equipped with 4 mainstream smartphones and equipment available in a rural school. Evaluation experiments show a sound screening accuracy among primary school students.

Methods

Principles of an Automated Hirschberg Test and Smartphone Photorefractory

Strabismus can be roughly defined as misalignment between the binocular axes, or 2 eyes appearing to look in different directions [1]. In a random photo, such gaze directions may rarely be accurately estimated; thus, a photo would not be reliable for determining whether the subject is at risk for strabismus. However, a simple point light source and the corneal luminous reflection it causes can make the problem noticeable.

Although the term "corneal luminous reflection" is technically inaccurate and the optical model is complex [21,22], the from-a-distance image of the corneal luminous reflection of a point light source can be approximated by a bright spot reflected off the surface of the cornea. Simultaneously, the gaze direction of that eye relative to the light source can be roughly reflected by the relative position of the center of the limbus [15], which is attached to the cornea, as illustrated in Figure 1. Combining both models, a simplified automated Hirschberg test can be achieved by taking a photo of the subject's face with a point flashlight at a sufficient distance and comparing the relative positions of the center of each iris and the bright spot on it [22], as illustrated in Figure 2.

Figure 1. Structure of the eye.

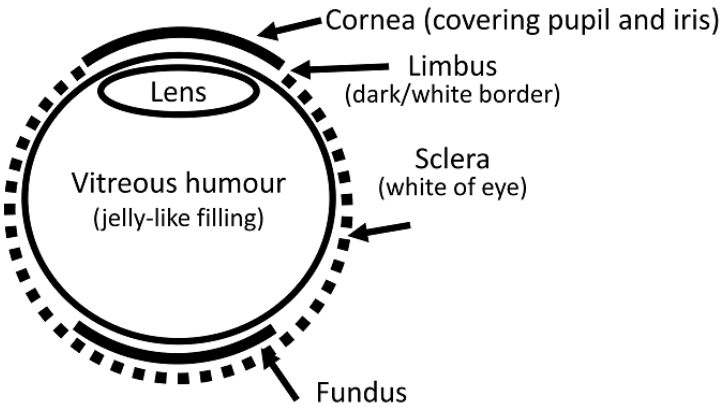
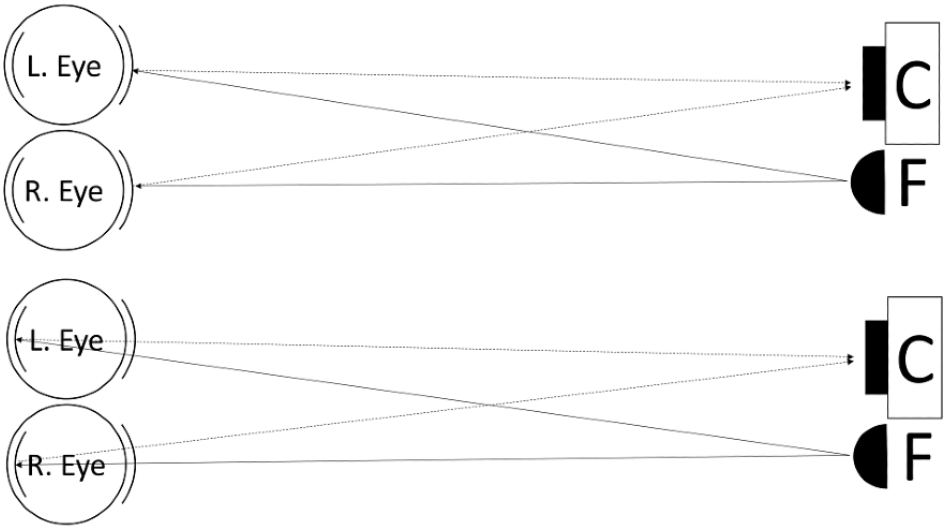


Figure 2. Simplified optical models. Top: Corneal luminous reflection. Bottom: Red reflex. C: camera; F: flash; L: left; R: right.



Smartphone photorefractory shares considerable similarity, in principle, with the automated Hirschberg test, but with a twist. They both work by analyzing the reflection pattern of a light source illuminating the eye. However, in smartphone photorefractory, the reflection occurs on the fundus instead of the cornea, and the pattern is primarily linked to the refractive error instead of the ocular axis, as illustrated in Figure 2. With no refractive error, the image of the light source on the fundus would be on its optically conjugate plane, and thus, it would be reflected back following exactly the same path and is almost invisible to the camera. Defocused by refractive error, the illumination would spread over the fundus and be reflected across directions and thus be partially captured by the camera. This is generally known as the red reflex [23].

Additionally, the size of the pupil affects the 2 phenomena. Although it does not interact with corneal luminous reflection, a dilated pupil admits more light into the vitreous chamber [24] and produces a wider viewing angle of the fundus [25]. As a result, when the pupil contracts to a minimal size, fundus reflection is subdued, for a high-quality automated Hirschberg test. When the pupil dilates widely, fundus reflection is prominent and can be better isolated from interfering with corneal luminous reflection. This is a critical consideration in designing the workflow that combines both tests.

The automated Hirschberg test can be numerically processed as follows. After extracting the positions of the limbus and the corneal luminous reflection spot from the captured image, the horizontal and vertical offsets of the center of the corneal luminous reflection spot relative to the center of the corresponding limbus are measured as Δx_L, Δy_L for the left eye and Δx_R, Δy_R for the right eye, respectively. These offsets are measured in pixels and signs following the computer vision convention, as illustrated in Figure 3 (top image). Assuming the frontal plane pixel scale is α mm per pixel, the eyeball diameter is *d* mm, and the gap between the camera and the flash is ignorable, the axial deviations of each eye from the camera/flash-ocular axis is as follows:

$$\Theta_{\text{horizontal, left}}=\tan^{-1}(2\alpha\Delta x_L/d)$$

$$\Theta_{\text{horizontal, right}}=\tan^{-1}(2\alpha\Delta x_R/d)$$

$$\Theta_{\text{vertical, left}}=\tan^{-1}(2\alpha\Delta y_L/d)$$

$$\Theta_{\text{vertical, right}}=\tan^{-1}(2\alpha\Delta y_R/d)$$

Where $\tan^{-1}()$ is the inverse tangent function. In this solution, *d* is unmeasurable and is thus set to an average of 20 mm [26]. In summary, the horizontal and vertical binocular axis deviations, measured in angles, are as follows:

$$\Theta_{\text{horizontal}}=\Theta_{\text{horizontal, left}}-\Theta_{\text{horizontal, right}}$$

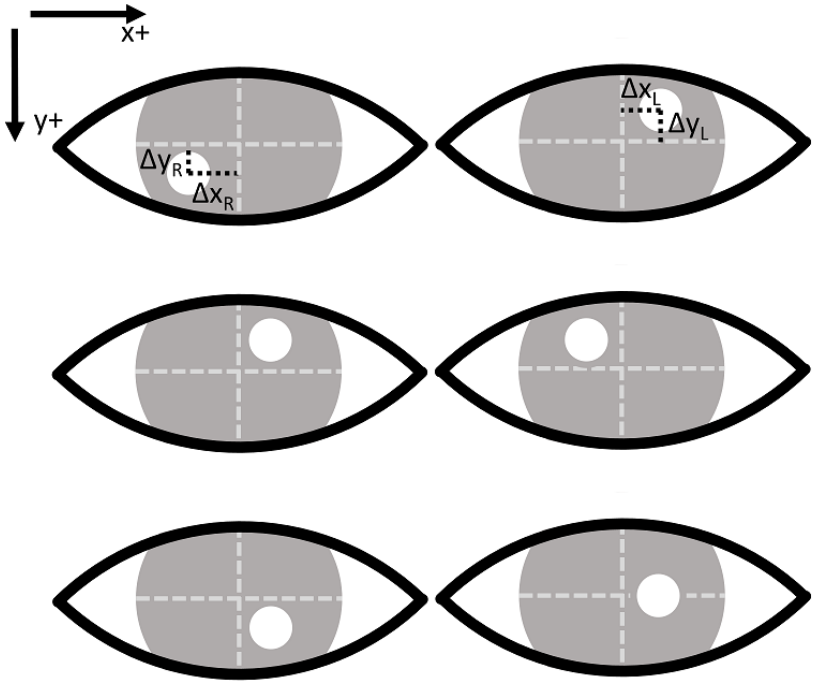
$$=\tan^{-1}(2\alpha\Delta x_L/d)-\tan^{-1}(2\alpha\Delta x_R/d)$$

$$\Theta_{vertical}=\Theta_{vertical,\ left}-\Theta_{vertical,\ right}$$

$$=\tan^{-1}(2\alpha\Delta y_L/d)-\tan^{-1}(2\alpha\Delta y_R/d)$$

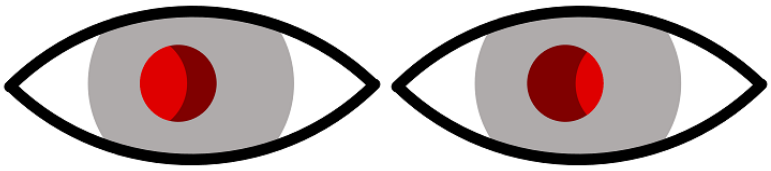
Intuitively, significantly positive and negative $\theta_{horizontal}$ indicate exotropia, as illustrated in Figure 3 (top image) and esotropia (middle image), respectively. A significantly nonzero $\theta_{horizontal}$ indicates vertical strabismus, as illustrated in Figure 3 (top and bottom images).

Figure 3. Examples of corneal luminous reflection. Top: Exotropia and vertical strabismus. Middle: Esotropia. Bottom: Vertical strabismus.



The red reflex phenomenon, based on which photorefractory is implemented, is hard to robustly quantify with a point light source and uncalibrated parameters [27]. From the examiner's perspective, assuming the light source is positioned to the left of the camera, a red crescent rising from left side of the pupil is associated with higher degrees of myopia and one from the right side indicates higher degrees of hyperopia, as illustrated in Figure 4. The presence of the crescent can be reasonably reliably detected as an objective and qualifiable signal for screening. The crescent width-severity relationship, however, requires smartphone model–specific calibration and may not hold across examinees; thus, it should be used only as a reference. Additionally, due to the eccentricity of the light source dot, no red reflex can be caused by either refractive error below threshold or a contracted pupil [25,27] (which is hard to detect without close examination). Further, both severe myopia and hyperopia lead to a fully lit-up pupil and are, thus, indistinguishable.

Figure 4. Typical red reflex pattern, with flash on the left. Left: Myopia. Right: Hyperopia.



Core Artificial Intelligence and Image-Processing Algorithms

Although the smartphone-based automated Hirschberg test and photorefraction solution share the same corneal luminous reflection and red reflex principles with specialized devices, it lacks important features, including nonvisible and structured light sources, distance sensors, and a calibrated imaging system. Despite the exam accuracy being inevitably compromised, the applicability of the solution to screening and measurement automation is still obtainable using software. Sequentially in the acquired images, the head tilting and eye locations are detected by facial landmark recognition, the centers of the limbus are estimated by contour detection, the patterns of corneal luminous reflection and red reflex are estimated through shape fitting, and the frontal plane scale of pixels is estimated using an image analysis of the quick response code. It should be noted that processing would be done first on the automated Hirschberg test images and then on the photorefractory images for the former's higher signal-to-noise ratio, which will be explained in the Discussion section.

Head Tilting and Eye Location Recognition

The proposed solution integrates the OpenFace [28] library for facial landmark detection. This deep learning–based library not only produces key point coordinates of each part of the face, it

also estimates head gestures and gaze directions, which may support automatic examinee attention analysis and other screening quality controls in later versions. These key points, however, are not positioned to pixel-level accuracy and thus require further processing.

With the key points on the oval contour from OpenFace, closed eyes with a height-to-width ratio below the threshold of η=0.25 can be rejected. In images where both eyes are open, local patches of the image are extracted for hard-coded image processing in later stages. The head tilting angle is recorded to correct coordinates in later processing.

Limbus Detection

Ideally, the limbus is a ring that is white on the outside and dark on the inside. In a photo, however, the transition may not be clear-cut, and part of it may be covered. The limbus detection algorithm in this solution addresses both issues with the following steps on each eye.

First, the sclera-limbus-iris region of interest is extracted from the background of nearby skin that matches the hue and saturation of other skin areas. Second, the IsoData automatic thresholding [29] is applied to a histogram of the region of interest to estimate the boundary between the bright sclera and the dark iris. Third, a classic Hough transform of curves [30] between ±45° from a horizontal line is performed to refine the estimated left and right boundary curves without considering the covered parts. Last, using the estimated color contrast between the sclera, iris, and the estimated limbus size from the automated Hirschberg test image, the limbus in the corresponding photorefraction image can be detected with the Hough transform of the same radius. Instead of covering the estimated boundaries, the transform is optimized to maximize contour contrast.

Corneal Luminous Reflection and Red Reflex Pattern Analysis

Within the detected limbus, the raw corneal luminous reflection pixels are detected as the cluster of pixels of maximal intensity across all channels, while the raw crescent pixels are filtered out as those with (a) the red channel brighter than green and blue combined, and (b) brightness more than twice the median of that of all pixels the limbus or red channel saturated. An inscribed circle is fitted among the raw corneal luminous reflection pixels, whose center is used to represent the point of corneal luminous reflection. A circumscribed oval is fitted among the raw red-flex pixels, whose width is used to estimate the severity of refractive error. In this experiment, a width over 1 mm is regarded as an indicator of refractive error. These choices of pattern fitting are slightly different from those in previous studies because preliminary research shows that they are more robust on the lower image quality of amateur users.

Frontal Plane Pixel Resolution Scale Estimation

In this solution, accurate interpupillary distance is measured for follow-up eyeglass frame selection. A quick response code displayed on a smartphone held close to the examinee's forehead with its physical size encoded is used as a reference scale. Since the application programming interface of the physical resolution of a screen is available on most smartphones, it offers the same functionality as dedicated tools such as a special glass frame with marks [17] but is universally available. Compared to other everyday-object tools, such as credit cards or rulers, the structured image pattern on a self-illuminating screen guarantees high contrast, which greatly improves accuracy and robustness.

Solution Design Guidelines and System Outline

As an alternative solution to the existing high-quality, full-service program with an automatic refractor exam and procedures by qualified optometrists, the proposed setup has been adapted to resource-limited rural schools with multiple modifications.

First, the fundamental technology may cover qualitative instead of quantitative analysis on the risk of strabismus, myopia, and anisometropia. Even though earlier research verified that such technology could produce reliable quantitative results, it is limited strictly to specific smartphone models, which are rarely available to the average user. Since the goal is to screen instead of diagnose, ease of use and accessibility dominate over a level of performance higher than necessary.

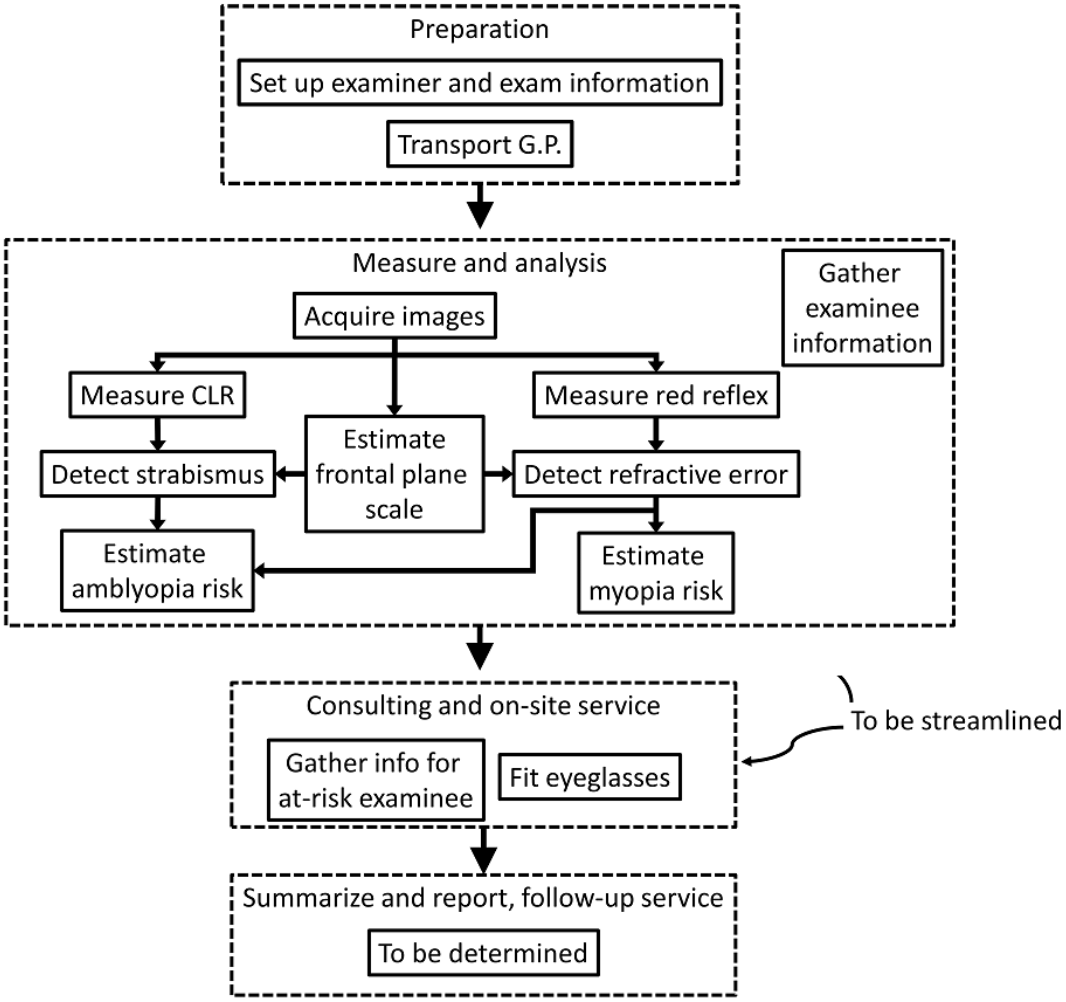
Second, the workflow should serve children as young as 4 years old. Although the application of smartphone photoscreening has been proven feasible to young children [18], their inability to understand instructions and reluctance to cooperate reduces the chances of successful measurement and thus prolongs measurement time. Therefore, pediatricians are consulted to optimize the workflow for young children's psychological characteristics.

Third, the procedure should be executable by school teachers under loose supervision of a general practitioner (GP) with only remote training, yet the exam time per child should be no more than 1 minute. Although expertise may not be required to carry out the procedures, legal concerns and local customs require a GP, who services several villages that may be hours of travel apart, to perform the job. Additionally, on-site consultation provided by the GP must be part of the procedure.

Last, data analysis should be mostly automatic and nearly instantaneous. In China, optometric teleconsultation is under-staffed, and few optometrists are qualified to interpret photoscreening images. Giving on-site face-to-face feedback on the screening results not only builds trust and reduces the burden of follow-up phone calls to farmer parents with only a few hours of free time in the evening, it also creates an opportunity for offering other services, such as distributing premade eyeglasses.

Following the guidelines and the smartphone-based automated Hirschberg test and photorefraction principles, the system is outlined in Figure 5. The core image-processing algorithms are discussed in the previous subsection. The streamlined workflow is explained in the next subsection.

Figure 5. System outline. CLR: corneal luminous reflection; G.P.: general practitioner.



Streamlined Workflow With One-Step Measurement

Figure 6 shows a typical screening room setup. Two children sit back-to-back with a teacher standing next to them (not illustrated). Dark curtains are hung between the children and behind the smartphones, the former to minimize interference between the 2 parallel tests and the latter to act as projector screens, described next.

Figure 6. Screening room setup.

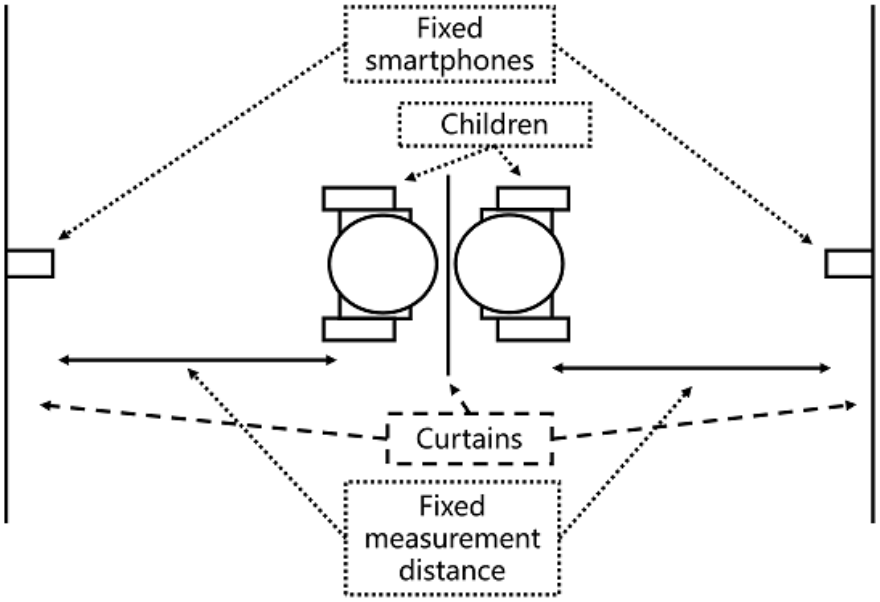


Figure 7 illustrates the exam timeline from an illumination and image acquisition perspective, specifically the coordination of the camera with flash on the smartphone and a video projected onto the curtain in front of the children. When called, the children walk into the room and get seated, then start watching the video projected onto the curtain in front throughout the whole exam, with varying brightness. After several seconds of normal intensity, the video is dimmed and smoothly adapts the children's eyes to the dark surroundings, thus dilating their pupils for more accurate measurement. After the video reaches its lowest brightness for 5 seconds to induce significant pupil dilation [31], the imaging procedure starts. With the smartphone camera, a short video is recorded during which a flash with maximum intensity is triggered. Followed by 5 seconds of the flash gradually brightening up from minimal in torch mode to contract the pupils in a controlled fashion, a still image is captured with another burst of a maximally intense flash, and the raw data acquisition for 1 child is completed. Since all equipment remains in place and the child sits still through the combined automated Hirschberg test and photorefraction with no intervention required, this is regarded as a one-step measurement. The first frame with maximal exposure of the video and the still image are used for red reflex and corneal luminous reflection pattern analysis, respectively, and the reasoning for this is explained in the Discussion section.

Figure 7. The exam timeline, from an illumination and image acquisition perspective. Max: maximum intensity.

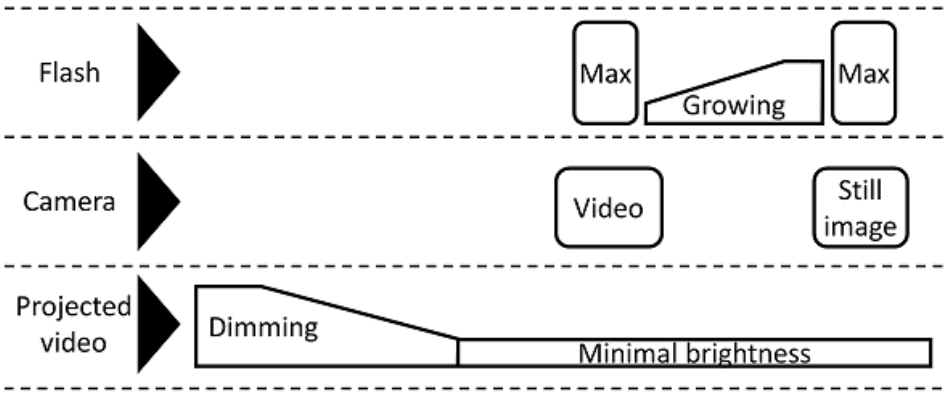
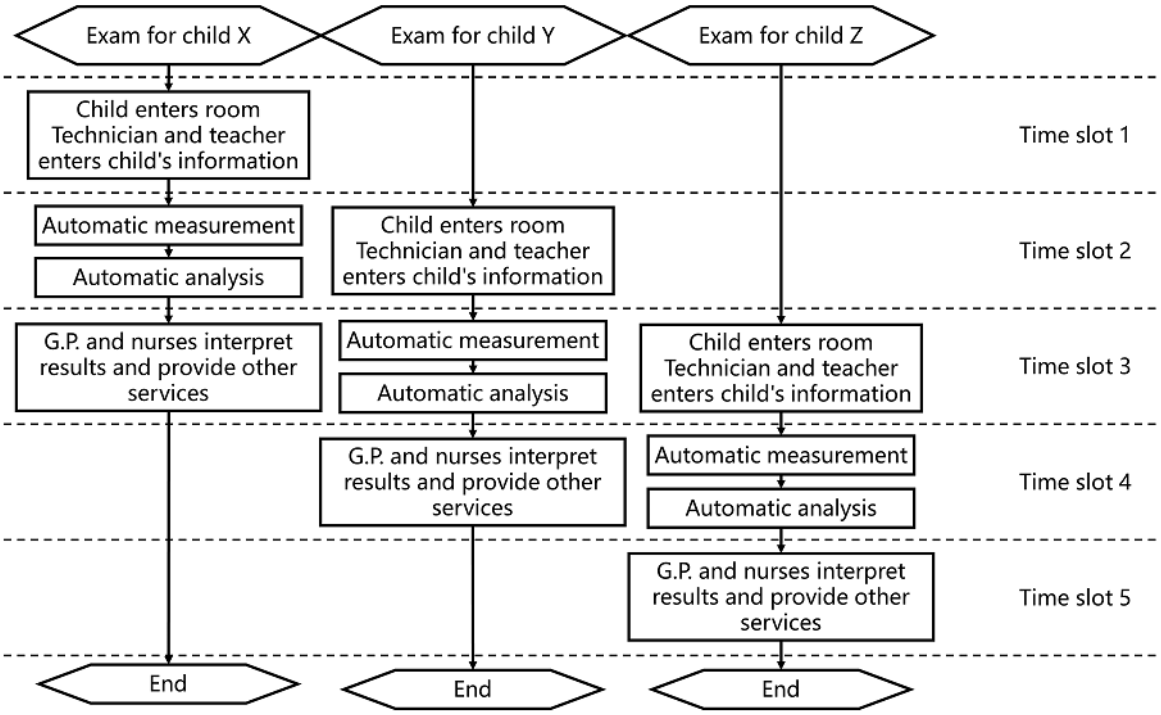


Figure 8 illustrates the workflow for serving 3 children in a streamlined fashion. On being called, a child enters the dark room and gets seated while a teacher enters the child's personal information on a smartphone. After the adaptation period, the child is given the smartphone and instructed to hold it directly in front of their forehead while the image acquisition process begins on the capturing smartphone, as described above. On receiving the automatically analyzed result, a GP or nurse interprets the results to the teacher.

Figure 8. Streamlined workflow for 3 children. G.P.: general practitioner.



Clinical Evaluation Setup

To evaluate the solution in real-world environments, experiments were conducted alongside physical exam programs in Luoyang, Henan, between September 1 and October 15. A total of 100 students aged evenly between 8 and 10 years from 2 schools were selected, 20 of whom were suspected by their teacher to have strabismus and the rest were randomly chosen among their classmates. The experiments were conducted in the school infirmaries under direct supervision of either the optometrists from the screening team or the school nurses. The smartphones used for capturing and analyzing data included 1 iPhone 5s (Apple Inc), 1 Honor 8 (Huawei Group Holdings Ltd), and 1 Mi 6 (XiaoMi Corp), all of which were low-end or outdated models at the time.

Results

Data Acquisition Efficiency

All 100 screened students successfully completed the data acquisition procedure within 2 rounds, and 87 of them passed with a single attempt. The failures were attributed to random eye blinking. Regardless of individual procedure success or failure, the streamlined workflow in Figure 8 was kept at an even pace, and each time slot took 19 seconds, on average. This translates into almost 200 students screened in 1 hour with just 1 supervising teacher and a consulting optometrist or nurse. With the projected cartoon playing, all students showed a high degree of compliance, and few complained about the exam duration or flashes. Outside the scheduled experiments, some 6-year-old students volunteered to participate, and none exhibited confusion or loss of concentration during the procedure.

Screening Experiment Results

After the smartphone-based exam, the screened students went through gold-standard optometry exams, and these results were recorded as the ground truth. The gold standard positive thresholds for strabismus, myopia, and anisometropia were chosen as at least 15Δ, at least –0.5D, and at least with a difference of 2D, respectively. The screening positives were defined as a crescent at least 1 mm wide, relative visual angle difference of 10°, and crescent width difference of at least 1 mm, respectively. Using the raw counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), the statistical metrics accuracy, sensitivity, and specificity were defined as follows:

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN)$$

$$\text{Sensitivity}=TP/(TP+FN)$$

$$\text{Specificity}=TN/(TN+FP)$$

The raw counts of the experiment are summarized in Figure 9. Since myopia is diagnosed on a per-eye basis, its total count is double the number of examinees. It should also be noted that all 4 false negatives in the strabismus tests were attributed to the weakness of the Hirschberg tests in detecting hidden strabismus. The statistical metrics are tabulated in Table 1.

Figure 9. Experiment results: counts. FN: false negatives; FP: false positives; TN: true negatives; TP: true positives.

Proposed / Ground truth	Strabismus	No strabismus
Strabismus	16 (TP)	4 (FN)
No strabismus	2 (FP)	78 (TN)
Proposed / Ground truth	**Myopia**	**No myopia**
Myopia	90 (TP)	18 (FN)
No myopia	0 (FP)	92 (TN)
Proposed / Ground truth	**Anisometropia**	**No anisometropia**
Anisometropia	4 (TP)	1 (FN)
No anisometropia	0 (FP)	95 (TN)

Table 1. Experiment results: accuracy, sensitivity, and specificity.

Metric	Risk factor		
	Strabismus	Myopia	Anisometropia
Accuracy	0.94	0.91	0.99
Sensitivity	0.80	0.83	0.80
Specificity	0.98	1.00	1.00

Discussion

Current Limitations

Although the solution produced satisfying accuracy with high efficiency in screening children in resource-limited areas, major improvements should be investigated in follow-up studies. First, the smartphone flash mechanism is unnecessarily counterproductive. The light-emitting diodes for flash used in modern smartphones are specifically designed to work in 2 modes: one consuming very high power in a short burst or limited power continuously. With a fixed exposure time to control motion blur, using the former mode reduces the camera's sensitivity to light or analog gain and thus improves the signal-to-noise ratio in acquiring both automated Hirschberg test and red reflex images. However, modern smartphones use preflash to induce pupil contraction to suppress the red reflex phenomenon, more generally known as the red-eye effect. To circumvent this, the implementation app replaced the still-image application programming interface with the first intensely illuminated frame extracted from the video. This frame, though, is usually not illuminated with the highest intensity flash, and thus it suffers a reduced signal-to-noise ratio. If a future version of the smartphone camera control provides more flexibility in such control, image quality may be improved.

Second, the implementation software uses a nongeneralized limbus detection algorithm. The proposed method worked robustly on the Chinese students partly because they have a black limbus that is in high contrast to a white sclera. However, on some of the older students who wore colored contact lenses, the low signal-to-noise ratio of the images occasionally caused detection error.

Potential Future Work

In addition to the factors known to limit the performance of the proposed solution, further exploration is planned to address other areas for improvement. First, on the coordination between the smartphone and the projector: In the current version, these run independently and rely on an operator to synchronize them. A future version may include automatic synchronization on the smartphone side by detecting a predefined pattern of the projector, such as information encoded in color variation.

A second issue of interest is the video played during the exam. Some clips seemed to contain too much activity for some of the younger students to be able to maintain a fixed gaze on a consistent part of the screen, introducing unnecessary change in the visual axes. Some background music seemed better at attracting the students' attention quickly. Detailed comparison and analysis, however, were beyond the scope of this preliminary study and will thus be left for follow-up studies.

The last area for future work is on carrying out the experiment on a larger scale and with different setups. The experiment reported in this paper was limited to 2 urban schools, and the equipment and personnel execution capability may be different from those in rural schools.

Conclusion

In this paper, we propose an effective and efficient smartphone-based children's vision screening solution for areas with limited resources, implemented and evaluated. In a dark room, the examinee watches cartoons projected onto a dark curtain screen with program-controlled intensity, while a smartphone in front of the curtain runs the measurement program. The smartphone's flash and camera are used as an eccentric point light source and image sensor, respectively, following the automated Hirschberg test and photorefractory principles. In the acquired images, the frontal plane resolution scale is automatically calibrated from a personal information–encoded quick response code displayed on another smartphone held in front of the examinee's forehead. The examinee's head gesture and eye location are estimated using a deep learning model. The limbus areas, interpupillary distance, corneal luminous reflection spots, and red reflex areas are detected by their corresponding algorithms to derive risk levels of strabismus, myopia, and anisometropia.

By splitting the workflow into 3 parts of inputting information, acquiring images, and automatically analyzing, interpreting, and consulting, the screening procedure is streamlined such that 1 GP supervising 1 teacher can screen, get an automatic diagnosis, and give on-the-spot consulting to 200 children per hour. The examiners need little training, and the equipment, including the model-neutral mainstream smartphones, is available to the average school. In the evaluation of 100 students between 8 and 10 years old, the high throughput and screening accuracy of the proposed solution were verified and recognized by participating optometrists.

Acknowledgments

SM is the primary designer of the proposed solution and the programmer implementing the core algorithms. YT is the engineering lead of the proposed solution and the related software services. TW provided clinical consultancy on the solution design and result analysis under YG's lead.

TerryDr Information Technology Co Ltd funded the preliminary research on the solution and is the owner of product.

Conflicts of Interest

SM, YT, and TW are the cofounders of TerryDr Information Technology Co Ltd, the company that developed the proposed solution, potentially as a commercial service.

References

1. Wong TY, Ferreira A, Hughes R, Carter G, Mitchell P. Epidemiology and disease burden of pathologic myopia and myopic choroidal neovascularization: an evidence-based systematic review. Am J Ophthalmol 2014 Jan;157(1):9-25.e12. [doi: 10.1016/j.ajo.2013.08.010] [Medline: 24099276]
2. von Noorden GK, Campos EC. Binocular Vision and Ocular Motility: Theory and Management of Strabismus. London, United Kingdom: Mosby; 2002.
3. Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. The Lancet Global Health 2017 Dec;5(12):e1221-e1234. [doi: 10.1016/S2214-109X(17)30393-5]
4. Barrett BT, Bradley A, Candy TR. The relationship between anisometropia and amblyopia. Prog Retin Eye Res 2013 Sep;36:120-158 [FREE Full text] [doi: 10.1016/j.preteyeres.2013.05.001] [Medline: 23773832]
5. Sanchez I, Ortiz-Toquero S, Martin R, de Juan V. Advantages, limitations, and diagnostic accuracy of photoscreeners in early detection of amblyopia: a review. Clin Ophthalmol 2016;10:1365-1373 [FREE Full text] [doi: 10.2147/OPTH.S93714] [Medline: 27555744]
6. Chia A, Chua W, Cheung Y, Wong W, Lingham A, Fong A, et al. Atropine for the treatment of childhood myopia: safety and efficacy of 0.5%, 0.1%, and 0.01% doses (Atropine for the Treatment of Myopia 2). Ophthalmology 2012 Feb;119(2):347-354. [doi: 10.1016/j.ophtha.2011.07.031] [Medline: 21963266]
7. Wan L, Wei CC, Chen CS, Chang CY, Lin CJ, Chen JJY, et al. The Synergistic Effects of Orthokeratology and Atropine in Slowing the Progression of Myopia. J Clin Med 2018 Sep 07;7(9) [FREE Full text] [doi: 10.3390/jcm7090259] [Medline: 30205439]
8. Repka MX, Kraker RT, Holmes JM, Summers AI, Glaser SR, Barnhardt CN, Pediatric Eye Disease Investigator Group. Atropine vs patching for treatment of moderate amblyopia: follow-up at 15 years of age of a randomized clinical trial. JAMA Ophthalmol 2014 Jul 01;132(7):799-805 [FREE Full text] [doi: 10.1001/jamaophthalmol.2014.392] [Medline: 24789375]
9. Wolffsohn JS, Calossi A, Cho P, Gifford K, Jones L, Li M, et al. Global trends in myopia management attitudes and strategies in clinical practice. Cont Lens Anterior Eye 2016 Apr;39(2):106-116. [doi: 10.1016/j.clae.2016.02.005] [Medline: 26895778]
10. Pan CW, Ramamurthy D, Saw SM. Worldwide prevalence and risk factors for myopia. Ophthalmic Physiol Opt 2012 Jan;32(1):3-16. [doi: 10.1111/j.1475-1313.2011.00884.x] [Medline: 22150586]
11. Bruce A. Re: Holden et al.: Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050 (Ophthalmology 2016;123:1036-1042). Ophthalmology 2017 Mar;124(3):e24-e25. [doi: 10.1016/j.ophtha.2016.06.066] [Medline: 28219507]
12. Spaide RF, Ohno-Matsui K, Yannuzzi LA, editors. Retinal detachment. In: Pathologic Myopia. New York: Springer-Verlag; 2014:273-282.
13. Ikuno Y. Overview of the complications of high myopia. Retina 2017;37(12):2347-2351. [doi: 10.1097/iae.0000000000001489]
14. Feng J, An L, Wang ZF, Zhan LL, Xu X. [Analysis on ophthalmic human resource allocation and service delivery at county level in Mainland China in 2014]. Zhonghua Yan Ke Za Zhi 2018 Dec 11;54(12):929-934. [doi: 10.3760/cma.j.issn.0412-4081.2018.12.011] [Medline: 30526793]
15. Wang T, Keller JM, Cibis GW. A fuzzy approach to find Hirschberg points and to determine fixation in digital images of infants. In: The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ '03.: IEEE; 2003 Presented at: The 12th IEEE International Conference on Fuzzy Systems; 25-28 May 2003; St Louis, MO, USA, USA. [doi: 10.1109/FUZZ.2003.1206560]
16. Almeida JDSD, Silva AC, Paiva ACD, Teixeira JAM. Computational methodology for automatic detection of strabismus in digital images through Hirschberg test. Comput Biol Med 2012 Jan;42(1):135-146. [doi: 10.1016/j.compbiomed.2011.11.001] [Medline: 22119221]
17. Kwok TCK, Shum NCM, Ngai G, Leong HV, Tseng GA, Choi H, et al. Democratizing Optometric Care: A Vision-Based, Data-Driven Approach to Automatic Refractive Error Measurement for Vision Screening. : IEEE; 2015 Presented at: 2015 IEEE International Symposium on Multimedia (ISM); 14-16 Dec. 2015; Miami, FL, USA. [doi: 10.1109/ISM.2015.55]
18. Arnold RW, O'Neil JW, Cooper KL, Silbert DI, Donahue SP. Evaluation of a smartphone photoscreening app to detect refractive amblyopia risk factors in children aged 1–6 years. OPTH 2018 Aug;Volume 12:1533-1537. [doi: 10.2147/opth.s171935]
19. Arnold RW, Arnold AW, Hunt-Smith TT, Grendahl RL, Winkle RK. The Positive Predictive Value of Smartphone Photoscreening in Pediatric Practices. J Pediatr Ophthalmol Strabismus 2018 Aug 29;55(6):393-396. [doi: 10.3928/01913913-20180710-01]
20. Peterseim MMW, Rhodes RS, Patel RN, Wilson ME, Edmondson LE, Logan SA, et al. Effectiveness of the GoCheck Kids Vision Screener in Detecting Amblyopia Risk Factors. Am J Ophthalmol 2018 Mar;187:87-91. [doi: 10.1016/j.ajo.2017.12.020] [Medline: 29305313]

21. Wright KW, Spiegel PH, Thompson L, editors. Handbook of Pediatric Strabismus and Amblyopia. New York, NY, USA: Springer; 2006.
22. Model D, Eizenman M. An Automated Hirschberg Test for Infants. IEEE Trans. Biomed. Eng 2011 Jan;58(1):103-109. [doi: 10.1109/tbme.2010.2085000]
23. Eventov-Friedman S, Leiba H, Flidel-Rimon O, Juster-Reicher A, Shinwell ES. The red reflex examination in neonates: an efficient tool for early diagnosis of congenital ocular diseases. Isr Med Assoc J 2010 May;12(5):259-261 [FREE Full text] [Medline: 20929074]
24. Atchison DA, Thibos LN. Optical models of the human eye. Clin Exp Optom 2016 Mar 10;99(2):99-106 [FREE Full text] [doi: 10.1111/cxo.12352] [Medline: 26969304]
25. Colicchia G, Wiesner H, Zollman D. Photorefraction of the Eye. The Physics Teacher 2015 Feb;53(2):103-105. [doi: 10.1119/1.4905810]
26. Bhardwaj V. Axial Length, Anterior Chamber Depth-A Study in Different Age Groups and Refractive Errors. JCDR 2013. [doi: 10.7860/jcdr/2013/7015.3473]
27. Bobier W. Eccentric photorefraction. Ophthalmic and Physiological Optics 1985;5(2):235. [doi: 10.1016/0275-5408(85)90101-2]
28. Tadas B, Amir Z, Yao Chong L, Louis-Philippe M. OpenFace 2.0: Facial Behavior Analysis Toolkit. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). 2018 Presented at: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); 2018; Xi'an, China. [doi: 10.1109/FG.2018.00019]
29. Ridler TW, Calvard S. Picture Thresholding Using an Iterative Selection Method. IEEE Transactions on Systems, Man, and Cybernetics 1978 Aug;8(8):630-632. [doi: 10.1109/tsmc.1978.4310039]
30. Ballard D. Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 1981 Jan;13(2):111-122. [doi: 10.1016/0031-3203(81)90009-1]
31. Lobato-Rincón LL, Cabanillas-Campos MDC, Bonnin-Arias C, Chamorro-Gutiérrez E, Murciano-Cespedosa A, Sánchez-Ramos Roda C. Pupillary behavior in relation to wavelength and age. Front Hum Neurosci 2014 Apr 22;8:221 [FREE Full text] [doi: 10.3389/fnhum.2014.00221] [Medline: 24795595]

Abbreviations

FN: false negative
FP: false positive
GP: general practitioner
TN: true negative
TP: true positive

Edited by G Eysenbach; submitted 12.02.20; peer-reviewed by S Zhang, E Chioma; comments to author 29.02.20; revised version received 15.04.20; accepted 16.04.20; published 13.07.20.

Please cite as:
Ma S, Guan Y, Yuan Y, Tai Y, Wang T
A One-Step, Streamlined Children's Vision Screening Solution Based on Smartphone Imaging for Resource-Limited Areas: Design and Preliminary Field Evaluation
JMIR Mhealth Uhealth 2020;8(7):e18226
URL: http://mhealth.jmir.org/2020/7/e18226/
doi:10.2196/18226
PMID:32673243

Original Paper

An App for Classifying Personal Mental Illness at Workplace Using Fit Statistics and Convolutional Neural Networks: Survey-Based Quantitative Study

Yu-Hua Yan[1,2*], PhD; Tsair-Wei Chien[3*], MBA; Yu-Tsen Yeh[4], BA; Willy Chou[5,6*], MD; Shu-Chen Hsing[7*], MBA

[1]Superintendent Office, Tainan Municipal Hospital (Managed by Show Chwan Medical Care Corporation), Tainan, Taiwan

[2]Department of Hospital and Health Care Management, Chia Nan University of Pharmacy and Science, Tainan, Taiwan

[3]Department of Medical Research, Chi Mei Medical Center, Tainan, Taiwan

[4]Medical School, St George's, University of London, London, United Kingdom

[5]Department of Physical Medicine and Rehabilitation, Chung Shan Medical University, Taichung, Taiwan

[6]Department of Physical Medicine and Rehabilitation, Chiali Chi Mei Hospital, Tainan, Taiwan

[7]Respiratory Therapy Unit, Chi Mei Medical Center, Tainan, Taiwan

[*]these authors contributed equally

Corresponding Author:
Shu-Chen Hsing, MBA
Respiratory Therapy Unit
Chi Mei Medical Center
No 901, Chung Hwa Road, Yung Kung Dist
Tainan, 710
Taiwan
Phone: 886 +886937399106
Email: rtlisa1965@gmail.com

Abstract

Background: Mental illness (MI) is common among those who work in health care settings. Whether MI is related to employees' mental status at work is yet to be determined. An MI app is developed and proposed to help employees assess their mental status in the hope of detecting MI at an earlier stage.

Objective: This study aims to build a model using convolutional neural networks (CNNs) and fit statistics based on 2 aspects of measures and outfit mean square errors for the automatic detection and classification of personal MI at the workplace using the emotional labor and mental health (ELMH) questionnaire, so as to equip the staff in assessing and understanding their own mental status with an app on their mobile device.

Methods: We recruited 352 respiratory therapists (RTs) working in Taiwan medical centers and regional hospitals to fill out the 44-item ELMH questionnaire in March 2019. The exploratory factor analysis (EFA), Rasch analysis, and CNN were used as unsupervised and supervised learnings for (1) dividing RTs into 4 classes (ie, MI, false MI, health, and false health) and (2) building an ELMH predictive model to estimate 108 parameters of the CNN model. We calculated the prediction accuracy rate and created an app for classifying MI for RTs at the workplace as a web-based assessment.

Results: We observed that (1) 8 domains in ELMH were retained by EFA, (2) 4 types of mental health (n=6, 63, 265, and 18 located in 4 quadrants) were classified using the Rasch analysis, (3) the 44-item model yields a higher accuracy rate (0.92), and (4) an MI app available for RTs predicting MI was successfully developed and demonstrated in this study.

Conclusions: The 44-item model with 108 parameters was estimated by using CNN to improve the accuracy of mental health for RTs. An MI app developed to help RTs self-detect work-related MI at an early stage should be made more available and viable in the future.

(JMIR Mhealth Uhealth 2020;8(7):e17857) doi:10.2196/17857

KEYWORDS

respiratory therapist; ELMI app; Rasch analysis; convolutional neural network; mental health; mobile phone

Introduction

Background

Mental illness (MI) is common among those who have come in contact with it at the workplace [1] and globally account for 32.4% of years living with a disability [2]. An estimated 264 million people suffer from depression, with many of these people having symptoms of anxiety [3]. The lost productivity cost for depression and anxiety disorders adds up to 1 trillion US dollars globally each year [3]. Not only are absence and the direct costs harmful to organizations but the effects are also concerned with workers who suffer from MI and still remain on the job. Accordingly, employers are increasingly paying attention to presenteeism—decreased productivity because of health problems—among employees who remain present at work [4]. This is because presenteeism might result in a higher economic cost than absenteeism or the medical costs paid by employers [5]. It is worth investing in mental health promotion and prevention (MHPP) programs for employees in the workplace. Every dollar invested in stepping up treatment (or prevention) for common MIs (eg, depression and anxiety) leads to a fourfold return in better health and the passion or the capability to work [6].

An App Required to Assess MI

Although there has been an increase in MHPP programs globally in recent years, only 7% of such initiatives are carried out at the workplace [1]. In 2015, a Cochrane systematic review [7] evaluated evidence on the effectiveness of interventions to prevent occupational stress in health care workers [8]. Most of them were restricted to measuring work-related stress and/or burnout by using validated tools, particularly lacking the classifications, such as true (or false) health (or illness). As such, a novel application of a more holistic assessment of workplace mental health, including a broader scope of mental health outcomes (eg, the classification of response patterns in visual displays), is required to explore the potential benefits that improve effective classifications of true or false (ie, with less confidence) MI.

MI App at Workplace

Mobile health interventions (ie, MI apps) are used to monitor mental health and are an increasingly popular approach available for both individuals and organizations [9]. However, at present, there is a lack of research on the effectiveness of mobile MI apps in pattern classifications for smartphone users. This is because many assessment tools are merely focused on the strata of mental health [10-12] instead of the classifications of response patterns (eg, answering questions carelessly, cheating, or guessing) [13] to verify the respondents' MI classification (false or true with confidence).

Although numerous studies have been conducted on MI apps [9,14-18], few have used the algorithm of artificial neural networks (ANNs) in classifying MI, particularly with convolutional neural networks (CNNs). Traditionally, cutoff scores are often used in the classification of MI [19,20].

Most of the classifications were based on the extent (or, say, strata) to which MI or other disorders are determined by summation scores or equivalent measures of a scale (eg, patients with a psychotic disorder [21] or those with 2 or 3 groups of nurse burnouts [or bullied victims] at the workplace [19,20]). To the best of our knowledge, none of the studies have applied the response pattern to classify the features of MI (or other disorders) with CNN modules to highlight less confident cases of MI characterized by the aberrant response pattern that deviated from normal cases.

Targeted Health Care Workers

Although we note that mental health has been a concern in physicians [22] and nurses [1], other health care professionals and hospital staff should also be looked after considering the similar working environment to physicians and nurses and the nature of responsibilities that put them at great risk of MI. Thus, it is worth studying how workplace-based organizational interventions can improve the mental health and well-being of health care workers, including respiratory therapists (RTs) who played a vital role in the frontline care of patients with COVID-19 in 2020.

Study Objectives

This study aimed to (1) determine featured variables used for CNN in the classification of MI, (2) differentiate MI patterns endorsed by participants, and (3) design an MI app for smartphones as a web-based assessment.

Methods

Data Source

A survey-based quantitative study was conducted to invite RTs in Taiwan hospitals to answer questions about MI at the workplace. A total of 107 institutes (ie, 88 regional hospitals and 19 medical centers) were targeted according to the hospital list of Taiwan National Health Insurance Administration, and 1521 (753 and 768 in regional hospitals and medical centers, respectively) RTs who were registered in the Taiwan Society for Respiratory Care in January 2019 were included in the study.

If the confidence level and the intervals were set at 0.05 and plus or minus 5%, respectively, and applied to the population of 1521 registered RTs, 307 are required for the sample size [23,24]. We estimated that the percentage of candidates' refusal to respond to the survey was 40%. Therefore, the minimal number of the study sample size was 530 (n=307/[1−0.4]).

We delivered 6 copies of the 44-item questionnaire (ie, including 2 sets of the emotional labor and mental health [ELMH] questionnaire [25,26]; Multimedia Appendix 1) to 107 targeted workplaces in hospitals. A total of 642 (6×107) RTs with at least three months of experience working in the hospital were randomly selected and invited to complete the ELMH survey in March 2019.

Taking into consideration patient rights, interest, and privacy, an informed consent form was included in the mail to each hospital. Candidates were allowed to decline to answer the anonymous questionnaire. The questionnaire was asked to be mailed back to our study clerk in an enclosed envelope. A total of 352 (>307 required sample designed in the first paragraph)

questionnaires were eligible, with a return rate of 54.8% (352/642). Data were deposited in Multimedia Appendix 2.

This study was monitored by the institutional review board of Show Chwan Memorial Hospital with the approval ID number (1080105) before data collection. All hospital and study participant identifiers were stripped.

Featured Variables and Factor Domains With Unsupervised Learning

Unsupervised learning indicates agnostic aggregation of unlabeled data sets yielding groups or clusters of entities with shared similarities that may be unknown before the analysis step [27,28]. Featured variables consist of 44 items including the 24-item ELMH subscales. Exploratory factor analysis (EFA) was applied to determine factors retained in the study based on the eigenvalue ≥1.0. Factor scores with the Bartlett approach [29] to yield weighted domain scores for each factor (ie, domain) were analyzed to obtain personal measures and fit statistics using the Rasch model with continuous responses [30-32].

Next, participants were plotted to cluster classes based on the criteria of outfit mean square errors (MNSQs; cutting at 2.0 [33]) and measures at zero logits (ie, log odds) in 4 quadrants on axes x and y, respectively.

The definition of measure in logit and outfit MNSQ is worth mentioning. The former is analogical to the traditional summation score on a scale. The latter (ie, person outfit MNSQ) can be defined by the equation  , whereas L denotes the item length, *Oj* is the observed score on the item j, and E_j stands for the expected value on the item j and is computed by the Rasch model [30-32,34-37]. The aberrant pattern for an individual response can be detected and identified using the outfit MNSQ cutoff of 2.0 [33].

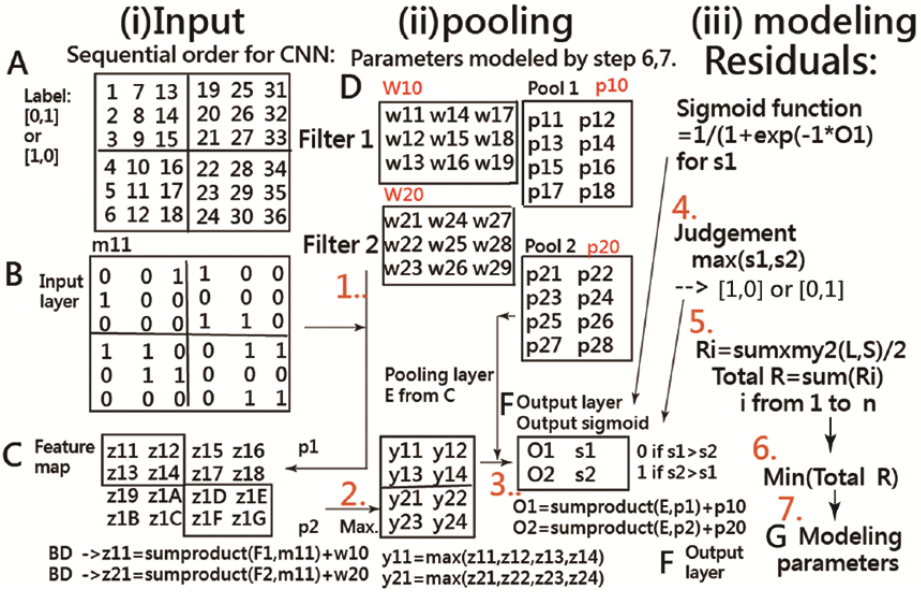
Factor scores yielded from all responses across all items were used to compare for an individual on each subscale because of each factor score following a normal distribution (ie, mean 0, SD 1).

The preliminary condition is required to determine how to obtain the factor score on a subscale when the respondent completes a survey. We performed a simple regression analysis on responses to predict the factor score for each score. The model parameters were used to compute the personal factor score on each subscale on an app.

Supervised Learning

Supervised learning employs *labeled* training data sets to yield a qualitative or quantitative output [27,38]. In this study, CNN was applied as supervised learning to build an ELMH prediction model to estimate 108 parameters (n=4×(10+17), with 4 sets of 10 parameters for featured maps and 17 parameters for pooled layers) because 4 categories are required in the CNN model (Figure 1 and Multimedia Appendix 3). Detailed information about CNN [38-40] is available in the literature [19,20].

Figure 1. Interpretation of the convolutional neural network algorithm in Microsoft Excel. CNN: convolutional neural network.



CNN Process

We illustrated the 2 categories of classification required for identification in the CNN process.

Data Arrangement

Three types of layers are included (eg, input, polling, and output) in Figure 1. Responses from 1 person assumed were placed into matrix A. Two sets of parameters were set (eg, 9 parameters and bias each in filter 1 and filter 2) to create 2 input layers of the feature map (eg, matrices C1 and C2 through panels B and D, respectively) based on the need of 2 (or more) categories in classification.

Step 1

Matrices C1 and C2 were created via a series of snapshots (eg, 9 elements on parameter sets in filter 1 and filter 2) of matrix B through step 1 using a sum-product function in Excel (Microsoft Corp; z11 and z12 at the bottom in Figure 1). Furthermore, the sigmoid function was performed to rescale all elements in matrices C1 and C2 to a range of 0 to 1.0.

Steps 2 and 3

The maximum value in the 4 elements as a convolutional snapshot in matrices C1 and C2 (refer to step 2 and the bottom max function in Figure 1) was selected. Two condensed pooling layers with 8 elements were constructed in matrix E via step 3; the bottom sum-product function with parameters in pooling layers (ie, P1 and P2 with 8 parameters and another bias) is illustrated in Figure 1.

Steps 4 and 5

The results (O1 and O2 in the output layer [step 4]) can be used to determine which category has a higher probability if the sigmoid function has been applied to rescale the 2 elements in matrices C1 and C2 into a range of 0 to 1.0 (step 5).

Tasks for Performing CNN

Task 1: Compute the CNN Prediction Accuracy Rate

CNN in Microsoft (MS) Excel [19,20] was performed to estimate model parameters and compute the prediction accuracy rate (1−the number of misclassification/352). Comparisons of prediction accuracy were evaluated using discrimination analysis on (1) factor scores and outfit MNSQ and (2) factor scores alone in each subscale.

Task 2: App Detecting MI for a Web-Based Assessment

A 44-item MI app was designed to predict RT mental health using the CNN algorithm and the model parameters. Summation scores were yielded for each domain and transformed into factor scores using the estimated parameters obtained by using the simple regression analysis (refer to the subsection *Featured Variables and Factor Domains With Unsupervised Learning* under the *Methods*).

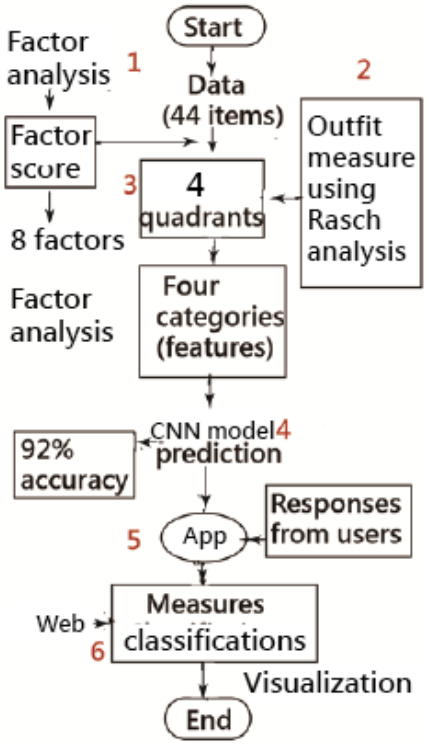
The resulting classification and domain scores will appear with visual displays on smartphones. The visual representations with the category characteristic curve and the category probabilities [35,36] were shown on a dashboard using Google Maps.

Statistical Tools and Data Analysis

IBM SPSS Statistics 18.0 for Windows (SPSS Inc) and MedCalc 9.5.0.0 for Windows (MedCalc Software) was used to perform descriptive statistics, EFA, discrimination analysis, the Fisher exact test or chi-square test on frequency distributions among groups, simple regression analyses, and then compute the CNN prediction accuracy rate. A significant level of type I error was set at 0.05.

A visual representation displaying the classification effect was plotted using 4 curves based on the Rasch rating scale model [35]. The study flowchart and the CNN modeling process are shown in Figure 2 [41] and Multimedia Appendix 3, respectively.

Figure 2. The study flowchart. CNN: convolutional neural network.



Results

Demographic Data of the 352 RTs

The demographic data of the RTs are shown in Table 1. We can see that 62.8% (221/352) are in medical centers and 37.2% (131/352) in regional hospitals. There were 88.4% (311/352) females and 11.6% (41/352) males with an average age of 37 (SD 9.5) years. The frequencies of participants' educational levels were 4.8% (17/352), 83.5% (294/352), and 11.6% (41/352) for college, university, and graduate school, respectively. The unmarried (single or divorced) and married status accounted for 61.4% (216/352) and 38.6% (136/352), respectively. Only these categories of age, marital status, and RT ability hierarchical level (ie, strata in a clinical RT ladder system) present statistically significant differences in a frequency distribution. Other detailed information about the sample is presented in Table 1.

Table 1. The demographical characteristics (n=352).

Characteristic	Medical center, n (%)	Regional hospital, n (%)	Total, n (%)	*P* value
Total	221 (62.8)	131 (37.2)	352 (100)	<.001[a]
Gender				.28
Female	193 (87.3)	118 (90.1)	311 (88.4)	
Male	28 (12.7)	13 (9.9)	41 (11.6)	
Age (years)				.003[a]
<30	81 (36.7)	26 (19.8)	107 (30.4)	
31-39	43 (19.5)	42 (32.1)	85 (24.1)	
40-49	74 (33.5)	45 (34.4)	119 (33.8)	
>50	23 (10.4)	18 (13.7)	41 (11.7)	
Education				.33
College	13 (5.9)	4 (3.1)	17 (4.8)	
Undergraduate degree	180 (81.4)	114 (87.0)	294 (83.5)	
Postgraduate degree	28 (12.7)	13 (9.9)	41 (11.7)	
Marital status				.003[a]
Single	143 (64.7)	62 (47.3)	205 (58.2)	
Married	74 (33.5)	62 (47.3)	136 (38.6)	
Others or divorced	4 (1.8)	7 (5.4)	11 (3.2)	
Job title				.30
Nonmanager	207 (93.7)	120 (91.6)	327 (92.9)	
Manager	14 (6.3)	11 (8.4)	25 (7.1)	
Work tenure (years)				.33
≤1	15 (6.8)	5 (3.8)	20 (5.7)	
1~3	45 (20.4)	19 (14.5)	64 (18.2)	
4-7	41 (18.6)	24 (18.3)	65 (18.5)	
8-10	17 (7.7)	15 (11.5)	32 (9.1)	
>10	103 (46.5)	68 (51.9)	171 (48.5)	
Ability hierarchical level				<.001[a]
RT0[b]	59 (26.7)	62 (47.3)	121 (34.4)	
RT1	91 (41.2)	44 (33.6)	135 (38.4)	
RT2	43 (19.5)	4 (3.1)	47 (13.4)	
RT3	11 (5)	1 (0.8)	12 (3.4)	
None	17 (7.6)	20 (15.2)	37 (10.4)	
Current preceptor				.33
No	63 (28.5)	41 (31.3)	104 (29.5)	
Yes	158 (71.5)	90 (68.7)	248 (70.5)	

[a]Denotes that the significant level using the chi-square test reaches the probability at 0.05.

[b]RT: respiratory therapist.

Sample Classifications Using 4 Quadrants

There were 8 factors that were extracted from the study data using EFA (Multimedia Appendix 4), including (1) mental health, (2) attitude to patients, (3) diversified, (4) adjustment, (5) persevering, (6) teamwork, (7) physical health, and (8) behavior, with 16, 5, 4, 4, 5, 4, 4, and 2 questions, respectively. The distributions of factor scores were drawn in box plots (Figure 3). The regression equations used for computing the factor score of each domain in the MI app are listed below:

Factor score (1) = −3.2516 + 0.08190 × sum (1) (**1**)

Factor score (2) = −7.7062 + 0.3711 × sum (2) (**2**)

Factor score (3) = −3.6067 + 0.3104 × sum (3) (**3**)

Factor score (4) = −4.7129 + 0.3349 × sum (4) (**4**)

Factor score (5) = −4.9833 + 0.2821 × sum (5) (**5**)

Factor score (6) = −5.4556 + 0.5632 × sum (6) (**6**)

Factor score (7) = −2.5251 + 0.2220 × sum (7) (**7**)

Factor score (8) = −4.2156 + 0.6377× sum (8) (**8**)

After performing the Rasch analysis with continuous responses [30-32], 4 types of characteristics were observed (n=6, 63, 265, and 18 in 4 quadrants; Figure 4) using the criteria of outfit MNSQ and the measure at 2.0 and zero logit.

Table 2 shows that 75% (265/352) of candidates fell under type III mental health (with confidence because of outfit MNSQ <2.0; Figure 4) [42], indicating that most RTs are mentally healthy at the workplace. No difference was found in frequency distribution among all groups. It is worth noting that the different colors of bubbles correspond to MI types, and the size corresponds to the MI scores. The 4 types are clearly illustrated on the dashboard with MI and MI-free on the left side and false MI and false MI–free on the right-hand side.

Figure 3. Eight domains using factor analysis to classification.

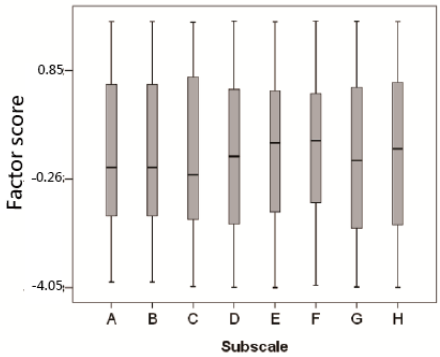


Figure 4. Four classes separated by 2 variables using the Rasch analysis. MI: mental illness.

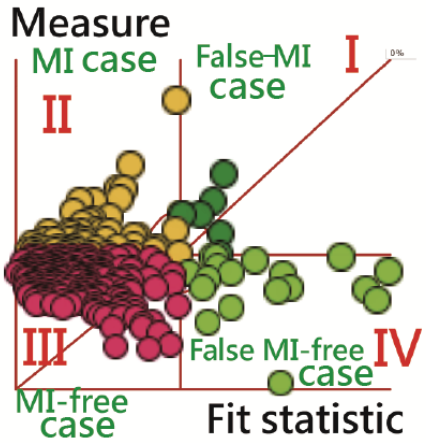


Table 2. Classifications for demographical characteristics (n=352).

Characteristic	I[a], n	II[a], n	III[a], n	IV[a], n	Total, n	*P* value
Gender						.53
Female	6	50	242	13	311	
Male	0	13	23	5	41	
Age (years)						.55
<30	3	22	80	2	107	
31-39	1	15	64	5	85	
40-49	2	19	89	9	119	
>50	0	7	32	2	41	
Education						.53
College	0	4	13	0	17	
Undergraduate degree	5	50	224	15	294	
Postgraduate degree	1	9	28	3	41	
Marital status						.52
Single	3	38	157	7	205	
Married	3	22	101	10	136	
Single or divorced	0	3	7	1	11	
Job title						.58
Nonmanager	5	57	247	18	327	
Manager	1	6	18	N/A[b]	25	
Work tenure (years)						.73
≤1	0	2	15	3	20	
1-3	3	13	48	0	64	
4-7	0	11	52	2	65	
8-10	0	9	22	1	32	
>10	3	28	128	12	171	
Ability hierarchical level						.38
RT0[c]	3	15	99	4	121	
RT1	2	23	102	8	135	
RT2	0	10	35	2	47	
RT3	0	5	6	1	12	
None	1	10	23	3	37	
Current preceptor						.28
No	3	18	80	3	104	
Yes	3	45	185	15	248	

[a]Quadrants from I to IV.

[b]N/A: not applicable.

[c]RT: respiratory therapist.

Tasks to Compute the Accuracy Rate in the Predictive Model

Comparisons of the prediction accuracy are shown in Table 3. We can see that the prediction accuracy rate (0.92 = [5 + 59 + 246 + 14] / 352) in the 9 variables (including outfit MNSQ) at the top panel is higher than that (0.86 = [5 + 53 + 210 + 12] / 352) in the 8 variables at the bottom panel.

The 44-item CNN model yields a higher accuracy rate (0.92), which is similar to the accuracy rate in the 9-variable model (including outfit MNSQ) at the top panel in Table 3. The 108

model parameters were embedded to create an MI app with the 44-item using the CNN model in the hope of identifying workplace MI for RTs.

Table 3. Comparison of prediction accuracy using a different number of variables.

Original or predicted	I[a], n	II[a], n	III[a], n	IV[a], n	Total, n
Eight-factor scores and outfit MNSQ[b] of Rasch analysis					
I	*5* [c]	1	0	0	6
II	3	*59*	1	0	63
II	0	17	*246*	2	265
II	4	0	0	*14*	18
Eight-factor scores					
I	*5*	1	0	0	6
II	10	*53*	0	0	63
II	2	18	*210*	35	265
II	3	1	2	*12*	18

[a]Quadrants from I to IV.

[b]MNSQ: mean square error.

[c]Italicization denotes the number of classification correction.

MI App Classifying ELMH for a Web-Based Assessment

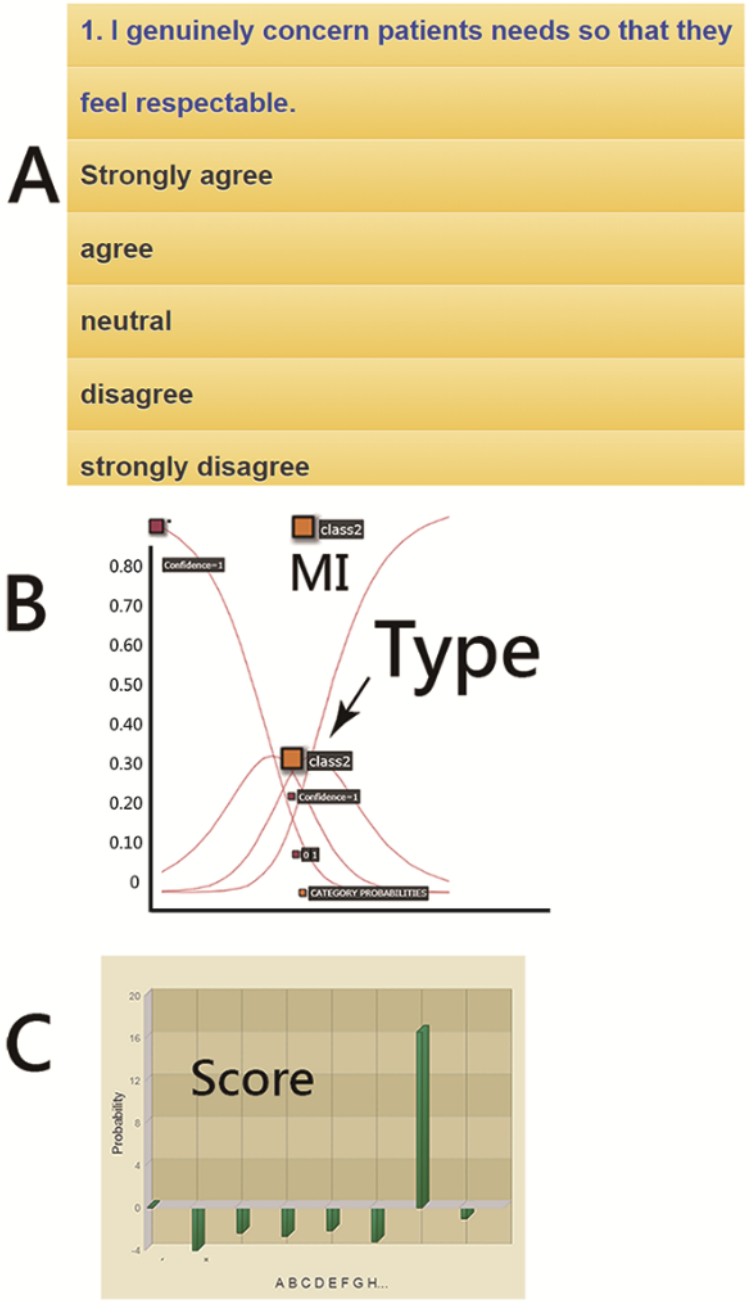
An available MI app for RTs predicting ELMH was developed and is shown in Figure 5. Readers are invited to click on the links [42] to experience the MI app on their own (Multimedia Appendix 5). It is worth noting that all 108 model parameters are embedded in the 44-item CNN model for classification of MI on 2 major personal abilities in dealing with (1) external changes at work and (2) internal skills in resilience to endure hardship using the ELMH for assessment [25,26].

One resulting example is illustrated in the middle panel in Figure 5. From this, we can see that the MI with a high probability (0.99) is shown above curve II, indicating that class II is classified. The feature of the category characteristic curve and the category probabilities [36,37] is that all the summation of probabilities on the intersections between any vertical line and the curves equals 1.0.

Eight domain scores are presented, and the MI resulting from the physical health problems most significantly displayed the highest factor score computed by equation (7), as illustrated in the highest bar at the bottom panel in Figure 4.

With the assessment on the app, the resulting plot along with the meaning of the 4 classes is displayed in the middle panel in Figure 5. After answering the questions on the app, the RT (1) can understand his or her MI class (ie, I, II, III, or IV) and (2) then needs a consultation with physicians specialized in mental health or occupational medicine at an early stage if the MI is class II. Informative messages with regard to the low confidence in MI assessment were given to us if the MI class is classified as either I or IV because of their response to questions significantly misfitting to the model's expectation.

Figure 5. Snapshot of the app assessment using convolutional neural network. MI: mental illness.



Discussion

Principal Findings

The 44-item CNN model yields a higher accuracy rate (0.92), which is similar to the accuracy rate in the 9-variable model (including outfit MNSQ) shown in Table 3. We observed that (1) 8 domains in ELMH were obtained by EFA, (2) 4 types of mental health (n=6, 63, 265, and 18 in 4 quadrants) were classified by the Rasch analysis, (3) the 44-item model yields a higher accuracy rate (0.92), and (4) an ELMI app is available and viable for RTs to predict mental health.

What This Knowledge Adds to What We Already Know

Articles on MI apps have been published in the literature [14-18,43,44]. Over 45 articles were found by searching the keywords (app[title] and “mental health”[title]) in PubMed Central (PMC) on January 10, 2020. However, none of the articles provided an acceptable scheme (eg, CNN algorithm) to classify the MI levels and aberrant patterns with a dashboard displayed on Google Maps.

As seen in the literature, a cutting point scheme for nurse burnouts was proposed using approximately equal sample sizes in each category [45]. The cutting point scheme was criticized because it arbitrarily assumed an equal sample size across the burnout levels (ie, high, moderate, and low) [46]. In addition,

the cases with false positive and false negative result in a lower prediction accuracy rate in the traditional cutting point scheme.

On the other hand, the CNN, a well-known deep learning method, can improve the prediction accuracy (up to 7.14%) [47]. With the discrimination analysis in Table 2, we can see that the prediction accuracy rate (0.92) is equal to the result from CNN. Applying the algorithm (eg, CNN or discrimination function) into an MI app used for classifying MI features is the feature of this study.

The reason for using the 44-item CNN model instead of the 9 variables (including outfit MNSQ) as shown in Table 3 is that the outfit MNSQ is not available in the CNN module. Accordingly, it was not possible to use outfit MNSQ if Rasch computerized adaptive testing was not performed on the web. Thus, the CNN was chosen in this study.

In recent years, the RT demand and health care quality have been particularly concerning in Taiwan because of the aging society and air pollution in the everyday environment exacerbating the COVID-19 outbreak in 2020 [48-51]. The RT-related service units include those in the critical care unit (as well as chronic respiratory care wards and home care) and the emergency room in hospitals. Respiratory care was newly established in Taiwan's clinical settings. The rising demand for RTs is indispensable and of importance to the quality of care in hospitals. As such, the ELMH [26] is worthy of introduction and application to health care professionals and support staff in hospitals [25].

Although a survey is required to understand the severity of MI in RTs at the workplace, the adjusted emotional labor and expression [26] at work and the skills in resilience to endure hardship are necessary to empower RTs to carry out their duties. A viable and suitable tool for MI assessment, such as the MI app introduced in this study, should also be applied in the nearest future.

What it Implies and What Should Be Changed

CNN can improve prediction accuracy (up to 7.14%) [47]. We conducted CNN in MS Excel (Multimedia Appendix 4), which is rare and unique in the literature. The major difference between a traditional ANN and CNN is that on CNN, only the last layer of a CNN is fully connected, whereas in ANN, all neurons are interconnected with each other [40]. The ANN algorithm in Excel in comparison with CNN is shown in Multimedia Appendix 6.

Over 1127 articles have been found using the keyword *convolutional neural network* [Title] searched in PMC on June 26, 2020 [19]. None of the articles used MS Excel to perform the CNN. The interpretations of the CNN concept and the process or even the parameter estimations are shown in Multimedia Appendix 3, which is another feature of this study.

The third feature is a breakthrough using the CNN approach to predict MI for RTs. The method used in this study can be mimicked by other health care professionals in the hospital.

Using 4 MI classifications in quadrants is another highlight of this study. We applied the Rasch analysis with continuous responses [30-32] and factor scores together for classifying the 4 classes of MI. In Table 2, we can see that the prediction accuracy rates are identical using either CNN or discrimination analysis (Table 3). Applying the classification to the MI app is challenging but worthy in the health care community, especially with the 4 categories (or classes) of MI, MI-free, false MI, and false MI–free (Figure 4). The latter 2 are based on the response pattern deviating from the normal (or model in mathematics).

Furthermore, the curves of category probabilities based on the Rasch rating scale model [35] are shown in Figure 5. The binary categories (eg, success and failure on an assessment in the psychometric field) have been applied to health-related outcomes [52-56]. However, no published article provided the animation-type dashboard with 4 categories that can be shown on Google Maps, as we have shown in Figure 5.

Strengths of This Study

It is easy to create an MI app if the designer only uploads items to the website. We applied the CNN algorithm along with the model's parameters to design the routine on an MI app that is used to classify MI risk for RTs in hospitals (Figure 5), which has never been seen before for ELMH implementation [24,25] on mobile phones.

As with all forms of web-based technology, advances in health communication technology are rapidly emerging [54]. Mobile MI apps are promising and worth considering for many health care professionals. A web-based MI app (Figure 5) can be modified to immediately inform users whether and when they should take actions or follow-up to see a psychiatrist and how to improve their behaviors and attitudes or strengthen their skills in resilience, given that their lifestyle remains unchanged.

The MI app is worth using to promote mental health of RTs using their smartphones. Interested readers are recommended to see Multimedia Appendices 5 and 7, one for the MP4 file and another for the app, and see (1) the details about responding to questions and (2) the real experience in answering the 44-item ELMH questionnaire as a web-based assessment.

The CNN module in MS Excel is unique and innovative (Multimedia Appendix 3). Users who are not familiar with the CNN software (eg, Python) can apply our Excel-Visual Basic for Applications module to conduct CNN-related research in the future. The module is not limited to the 4 classifications we used in this study. The multiclassification module can be performed by adding the layers on the CNN. Any other types of self-assessment, such as work bully, depression, and dengue fever, can apply the CNN model to predict and classify the levels of harm of diseases in the future.

Limitations and Suggestions

There are limitations to our study. First, although the psychometric properties of the 44-item ELMH have been validated for measuring MI for RTs, as shown in Multimedia Appendix 1, there is no evidence to support that the 44-item ELMH is suitable for other health care professionals or RTs in other regions. We recommend additional studies using their own approaches and the CNN model to estimate the parameters to compare and contrast with this study.

Second, we did not explore the possibility of any improvement in predictive accuracy. For instance, whether other featured variables (eg, mean, SD, and LZ [defined as the standardized log-likelihood of the respondent's response vector] index or the addition of sociodemographic information) applied to the CNN model can increase the accuracy rate is worthy of further study. However, the disadvantage of inputting demographical data is that it will take more time to complete and lead to reluctance in response as well as concerns regarding personal privacy. A small study (Multimedia Appendix 7) was conducted. A greater number of variables involving sociodemographic information (even with featured variables) cannot be guaranteed to have a higher accuracy rate in classification. Nonetheless, more studies are required to verify this in the future.

Third, the classification scheme using 4 quadrants with the Rasch outfit MNSQ and MI measure constructed by 8-factor scores is challenging. All these factor scores were independent with normal distribution (ie, ~$N(0,1)$). Whether using the original summation score for each domain is available and simple in classification with the Rasch model warrants further studies in the future.

Fourth, the ELMH is an 8-dimensional construct. The CNN model may ignore the issue of dimensionality and have a favorable prediction effect that should be examined and verified in the future.

Finally, the study sample was taken from Taiwanese RTs in a survey. The model parameters estimated for the ELMH version are only suitable for the Chinese (particularly for Taiwanese) in health care settings. Generalizing this MI app (eg, easy-to-use on the web) might be somewhat limited and constrained because the app is merely a prototype instead of being fully designed for internet use. Additional improvements are needed to redesign the features of the MI app in use for RTs in the future.

Conclusions

We demonstrated the features and contributions of this study as follows: (1) CNN performed in MS Excel, (2) ELMH applied to assess MI for RTs, (3) a web-based MI app demonstrated display results using the visual dashboard on Google Maps, and (4) the category probability curves based on the Rasch rating scale model along with the CNN prediction model. The novelty of the MI app with the CNN algorithm improves the predictive accuracy of MI for RTs. It is expected to help RTs self-assess and detect work-related MI at an early stage.

Authors' Contributions

YH conceived and designed the study, SC performed the statistical analyses, and YT was in charge of recruiting study participants. TC helped design the study, collected information, and interpreted the data. SC monitored the research. All authors have read and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 17 KB - mhealth_v8i7e17857_app1.docx]

Multimedia Appendix 2
Study data set (Microsoft Excel).
[XLSX File (Microsoft Excel File), 2403 KB - mhealth_v8i7e17857_app2.xlsx]

Multimedia Appendix 3
Convolutional neural network using Microsoft Excel to interpret the process.
[DOCX File , 893 KB - mhealth_v8i7e17857_app3.docx]

Multimedia Appendix 4
Eight factors extracted from the data.
[XLSX File (Microsoft Excel File), 98 KB - mhealth_v8i7e17857_app4.xlsx]

Multimedia Appendix 5
App on the web assessing mental illness.
[DOCX File , 13 KB - mhealth_v8i7e17857_app5.docx]

Multimedia Appendix 6
Artificial neural network model in Excel.
[DOCX File , 13 KB - mhealth_v8i7e17857_app6.docx]

Multimedia Appendix 7
Whether the more variables are better in the convolutional neural network module with a small study.
[DOCX File , 13 KB - mhealth_v8i7e17857_app7.docx]

References

1. Gray P, Senabe S, Naicker N, Kgalamono S, Yassi A, Spiegel JM. Workplace-based organizational interventions promoting mental health and happiness among healthcare workers: a realist review. Int J Environ Res Public Health 2019 Nov 11;16(22):4396 [FREE Full text] [doi: 10.3390/ijerph16224396] [Medline: 31717906]
2. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
3. Mental Health in the Workplace. World Health Organization. URL: http://www.who.int/mental_health/in_the_workplace/en/ [accessed 2010-01-10]
4. Schultz AB, Edington DW. Employee health and presenteeism: a systematic review. J Occup Rehabil 2007 Sep;17(3):547-579. [doi: 10.1007/s10926-007-9096-x] [Medline: 17653835]
5. Ammendolia C, Côté P, Cancelliere C, Cassidy JD, Hartvigsen J, Boyle E, et al. Healthy and productive workers: using intervention mapping to design a workplace health promotion and wellness program to improve presenteeism. BMC Public Health 2016 Nov 25;16(1):1190 [FREE Full text] [doi: 10.1186/s12889-016-3843-x] [Medline: 27884132]
6. Mental Health: Massive Scale-Up of Resources Needed if Global Targets Are to Be Met. World Health Organization. URL: http://www.who.int/mental_health/evidence/atlas/atlas_2017_web_note/en/ [accessed 2019-01-10]
7. About Cochrane Reviews: What is a Systematic Review? Cochrane Library. URL: https://www.cochranelibrary.com/about/about-cochrane-reviews [accessed 2019-01-10]
8. Ruotsalainen JH, Verbeek JH, Marine A, Serra C. Cochrane Database of Systematic Reviews. Hoboken, NJ: John Wiley & Sons; 2015.
9. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019;10:2745 [FREE Full text] [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
10. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
11. Milne-Ives M, Lam C, de Cock C, van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
12. Berghöfer A, Martin L, Hense S, Weinmann S, Roll S. Quality of life in patients with severe mental illness: a cross-sectional survey in an integrated outpatient health care model. Qual Life Res 2020 Mar 13:- epub ahead of print. [doi: 10.1007/s11136-020-02470-0] [Medline: 32170584]
13. Rudner L, Wright B. Diagnosing person misfit. Rasch Measurement Transactions 1995;9(2):430 [FREE Full text]
14. Renfrew ME, Morton DP, Morton JK, Hinze JS, Beamish PJ, Przybylko G, et al. A web- and mobile app-based mental health promotion intervention comparing email, short message service, and videoconferencing support for a healthy cohort: randomized comparative study. J Med Internet Res 2020 Jan 6;22(1):e15592 [FREE Full text] [doi: 10.2196/15592] [Medline: 31904578]
15. Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE. Patient interest in mental health mobile app interventions: demographic and symptom-level differences. J Affect Disord 2020 Feb 15;263:216-220. [doi: 10.1016/j.jad.2019.11.083] [Medline: 31818779]
16. Kenny R, Fitzgerald A, Segurado R, Dooley B. Is there an app for that? A cluster randomised controlled trial of a mobile app-based mental health intervention. Health Informatics J 2019 Nov 8:1460458219884195. [doi: 10.1177/1460458219884195] [Medline: 31702409]
17. Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CR, Amaro E, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res 2019 Nov 7;21(11):e14269 [FREE Full text] [doi: 10.2196/14269] [Medline: 31697244]
18. McEwan K, Richardson M, Sheffield D, Ferguson FJ, Brindley P. A smartphone app for improving mental health through connecting with urban nature. Int J Environ Res Public Health 2019 Sep 12;16(18):3373 [FREE Full text] [doi: 10.3390/ijerph16183373] [Medline: 31547286]
19. Lee Y, Chou W, Chien T, Chou P, Yeh Y, Lee H. An app developed for detecting nurse burnouts using the convolutional neural networks in Microsoft Excel: population-based questionnaire study. JMIR Med Inform 2020 May 7;8(5):e16528 [FREE Full text] [doi: 10.2196/16528] [Medline: 32379050]
20. Ma S, Chou W, Chien T, Chow JC, Yeh Y, Chou P, et al. An app for detecting bullying of nurses using convolutional neural networks and web-based computerized adaptive testing: development and usability study. JMIR Mhealth Uhealth 2020 May 20;8(5):e16747 [FREE Full text] [doi: 10.2196/16747] [Medline: 32432557]

21. Evans L, Wewiorski NJ, Ellison ML, Ni P, Harvey KL, Hunt MG, et al. Development and validation of an instrument to measure staff perceptions of recovery climate and culture in mental health programs. Psychiatr Serv 2020 Jun 1;71(6):570-579. [doi: 10.1176/appi.ps.201900181] [Medline: 32151213]
22. Carrieri D, Briscoe S, Jackson M, Mattick K, Papoutsi C, Pearson M, et al. 'Care under pressure': a realist review of interventions to tackle doctors' mental ill-health and its impacts on the clinical workforce and patient care. BMJ Open 2018 Feb 2;8(2):e021273 [FREE Full text] [doi: 10.1136/bmjopen-2017-021273] [Medline: 29420234]
23. Dillman D. Constructing the Questionnaire, Mail and Internet Surveys. New York, USA: John Wiley & Sons; 2000.
24. Sample Size Calculator. The Survey System. 2019. URL: https://www.surveysystem.com/sscalc.htm [accessed 2019-12-01]
25. Lee CY, Hsieh PC, Su HF. The relationship between emotional labor and mental health among nurses in Catholic hospitals in Taiwan. Taiwan Gong Gong Wei Sheng Za Zhi 2013;32(2):140.
26. Brotheridge CM, Grandey AA. Emotional labor and burnout: comparing two perspectives of 'people work'. J Vocat Behav 2002 Feb;60(1):17-39. [doi: 10.1006/jvbe.2001.1815]
27. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. Acad Pathol 2019;6:2374289519873088 [FREE Full text] [doi: 10.1177/2374289519873088] [Medline: 31523704]
28. Buehler L, Rashidi H. Bioinformatics Basics: Applications in Biological Science and Medicine. Second Edition. New York, USA: Taylor and Francis Group; 2005.
29. Hershberger S. Factor scores. In: Everitt D, Howell C, editors. Encyclopedia of Statistics in Behavioral Science. New York, USA: John Wiley; 2005.
30. Chien T, Lee Y, Wang H. Detecting hospital behaviors of up-coding on DRGs using Rasch model of continuous variables and online cloud computing in Taiwan. BMC Health Serv Res 2019 Sep 4;19(1):630 [FREE Full text] [doi: 10.1186/s12913-019-4417-2] [Medline: 31484551]
31. Chien T, Shao Y, Kuo S. Development of a Microsoft Excel tool for one-parameter Rasch model of continuous items: an application to a safety attitude survey. BMC Med Res Methodol 2017 Jan 10;17(1):4 [FREE Full text] [doi: 10.1186/s12874-016-0276-2] [Medline: 28068901]
32. Chien T, Shao Y. Rasch analysis for continuous variables. Rasch Meas Trans 2016;30(1):1574-1576 [FREE Full text]
33. Linacre J. What do Infit and Outfit, Mean-Square and Standardized Mean? Institute for Objective Measurement. 2003. URL: https://www.rasch.org/rmt/rmt162f.htm [accessed 2020-06-30]
34. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, US: University of Chicago Press; 1980.
35. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978 Dec;43(4):561-573. [doi: 10.1007/bf02293814]
36. Wu HM, Shao Y, Chien TW. Student's performance shown on google maps using online Rasch analysis. J Appl Measure 2019;21(2):1-10.
37. Bond T, Fox C. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Second Edition. Mahwah, NJ: Erlbaum; 2007.
38. Caruana R, Niculescu-Mizil A. An Empirical Comparison of Supervised Learning Algorithms. In: Proceedings of the 23rd international conference on Machine learning. 2006 Presented at: ICML'06; June 25-29, 2006; Pittsburgh, Pennsylvania, USA. [doi: 10.1145/1143844.1143865]
39. Guzmán MG, Kourí G. Dengue: an update. Lancet Infect Dis 2002 Jan;2(1):33-42. [doi: 10.1016/s1473-3099(01)00171-2] [Medline: 11892494]
40. Islam KT, Raj R. Performance of SVM, CNN, and ANN with BoW, HOG, and Image Pixels in Face Recognition. In: 2nd International Conference on Electrical & Electronic Engineering. 2017 Presented at: ICEEE'17; December 27-29, 2017; Rajshahi, Bangladesh. [doi: 10.1109/ceee.2017.8412925]
41. Bengio Y. Learning deep architectures for AI. FNT in Mach Learn 2009;2(1):1-127. [doi: 10.1561/2200000006]
42. Chien T. Kano Model Used for Display Mental Health Status for Respiratory Therapists. Health Up. URL: http://www.healthup.org.tw/gps/rtkano.htm [accessed 2020-05-14]
43. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019;10:2745 [FREE Full text] [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
44. Chien T. CNN Used for Predicting Mental Illness for Respiratory Therapists. YouTube. URL: https://youtu.be/PGKT9HIz53s [accessed 2020-05-14]
45. Maslach C, Jackson S. Maslach Burnout Inventory Manual. Second Edition. Palo Alto, CA: Consulting Psychologists Press; 1986.
46. Schaufeli WB, van Dierendonck D. A cautionary note about the cross-national and clinical validity of cut-off points for the Maslach burnout inventory. Psychol Rep 1995 Jun;76(3 Pt 2):1083-1090. [doi: 10.2466/pr0.1995.76.3c.1083] [Medline: 7480470]

47. Sathyanarayana A, Joty S, Fernandez-Luque L, Ofli F, Srivastava J, Elmagarmid A, et al. Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth 2016 Nov 4;4(4):e125 [FREE Full text] [doi: 10.2196/mhealth.6562] [Medline: 27815231]
48. Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, et al. The rate of underascertainment of novel coronavirus (2019-nCoV) infection: estimation using Japanese passengers data on evacuation flights. J Clin Med 2020 Feb 4;9(2):419 [FREE Full text] [doi: 10.3390/jcm9020419] [Medline: 32033064]
49. Zhao S, Lin Q, Ran J, Musa S, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020 Mar;92:214-217 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.050] [Medline: 32007643]
50. Jin Y, Cai L, Cheng Z, Cheng H, Deng T, Fan Y, for the Zhongnan Hospital of Wuhan University Novel Coronavirus ManagementResearch Team, Evidence-Based Medicine Chapter of China International ExchangePromotive Association for MedicalHealth Care (CPAM). A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Mil Med Res 2020 Feb 6;7(1):4 [FREE Full text] [doi: 10.1186/s40779-020-0233-6] [Medline: 32029004]
51. Are Coronavirus Diseases Equally Deadly? NBC news. URL: https://www.nbcnews.com/health/health-news/coronavirus-diseases-comparing-covid-19-sars-mers-numbers-n1150321 [accessed 2020-03-05]
52. Wylie L, Corrado AM, Edwards N, Benlamri M, Murcia Monroy DE. Reframing resilience: strengthening continuity of patient care to improve the mental health of immigrants and refugees. Int J Ment Health Nurs 2020 Feb;29(1):69-79. [doi: 10.1111/inm.12650] [Medline: 31478332]
53. Lee Y, Lin K, Chien T. Application of a multidimensional computerized adaptive test for a clinical dementia rating scale through computer-aided techniques. Ann Gen Psychiatry 2019;18:5 [FREE Full text] [doi: 10.1186/s12991-019-0228-4] [Medline: 31131014]
54. Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014 Jan 13;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]
55. Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the negative acts questionnaire-revised: Rasch analysis of workplace bullying. J Med Internet Res 2014 Feb 17;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]
56. Chien T, Lin W. Improving inpatient surveys: web-based computer adaptive testing accessed via mobile phone QR codes. JMIR Med Inform 2016 Mar 2;4(1):e8 [FREE Full text] [doi: 10.2196/medinform.4313] [Medline: 26935793]

Abbreviations

ANN: artificial neural network
CNN: convolutional neural network
EFA: exploratory factor analysis
ELMH: emotional labor and mental health
MHPP: mental health promotion and prevention
MI: mental illness
MNSQ: mean square error
MS: Microsoft
PMC: PubMed Central
RT: respiratory therapist

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by T Muto, 翔 闫, J Rahman; comments to author 10.03.20; revised version received 24.03.20; accepted 03.06.20; published 31.07.20.

Please cite as:
Yan YH, Chien TW, Yeh YT, Chou W, Hsing SC
An App for Classifying Personal Mental Illness at Workplace Using Fit Statistics and Convolutional Neural Networks: Survey-Based Quantitative Study
JMIR Mhealth Uhealth 2020;8(7):e17857
URL: https://mhealth.jmir.org/2020/7/e17857
doi:10.2196/17857
PMID:32735232

Original Paper

User Engagement Among Diverse Adults in a 12-Month Text Message–Delivered Diabetes Support Intervention: Results from a Randomized Controlled Trial

Lyndsay A Nelson[1,2], PhD; Andrew Spieker[3], PhD; Robert Greevy[3], PhD; Lauren M LeStourgeon[1,2], MPH; Kenneth A Wallston[4], PhD; Lindsay S Mayberry[1,2,5,6], MS, PhD

[1]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Center for Health Behavior and Health Education, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

[4]Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, United States

[5]Center for Diabetes Translation Research, Vanderbilt University Medical Center, Nashville, TN, United States

[6]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Lyndsay A Nelson, PhD
Department of Medicine
Vanderbilt University Medical Center
2525 West End Ave
Suite 450
Nashville, TN, 37203
United States
Phone: 1 6158757621
Email: lyndsay.a.nelson@vumc.org

Abstract

Background: Text message–delivered interventions are a feasible and scalable approach for improving chronic disease self-care and reducing health disparities; however, information on long-term user engagement with these interventions is limited.

Objective: The aim of this study is to examine user engagement in a 12-month text message–delivered intervention supporting diabetes self-care, called REACH (Rapid Education/Encouragement And Communications for Health), among racially and socioeconomically diverse patients with type 2 diabetes (T2D). We explored time trends in engagement, associations between patient characteristics and engagement, and whether the addition of a human component or allowing patients to change their text frequency affected engagement. Qualitative data informed patients' subjective experience of their engagement.

Methods: We recruited patients with T2D for a randomized trial evaluating mobile phone support relative to enhanced treatment as usual. This analysis was limited to participants assigned to the intervention. Participants completed a survey and hemoglobin A1c (HbA1c) test and received REACH text messages, including self-care promotion texts, interactive texts asking about medication adherence, and adherence feedback texts. For the first 6 months, texts were sent daily, and half of the participants also received monthly phone coaching. After 6 months, coaching stopped, and participants had the option to receive fewer texts for the subsequent 6 months. We defined engagement via responses to the interactive texts and responses to a follow-up interview. We used regression models to analyze associations with response rate and thematic and structural analysis to understand participants' reasons for responding to the texts and their preferred text frequency.

Results: The participants were, on average, aged 55.8 (SD 9.8) years, 55.2% (137/248) female, and 52.0% (129/248) non-White; 40.7% (101/248) had ≤ a high school education, and 40.7% (101/248) had an annual household income <US $25,000. The median response rate to interactive texts was 91% (IQR 75%-97%) over 12 months. Engagement gradually declined throughout the intervention but remained high. Engagement did not differ by age, gender, education, income, diabetes duration, insulin status, health literacy, or numeracy. Black race and worse baseline medication adherence and HbA1c were each associated with lower engagement, although the effects were small. Nearly half of the participants chose to continue receiving daily texts for the last 6 months of the intervention. Participants who continued daily text messages said they wanted to continue experiencing benefits to their health, whereas those who chose fewer texts said that the daily texts had helped them create routines and they no longer needed them as often. Engagement was not impacted by receiving coaching or by participants' chosen text frequency.

Conclusions: Well-designed interactive text messages can engage diverse patients in a self-care intervention for at least 1 year. Variation in and reasons for frequency preference suggest that offering a frequency choice may be important to users' engagement.

(JMIR Mhealth Uhealth 2020;8(7):e17534) doi:10.2196/17534

KEYWORDS

engagement; text messaging; mobile health; mHealth; mobile phone; technology; diabetes mellitus, type 2; self-management; self-care; medication adherence

Introduction

Background

For many patients with chronic conditions, including type 2 diabetes (T2D), consistent, daily self-care (eg, taking medications, eating healthy, and exercising) is critical to avoid adverse health outcomes. However, patients with T2D face many barriers to self-care, including a lack of diabetes-related education and information, social barriers (eg, lack of motivation and caregiver stress), and difficulties with finances and transportation [1]. Racial/ethnic minorities and patients with low socioeconomic status (SES) tend to experience more barriers, which leads to worse self-care and health outcomes compared with non-Hispanic white patients and those with a high SES [2-5]. Text message–delivered interventions may offer an opportune pathway to extend self-care support to the hardest to reach and most vulnerable populations. More than 96% of adults in the United States use cell phones and find them to be essential for daily functioning [6,7]. Although smartphone ownership is growing and the digital divide has narrowed, individuals who have lower incomes, have less education, or live in rural areas are still less likely to have internet access or to own a smartphone [6,8,9]. Text messaging does not require internet access and is the most popular cell phone activity among all mobile phone users [10]. Evidence is growing for the efficacy of text message–delivered interventions to improve self-care for T2D and other chronic conditions [11-13]; however, there is a paucity of research examining how users engage with these interventions, particularly in the long term [14,15].

User engagement is essential to the ultimate success of any intervention. If target users are not attending to the intervention or lose interest quickly, potential effects may be attenuated or nullified. Approaches to measuring user engagement with technology-delivered interventions vary considerably across studies, making it a challenge to identify patterns and generate predictions. To help address this issue, Perski et al [16] synthesized past work on engagement with digital behavior change interventions to develop an integrative definition of engagement. They proposed engagement to be a multidimensional construct defined by both the extent of usage (often measured objectively) and a subjective experience characterized by attention, interest, and affect (often captured through qualitative data such as interviews) [16]. In many text message–delivered intervention studies, engagement is either not reported at all or reported using only a single metric (eg, average text message response rate) [15], which limits understanding of why participants were engaged, how engagement may change throughout an intervention, and factors contributing to engagement. Furthermore, among the studies reporting on engagement, intervention duration is typically 6 months or less [15,17] and, therefore, not representative of how users might engage in a longer experience. In these studies, text message response rates vary widely, but are often low, with some studies reporting rates as low as 17% [15,18,19]. Variability is likely due to variation in user attributes (eg, health literacy status and diabetes self-efficacy) and/or intervention characteristics across studies, but associations (or lack thereof) are often not reported [15]. In the few studies reporting associations, which span texting, interactive voice response, and internet-based interventions, patient characteristics such as older age [20], being nonwhite [20,21], and lower health literacy [20,22,23] have been linked to lower engagement.

Closely examining the features of the intervention may inform the conditions under which participants are more likely to engage. In text message interventions, specifically, the ideal frequency (ie, dose) of sending text messages to sustain engagement is unclear [12]. Suffoletto [24] referred to this as the Goldilocks problem: sending too few texts may not produce a strong enough effect, whereas sending too many texts may lead to participant fatigue and burnout. Furthermore, text messaging interventions often include additional modalities for delivering content, such as a human support component [25]. However, few studies have compared engagement between a condition receiving only automated text messages and a condition receiving texts in conjunction with human support.

Exploring engagement in long-term text messaging interventions is a critical step in understanding whether text messages are a viable option for improving and sustaining health outcomes. Identifying the determinants of intervention engagement can inform the design of interventions to optimize engagement [12,26]. Studies reporting on engagement with text messaging interventions tend to be short term, based on small sample sizes, and involve predominantly non-Hispanic white patients, limiting the understanding of long-term, generalizable results [17]. Important gaps in knowledge must be addressed to enhance the expanding use of text messaging technology to support chronic disease self-care in vulnerable groups at the greatest risk for worse outcomes.

Objectives

We used mixed methods to examine user engagement in a 12-month text message–delivered intervention designed to support diabetes self-care, called REACH (Rapid Education/Encouragement And Communications for Health) [27,28]. We sought to explore how engagement changed over the course of the intervention, patient characteristics associated with engagement, whether the addition of a human component or letting participants choose their text frequency affected

engagement, and patients' reasons for their preferred text message frequency.

Methods

Study Design and Eligibility

This research was conducted as part of a larger randomized controlled trial (RCT), evaluating the effects of mobile phone–based support on diabetes self-care and hemoglobin A_{1c} (HbA_{1c}) [27]. The trial and intervention details have been previously described in our development and protocol papers [27-29]. For this study, we analyzed data for participants who were randomly assigned to the intervention.

We recruited patients from federally qualified health centers (FQHCs) and Vanderbilt University Medical Center (VUMC) primary care clinics in and around Nashville, Tennessee. Eligible patients were aged ≥18 years, diagnosed with T2D, prescribed a daily diabetes medication, and responsible for administering their diabetes medications. In addition, patients were required to own a cell phone with text messaging capability and speak and read in English. We excluded participants whose most recent HbA_{1c} value within 12 months was <6.8% and who had auditory limitations or an inability to orally communicate, as determined by trained research assistants (RAs). We also excluded patients who failed a brief cognitive screener [30] to help ensure the accuracy of the measures and data integrity. Finally, due to the requirements of the intervention, we excluded patients who were unable to receive, read, or send text messages after a demonstration by an RA.

Procedure

The Vanderbilt University institutional review board approved all study procedures. Interested and eligible patients completed informed consent, a baseline survey, and an HbA_{1c} test. RAs collected additional information from the participants' electronic health record. RAs entered participants' data into Research Electronic Data Capture (REDCap) [31]. Participants' relevant survey responses were transferred from REDCap to a digital health platform called MEMOTEXT, via an automated application programming interface. MEMOTEXT used participant information to tailor, schedule, and send text messages to participants. When participants completed follow-up assessments at 3 and 6 months, updated patient information was used to retailor the text message content. When participants completed their 6-month assessment, they were given the option to continue receiving daily text messages or to receive fewer text messages, for the remaining 6 months of the intervention (described in *The Intervention* below). Participants could earn a total of US $210 for completing all study measures (ie, through 15 months for the larger RCT), with the payment schedule increasing for longer-term follow-ups.

We used strategic purposeful sampling to invite a subset of intervention participants to complete a follow-up interview after they finished their participation in the trial. Interviews were designed to assess perceptions of the intervention and understand dose choice and potential for implementation. Factors that informed our sampling approach included assigned condition, age, gender, race, education, income, clinic site, whether they owned a basic phone or smartphone, and their chosen text frequency. Interviews were conducted either in person or by phone and took approximately 20 min to complete (range 11-40 min). All interviews were audio recorded and transcribed verbatim. Participants were paid an additional US $40 for completing the interview. Herein, we report specifically on the participants' comments about responding to text messages and reasons for their text message frequency choice after 6 months of receiving daily texts.

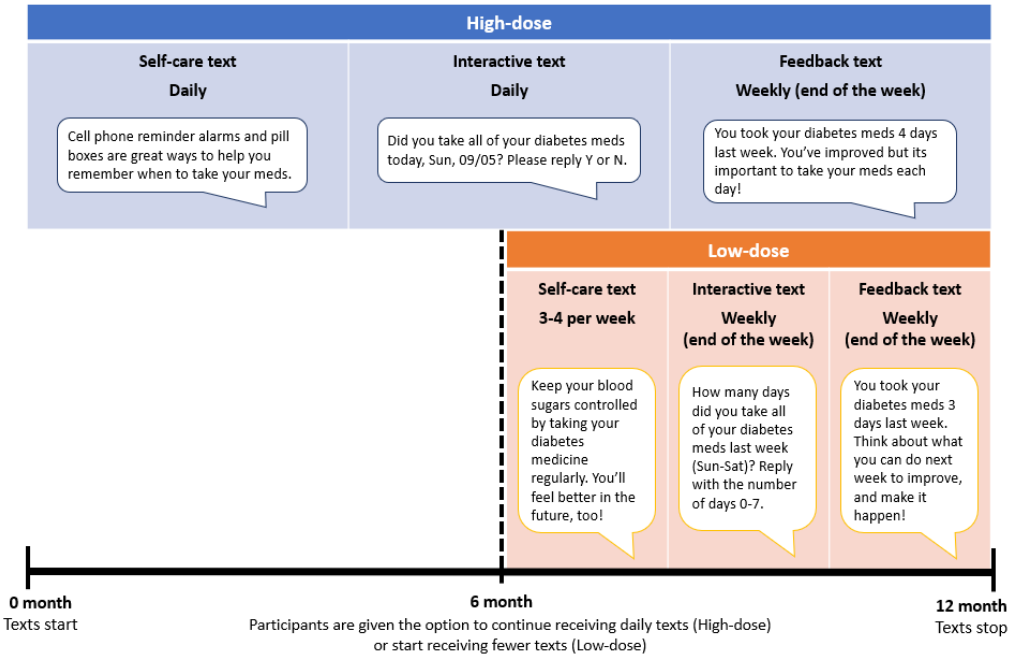
The Intervention

All participants received self-care promotion text messages. Each message was either tailored to address user-specific barriers to medication adherence based on the Information-Motivation-Behavioral Skills model [32,33] or addressed other self-care behaviors (ie, diet, exercise, and self-monitoring of blood glucose). Participants chose a window of time to receive this text (eg, between 3 PM and 6 PM). Participants also received interactive text messages asking about medication adherence. We scheduled these to be sent after participants took their last dose of diabetes medication, but before their bedtime, to allow time to respond. At the end of each week, participants received an adherence feedback message based on their responses to the interactive text; feedback included an encouraging statement tailored to whether the participant's adherence improved, stayed the same, or declined relative to the week prior.

For the first 6 months, all participants received daily self-care texts and daily interactive texts. After 6 months, participants had the option to receive fewer text messages for the remaining 6 months of the intervention (ie, low-dose) or continue to receive daily text messages (ie, high-dose; Figure 1). If we could not reach participants for their 6-month assessment and therefore were not able to present this option, they continued to receive daily text messages (high-dose).

Participants assigned to receive REACH text messages were randomly assigned to also receive monthly phone coaching to set diabetes self-care goals for the first 6 months at a ratio of 1:1 [27,29]. After 6 months, phone coaching ended for those so assigned, all participants were offered the low-dose option described earlier, and all continued to receive REACH text messages for the next 6 months.

Figure 1. High-dose and low-dose REACH text message frequencies and content examples. REACH: Rapid Education/Encouragement And Communications for Health.



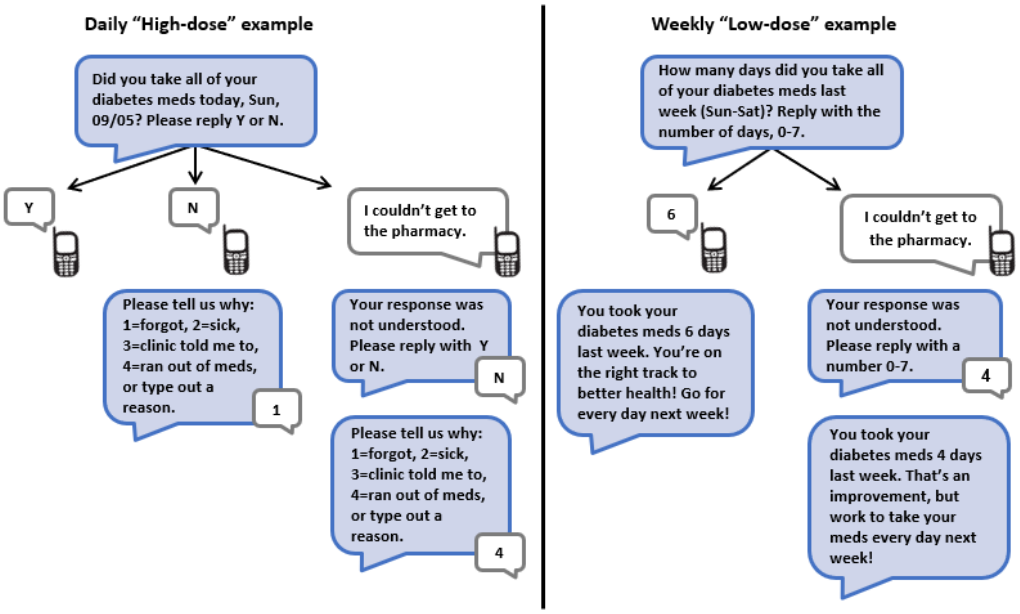
Functionality and Monitoring of Interactive Text Messages

When sent daily, the interactive text message asked participants if they took all their diabetes medication that day, requesting a yes or no response (Figure 2). On the basis of usability testing results [28], we allowed a range of acceptable response options representing yes and no (eg, yeah, Y, yep, N, and nope). If participants provided a no response, a follow-up text was sent asking why with various response options. Participants received their adherence feedback text based on their daily responses at the end of the week. When sent weekly for participants who chose low-dose, the interactive text asked participants how many days that week they had taken all their diabetes medication, requesting a 0 to 7 response. After participants provided a response, they received their adherence feedback text message (Figure 2).

If the participant responded to the interactive text message with something other than a solicited response, error messages were triggered (Figure 2). The system accepted any response until the next text message was sent the subsequent day.

We monitored participants' responses to the interactive texts weekly throughout the trial. If a participant did not respond to the interactive message for 2 consecutive weeks, an RA called the participant to identify any technical problems. To avoid pressuring the participants to respond, the RA only asked whether they were experiencing technical problems with receiving or responding to text messages and would troubleshoot as needed. If a participant remained nonresponsive after we confirmed that they were not having problems, we did not make repeat calls.

Figure 2. Comparison of functionality for the interactive text message based on whether it was sent daily (high-dose) or weekly (low-dose). Blue messages are intervention texts, and white messages are examples of participant responses.



Measures

Demographics and Clinical Characteristics

We collected self-reported age, gender, race, ethnicity, education (ie, years in school), income, diabetes duration (ie, years since a diabetes diagnosis), and insulin status.

Psychosocial Factors

Participants completed self-reported measures of health literacy (Brief Health Literacy Screen [34]), numeracy (Subjective Numeracy Scale, 3-item version [35]), and stressors (Tool for Assessing Patients' Stressors, 8-item version [36]).

Diabetes Self-Care

Participants also completed a self-reported measure of diabetes self-efficacy (Perceived Diabetes Self-Management Scale, 4-item version [37]) and diabetes medication adherence (Adherence to Refills and Medications Scale for Diabetes [38]). In addition, all participants completed an HbA_{1c} test via venipuncture or point of care by the patient's clinic or using a mail-in HbA_{1c} kit provided and analyzed by CoreMedica Laboratories.

Engagement

We defined engagement as *any* response to the interactive text messages depicted in Figure 2. If a participant did not respond, it was coded as nonengagement for that day. We calculated text message response rates by dividing each participant's number of responses by the total number of interactive texts sent to that participant.

To provide a more comprehensive understanding of engagement, beyond average response rate, we assessed response rates over time to explore change in engagement and examined both patient and intervention characteristics associated with response rate to identify factors that may impact engagement. In addition, to understand the more nuanced reasons about why participants chose to engage, we supplemented our objective measure of engagement (ie, response rate) with interview data to inform patients' subjective experience of their engagement and reasons for their chosen text frequency. This is consistent with the conceptualization of engagement proposed by Perski et al [16].

Analyses

All statistical analyses were performed using R version 3.5.1. We described patient characteristics using means and SDs or frequencies and percentages as appropriate. We computed the following descriptive statistics regarding text message response rates: means, medians, and the first and third quartiles (interquartiles).

Long-Term Engagement

Response rates were examined in 3 ways: during the full 12 months of the intervention, the first 6 months (predose choice), and the last 6 months (postdose choice). There was variation in when participants were offered the low-dose option due to variation in when they completed their 6-month assessment; therefore, we defined the period of prechoice engagement as ≤160 days (the period where no participants had yet been given the choice) and postchoice engagement as ≥240 days (the point at which all participants had made the choice). We excluded the first 30 days to mitigate the impact of participants' acclimation to the intervention (ie, learning curve).

We used logistic regression to examine whether there was a time trend in the odds of responding during the first 6 months. If participants withdrew during this period, we coded them as having a nonresponse from the day they withdrew through 160 days. We also used logistic regression to examine whether there was a time trend in the last 6 months of the intervention. If participants withdrew after making their choice, we coded them as having a nonresponse from the time they withdrew through 365 days, either daily or weekly, depending on their choice. Both time-trend analyses used generalized estimating equations

(GEE) to fit the model under a working exchangeable correlation structure to account for repeated outcome measures [39].

Patient Characteristics and Engagement

We used simple linear regression to examine associations between patient characteristics and 6-month response rate and between patient characteristics and change in response rate during the first 6 months. We defined change as each participant's slope as estimated from a linear fit over time (30 days to 160 days). Patient characteristics included demographics and clinical characteristics (age, gender, race, education, income, diabetes duration, and insulin status), baseline psychosocial factors (health literacy, numeracy, and stressors), baseline diabetes management (diabetes self-efficacy, medication adherence, and HbA_{1c}), and clinic site (FQHC vs VUMC). In addition, we used unadjusted odds ratios to examine associations between patient characteristics and dose choice (ie, high-dose or low-dose).

Intervention Features and Engagement

To assess whether a human component (ie, monthly phone coaching) enhanced engagement with text messages, we used unadjusted linear regression with the condition (text messages only or text messages plus phone coaching) as the predictor and the 6-month response rate as the outcome. We further used unadjusted logistic regression to examine whether a time trend in the odds of responding during the first 6 months differed based on the assigned condition.

To evaluate whether there was an association between dose choice and postchoice response rate, we used GEE with a working exchangeable correlation structure to account for repeated outcome measures. We used logistic models to examine (1) whether a time trend in the odds of responding during the last 6 months differed based on dose choice, (2) whether there was an association between dose choice and the odds of responding postchoice, and (3) whether there was an association between dose choice and change in the odds of responding from before to after making the choice. Covariates for the latter 2 models included age, gender, and race.

Subjective Engagement and Reasons for Dose Choice

We used thematic analysis with NVivo version 11 to identify, organize, and interpret themes in the follow-up interview transcripts [40]. First, we used an inductive approach to construct a codebook based on coders' preliminary read of the transcripts. The codebook indicated the themes identified from the data. We then applied the initial codebook to a subset of the transcripts to clarify the definitions and resolve discrepancies. All transcripts were then coded independently, with one-third coded by both reviewers to evaluate interrater reliability (κ=0.89). We then used deductive analysis on content within relevant themes or structural analysis to answer specific research questions. Specifically, to understand reasons for engagement (responding to texts), we identified subthemes and described content within the themes *medication reminders* and *accountability*. To understand participants' dose choice, we conducted structural coding of all responses to interview questions about that choice (ie, "After six months in REACH, you chose to receive 3-4 text messages per week/to receive daily text messages for the rest of the program. Why did you choose that option?").

Results

Participants

In the trial, 256 participants were assigned to receive text messages. Of those enrolled, 8 participants were not included in these analyses due to early withdrawal (n=5), a technical error resulting in text messages not being sent (n=2), or opting out of text messages during the first 30 days (n=1). The remaining 248 participants were included in the quantitative analyses. Participants were, on average, aged 55.8 (SD 9.8) years, 55% (137/248) were female, 52% (129/248) were nonwhite, 39% (97/248) were black, 41% (101/248) had a high school degree or less education, and 41% (101/248) had an annual household income of <US $25,000 (Table 1).

Of the 46 participants invited to complete a follow-up interview, 36 (78%) did so. The characteristics of the interviewed participants were similar to those of the larger sample: aged 51.5 (SD 11.0) years, 56% (20/36) female, 67% (24/36) nonwhite, 53% (19/36) black, 44% (16/36) with a high school degree or less education, and 53% (19/36) with an annual household income of <US $25,000.

Table 1. Patient characteristics (N=248).

Patient characteristics	Values
Age (years), mean (SD)	55.8 (9.8)
Gender, n (%)	
Male	111 (44.8)
Race/ethnicity, n (%)	
Non-Hispanic White only	119 (48.0)
Non-Hispanic Black only	97 (39.1)
Hispanic	15 (6.0)
Other and multiracial	17 (6.9)
Education (years), mean (SD)	14.1 (3.0)
Annual household income (US $), n (%)	
<25,000	101 (40.7)
≥25,000	125 (50.4)
Refused/do not know	22 (8.9)
Diabetes duration (years), mean (SD)	10.9 (7.5)
Insulin status (% prescribed), n (%)	123 (49.6)
Health literacy (BHLS[a]), mean (SD)	13.0 (2.6)
Numeracy (SNS-3[b]), mean (SD)	4.4 (1.3)
Stressors (TAPS-8[c]), mean (SD)	3.8 (2.0)
Diabetes self-efficacy (PDSMS-4[d]), mean (SD)	13.8 (3.5)
Medication adherence (ARMS-D[e]), mean (SD)	39.8 (3.8)
HbA_{1c}[f] (%), mean (SD)	8.6 (1.8)
Clinic site (FQHC[g]), n (%)	103 (41.5)
Assigned condition (text messages plus phone coaching), n (%)	123 (49.6)

[a]BHLS: Brief Health Literacy Scale; Possible score range: (3-15).

[b]SNS-3: Subjective Numeracy Scale; Possible score range: (1-6).

[c]TAPS-8: Tool for Assessing Patient Stressors; Possible score range: (0-8).

[d]PDSMS: Perceived Diabetes Self-Management Scale; Possible score range: (4-20).

[e]ARMS-D: Adherence to Refills and Medications Scale for Diabetes; Possible score range: (11-44), items are reserve-scored so higher scores indicate greater adherence.

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

[g]FQHC: federally qualified health center.

Long-Term Engagement

Participants responded to a mean of 81% (SD 23%; median 91%; IQR 75%-97%) of the interactive text messages over the 12-month intervention. Less than 1% of participants' responses to these texts were unexpected responses (ie, a response other than yes or no to the daily text message or 0 through 7 to the weekly text message). Response rates were negatively skewed such that over two-thirds of participants' response rates were greater than 75%. Participants' average response rate decreased over time, although the predicted rate never dropped below 70% (Figure 3).

We also examined engagement separately during the first and last 6 months of the intervention. During the first 6 months, participants responded to a mean of 85% (SD 20%; median 94%; IQR 80%-98%) of the interactive text messages, and during the last 6 months, participants responded to a mean of 81% (SD 22%; median 92%; IQR 75%-97%). For both time periods, we found statistically significant time trends such that participants had decreased odds of responding over time, although there was little substantive change (Table 2). We estimated the odds of responding to be about 0.7% lower on a given day as compared with the previous day during the first 6 months and to be about 0.4% lower on a day-by-day basis during the latter 6 months (Table 2).

Figure 3. Predicted response rates over the 12-month intervention using a locally weighted scatterplot smoothing (LOWESS) curve.

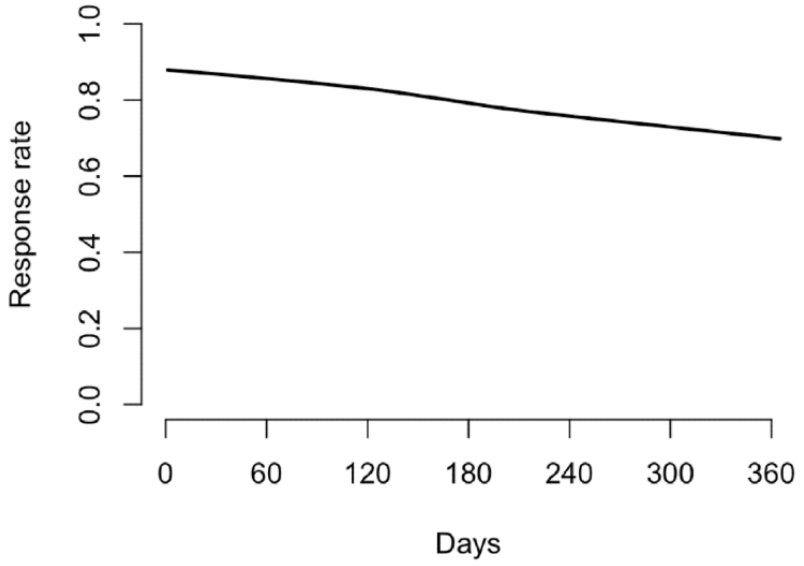


Table 2. Results for logistic regression models examining time trends in responding.

Time trends and dose choice associations	Estimate	95% CI	*P* value
Time trends in response rates before dose choice (first 6 months)			
Day	0.993	0.988-0.998	.008[a]
Condition (text messages plus phone coaching)[b]	0.840	0.527-1.337	.46
Day×condition (text messages plus phone coaching)	1.002	0.999-1.006	.09
Time trends in response rates after dose choice (latter 6 months)			
Day	0.996	0.995-0.998	<.001[a]
Low-dose choice[c]	0.613	0.364-1.032	.07
Low-dose choice×day	1.001	0.999-1.002	.16
Association between dose choice and (subsequent) response rate[d]			
Low-dose choice	0.842	0.612-1.157	.29
Association between dose choice and (subsequent) change in response rate[d]			
Low-dose choice	0.745	0.538-1.032	.08
Postchoice responding	0.439	0.334-0.579	<.001[a]
Low-dose choice×postchoice responding	1.208	0.931-1.567	.16

[a]Significant based on $P<.05$.

[b]Reference group is texts only.

[c]Reference group is choosing high-dose (high-dose choice).

[d]Covariates included age, gender, and race.

Participant Characteristics and Engagement

Patients' clinic site, age, gender, education, income, diabetes duration, insulin status, health literacy, numeracy, and diabetes self-efficacy were not associated with 6-month engagement (Table 3). However, black race (compared with white), worse baseline medication adherence and worse HbA_{1c} were associated with lower engagement, although the effects were small (2%-7% lower; Table 3). We did not find evidence of an association between any measured patient characteristics and change in 6-month engagement, and we did not find evidence of an association between any measured patient characteristics (including prechoice response rate) and dose choice (Table 3).

Table 3. Associations between patient characteristics and engagement.

Patient characteristics	6-month response rate, estimate (95% CI)	Change in 6-month response rate, estimate (95% CI)	Low-dose choice, OR[a] (95% CI)
Age (years)	–0.001 (–0.003 to 0.003)	–0.001 (–0.004 to 0.002)	0.989 (0.961-1.017)
Gender			
Male	–0.015 (–0.065 to 0.039)	0.030 (–0.027 to 0.085)	0.733 (0.406-1.295)
Race/ethnicity[b]			
Non-Hispanic Black vs non-Hispanic White	–0.070[c] (–0.129 to –0.012)	–0.013 (–0.073 to 0.049)	0.845 (0.432-1.545)
Education (years)	0.002 (–0.008 to 0.013)	0.002 (–0.008 to 0.012)	1.074 (0.980-1.189)
Annual household income ≥US $25,000	0.040 (–0.013 to 0.098)	0.021 (–0.100 to 0.187)	1.055 (0.600-1.881)
Diabetes duration (years)	–0.001 (–0.004 to 0.002)	0.001 (–0.003 to 0.004)	0.989 (0.951-1.028)
Prescribed insulin	–0.040 (–0.092 to 0.013)	–0.015 (–0.069 to 0.041)	0.808 (0.451-1.417)
Health literacy (BHLS[d])	–0.000 (–0.010 to 0.009)	–0.000 (–0.013 to 0.013)	0.994 (0.890-1.123)
Numeracy (SNS-3[e])	–0.001 (–0.023 to 0.021)	0.002 (–0.021 to 0.026)	1.055 (0.841-1.322)
Stressors (TAPS-8[f])	–0.006 (–0.019 to 0.007)	–0.008 (–0.022 to 0.004)	0.945 (0.821-1.080)
Diabetes self-efficacy (PDSMS-4[g])	0.008 (–0.000 to 0.016)	0.004 (–0.003 to 0.011)	1.004 (0.924-1.096)
Medication adherence (ARMS-D[h])	0.014[c] (0.006 to 0.023)	0.005 (–0.003 to 0.013)	0.997 (0.925-1.084)
HbA_{1c}[i] (%)	–0.019[c] (–0.036 to –0.004)	0.001 (–0.015 to 0.018)	0.869 (0.728-1.006)
Clinic site (FQHC[j])	0.005 (–0.045 to 0.058)	–0.016 (–0.072 to 0.038)	0.680 (0.378-1.188)
Assigned condition (text messages plus phone coaching)	0.008 (–0.044 to 0.060)	0.024 (–0.031 to 0.081)	1.438 (0.847-2.583)
Prechoice response rate	N/A[k]	N/A	1.247 (0.262-7.288)

[a]OR: odds ratio.

[b]Due to the small number of participants who identified as either Hispanic or multiracial, we did not report associations for these groups.

[c]Significant association based on 95% CI.

[d]BHLS: Brief Health Literacy Scale.

[e]SNS: Subjective Numeracy Scale.

[f]TAPS-8: Tool for Assessing Patient Stressors.

[g]PDSMS: Perceived Diabetes Self-Management Scale.

[h]ARMS-D: Adherence to Refills and Medications Scale for Diabetes.

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

[j]FQHC: federally qualified health center.

[k]N/A: not applicable.

Intervention Features and Engagement

Participants' assigned condition (ie, whether they received text messages only or text messages plus phone coaching) was not associated with their 6-month response rate (Table 2). In addition, the odds of responding over time did not differ by condition (Table 2).

Among participants given the option to receive fewer text messages at 6 months (n=213), 55.9% (119/213) of participants chose to receive fewer texts and 44% (94/213) chose to continue receiving daily texts. The remaining 35 participants were not asked this question because of not completing the 6-month assessment, being unreachable by phone, or an RA error. Dose choice was not associated with the odds of responding postchoice (Table 2). When we examined whether dose choice was associated with a change in the odds of response, from before to after making the choice, there was no association. Odds of responding were lower over time in both groups but did not differ by choice (Table 2).

Subjective Engagement

Reasons for Engaging With Text Messages

The most common theme about why participants responded to the interactive text is because they used it as a reminder for taking their medicine (n=19). For instance, participants described how receiving and responding to the text served to routinize their medication taking and helped increase their awareness:

> *My little reminder I received on the cell phone, the text, would, you know, help me remember to take my medicine. Helped me get on a regular schedule.* [Hispanic white male, aged 45 years]

Participants explained how they came to anticipate the text messages which helped them remember to take their medications before they received the message. Sometimes, the text message was a cue to action if they had not yet taken their medication:

> *I was forgetting to take my medications. Once I started getting the messages...I would go get it and take it. And it would remind me...I knew it was coming, because I picked the evening hours to make sure that, at eleven o'clock, I would get the message...If I didn't, I know I had enough time to still take it for that day.* [Hispanic white female, aged 50 years]

The other, less common theme, explaining participants' responses to the interactive text messages, was feeling accountable—either to themselves or to the REACH team (n=9). Some participants specifically described how they looked forward to being able to respond *yes* to the text message when they had taken their medication, whereas others talked about negative feelings if they had not taken their medication and not wanting to respond with a *no*:

> *Each day I got a message. If I didn't take my medicine, I was taking it. And if I did take my medicine, I was waiting on that message to feel like I did something right.* [Non-Hispanic black female, aged 26 years]

> *So, it was like when I didn't want to respond, it was because I hadn't taken my medicine. Then, I felt guilty and I felt that I was cheating myself.* [Non-Hispanic black female, aged 31 years]

In contrast, some described the importance of being honest in their responses and not wanting to lie if they had not taken all their diabetes medications that day. Finally, a few participants described feeling like someone was keeping tabs on them, one referring to us as "big brother":

> *It kept you on track, gave you reminders, and it was like you having like this little voice in the phone telling you, "Make sure you take your medicine," and, you know, do what you're supposed to do. Stay on track.* [Black female, ethnicity not reported, aged 60 years]

Reasons for Dose Choice

Of the 36 participants who completed a follow-up interview, 64% (23/36) chose the low-dose option, 25% (9/36) chose to keep the high-dose option, and 11% (4/36) were not asked because of the reasons listed earlier. Common themes in participants' reasons for choosing the low-dose option included that they had established a routine (n=6), the text messages were interfering with their schedule (n=4), and they were becoming frustrated with the text messages (n=6). Participants who mentioned having a routine shared how the text messages had helped them to create a habit, and therefore, they felt they no longer needed the text messages as often:

> *I think I had reached the point where the text messages had the desired effect on me, and so I was sort of trained at that point to anticipate them and to be aware of their purpose...so fewer text messages still had the same purpose.* [Non-Hispanic white male, aged 60 years]

Some participants who switched to low-dose felt that their schedules were too busy to attend to and keep up with the messages on a daily basis, whereas others shared how they were frustrated with the text messages either because the content felt redundant or because they did not like how many texts were being sent:

> *I find you're getting too much. It's too much. It's too much. And a reminder here or there is fine. But to be getting constant messages like that...I personally find it annoying.* [Non-Hispanic black female, aged 60 years]

Less commonly mentioned reasons included wanting more autonomy to manage diabetes without the assistance of the text messages (n=2) and issues pertaining to their cell phone or plan (n=3; eg, limited plans made it difficult to receive daily text messages and still be able to send text messages toward the end of the month).

The most common reason participants provided for wanting to continue receiving high-dose was the helpfulness of the daily text messages (n=9). Specifically, participants stated that the text messages benefited their health, and they wanted to continue receiving them daily to keep experiencing that benefit. Many mentioned how the interactive texts made them feel more engaged in their diabetes self-care and reminded them to take their medications each day:

> *I'm very forgetful...It's going to sound silly, but [it's] sort of like a security blanket. You know, by receiving that message, it's like, okay, I felt like I was more involved in taking care of my diabetes. I felt like I was participating more in something.* [Non-Hispanic white male, aged 45 years]

> *If for some reason I hadn't taken my medication that day, because I got busy doing other things, when a message comes in reminding me, [I think] "Oh, I haven't taken the medication." So I will go ahead and take it, then answer them. If that were to happen every other two days or whatever, then maybe I would have missed taking my medication...It brought you back, even when you're on a really busy day. You're able to get that message and say, "Oh yeah" and take your medicine.* [Hispanic female, aged 65 years]

Other participants specifically mentioned the self-care promotion text messages and appreciated the tips and information included in that content. Finally, some emphasized improvements in their health, which they attributed to the daily text messages:

> *I guess, you know, texting is a part of life now...it's just kind of a daily thing, and it's information I can share with somebody else. "Oh, look what I got." Or if you're somewhere all you have to do is show your phone...I just liked the daily input, and it was good*

information. [Non-Hispanic black female, aged 47 years]

When I noticed my A1c had went down, I believe in that period it was just like, "Oh, it worked." So, I think this is what made me [say] "Nah, I need it because some progress is better than none." [Non-Hispanic black female, aged 31 years]

Discussion

Principal Findings

We examined user engagement in a 12-month text messaging intervention designed to support diabetes self-care in a sample overrepresenting adults with low SES and racial/ethnic minorities. We wanted to understand how intervention characteristics and patient characteristics affected engagement with the text messages, with particular attention to the addition of a human component and participants' opportunity to choose their text message frequency after 6 months of daily texts. Engagement gradually decreased throughout the intervention but remained relatively high over 12 months. We did not find any differences in engagement by participants' clinic site, age, gender, education, income, diabetes duration, insulin status, health literacy, or numeracy. Engagement was lower among participants who were black and who had worse medication adherence and HbA_{1c} at baseline; however, the magnitude of these differences was relatively small. When given the option at 6 months to receive fewer text messages, nearly half of the participants chose to continue receiving daily text messages for the remaining 6 months. We did not find differences in engagement based on participants' receipt of phone coaching or their selected text message frequency.

The literature reporting on engagement with mobile health (mHealth) interventions is growing, although few studies report on interventions lasting more than 6 months [17]. Our findings support the use of text messages to engage participants long term. Before starting the RCT, we conducted iterative usability testing with participants from the target population; we addressed participant concerns regarding content and functionality, which may have helped sustain engagement in the trial [26,28]. In addition, the personalization used in REACH (eg, tailored content and message timing) may have promoted engagement, which is consistent with other studies employing individualized content [12]. Black participants and those with worse baseline medication adherence and HbA_{1c} had lower engagement. Notably, although these differences were statistically significant, average engagement was still relatively high across all participants (eg, the average response rate among black participants was 80%, whereas for white participants, it was 87%). On the basis of the qualitative data, participants' engagement appeared to be influenced by the perceived helpfulness of the text messages with reminding them to take their medication and feelings of accountability engendered by the texts.

Studies exploring the role of human support in engagement with automated technologies have shown inconsistent findings [26,41]. Variation in results are likely due to variation in the type of human involvement (eg, remote vs in person, familiar vs new health care professional, and providing structured support vs following up as issues arise), the type of technology, the engagement measure, and users' needs and preferences [41]. In this study, there was no difference in text message response rates between users who received text messages only and those who received both text messages and monthly phone coaching. Mohr et al [42] proposed a model for how human support increases adherence to electronic health interventions, such that patients experience accountability to a person they view as trustworthy, benevolent, and having expertise. Although the text messages in this study were automated, the content contained encouragement, information, and bidirectional communication, which may have been perceived as human support, thereby increasing engagement, regardless of assignment to phone coaching. In a study that explored perceptions of a similar automated text messaging program for diabetes self-care, black patients with T2D shared how they felt high levels of social support and cared for by a person or friend despite knowing that the program was automated [43]. Furthermore, in this study, all participants experienced a human component from interacting with the study staff for other elements of the study (eg, completing surveys), which may have dampened our ability to detect additive effects of the monthly coaching.

The ideal dose or frequency of text messages to sustain engagement in an intervention is unclear and likely varies by individual user. We did not find any patient characteristics associated with frequency choice, which suggests that preference was idiosyncratic. We included the low-dose option to sustain engagement among participants who may prefer receiving text messages less often than daily. The variability in choice and the specific reasons for making the choice suggest it was valuable to participants. The choice may also have kept engagement high for some participants who otherwise would have stopped responding. Reconciling dose preference with the dose needed to impact patient outcomes will be integral for the design of future text messaging interventions [13].

Limitations

This study recruited patients with T2D from a specific region in middle Tennessee. Therefore, we acknowledge that the findings may not be generalizable to patients with T2D in other locations or to other patient populations. In addition, because patients who did not want to respond to text messages were likely those who chose not to participate in this study, this may have led to a self-selection bias among our sample. As mentioned earlier, interactions with study staff may have reduced our ability to detect the effects of human coaching on engagement. We did not find evidence that patient characteristics were associated with dose choice; however, it is possible that characteristics not assessed in this study may have impacted the decision. For example, in the case of employment status, a person who is retired may be more receptive to receiving a higher frequency of messages than someone who is working full time. Finally, we were unable to compare how engagement would have been impacted if we did not give a choice for frequency preference.

Conclusions

There is growing evidence for the efficacy of text message–delivered interventions to improve adherence and clinical outcomes [11], and this study addresses important gaps by examining long-term engagement and associations between engagement and participant and intervention characteristics. As the literature on engagement is growing, there have been calls for more standardized engagement reporting to compare results across studies and draw more concrete conclusions about the role of engagement [44,45]. Kelders and Kip [46] recently developed a self-report scale to capture multiple components of engagement with health technologies (ie, behavior, cognition, and affect). Similarly, we encourage reporting on usage that goes beyond the average engagement level, but rather examines how engagement changes over the course of an intervention, and factors that influence engagement. In any intervention, users will become fatigued over time, so testing engagement promotion strategies (including new content to sustain interest over time and/or gamification) will also provide insights on how to improve and sustain engagement [47,48]. Finally, robust measurement of engagement will also help us explore the effect of engagement on outcomes to enhance our understanding of how and for whom mHealth interventions improve health. As a next step to this study, we plan to evaluate how engagement with REACH impacts outcomes, including HbA_{1c} and medication adherence. Specifically, we will explore whether there is a minimum engagement threshold (ie, response rate) needed for patients to experience improvements in outcomes.

As patients with low SES, who are less likely to have smartphones and internet access, also tend to have worse health outcomes, the spread of apps and other internet-dependent technologies may unintentionally widen health disparities [49]. Compared with other forms of mHealth technology, text messages offer key advantages given their ubiquity and potential for scalability [12,50]. This study contributes to evidence for the case that these interventions may help create equity if implemented in clinical care.

Acknowledgments

This research was funded by the National Institutes of Health NIH/NIDDK R01DK100694 and NIH/NIDDK Center for Diabetes Translation Research Pilot and Feasibility Award P30DK092986. LM was also supported by a career development award from NIH/NIDDK (K01DK106306), and LN was supported by a career development award from NIH/NHBLI (K12HL137943). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the REACH team, the partnering clinics (ie, Faith Family Medical Center, The Clinic at Mercury Courts, Connectus Health, Shade Tree Clinic, United Neighborhood Health Services, and Vanderbilt Adult Primary Care), and the participants for their contributions to this research.

Authors' Contributions

LN led the study design and wrote the paper. AS conducted the analyses and edited the paper. RG planned and oversaw the analyses and edited the paper. LL created the tables and figures, helped clean the data, and edited the paper. KW helped plan the analyses and reviewed and edited the paper. LM oversaw the parent study, planned analyses, and reviewed and edited the paper.

Conflicts of Interest

None declared.

References

1. Munshi MN, Segal AR, Suhl E, Ryan C, Sternthal A, Giusti J, et al. Assessment of barriers to improve diabetes management in older adults: a randomized controlled study. Diabetes Care 2013 Mar;36(3):543-549 [FREE Full text] [doi: 10.2337/dc12-1303] [Medline: 23193208]
2. Nelson LA, Ackerman MT, Greevy RA, Wallston KA, Mayberry LS. Beyond race disparities: accounting for socioeconomic status in diabetes self-care. Am J Prev Med 2019 Jul;57(1):111-116. [doi: 10.1016/j.amepre.2019.02.013] [Medline: 31130463]
3. Mayberry LS, Bergner EM, Chakkalakal RJ, Elasy TA, Osborn CY. Self-care disparities among adults with type 2 diabetes in the USA. Curr Diab Rep 2016 Nov;16(11):113 [FREE Full text] [doi: 10.1007/s11892-016-0796-5] [Medline: 27671320]
4. Sabaté E. Adherence to Long-term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
5. Kazemian P, Shebl FM, McCann N, Walensky RP, Wexler DJ. Evaluation of the cascade of diabetes care in the United States, 2005-2016. JAMA Intern Med 2019 Aug 12:- epub ahead of print. [doi: 10.1001/jamainternmed.2019.2396] [Medline: 31403657]
6. Mobile Fact Sheet. Pew Research Center. 2019. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-01-27]
7. Smith A. The Best (and Worst) of Mobile Technology. Pew Research Center. 2012. URL: https://www.pewresearch.org/internet/2012/11/30/the-best-and-worst-of-mobile-connectivity/ [accessed 2019-01-27]
8. Anderson M. Mobile Technology and Home Broadband 2019. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2019-01-27]

9. Perrin A. Digital Gap Between Rural and Nonrural America Persists. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/ [accessed 2019-01-27]
10. Cell Phone Activities. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/chart/cell-phone-activities/ [accessed 2019-01-20]
11. Saffari M, Ghanizadeh G, Koenig HG. Health education via mobile text messaging for glycemic control in adults with type 2 diabetes: a systematic review and meta-analysis. Prim Care Diabetes 2014 Dec;8(4):275-285. [doi: 10.1016/j.pcd.2014.03.004] [Medline: 24793589]
12. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
13. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
14. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
15. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016 Jul;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
16. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
17. 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]
18. 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]
19. Hanauer DA, Wentzell K, Laffel N, Laffel LM. Computerized automated reminder diabetes system (CARDS): e-mail and SMS cell phone text messaging reminders to support diabetes management. Diabetes Technol Ther 2009 Feb;11(2):99-106 [FREE Full text] [doi: 10.1089/dia.2008.0022] [Medline: 19848576]
20. Nelson LA, Mulvaney SA, Gebretsadik T, Ho Y, Johnson KB, Osborn CY. Disparities in the use of a mHealth medication adherence promotion intervention for low-income adults with type 2 diabetes. J Am Med Inform Assoc 2016 Jan;23(1):12-18 [FREE Full text] [doi: 10.1093/jamia/ocv082] [Medline: 26186935]
21. Quinn CC, Butler EC, Swasey KK, Shardell MD, Terrin MD, Barr EA, et al. Mobile diabetes intervention study of patient engagement and impact on blood glucose: mixed methods analysis. JMIR Mhealth Uhealth 2018 Feb 2;6(2):e31 [FREE Full text] [doi: 10.2196/mhealth.9265] [Medline: 29396389]
22. Piette JD, Valverde H, Marinec N, Jantz R, Kamis K, de la Vega CL, et al. Establishing an independent mobile health program for chronic disease self-management support in Bolivia. Front Public Health 2014;2:95 [FREE Full text] [doi: 10.3389/fpubh.2014.00095] [Medline: 25165687]
23. Glasgow RE, Strycker LA, King DK, Toobert DJ. Understanding who benefits at each step in an internet-based diabetes self-management program: application of a recursive partitioning approach. Med Decis Making 2014 Feb;34(2):180-191. [doi: 10.1177/0272989X13498156] [Medline: 23913917]
24. Suffoletto B. Text message behavioral interventions: from here to where? Curr Opin Psychol 2016 Feb 1;9:16-21 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.012] [Medline: 26665157]
25. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in ehealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 1;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
26. 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]
27. Nelson LA, Wallston KA, Kripalani S, Greevy RA, Elasy TA, Bergner EM, et al. Mobile phone support for diabetes self-care among diverse adults: protocol for a three-arm randomized controlled trial. JMIR Res Protoc 2018 Apr 10;7(4):e92 [FREE Full text] [doi: 10.2196/resprot.9443] [Medline: 29636319]
28. Nelson LA, Mayberry LS, Wallston K, Kripalani S, Bergner EM, Osborn CY. Development and usability of REACH: a tailored theory-based text messaging intervention for disadvantaged adults with type 2 diabetes. JMIR Hum Factors 2016 Sep 8;3(2):e23 [FREE Full text] [doi: 10.2196/humanfactors.6029] [Medline: 27609738]
29. Mayberry LS, Berg CA, Harper KJ, Osborn CY. The design, usability, and feasibility of a family-focused diabetes self-care support mhealth intervention for diverse, low-income adults with type 2 diabetes. J Diabetes Res 2016;2016:7586385 [FREE Full text] [doi: 10.1155/2016/7586385] [Medline: 27891524]
30. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002 Sep;40(9):771-781. [doi: 10.1097/00005650-200209000-00007] [Medline: 12218768]

31. Harris PA. Research electronic data capture (REDCap) - planning, collecting and managing data for clinical and translational research. BMC Bioinformatics 2012 Jul 31;13(S12):-. [doi: 10.1186/1471-2105-13-s12-a15]
32. Mayberry LS, Osborn CY. Empirical validation of the information-motivation-behavioral skills model of diabetes medication adherence: a framework for intervention. Diabetes Care 2014;37(5):1246-1253 [FREE Full text] [doi: 10.2337/dc13-1828] [Medline: 24598245]
33. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol 2006 Jul;25(4):462-473. [doi: 10.1037/0278-6133.25.4.462] [Medline: 16846321]
34. Wallston KA, Cawthon C, McNaughton CD, Rothman RL, Osborn CY, Kripalani S. Psychometric properties of the brief health literacy screen in clinical practice. J Gen Intern Med 2014 Jan;29(1):119-126 [FREE Full text] [doi: 10.1007/s11606-013-2568-0] [Medline: 23918160]
35. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a short, 3-item version of the subjective numeracy scale. Med Decis Making 2015 Nov;35(8):932-936 [FREE Full text] [doi: 10.1177/0272989X15581800] [Medline: 25878195]
36. Osborn CY, Mayberry LS, Wagner JA, Welch GW. Stressors may compromise medication adherence among adults with diabetes and low socioeconomic status. West J Nurs Res 2014 Oct;36(9):1091-1110 [FREE Full text] [doi: 10.1177/0193945914524639] [Medline: 24569697]
37. Wallston KA, Rothman RL, Cherrington A. Psychometric properties of the perceived diabetes self-management scale (PDSMS). J Behav Med 2007 Oct;30(5):395-401. [doi: 10.1007/s10865-007-9110-y] [Medline: 17522972]
38. Mayberry LS, Gonzalez JS, Wallston KA, Kripalani S, Osborn CY. The ARMS-D out performs the SDSCA, but both are reliable, valid, and predict glycemic control. Diabetes Res Clin Pract 2013 Nov;102(2):96-104 [FREE Full text] [doi: 10.1016/j.diabres.2013.09.010] [Medline: 24209600]
39. Diggle P, Diggle P, Heagerty P, Liang K, Heagerty P, Zeger S. Analysis of Longitudinal Data. Oxford, UK: Oxford University Press; 2002.
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
42. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
43. Nundy S, Dick JJ, Solomon MC, Peek ME. Developing a behavioral model for mobile phone-based diabetes interventions. Patient Educ Couns 2013 Jan;90(1):125-132 [FREE Full text] [doi: 10.1016/j.pec.2012.09.008] [Medline: 23063349]
44. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in ehealth and mhealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
45. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A framework for analyzing and measuring usage and engagement data (AMUsED) in digital interventions: viewpoint. J Med Internet Res 2019 Feb 15;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]
46. Kelders S, Kip H. Development and Initial Validation of a Scale to Measure Engagement with eHealth Technologies. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA'19; May 4-9, 2019; Glasgow, Scotland UK URL: https://doi.org/10.1145/3290607.3312917; [doi: 10.1145/3290607.3312917]
47. 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 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
48. 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]
49. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
50. Mayberry LS, Jaser SS. Should there be an app for that? The case for text messaging in mHealth interventions. J Intern Med 2018 Feb;283(2):212-213 [FREE Full text] [doi: 10.1111/joim.12687] [Medline: 28881402]

Abbreviations

FQHC: federally qualified health center
GEE: generalized estimating equations
HbA1c: hemoglobin A1c
mHealth: mobile health
RA: research assistant

RCT: randomized controlled trial
REACH: Rapid Education/Encouragement And Communications for Health
REDCap: Research Electronic Data Capture
SES: socioeconomic status
T2D: type 2 diabetes
VUMC: Vanderbilt University Medical Center

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by R Cruz Martínez, L Hynes, S Baptista; comments to author 11.02.20; revised version received 06.04.20; accepted 03.06.20; published 21.07.20.

Please cite as:
Nelson LA, Spieker A, Greevy R, LeStourgeon LM, Wallston KA, Mayberry LS
User Engagement Among Diverse Adults in a 12-Month Text Message–Delivered Diabetes Support Intervention: Results from a Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e17534
URL: http://mhealth.jmir.org/2020/7/e17534/
doi:10.2196/17534
PMID:32706738

Original Paper

Patient and Provider Cocreation of Mobile Texting Apps to Support Behavioral Health: Usability Study

Armen C Arevian[1], MD, PhD; Jennifer O'Hora[1], PhD; James Rosser[1], LCSW; Joseph D Mango[1], MFA; David J Miklowitz[1], PhD; Kenneth B Wells[1], MD, MPH

Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Armen C Arevian, MD, PhD
Jane and Terry Semel Institute for Neuroscience and Human Behavior
University of California Los Angeles
10920 Wilshire Blvd, Suite 300
Los Angeles, CA,
United States
Phone: 1 310 794 3732
Fax: 1 310 794 3724
Email: AArevian@mednet.ucla.edu

Abstract

Background: Mobile technologies hold potential for improving the quality of care and engagement of patients. However, there are considerable challenges in ensuring that technologies are relevant, useful, and engaging. While end users such as patients and providers are increasingly involved in the design of health technologies, there are limited examples of their involvement in directly creating technologies for their personal use.

Objective: We aim to evaluate the feasibility and acceptability of patients and providers creating mobile texting apps to support treatment goals.

Methods: In an 11-month usability study, we enrolled 4 providers and 28 patients in an intensive outpatient program for obsessive-compulsive disorder. Patients and providers created their own mobile texting apps using a visual app development platform. A subsample of 10 patients and 4 providers completed a usability measure.

Results: Participants created a total of 360 unique mobile text messages (1787 total messages sent). There were 4 types of messages identified, including personalized reminders, clinical exposures, interactive prompts, and encouraging/informational messages. A total of 9 out of 10 (90%) patients agreed that the messages were relevant to their recovery, and 8 out of 10 (80%) agreed that the messages were effective at helping complete treatment plans.

Conclusions: Enabling patients and providers to cocreate apps for their own use by using a visual application platform is feasible and holds potential for increasing the relevance, sustainability, and effectiveness of digital health technologies.

(JMIR Mhealth Uhealth 2020;8(7):e12655) doi:10.2196/12655

KEYWORDS

mobile health; community-based participatory research; app development; technology platforms; personalized medicine; behavioral health; mobile phone

Introduction

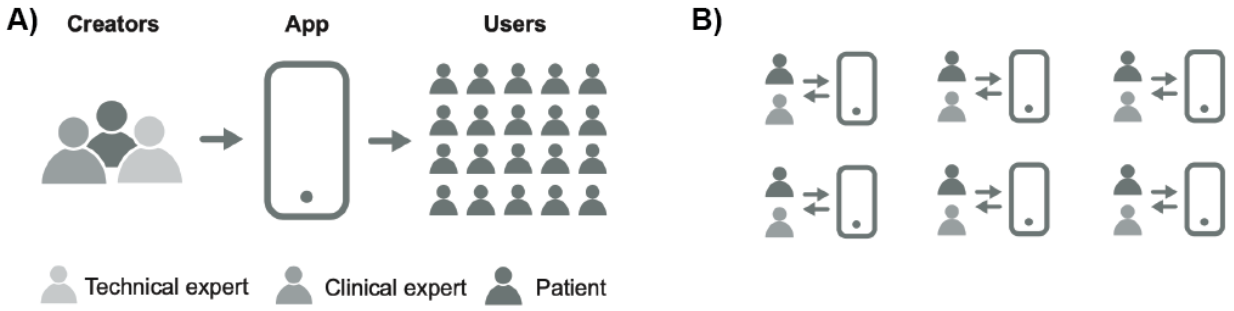
Digital technologies such as mobile apps are increasingly used to improve the quality of care and engagement of patients [1], especially for chronic disease management, in which engagement of individuals over time is challenging [2]. There are considerable barriers in ensuring that technologies are relevant, useful, and lead to sustained use at the patient, provider, and institutional levels [3,4]. Involving broader stakeholders, including patients and providers, is increasingly seen as critical in addressing these challenges. However, their involvement in the development process is typically limited to the design of apps, often providing feedback on the content, functionality, and visual appearance. Translating these designs into functioning apps is often undertaken separately by computer programmers, resulting in the removal of the end users from this aspect of development, as well as additional cost, complexity, and time. There have not been previous reports of

a systematic approach to engaging patients and providers as the direct creators of mobile health apps that they use to support their health needs [5].

The technology adoption model identifies several intrinsic and extrinsic factors related to technology use, including perceived ease of use and relevance [6]. Both of these factors may be negatively impacted when apps are not created for specific individuals or local groups but rather made to address broader diseases (eg, diabetes management), conditions (eg, weight loss, stress reduction), or approaches (eg, cognitive behavioral therapy). At the same time, direct engagement by patients has been recognized as a key determinant of health behavior change, better health outcomes, and satisfaction [7].

Participatory approaches, including participatory design and community-partnered participatory research [8,9], were created to involve stakeholders such as patients, providers, and community leaders, and are increasingly used in a variety of settings, including those involving digital health [5]. However, the translation of designs to app creation is often undertaken separately by technical individuals such as computer programmers. In addition, existing participatory methods often aim to create an app for a population whose users were not involved in the process. This is different than a systematic approach to involving individuals in the creation of an app for their personal use, as seen in Figure 1.

Figure 1. Comparison of two approaches to the participatory technology development process for mobile apps. A) Stakeholders cocreate an app for a group of end users. B) Patients and providers cocreate apps to support themselves.



A major assumption of the present study is that when individuals are involved in creating the apps that they will personally use, their engagement and the perceived usefulness of the apps will be enhanced. We created the Chorus application platform (Chorus Innovations Inc) [10] to facilitate this process of participatory technology development, in which stakeholders can be directly involved in both designing and creating mobile apps. An application platform provides modules and functions that can be reused to create specific applications to reduce development time, lower cost, and improve scalability [11]. Furthermore, Chorus is a no-code application platform that uses a visual interface to configure apps without the need for computer programming, with the objective of supporting a broad population of stakeholders as app creators [12]. Chorus has been used to create mobile web, text messaging, and interactive voice apps for more than 50 research and clinical projects, including ongoing studies of automated text messaging interventions for asthma [13] and healthy lifestyle [14].

Our group previously used the Chorus platform to facilitate a participatory development process to cocreate a mobile texting app to support resiliency (B-RESILIENT) [15]. We conducted a series of partnered workgroups with stakeholders, including patients, community leaders, and academic researchers. In these workgroups, key needs of the community, selected and adapted content to include in the app, and privacy and other concerns related to trust of the technology were discussed. The workgroups then created an app that addressed these needs.

It remains an open question to what extent it is feasible for individual patients and providers to directly create personal mobile apps for their own use in real-world clinical settings. The goal of this pilot study was to evaluate the usability of the Chorus platform and the feasibility of a participatory development process in the creation of mobile texting apps by patients and providers, as well as to describe the kinds of mobile texting apps created by these participants. The study was conducted with patients in an intensive outpatient treatment program and their providers.

Methods

Setting and Participants

This study was conducted in the Obsessive-Compulsive Disorder (OCD) Intensive Outpatient Program (IOP) of University of California Los Angeles (UCLA) Health. Patients with OCD attended the IOP program 3 to 5 weekdays per week. The IOP uses cognitive behavioral therapy (CBT) techniques, including exposure to situations and stimuli that trigger anxiety symptoms, known as Exposure and Response Prevention (ERP) [16]. The core of ERP is intentional and planned exposure to stimuli that trigger anxiety responses, with progressively increasing levels of stimulus threat as treatment progresses.

Key clinical challenges were identified in a workgroup meeting with IOP staff and research personnel at the beginning of the study. The IOP staff identified low patient satisfaction with support after clinic hours and the difficulty of completing exposure tasks outside of the clinic setting, a common barrier in OCD treatment [17].

The Chorus platform was implemented by the IOP as part of clinical care rather than as part of a research study. Patients who enrolled in the IOP were consented by clinic staff for use of text

messaging as part of their care. Those that agreed were then consented by research staff to participate in this study. The scope of the research study was to evaluate the clinic's implementation and use of the Chorus platform by analyzing text messages, administering usability measures, and conducting workgroups. All research methods were approved by the UCLA Institutional Review Board.

Inclusion criteria for study participation included being an active patient of the IOP clinic, having a mobile phone (smartphone not required), and being willing to receive text messages. Patients used their personal mobile phones and interacted with their app through text messaging. There was no app to download. There were no exclusion criteria. We chose these broad criteria to reflect the typical population from this clinic. Technical skills such as computer programming were not a requirement for participation.

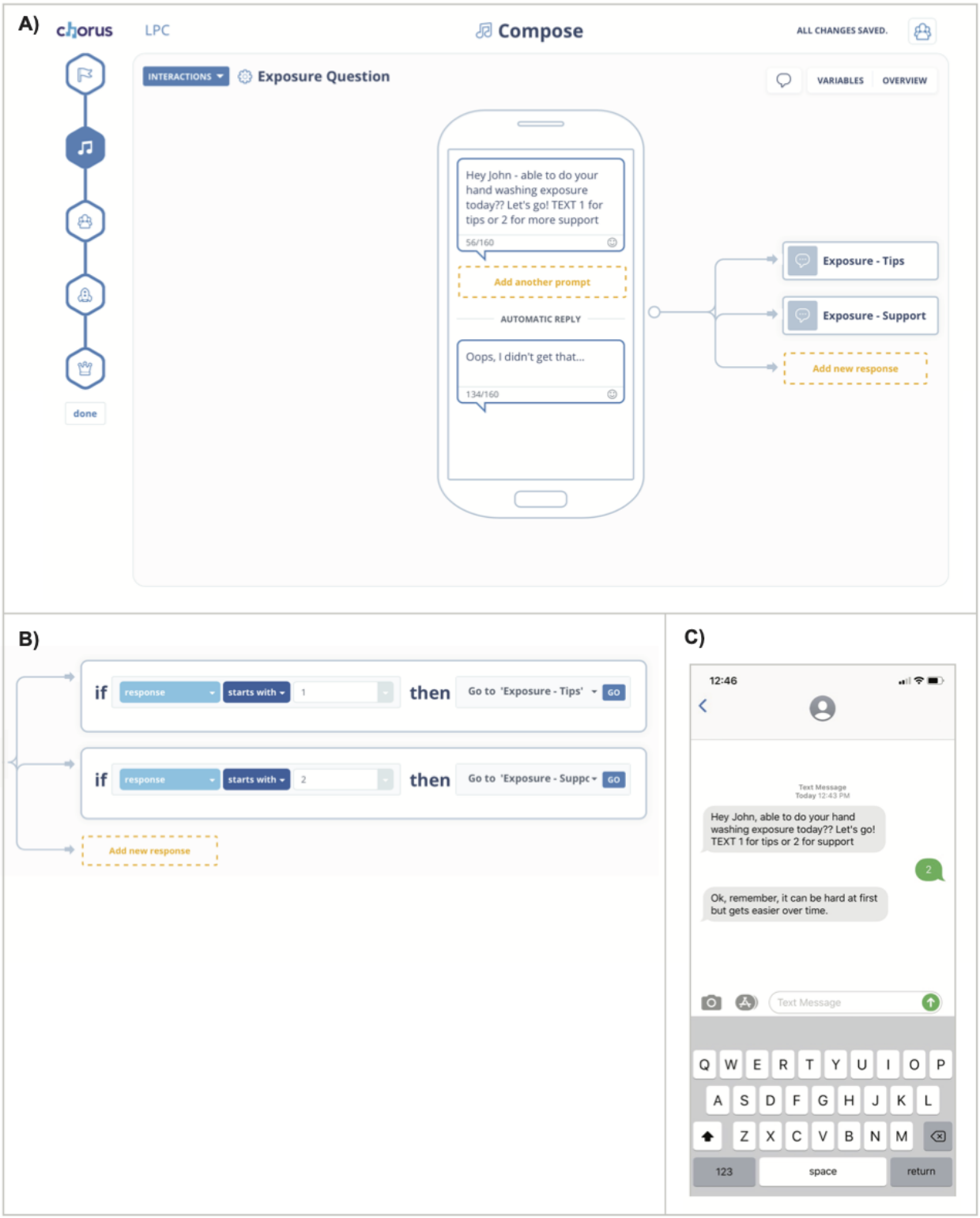
During the pilot period, there were 4 PhD psychology providers who participated in the study. The number of patients active in the clinic at any given time varied between 6 and 8 patients. Patients participated in the study for as long as they were enrolled in the IOP (average stay of 6-8 weeks). No participants disenrolled early from the study.

Participatory Technology Development

The clinic implemented a process of participatory technology development (Figure 1) to facilitate patients and providers to cocreate personal mobile text messaging apps using the Chorus platform created by Arevian et al [15]. Providers were given log-in access to the platform and used the visual web interface with their patients to cocreate the texting apps, as seen in Figure 2. This included specifying the content of messages, logic to handle responses from the patient if needed, and scheduled times to send the messages to the patient from the platform. The content and functionality of the apps were entirely driven by the patient and provider. The messaging did not facilitate direct communication between patients and providers. Instead, the app was configured to automatically send the messages to the patient at a later time. If the patient replied to the message, they would receive an automated preconfigured (by the patient and provider) response from the app.

Initial training for clinic psychotherapists to use Chorus was a single 1-hour session. One therapist requested an additional training session. Therapists were not provided specific guidance on the types of apps to create other than to address the clinical needs of their patients. Therapists reviewed the app's functionality with patients as needed as part of the cocreation process. The Chorus platform was provided through the Innovation Lab at the Jane and Terry Semel Institute for Neuroscience and Human Behavior at UCLA.

Figure 2. Visually creating apps with Chorus. A) Creation of text messaging content visually with a simulated phone. B) Visual interface to configure the logic that guides subsequent messages to send based on user's response. C) Screenshot of an example mobile texting interaction as seen by patients.



Measures and Analysis

Demographic details (age, sex, race) and clinical outcomes were extracted from patients' medical records. We report the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) score to describe the clinical severity of this population. Y-BOCS is a measure of OCD symptom severity and was assessed by study staff on admission as part of routine care [18].

Feasibility and acceptability were evaluated through a written questionnaire that study staff administered to patients. The measure included items of perceived ease of use, perceived usefulness, self-efficacy, and patient-centeredness (see Multimedia Appendix 1) [6,19]. Participants indicated their agreement with statements in the survey using a Likert scale ranging from 1 to 7, where 1 indicated they strongly disagreed, 4 indicated they were neutral, and 7 indicated they strongly

agreed with the statement. The survey also included 2 open-ended prompts to list the most positive and negative aspects of their apps, the platform, or the process of participatory development. The survey was administered to 14 participants (10 patients and 4 therapists) during a 2-month evaluation period of the pilot.

We conducted a total of 7 workgroups with patients and 7 with providers to discuss experiences related to the app development and participatory process. Each workgroup typically had 2 to 7 participants and lasted 30 minutes to 1 hour. Workgroup sessions were audiorecorded and reviewed by study staff. Qualitative analyses were conducted by reviewing transcripts (for workgroups) and message logs (for app content) to group phrases or messages together and identify themes [20]. Themes were cross-checked by 2 study staff and reviewed with clinic staff for validity. Representative quotes and messages are tabulated in the results below. App use metrics were extracted from the Chorus activity logs for all participants.

Retention rates for the use of the mobile texting apps were calculated using the date of the last message sent or received by the patient from the Chorus logs and comparing this to the date of termination from the clinical program.

Results

We conducted a rolling enrollment over an 11-month period between July 2015 and May 2016. A total of 28 patients and 4 providers were enrolled in the study during this period. The providers were all PhD-level psychotherapists specializing in the treatment of anxiety disorders. Of the 28 patients, 11 were female (39%). Mean age was 33 years (SD 14.2; range 18-69). Patients were mostly white (26/28, 93%) and had severe OCD symptom burden on admission (Y-BOCS mean score of 38 out of 40, indicating extreme level of symptoms).

A total of 1787 messages were sent to patients during the evaluation period (360 unique messages) and 80 responses were received from patients. Patients received an average of 51 messages (SD 55). All patients responded at least once to a text message. The number of messages sent did not vary by day of the week. Most text messages were sent after clinic hours (between 2 PM and 10 PM).

We identified 4 basic types of messages created, including personalized reminders (ie, prompting of a clinical goal to be achieved at home), messages serving as clinical exposures (ie, intended to evoke feared stimuli), interactive prompts (messages that requested a response from the patient), and encouraging or informational messages (Textbox 1). A common theme of the messages was to use personalization and humor to create messages related to treatment engagement.

Textbox 1. Content of mobile texting apps.

Interactive

- "Were you able to sit with the anxiety and contamination without washing today? Press 1 for yes and 2 for no"
- "Hey there {patient name}, did you meet your goal today? TEXT 1 for yes or 2 for no"
- "Don't forget to take the Namenda tonight. Press 1 if you took the medication"
- "Hey, did you go to the store today?! TEXT: 1 for yes, 2 for no"

Personalized reminders

- "Hi {patient name}, No more showering. Life is too short. Lean into the grubbiness. You can do this"
- "It's 10pm, time to give up your devices. They turn into pumpkins!"
- "Good morning {patient name}, here is your sleep diary reminder! {link to diary website}"
- "Ready, set, go - shower in under 12. Otherwise get ready to shower at [University of California Los Angeles] in 6 minutes"
- "Leave those faucets alone {patient name}. You don't need them"
- "C'mon tin ribs, get to the gym and feel the burn!"

Message as the exposure (exposure type)

- "Remember, Typhoid {patient name}, No handwashing. You are a walking petri dish!" (contamination)
- "Showering once a day means you have some pretty gross stuff on you" (contamination)
- "gr8t job this week - don't forget two exercise this weekend :)" (spelling)
- "{expletive}" (unwanted thoughts)

Encouraging/informational

- "Gold star Mr. Jelly Legs. Be proud of yourself"
- "Remember to try to go as long as possible without washing, even if you feel contaminated. You are stronger than your [obsessive-compulsive disorder]!"
- "Very good padawan"
- "Exercise combats health conditions and diseases. It can also improve mood"
- "That's awesome!! Great Job!"

Results from the usability survey for patients and providers are presented in Table 1. Individuals responded to how much they agreed with the statements listed in Table 1 using a Likert scale, where 1=strongly disagree, 4=neutral, and 7=strongly agree. Representative open-ended responses for positive and negative features of the system are included in Textbox 2. Regarding the use of the Chorus platform, 11 out of 14 (79%) participants (10 patients and 4 providers) agreed that Chorus was simple to use, 12 out of 14 (86%) felt comfortable using Chorus, 13 out of 14 (93%) felt comfortable using the text messaging app, and 11 out of 14 (79%) felt they (or their patients) were more engaged as a result. All providers (4/4, 100%) and most patients (8/10, 80%) agreed or responded neutrally to the statement "I can effectively develop messages with my provider (patient) using this messaging application." Of the 10 patients surveyed, 9 (90%) agreed that the messages were relevant to them and their recovery, and 8 (80%) agreed that the messages were effective in helping them complete treatment exercises at home.

Overall, engagement with the cocreated apps was sustained over the time the patient was engaged in the clinic (Figure 3). The average duration of mobile app use was 25.1 days (SD 15.1), which was 6.7 days (SD 8.8) shorter than the average length of stay in the clinical program (mean 31.8 days, SD 13.3).

Key themes that emerged from provider workgroups were related to the effects of the app development process on engagement and between-session homework completion (eg, "Follow through on homework is better"; "One of the patients I had, he really loved it. He said he felt more accountable. And it involved his wife because he showed his wife so she was kind of onboard with the whole treatment. And he said it felt like [my provider] was in the room..."). Themes emerging from patient workgroups included discussion of personalization of treatment (eg, "I tell [my provider] what kind of topic I want. And then we talk about what times are good. But then she makes up the specifics"), motivation (eg, "Helps the motivation continue"; "It helps remind you on what you need to be doing"), and connection after hours (eg, "You're [in the IOP clinic] for an allotted amount of time and then you have to be out into the real world… This supplements [the time in the IOP clinic]"). Providers stated that during the final week that patients are in the IOP, providers are focused on discharge planning and therefore less focused on app use by the patients.

Table 1. Usability and satisfaction with the Chorus platform and mobile texting apps.

Measure	Score, mean (SD)	
	Patients (n=10)	Providers (n=4)
Perceived ease of use		
It is simple to use this messaging application.	6.0 (1.5)	5.8 (0.5)
Perceived usefulness		
This messaging application has all the functions and capabilities I expect it to have.	5.7 (1.6)	5.3 (1.0)
I feel (I am/my patients are) more engaged in my treatment as a result of using this messaging application.	5.8 (1.2)	6.0 (1.4)
Self-efficacy		
I can effectively develop messages with my provider (patient) using this messaging application.	6.0 (1.3)	6.8 (0.5)
I (my patient) was directly involved in creating the text messages I would receive.	4.8 (2.2)	5.5 (3.0)
Receiving these text messages is effective in helping me (my patient) complete the treatment plans at home.	5.7 (1.1)	7.0 (0)
Patient-centeredness		
Information provided by the text messages is relevant to me (my patient) and my (patient's) recovery.	6.2 (1.0)	7.0 (0)
Overall, I am satisfied with this messaging application.	5.8 (1.4)	5.5 (0.6)

Textbox 2. Illustrative quotes of positive and negative aspects of Chorus.

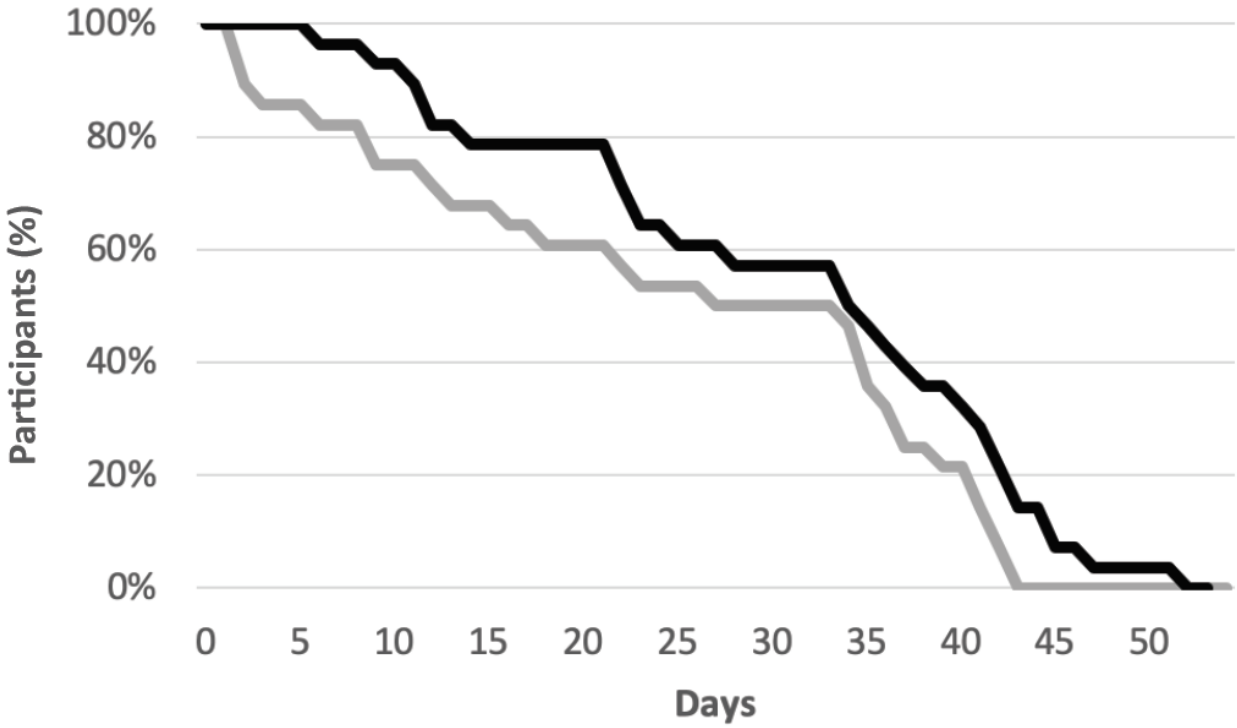
Positive aspects

- "It's nice to have it remind me of all the things I need to accomplish"
- "Helps the motivation continue"
- "Get out of comfort zone"
- "I feel more encouraged in treatment"
- "It is simple"
- "Very straightforward"
- "Simple to create"
- "Fun texts"
- "Patients love it"
- "Patients like interaction"

Negative aspects

- "Harder to use if you're a bad texter"
- "Some bugs"
- "Disappointment it's not a text from a friend"
- "Folders needed [to organize messages]"

Figure 3. Patient engagement. Percentage of participants enrolled in the Intensive Outpatient Program clinic (black line) compared with the percentage of participants actively using their mobile app (grey line) by number of days in the program.



Discussion

This is the first study of a systematic approach for patients and their providers to directly cocreate mobile texting apps for their own use as part of clinical care. We implemented a web-based application development platform (Chorus) and participatory development process in an outpatient clinical setting for patients with severe OCD. Providers required minimal training and were able to cocreate and use apps the same day that they were trained on how to use the platform. Providers and patients determined the content, frequency, and timing of notifications. Despite the high symptom burden experienced by patients, both patients and providers reported that the Chorus platform was easy to use and that it helped patients engage in treatment goals. A total of 9 out of 10 patients agreed that the messages were relevant to them. We observed sustained use of the cocreated apps throughout patients' time in the IOP. In addition, while the patients and providers were not given specific direction as to the types of messages to create, the 4 categories of messages identified were consistent with key principles of CBT and ERP implemented by the clinic.

There is increasing interest in including patients and providers at various levels of technology development. The Nightscout project is an example of an app first created by a patient's parent for diabetes management [21]. The parent, who was a computer programmer, identified a need and created a technology to address it, resulting in an open source app for monitoring glucose levels. While this project involved an end user in its development, our approach differs in that computer programming is not required so that apps can be created by individuals without technical backgrounds, including for their own use. Torous and Roux [22] describe an individual patient with schizophrenia working with their provider to create a custom symptom-tracking system. Though the patient and provider were not creating a mobile app, they used spreadsheets to graph symptom counts recorded with a manual tally counter by the patient.

Part of the challenge of involving patients in creating apps for their own use is the high technical and financial barriers to creating apps. The use of application platforms may offer several advantages. Visual development interfaces mean that individuals without knowledge of computer programming can create apps. In addition, compliance, auditing, and security requirements, such as those in the Health Insurance Portability and Accountability Act, can be handled centrally by the platform without the need to repeat these functions and organizational review processes for each individual app. There are several examples of no-code platforms used in health care and research outside of mobile app development. For example, REDCap (Vanderbilt University) [23] and Qualtrics (SAP SE) [24] are commonly used visual development platforms for creating online surveys. Squarespace (Squarespace Inc) [25] and Wix (Wix.com Inc) [26] are platforms to visually create and host websites. This is in addition to platforms for use by developers and institutions with technical expertise, including platforms to facilitate device sensing [27], data collection [28], and data storage and analysis [29].

The direct development of technologies by patients, providers, and other stakeholders has several implications. First, this participatory process aims to increase the level of patient involvement and the relevance of health technologies, which is consistent with recommendations from behavior change theories and guidelines for patient-centered care [7,8,30]. Second, this

process also has implications for equity and power sharing, a key principle from participatory approaches such as community-partnered participatory research. If patients are primarily involved at the design stage only (as is the current state of practice), the ability to create and maintain technologies still rests with computer programmers and server infrastructure staff. When patients and providers directly create and maintain their own apps, they may be better able control what is created, change the apps over time as their needs and priorities change, and directly benefit from their own apps.

Our approach is consistent with recent recommendations regarding how health information technologies may be evaluated more effectively within health systems. Learning health systems [31], agile science [32], and responsive research [33] each aim to support iterative intervention development, learning through implementation in real-world settings, and flexible evaluation options such as N-of-1 and pragmatic trials. End users are an untapped talent pool. According to the US census, only 0.1% of the US population are computer programmers [34]. By exploring approaches and technology platforms that do not require programming skills, we are better able to tap into the expertise and capacity of individuals who have lived experiences. An important future consideration is how clinical training programs may be modified to prepare clinicians and other staff for increased patient involvement in health technology development [35].

This study has several limitations. It was conducted at a single site with a limited number of patients, most of whom were white, young (mean 33 years), and had severe OCD symptoms. The clinical program was also suited to texting interventions, given its focus on CBT and exposure-driven treatment. Implementation in larger samples, at other sites, and in other clinical conditions would allow further evaluation of the generalizability of this approach, including how it may be adapted to other technology platforms [11]. This study did not evaluate the impact of the app development and cocreation process on clinical outcomes, a topic to be explored in randomized controlled trials. While patients engaged with the app for the majority of their length of stay in the IOP, we did observe a discontinuation of the app an average 6.7 days prior to discharge from the IOP. This may be due to the reduced focus on exposures and interventions in the final week of the IOP, with a transition to discharge planning reported by providers.

Technology approaches that use flexible, user-driven platforms to engage a broader set of individuals in development may hold potential for increasing the relevance and sustainability of technology interventions, which may in turn lead to improved patient engagement and outcomes.

Acknowledgments

The authors would like to thank the patients and staff of the Resnick Neuropsychiatric Hospital Intensive Outpatient Programs for their engagement and support of this project. ACA, KBW, and JO were supported by California State Behavioral Health Center of Excellence (SB-852). ACA and KBW were supported by NIMH P30 MH082760-05S1. ACA was supported by NIMH R01 MH093676-02S1.

Conflicts of Interest

ACA is the founder of Chorus Innovations Inc, Arevian Technologies, and Open Science Initiative. ACA developed the Chorus platform, which is licensed to Chorus Innovations Inc from the University of California Los Angeles.

Multimedia Appendix 1
Patient Usability Measure.
[PDF File (Adobe PDF File), 47 KB - mhealth_v8i7e12655_app1.pdf]

References

1. 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]
2. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
3. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three Problems With Current Digital Mental Health Research . . . and Three Things We Can Do About Them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
4. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
5. Unertl KM, Schaefbauer CL, Campbell TR, Senteio C, Siek KA, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016 Jan;23(1):60-73 [FREE Full text] [doi: 10.1093/jamia/ocv094] [Medline: 26228766]
6. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. International Journal of Man-Machine Studies 1993 Mar;38(3):475-487. [doi: 10.1006/imms.1993.1022]

7. Epstein R, Street R. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
8. Jones L, Wells K. Strategies for academic and clinician engagement in community-participatory partnered research. JAMA 2007 Jan 24;297(4):407-410. [doi: 10.1001/jama.297.4.407] [Medline: 17244838]
9. Wells KB, Jones L, Chung B, Dixon EL, Tang L, Gilmore J, et al. Community-partnered cluster-randomized comparative effectiveness trial of community engagement and planning or resources for services to address depression disparities. J Gen Intern Med 2013 Oct;28(10):1268-1278 [FREE Full text] [doi: 10.1007/s11606-013-2484-3] [Medline: 23649787]
10. Arevian AC. Chorus Participatory Mobile Framework. URL: https://chorus.semel.ucla.edu/ [accessed 2020-06-01]
11. Reviews for Enterprise High-Productivity Application Platform as a Service. Gartner. 2019. URL: https://www.gartner.com/reviews/market/enterprise-high-productivity-application-paas [accessed 2020-06-01]
12. Bloomberg J. The Low-Code/No-Code Movement: More Disruptive Than You Realize. Forbes. 2017 Jul 20. URL: https://www.forbes.com/sites/jasonbloomberg/2017/07/20/the-low-codeno-code-movement-more-disruptive-than-you-realize/#6c0e2dd9722a, [accessed 2020-06-01]
13. Teach S. Breathe With Ease: A Unique Approach to Managing Stress (BEAMS). clinicaltrials.gov. 2019. URL: https://clinicaltrials.gov/ct2/show/NCT02374138, [accessed 2020-03-01]
14. Brown A. Financial Incentives and Text Messaging to Improve African American Womens' Glycemic Control (FRIENDS-Text). clinicaltrials.gov. 2015. URL: https://clinicaltrials.gov/ct2/show/NCT02384265 [accessed 2020-03-01]
15. Arevian AC, O'Hora J, Jones F, Mango J, Jones L, Williams PG, et al. Participatory Technology Development to Enhance Community Resilience. Ethn Dis 2018;28(Suppl 2):493-502 [FREE Full text] [doi: 10.18865/ed.28.S2.493] [Medline: 30202203]
16. Koran L, Hanna G, Hollander E, Nestadt G, Simpson H, American Psychiatric Association. Practice guideline for the treatment of patients with obsessive-compulsive disorder. Am J Psychiatry 2007 Jul;164(7 Suppl):5-53. [Medline: 17849776]
17. Abramowitz JS, Franklin ME, Zoellner LA, DiBernardo CL. Treatment compliance and outcome in obsessive-compulsive disorder. Behav Modif 2002 Sep;26(4):447-463. [doi: 10.1177/0145445502026004001] [Medline: 12205821]
18. Koen N, Stein D. Obsessive-Compulsive Disorder. In: Zigmond MJ, Rowland LP, Coyle JT, editors. Neurobiology of Brain Disorders: Biological basis of neurological and psychiatric disorders. Amsterdam, Netherlands: Elsevier; 2015:621-638.
19. Bevan N, Kirakowski J, Maissel J. What is Usability? In: Human Aspects in Computing: Design and use of interactive systems and information management. Amsterdam, Netherlands: Elsevier Science Ltd; 2001.
20. McQuarrie EF, Marshall C, Rossman GB. Designing Qualitative Research. Journal of Marketing Research 1990 Aug;27(3):370. [doi: 10.2307/3172595]
21. Lee JM, Hirschfeld E, Wedding J. A Patient-Designed Do-It-Yourself Mobile Technology System for Diabetes: Promise and Challenges for a New Era in Medicine. JAMA 2016 Apr 12;315(14):1447-1448. [doi: 10.1001/jama.2016.1903] [Medline: 27115262]
22. Torous J, Roux S. Patient-Driven Innovation for Mobile Mental Health Technology: Case Report of Symptom Tracking in Schizophrenia. JMIR Ment Health 2017 Jul 06;4(3):e27 [FREE Full text] [doi: 10.2196/mental.7911] [Medline: 28684386]
23. REDCap.: REDCap; 2004. URL: http://projectredcap.org [accessed 2020-02-29]
24. Qualtrics.: Qualtrics URL: http://qualtrics.com [accessed 2020-02-29]
25. Squarespace.: Squarespace URL: http://squarespace.com [accessed 2020-03-01]
26. Wix.: Wix URL: http://wix.com [accessed 2020-03-01]
27. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
28. Tangmunarunkit H, Hsieh CK, Longstaff B, Nolen S, Jenkins J, Ketcham C, et al. A general extensible end-to-end participatory sensing platform. ACM Trans Intell Syst Technol 2015;6(3):1-21 [FREE Full text] [doi: 10.1145/2717318]
29. Chen C, Haddad D, Selsky J, Hoffman JE, Kravitz RL, Estrin DE, et al. Making sense of mobile health data: an open architecture to improve individual- and population-level health. J Med Internet Res 2012 Aug 09;14(4):e112 [FREE Full text] [doi: 10.2196/jmir.2152] [Medline: 22875563]
30. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
31. Patrick K, Hekler E, Estrin D, Mohr DC, Riper H, Crane D, et al. The Pace of Technologic Change: Implications for Digital Health Behavior Intervention Research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
32. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
33. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]

34. Labor Force Statistics. US Census Bureau. URL: https://www.census.gov/topics/employment/labor-force.html [accessed 2020-07-04]
35. Torous J, Chan S, Luo J, Boland R, Hilty D. Clinical Informatics in Psychiatric Training: Preparing Today's Trainees for the Already Present Future. Acad Psychiatry 2018 Oct;42(5):694-697 [FREE Full text] [doi: 10.1007/s40596-017-0811-4] [Medline: 29047074]

Abbreviations

CBT: cognitive behavioral therapy
ERP: Exposure and Response Prevention
IOP: intensive outpatient program
OCD: obsessive-compulsive disorder
UCLA: University of California Los Angeles
Y-BOCS: Yale-Brown Obsessive-Compulsive Scale

Edited by G Eysenbach; submitted 30.10.18; peer-reviewed by D Chao, T Irizarry, A Pratap, K Usop; comments to author 31.03.19; revised version received 05.08.19; accepted 29.03.20; published 29.07.20.

Please cite as:
Arevian AC, O'Hora J, Rosser J, Mango JD, Miklowitz DJ, Wells KB
Patient and Provider Cocreation of Mobile Texting Apps to Support Behavioral Health: Usability Study
JMIR Mhealth Uhealth 2020;8(7):e12655
URL: http://mhealth.jmir.org/2020/7/e12655/
doi:10.2196/12655
PMID:32723714

Original Paper

A Community-Engaged Approach to Creating a Mobile HIV Prevention App for Black Women: Focus Group Study to Determine Preferences via Prototype Demos

Rasheeta Chandler[1], RN, PhD, FNP-BC, FAANP, FAAN; Natalie Hernandez[2], MPH, PhD; Dominique Guillaume[1], MSN, APRN, AGPCNP-BC, ACRN; Shanaika Grandoit[3], MPH, CHES; Desiré Branch-Ellis[3], BSc; Marguerita Lightfoot[4], PhD

[1]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States

[2]Department of Community Health and Preventive Medicine, Morehouse School of Medicine, Atlanta, GA, United States

[3]Rollins School of Public Health, Emory University, Atlanta, GA, United States

[4]Center for AIDS Prevention Studies, Division of Prevention Science, Department of Medicine, University of California San Franciso, San Francisco, CA, United States

Corresponding Author:
Rasheeta Chandler, RN, PhD, FNP-BC, FAANP, FAAN
Nell Hodgson Woodruff School of Nursing
Emory University
1520 Clifton Rd
Atlanta, GA, 30322
United States
Phone: 1 404 727 8164
Fax: 1 678 668 8050
Email: r.d.chandler@emory.edu

Abstract

Background: Black women are an important but relatively overlooked at-risk group in HIV prevention efforts. Although there is an aggregate decline of HIV diagnoses among women in the United States, there are persistent disparate rates of new HIV infections among Black women compared to any other cisgender female subgroup. Strategies to end the HIV epidemic—as outlined in the Ending the HIV Epidemic initiative—for all communities must consider HIV prevention messaging and message delivery mediums that are created with community input. Although mobile health (mHealth) is a popular platform for delivering HIV interventions, there are currently no mobile apps that consider cisgender Black women with the goal of promoting a comprehensive women's reproductive health and HIV prevention lifestyle. Previous research recommends inclusion of the target population from project inception and iteratively throughout development, to promote use of the intervention.

Objective: The purpose of this study is to understand cisgender Black women's preferences for functionality, format, and design of a mobile HIV prevention app and to examine their willingness to use an app for HIV prevention.

Methods: We conducted a series of four focus groups with 23 Black cisgender women. Focus groups included discussion and demonstration elements to address cisgender women's general preference for apps, HIV prevention content that would be useful in an app, and preferred app features that would promote use of an HIV-centric app. During focus group discussions, participants were shown narrated, custom wireframes of HIV prevention app prototypes to demonstrate potential app function.

Results: Findings indicated the presence of eight subthemes within the coding structure of three overall themes: (1) health content within the mobile app, (2) mobile app functionality, format, and design, and (3) other suggested features. Specifically, participants detailed preferred educational content, content distribution, app aesthetics, privacy considerations, and marketing of the app.

Conclusions: Findings suggest that Black cisgender women preferred an app that integrated HIV prevention and optimal sexual health promotion. Participants provided a range of preferences for content integration and facilitators of app engagement with an HIV prevention app. Preferences centered on gender and cultural congruency of information and content, evidenced by visuals, language, and resources. Black cisgender women are viable consumers for a mobile app–based HIV prevention intervention.

(JMIR Mhealth Uhealth 2020;8(7):e18437) doi:10.2196/18437

KEYWORDS

mHealth app; mobile technology; Black women; HIV prevention; reproductive health; women's health

Introduction

Nearly 40,000 Americans were newly diagnosed with HIV in 2017 [1]. Although there has been an 8% decline in new HIV diagnoses, individuals of color—particularly Black women—continue to be disproportionally affected by the current HIV epidemic, trailing only men who have sex with men (MSM) [2,3]. Black individuals are the most affected racial or ethnic group, with Black men having a lifetime HIV risk of 1 in 20 and Black women having a risk of 1 in 48, compared to 1 in 132 and 1 in 880 for White men and White women, respectively [4]. In 2017, Black women accounted for almost 60% of newly diagnosed HIV infections among women and comprised 26% of HIV incidences among the Black population in the United States [3,5].

Black heterosexual women have a higher rate of HIV and sexually transmitted infections (STIs) compared to any other female group in the United States [1]. There are several social determinants of health that contribute to this increased HIV risk among this population [6]. These include poverty, lack of access to quality health care, unstable housing, sex-based power differentials in couple relationships that limit women's ability to negotiate HIV protective actions with their regular male sex partners, limited HIV prevention education, the increased prevalence of other STIs, and alcohol and drug use [6,7]. In addition to these factors, it is also important to consider that many Black women are at an increased risk for HIV acquisition due the mere concentration of people living with HIV within their geographical area or communities, coupled with their sexual networks [8]. Black women themselves may have few sexual partners; however, they are more likely to have sexual contact with individuals who have multiple sexual partners [8,9]. Thus, Black women may remain at high risk for contracting HIV even when they do not exhibit risky behaviors [8].These factors are important to recognize when addressing HIV prevention interventions for Black women in high-risk areas, such as the southern United States.

Evidence-based, face-to-face behavioral interventions (eg, SiHLE [Sisters, Informing, Healing, Living, Empowering] and Horizons) [10] that involve educational sessions led primarily by Black women have been effective in reducing disparities in HIV transmission [7,11]. Although effective, such interventions have been greatly limited by factors related to cost, scalability, sustainability, and reproducibility [7,12]. For example, barriers commonly associated with traditional interventions are limitations related to ease of access (eg, location of study activities), cost (eg, childcare), transportation difficulties, and time constraints, which can prevent study enrollment and engagement and/or lead to high study attrition [13]. Technologically delivered behavioral interventions driven by mobile health (mHealth) have the ability to overcome such barriers while offering several advantages compared to traditional face-to-face interventions [13]. Promising implications for the use of mobile technology to facilitate the delivery of health care rely on general mobile phone usage by the target population.

Among Black youth and residents of low-income households, smartphone ownership exceeds 80% and plays a critical role in providing internet access, compared to laptops or desktop computers [9,14-16]. As a result, Black youth and adults are more likely than White people to rely on their smartphones for certain activities, including seeking health care information [14]. The increasing popularity of smartphones along with smartphone apps have made the possibility of employing mobile phones and apps as a platform to provide HIV prevention information for Black women highly feasible [17].

HIV prevention interventions delivered through mobile apps have been largely aimed at MSM [18,19]; yet, they have failed to attract the attention of, and positive responses from, other ideal audiences, particularly Black women [18]. Studies that have focused on mHealth for HIV prevention among Black women have primarily utilized telephone calling and text-based counseling interventions, which have demonstrated marked efficacy in reducing risk behaviors and incident STIs among women compared to face-to-face counseling sessions [10,12]. Although interventions that have assessed texting and calling as HIV prevention interventions have been effective, there is a need for the development of mHealth interventions that specifically target Black women and use innovative and resourceful approaches that promote user engagement and sustainability. The few research studies that have taken on this approach have shown promising results. Jones and Lacroix [9] employed an mHealth intervention for Black women that used smartphones to deliver culturally tailored, soap opera videos that focused on HIV prevention for young Black women to encourage a reduction of high-risk behaviors. The video intervention was compared to weekly text-based HIV risk–reduction messages, which tend to be more commonplace in mHealth HIV prevention efforts. The intervention arm showed a dramatic reduction in unprotected sexual encounters postintervention when compared to baseline [9]. Gonzalez Gladstein [20] conducted a pilot randomized controlled trial that evaluated the feasibility and efficacy of a web-based app to increase HIV and STI knowledge and use among Black and Latina young adult women. Study participants in the intervention arm found the app to be trustworthy and useful, and results revealed that participants in the intervention arm had high levels of engagement and retention [20]. Currently, Browne et al [21] is conducting a randomized controlled trial that involves trialing a mobile app to provide HIV prevention and risk-reduction education to young African American women living in North Carolina who are sexually active and use drugs. If efficacious, this mHealth app can be useful in expanding the reach of HIV risk–reduction interventions for Black women and increasing accessibility for Black women who use mobile devices [21].

The etiology of HIV transmission among Black women largely differs from that of other high-risk populations, further emphasizing the need for more tailored interventions to be developed for Black women [7,11]. While HIV prevention efforts targeting MSM are necessary, it is crucial that we engage other vulnerable populations in prevention efforts in order to successfully meet both national and global HIV elimination goals. Black women are not likely to engage with mHealth apps developed for MSM, due to lack of relatability and contextual relevance as it relates to their intersecting HIV prevention and women's health needs. More specifically, health promotion interventions for Black women focusing on HIV prevention should promote self-empowerment, gender and ethnic pride, self-efficacy, and skills building [7]. Mobile apps developed for Black women have the potential to deliver HIV prevention information and skills in interactive, useful, nonstigmatizing, and discrete ways [7]. Furthermore, Black women are willing to participate in mHealth research that promotes the prevention and management of chronic illnesses, especially when such research is culturally tailored and comprehensively considers their reproductive and sexual health concerns. mHealth purports care continuity and constant accessibility, thus offering users flexibility and convenience with no limitation of time nor space [17,22-26]. In addition, mHealth provides the ability to deliver highly engaging HIV prevention information to populations that have been typically hard to reach, while offering user privacy and anonymity [17]. Considering Black women's use of mobile apps for the prevention of other chronic diseases, we are optimistic that our proposed HIV prevention mobile app will have similar outcomes [13,27].

This study was guided by the social cognitive theory of mass communication, which postulates four constructs from the original theory—self-efficacy, use of incentive motivation, social environment, and reciprocal determinism—that impact behavior, but adds that messaging to influence these behaviors can be effectively delivered through media and technological sources [28]. In order to understand vulnerable, cisgender, Black women's preferences for functionality, format, and design of a mobile HIV prevention app, and to examine their willingness to use an app for HIV prevention, we conducted formative qualitative research with Black cisgender women who live in communities that are geographically affirmed to have the highest HIV rates in Atlanta, Georgia [11].

Methods

Study Population and Recruitment

This study was approved by the Emory University Institutional Review Board. This qualitative study was implemented from February to March 2019 with cisgender Black women residing in metro Atlanta, specifically Fulton County, one of the targeted counties from the Ending the HIV Epidemic initiative [29]. Participants were recruited for focus group discussions (FGDs) via flyer distribution and community-based organization outreach. Flyers were distributed in venues that Black women frequent, such as beauty salons, churches, community-based organizations, and community events like health fairs. The flyers provided a link to a survey via the online survey tool SurveyMonkey, which assessed participant eligibility. Inclusion criteria for this research study required participants to be (1) English speaking, (2) cisgender female (ie, assigned female sex at birth and identified as female), (3) 18 years of age or older, (4) self-identified as Black, African American, and/or Hispanic or Latina, and (5) sexually active within the previous 3 months during study enrollment. Additionally, it was required that participants owned a smartphone and had never tested positive for HIV per self-report. Eligible women were contacted by phone to participate in a prescheduled FGD.

Study Procedures

In total, there were 23 participants who were divided among the four in-person FGDs held. We concluded that our sample size was adequate based on data saturation and empirical evidence of sufficiency [30]. Each FGD lasted approximately 90 minutes and took place at a partnering community-based organization or academic institution. FGDs were conducted by two trained moderators who were knowledgeable about the objectives of the HIV prevention mobile app. We employed an FGD guide (see Multimedia Appendix 1) during each session, and both moderators were present during all sessions to ensure the specific aims of the study were met and to ensure that there was consistency across the data collection procedures. The FGDs covered four main topics: (1) HIV prevention app usability, (2) features of an HIV prevention app, (3) app content to include barriers and facilitators to HIV testing and pre-exposure prophylaxis (PrEP) initiation, and (4) mobile app commodity ordering (eg, condoms and at-home HIV testing kits). During the FGD, we demonstrated four HIV prevention–focused mobile app wireframes—digital depictions of app content and functionality—that were developed by and for Black women. Specific mobile app functions included the following: (1) information about HIV and women's reproductive health (eg, via videos), (2) location-based HIV testing and PrEP clinics, (3) use of an in-app calendar for reproductive health (eg, ovulation and sexual acts diary) and HIV-specific notifications (eg, testing reminders), (4) commodity ordering for HIV prevention efforts (eg, condoms), (5) sexual behavior tracking, (6) frequently asked questions repository, (7) prevention navigator and/or provider communication, and (8) community connectivity (eg, peer chat group). Participants were also encouraged to provide suggestions for how to improve app function and recommend additional app features that should be integrated. All interviews were digitally recorded with the consent of each participant. Field notes were drafted in real time by the research assistant and later transcribed and appended to the FGD transcripts. Participants were given US $30 in compensation for their time.

Data Analysis

All FGD audio files were transcribed verbatim using a professional transcription service. Thematic analysis, combining inductive and deductive approaches, was completed using MAXQDA software, version 18 (VERBI GmbH). A codebook was compiled in close coordination between researchers (RC, NH, SG, and DBE) using existing literature, the research objective and aims, along with themes that emerged during the FGDs. The researchers then evaluated the FGD transcripts to

ensure congruency with the extracted themes using MAXQDA software. Following this process, the researchers discussed and compared their findings. Transcribed text and field-note data were then reviewed for overall impressions; finally, a line-by-line review for extraction of significant statements occurred [31].

Results

Overview

A total of 23 cisgender Black women participated in the FGDs; they ranged in age from 18 to 45 years, with a mean age of 30 years (SD 8). Demographic content was recorded for 17 of the 23 participants (74%), as 6 participants opted not to complete the demographic form (see Table 1). Most of the participants in the study were not married (12/16, 75%), did complete high school (10/16, 63%), had health insurance (15/17, 88%), and had a regular health care provider (15/17, 88%). Over half of the participants reported they had previously heard of PrEP (12/17, 71%) and less than half reported that they would not use PrEP (7/16, 44%). Broadly, participants reported their main health concerns as chronic diseases, sexual and reproductive health issues, along with inadequate acute and chronic mental health services.

Results were categorized into three overarching themes and eight related subthemes. The first overarching theme, *health content and communication*, included two subthemes: *comprehensive information* and *health provider profiles*. The second theme, *functionality, format, and design of the mobile app*, included three subthemes: *customizability, layout, colors, easy navigation, and simplistic design; safety and privacy concerns;* and *visual content*. The third theme, *other suggested features*, included two subthemes: *peer chat room(ie, community building)* and *"tell me where, get me there" transportation*.

Table 1. Participant demographics.

Category and description	Value, n (%)[a,b]
Race (n=17)	
Black or African American	16 (94)
West Indian	1 (6)
Marital status (n=16)	
Married	2 (13)
Never married or single	12 (75)
Not married but living with a sexual partner	2 (13)
Currently in school (n=17)	
Yes: full time	6 (35)
Yes: part time	0 (0)
No	11 (65)
Highest level of schooling (n=16)	
Elementary or middle school	1 (6)
High school	10 (63)
Trade or technical college	1 (6)
College or university	3 (19)
Not reported	1 (6)
Employment status (n=16)	
Full time	3 (19)
Part time	7 (44)
Unemployed	6 (38)
Yearly household income (US $; n=16)	
0-9999	7 (44)
10,000-19,999	2 (13)
20,000-29,999	2 (13)
30,000-39,999	3 (19)
40,000-49,999	2 (13)
Current health insurance (n=17)	
Yes	15 (88)
No	1 (6)
Unsure	1 (6)
Access to a regular health care provider (n=17)	
Yes	15 (88)
No	2 (12)
Have you heard of pre-exposure prophylaxis (PrEP; n=17)	
Yes	12 (71)
No	5 (29)
If you have heard of PrEP, would you use it (n=16)	
Yes	7 (44)
No	7 (44)
Maybe	2 (13)
Where do you obtain health information (n=17)	

Category and description	Value, n (%)[a,b]
Family and/or friends	6 (35)
Google	10 (59)
Health apps	7 (41)
Health care provider	15 (88)
Social media	2 (12)

[a]Out of 23 participants, 6 opted out of completing the demographic document.

[b]Not all category percentages add up to 100 due to rounding.

Health Content and Communication

Overview

The participants expressed the need for a comprehensive health app that included various topics in connection to a woman's overall health. Along with HIV, participants communicated that the mobile app would be most beneficial and would foster continued use if it included resources on additional health topics that were of concern to them, such as mental health. A directory of providers was a distinct feature desired by participants. More specifically, Black women requested having the app generate health care provider profiles based on geographic location (ie, geofencing) that would locate Black female providers. In regard to communication facilitated by the app, there was a consensus among the participants on having a virtual communication option to access their health care providers using either a face-to-face feature or a private chat room forum. Some participants were interested in having a peer-support option and wanted a monitored peer chat room available. The risk assessment questionnaire that was incorporated in three of four of the mobile app prototypes garnered reactions from participants that should govern how to solicit sensitive information via electronic devices from Black women. Existing app functions were desired; however, participants explicitly expressed wanting all features to be included in one mobile app, thereby minimizing the need to download multiple apps or the need to exit the app (eg, period or mood trackers). Table 2 lists and describes the codes for the themes and subthemes.

Table 2. Code definitions.

Code	Definition
Health concerns	Primary health concerns mentioned by participants, including those related to sexual and reproductive health; nutrition, exercise, or weight management; mental health; chronic disease; cancer; and accessibility to care
HIV	Discussions of HIV or HIV-related topics by participants, including HIV prevention methods and strategies they are aware of; experience, utilization, discussion, or knowledge regarding pre-exposure prophylaxis (PrEP); perceptions and preconceived notions regarding HIV; discussions related to HIV, HIV treatment and care, HIV-related illness, and risk factors related to HIV
Information delivery preference	Discussions regarding participants' preferred methods of receiving health-related information, including those related to the utilization of mobile apps or social media apps to receive health information, seeking health information on the web, and seeking health information from television commercials
Health app features: reliable digital dialogue	Discussions regarding participant preferences in app features and usability, including the following preferences: receiving health information through the app, interacting with health care providers through the app, interacting with other women in an app chat room, and personalized health information received by participants through the risk assessment
Comprehensive app content	Discussions of the app that include comprehensive information pertaining to all aspects of women's health, discussions about participant ability to search the app for information by typing in specific symptoms, and ability to access provider profiles that include information regarding their services
Convenient consumerism	Discussions regarding participant ability to order health kits, tests, etc, directly from the app
Customizability	Discussions regarding customizing the app and the display of features, including, but not limited to, color, font, layout, background, and music
Visual content	Discussions regarding nontraditional delivery methods of health care information through videos, short clips, and representative imagery
"Tell me where, get me there" transportation	Discussions of transportation and accessibility of services presented in the app, including health care providers within the participant's geographic area, links and accessibility to Lyft and Uber, and information about community carpools to health care services
Other suggested features	Discussions regarding additional suggestions voiced by participants to ensure efficiency and overall satisfaction with the app (ie, iOS and Android compatibility, journal component, and panic button for emergencies)
Privacy and confidentiality	Discussions regarding security measures to ensure participant confidentiality while using the app

Comprehensive Information

Topics that participants voiced wanting to see included in the mobile app consisted of the following: location of health screening services for women (eg, mammograms), mental health coping strategies and local resources, information on chronic health conditions that are most prevalent within the Black community (eg, diabetes and hypertension), holistic wellness, and information on sexual and reproductive health that appends to HIV prevention, such as other STIs.

> *My three concerns are STDs [sexually transmitted diseases], mental health, and chronic illnesses like high blood pressure, diabetes, and heart issues.* [Focus group #3 participant]

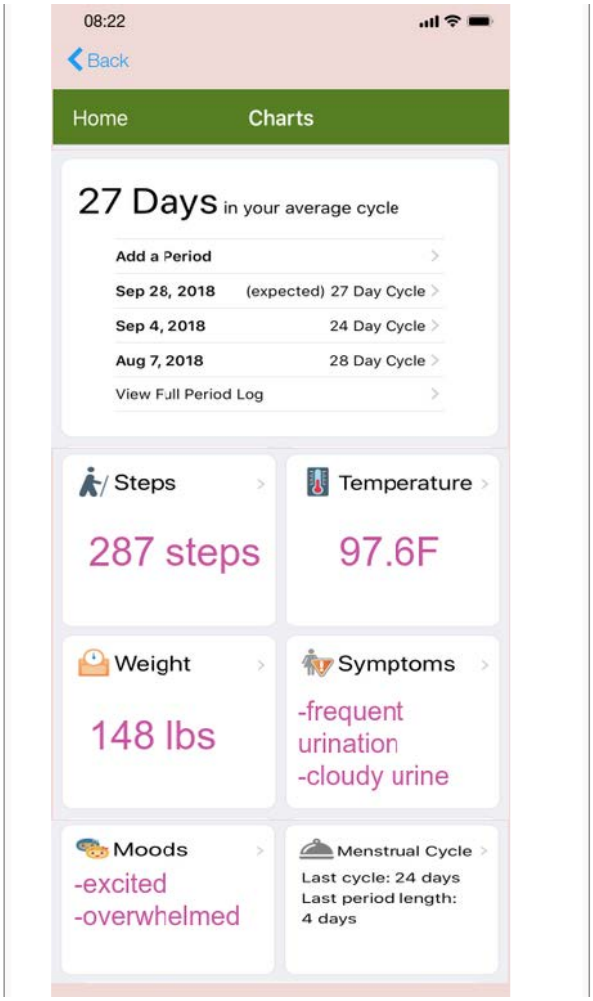
> *My three [health concerns] are, well I'd say asthma, I haven't got a sexually transmitted disease, but I'm just saying it for other women. So, I would say sexually transmitted diseases and I'd say anxiety too.* [Focus group #2 participant]

Participants also expressed an interest in a search engine function. With this feature, women could receive quick, reliable, and relevant information pertaining to common reproductive health trepidations, such as STIs and birth control methods. Additionally, participants recommended the inclusion of a period tracker feature within the app. Several participants already had period tracker apps on their phones and stated that it would be most convenient for them if the app afforded period initiation, duration, and symptom monitoring so that they could refer to just one app for all of their reproductive health needs (see Figure 1). Participants also liked the option of having a journal feature within the app to log journal entries and notes that are relevant to their health.

> *I like the multipurpose features because if I'm not necessarily looking for HIV information, if it's a well-rounded site and I'm able to get the period tracker, count my steps, and the other features like the news section it's well-rounded for me.* [Focus group #2 participant]

> *I think if there's a way to search quick things like yeast infections. A quick way to search things you know about common female health issues. So, you can just go ahead and see what you need to do. Instead of jumping through hoops for something basic you can get out of the way quickly.* [Focus group #3 participant]

Figure 1. Comprehensive health features included in the sample mobile app prototype.



Health Providers' Profiles

The participants reported an interest in having the mobile app accommodate a comprehensive directory of female health care providers of color within their communities. For the participants in this study, having access to health care providers they could relate to—considering intersections like race and ethnicity as well as gender—was influential when seeking reproductive and general health services. Participants also preferred having the option of directly communicating with a certified health care provider or a trained health care surrogate (eg, patient navigator) via either in-app video conference or chat room in case they have any immediate health concerns or questions (see Figure 2).

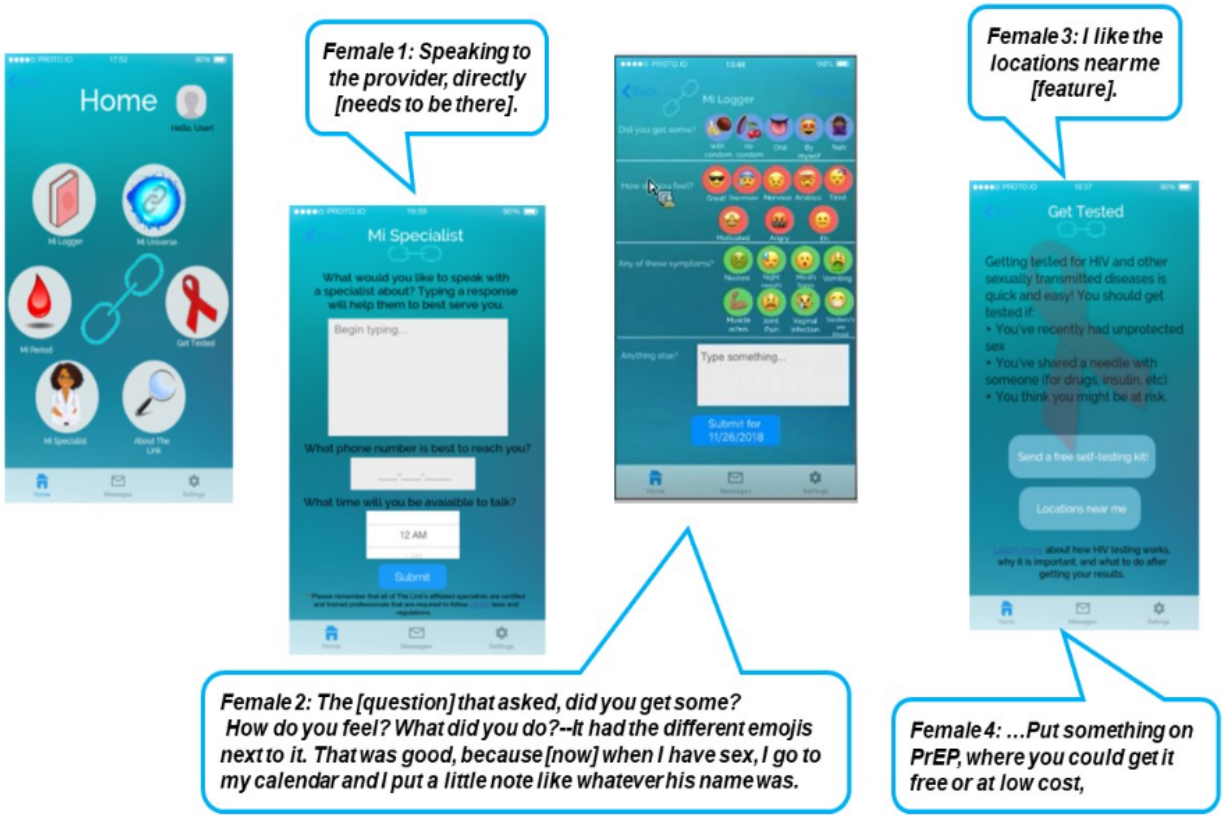
> *...Or even something where you could literally get a directory of female doctors in your area. Female Black doctors in your area. I think that would be very, very helpful because I have no idea of any Black women doctors in Atlanta. That would be very helpful. [Focus group #2 participant]*

A risk assessment that would collect health information about participants was a proposed component of most mobile app prototypes that were presented to the focus group participants. Components of the risk assessment were as follows: sexual encounters, use of contraception, and demographic information, such as age, race, and gender. The participants noted that the placement of the risk assessment was crucial for the overall acceptability of the app. There were differences in opinions among FGD participants as to where the risk assessment should be placed, along with the content of the risk assessment. Some participants stated that having the risk assessment appear immediately after signing in was a deterrent to using the app because the questions being asked were perceived as intimate or invasive. One participant stated that the risk assessment included too many questions, which caused her to become disengaged with the app. Multiple participants expressed that if the risk assessment were to be included in the app, then it should be placed in a different section entirely and serve a purpose (eg, usefulness to the health care provider for giving feedback to the patient regarding a particular health concern) for collecting the type of information being queried. However, there were also some participants, specifically younger participants, who did not have a problem with the placement of the risk assessment nor the content of the questions asked in the risk assessment.

> *It was for me [risk assessment], right off the bat, I was like, ya know...I was over it. But now what I will say, I think that those questions can be asked, but you would probably do it when they get to certain levels [within the app].* [Focus group #1 participant]

> *I feel like even the health tracker thing on your iPhone tells you to put in your weight and your height. I know my BMI is 39, but I also know I'm active as heck. So, yes, you can tell me this but my lifestyle says something different. I feel like I would expect a health app to ask me that. I feel like that would be part of the sign-in process. But I guess I'd maybe give the option to skip [the risk assessment].* [Focus group #3 participant]

Figure 2. Sample mobile app prototype features and pages. PrEP: pre-exposure prophylaxis.



Functionality, Format, and Design of the Mobile App

Overview

Participants were adamant about helping to create a mobile app that would not only be functional but also highly competitive with, and more culturally centric than, all other pre-existing mobile apps concerning reproductive health. These suggestions ranged from design features, layout, navigation within the mobile app, and concerns regarding safety and privacy to visual content and representation.

Customizability, Layout, Colors, Easy Navigation, and Simplistic Design

Across all focus groups, the participants discussed the various customization elements they wanted to see incorporated into the mobile app; for example, compatibility with both iOS and Android systems, limited ads within the app, volume options, and notification alerts. They also stressed that the app should be user friendly. This included a color layout with simplistic designs. One participant did note that she would like to see a reproductive app for women that was not "pink and flowery." Participants also expressed how they liked that the pictures, which were of Black women, within the app felt relatable to them. There was not a clear consensus regarding the use of emojis in the app; as some mentioned, it would be appealing to younger women, while others noted that it would come off as less mature.

> *I feel like just stay away from the color pink. I feel like I'm very tired of seeing that color associated with women.* [Focus group #3 participant]

> *I think a lot of the ads come from funding. I want to talk about my health! If the ads pop up, make it relevant to what you're looking for. Just no random ads...* [Focus group #4 participant]

Safety and Privacy Concerns

When using technology, privacy is oftentimes one of the major concerns for an individual, and the same could be said of our participants. Many were concerned that their private medical records and HIV status could be inadvertently disclosed. With web-based and smartphone apps allowing multiple ways to sign up and sign in, participants were concerned with their personal information being shared. The majority of participants emphasized not wanting to sign in to the app using social media platforms, such as Facebook or Instagram, due to the fear of their information potentially becoming compromised. If personal health information was required in order to sign up for the app, many participants expressed that they would be less inclined to complete the form due to privacy concerns. Participants also suggested that they would prefer a name for the app that was discrete and did not imply that the app pertained to HIV.

> *With so much stuff that's going on with these different social [media] sites, that's why a lot of people don't [share information online using these platforms] because they fear that this stuff can get linked back.* [Focus group #2 participant]

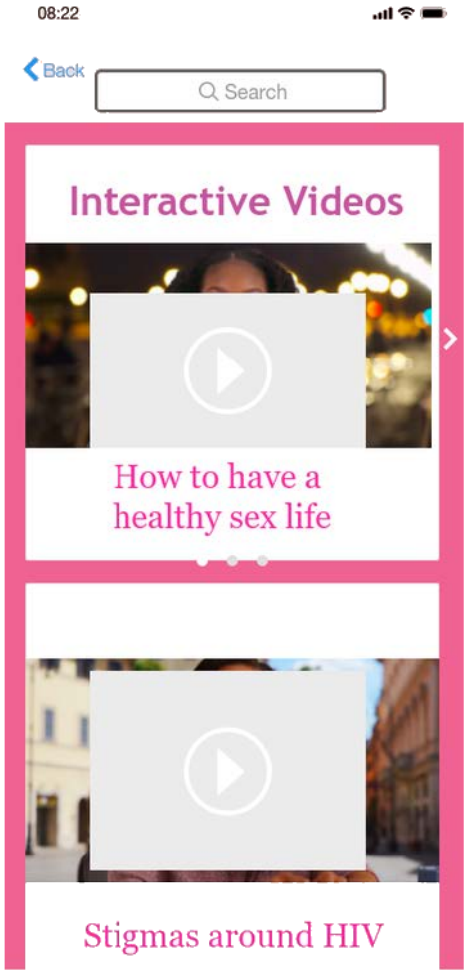
> *I would rather not do it with anything social media. Especially anything owned by the company Facebook, which includes Instagram. Just because they track and keep everything.* [Focus group #3 participant]

Visual Content

Participants were especially interested in receiving their health care information in new, innovative ways. Some participants described including health education resources in the form of short video clips of real-life scenarios (see Figure 3). Others discussed the incorporation of animations throughout the app as a way of providing relevant information. These alternatives to traditional methods of health information delivery were presented as a means of reaching all potential users. Participants suggested that given that literacy levels may vary among users, along with young adults primarily obtaining news and information through videos and clips, health education in the form of videos may be easier for some to understand.

> *I think in this day and age, people are visual people because of social media. A visual, I was thinking something more like real-life scenarios, whether that be a 60-second video, but something that's real, something that's honest, a real story.* [Focus group #1 participant]

Figure 3. Interactive videos on the mobile app featuring Black women.



Other Suggested Features

Overview

In addition to the themes and subthemes related to the app health content, functionality, format, and design, participants also provided additional suggested features they would like to see present in the app. This included a peer chat room and information regarding transportation and location of services. Regular news updates on important health topics were also suggested, along with the incorporation of a panic button for emergency situations.

Peer Chat Room

There was no clear consensus among the participants regarding a peer chat room. On one hand, some participants were intrigued by this feature, as it would allow an opportunity for community building and an opportunity to learn from others who may be experiencing similar circumstances. Much of the opposition regarding the chat room was related to the potential for sharing of misinformation relating to HIV and STI risks. In addition, participants were concerned with potential misuse of the chat room. They suggested security and privacy measures should be considered, as individuals may be sharing personal information relating to their reproductive health. Some participants preferred having chat rooms solely with health care providers as opposed to the peer chat rooms.

> *It [the peer chat room] just depends on how compelling it would be. Also, if it's regarding health, I would like a doctor in the chat room.* [Focus group #3 participant]

> *...The fact that they're having these chat rooms and avatars is too...childish. That would push me away because if we're talking about something serious, I don't want to make an avatar of myself and I don't want to be in a chat room talking to a whole bunch of people that may potentially have something that I'm concerned about. Because to me it tends to turn into something more fun than something more serious.* [Focus group #2 participant]

"Tell Me Where, Get Me There" Transportation

Participants stated that a GPS locator for HIV testing and PrEP clinics was a feature they would be interested in seeing incorporated within the app. For some of the participants, although they had heard of PrEP, they did not know where to obtain it. Thus, they stated that having this information accessible would make the process of utilizing such services much easier. Some participants even mentioned partnering with transportation services, such as Uber and Lyft, in order to provide transportation to promote accessibility of health care facilities.

> *...if [you] could put something on PrEP, where to get it free or at a low cost, because that's the question I get most of the time, because PrEP is very expensive.* [Focus group #2 participant]

Discussion

Overview

Aims of app development and definitions of success may differ between researchers and programmers, and each may be responding to different perceived needs of end users. Teams can address some of these differences through continuous formative research with the target population during intervention development and creating very explicit deliverables, timelines, and division of labor within the development team [32]. These first and second points emphasize the importance of designing technology-based interventions with the end user constantly in mind to create something that is intuitive, useful, engaging, and fun [32].

Black Women and the Use of Mobile Technology for Sexual and Reproductive Health

In this study, researchers solicited feedback from focus groups consisting of Black women residing in a high-risk area for HIV acquisition, in order to obtain qualitative data for the development of an HIV prevention mobile app that specifically targets Black women. Black women remain disparate in HIV acquisition and in other adverse sexual and reproductive health outcomes. Features of an HIV prevention and sexual and reproductive mobile health app informed by Black women could help combat health outcome inequalities. The findings obtained from this study indicate that, overall, participants strongly desire a mobile app that incorporates comprehensive health information that discusses HIV prevention information. Participants also strongly desire a mobile app that incorporates information pertaining to other aspects of women's sexual and reproductive health and information on other health conditions that have a high prevalence within the Black community. One particular aspect of the app that was highly regarded by study participants was the ability for participants to be linked with health care providers within their geographical area who were Black women. Representation oftentimes can play a major role in an individual's health decision making. Research has demonstrated that when Black patients have a health care provider of the same racial background, there tends to be a more patient-positive effect involving patient-centered communication, along with increased satisfaction [33,34]. Thus, by having access to Black providers, Black women may become more engaged in their overall health care, which can facilitate access to more health promotion initiatives, including HIV prevention efforts. Other desired app features may currently exist in some form implicating feasibility, either in related apps (eg, HIV prevention apps for MSM) or unrelated apps (eg, period tracker), but distinction will be demonstrated by culturally and contextually relevant content that Black women will deem useful regarding their sexual and reproductive health. Black women who are skeptical about health resources or who cannot always decipher health content will have a reliable source that will share evidence-based content in a manner that considers their literacy level and preference for information delivery. We also confer that the app will have to be malleable and inclusive as it relates to health issues that disproportionately impact Black women in order to remain relevant and be sustainable. Our participants were enthusiastic about being partners in mobile app development efforts and ensuring that the app was an authentic reflection of their needs.

HIV prevention delivered through mHealth that is specifically developed for Black women can offer a promising strategy to curtail HIV incidence rates within this group.

The data obtained from focus groups will be used to adjust and modify the mobile app we developed to ensure that the app includes the needs and concerns of Black women and to ensure that the app is culturally tailored to the population of interest. Cultural tailoring is a crucial component of HIV prevention interventions, and research has demonstrated that Black youth are more likely to reduce high-risk behaviors and increase condom use when they are able to identify with, and find meaning behind, the education that is provided [7,35,36]. Solely possessing HIV or STI prevention information has shown to not effectively reduce high-risk sexual behaviors [22]; however, greater message frequency and individually tailored messages have been effective in sustaining new health promotion behaviors and reducing high attrition rates [10,37].

Aptitude When Considering Prior Work

Black women continue to be disproportionately affected by HIV, and there is a great need for strategic interventions that are targeted, scalable, and sustainable in order to help reduce HIV rates among this group [7]. Through mHealth, HIV prevention efforts can be specifically targeted toward the needs of Black women, and they can also be scalable so that more participants are reached compared to more traditional HIV prevention interventions. Compared to traditional interventions, such as group counseling, mHealth reduces the amount of resources needing to be expended, making the sustainability of mHealth more feasible [7]. Although research has demonstrated that Black women are willing to utilize culturally tailored mHealth interventions, there has been a dearth of research assessing mHealth interventions—specifically the use of mobile apps—as a means of providing HIV prevention and risk-reduction education for Black women [25]. Studies that have assessed mHealth as a means of HIV prevention among Black women have focused primarily on web-based and telephone-based (eg, calling or texting) interventions [10,22]. In our review of the literature, we did not identify any articles that speak to the use of mobile apps as a means of HIV prevention for Black women. Although calling and texting interventions have proven to be effective, with the increase in smartphone ownership and mobile app use among Black women, a mobile app will offer a more expansive strategy that relies minimally on human resources for the delivery of health promotion content for this target population. Traditional HIV risk-reduction interventions targeting Black women have demonstrated marked challenges, including maintaining sustained intervention effects over time, along with the feasibility of expanding such interventions to reach more participants [7]. Thus, there is ample potential for the utilization

of mobile apps in order to overcome challenges in the delivery of HIV prevention content for Black women.

Mobile App Considerations

We are not oblivious to the fact that mobile apps have to be maintained and will require some resources for sustainability. We propose integrating this app into an existing health care system, whose administrators can assume ownership of protecting and storing the data, mobile app updates, and tracking patient usage. Our efforts would be to provide a general app that can be assumed by health care agencies (eg, Healthy Start) that deal primarily with female patients of color who oftentimes are not afforded digital services [38]. A recent systematic review revealed the limited representation of Black people and African Americans in health intervention research, even though they experience the greatest burden of health inequities [39]. Additionally, the review highlighted the increased willingness by Black women to participate in mHealth studies [39-41]. It is crucial for HIV prevention interventions delivered through mHealth to expand beyond recurrent themes and groups, such as MSM and medication adherence [40]. Other key populations such as Black women, along with themes such as HIV prevention and care initiation, need to be of focus in mHealth interventions [40]. In developing an mHealth app for Black women, it is important for the app to be revered as essential in order to incentivize Black women to keep it on their phones and permit the app to occupy valuable data space. We deduced that the health concerns of Black women will motivate their use of the mobile app if they can receive some of the same health benefits that are ordinarily afforded to them through traditional engagement with a health care provider or the health care system. There are commercial apps, such as Maven [42], that are available but are not within the purview of our population—the most susceptible to poor reproductive health outcomes—because they cannot afford the services. We, however, plan to integrate the mobile app into an existing health care system that will afford our target population an opportunity to engage with this mobile technology and with their health care providers.

Limitations

We identified some limitations to this study. One of these included the fact that our sample represented Black women residing in a large metropolitan community. Therefore, the results of this study can only be generalized to women of a similar demographic. Additionally, of the four focus groups, only one targeted the ideal age group for the proposed mobile app, due to recruitment methods and interest in the research project.

Conclusions

This research project sought to understand the interest in an HIV prevention app by Black women of reproductive age. In eliciting the opinions of Black women through focus groups, the information obtained can be used to ensure that the app that is developed will be relevant to the concerns and needs of Black women. In doing so, we can ensure that the app we develop is both culturally and contextually relevant so that users can be heavily engaged with the goal of promoting positive behavior change and risk reduction. Moving forward, the researchers would like to work alongside an advisory committee of Black women, a research team, and a technology company in order to develop the app prototype.

Acknowledgments

RC has received research support from the National Institute of Mental Health (grant No. R25MH067127).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group discussion guide.
[DOCX File , 174 KB - mhealth_v8i7e18437_app1.docx]

References

1. Basic statistics. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/basics/statistics.html [accessed 2020-01-29]
2. Maciak B, Krueger AL, Harris N, Division of HIV/AIDS Prevention. CDC HIV Prevention Progress Report, 2019. Atlanta, GA: Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services; 2019. URL: https://www.cdc.gov/hiv/pdf/policies/progressreports/cdc-hiv-preventionprogressreport.pdf [accessed 2020-01-28]
3. HIV and African Americans. Centers for Disease Control and Prevention. 2020 Jan. URL: https://www.cdc.gov/hiv/pdf/group/racialethnic/africanamericans/cdc-hiv-africanamericans.pdf [accessed 2020-01-28]
4. Lifetime risk of HIV diagnosis. Centers for Disease Control and Prevention. 2016 Feb 23. URL: https://www.cdc.gov/nchhstp/newsroom/2016/croi-press-release-risk.html [accessed 2020-01-28]
5. HIV and women. Centers for Disease Control and Prevention. 2020 Mar. URL: https://www.cdc.gov/hiv/pdf/group/gender/women/cdc-hiv-women_fact_sheet.pdf [accessed 2020-07-01]

6. Sharpe TT, Voûte C, Rose MA, Cleveland J, Dean HD, Fenton K. Social determinants of HIV/AIDS and sexually transmitted diseases among black women: Implications for health equity. J Womens Health (Larchmt) 2012 Mar;21(3):249-254. [doi: 10.1089/jwh.2011.3350] [Medline: 22196231]
7. Hendrick CE, Canfield C. HIV risk-reduction prevention interventions targeting African American adolescent women. Adolesc Res Rev 2017 Jun;2(2):131-149 [FREE Full text] [doi: 10.1007/s40894-016-0036-x] [Medline: 28626791]
8. Adimora AA, Schoenbach VJ, Floris-Moore MA. Ending the epidemic of heterosexual HIV transmission among African Americans. Am J Prev Med 2009 Nov;37(5):468-471 [FREE Full text] [doi: 10.1016/j.amepre.2009.06.020] [Medline: 19840704]
9. Jones R, Lacroix LJ. Streaming weekly soap opera video episodes to smartphones in a randomized controlled trial to reduce HIV risk in young urban African American/black women. AIDS Behav 2012 Jul;16(5):1341-1358. [doi: 10.1007/s10461-012-0170-9] [Medline: 22430640]
10. DiClemente RJ, Wingood GM, Sales JM, Brown JL, Rose ES, Davis TL, et al. Efficacy of a telephone-delivered sexually transmitted infection/human immunodeficiency virus prevention maintenance intervention for adolescents: A randomized clinical trial. JAMA Pediatr 2014 Oct;168(10):938-946 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.1436] [Medline: 25155070]
11. Hess K, Johnson S, Hu X, Li J, Wu B, Yu C, HIV Incidence and Case Surveillance Branch, Data Management Team of the Quantitative Sciences and Data Management Branch. HIV Surveillance Report (Volume 29): Diagnoses of HIV Infection in the United States and Dependent Areas, 2017. Atlanta, GA: Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services; 2018 Nov. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2017-vol-29.pdf [accessed 2019-11-04]
12. Blackstock OJ, Patel VV, Cunningham CO. Use of technology for HIV prevention among adolescent and adult women in the United States. Curr HIV/AIDS Rep 2015 Dec;12(4):489-499 [FREE Full text] [doi: 10.1007/s11904-015-0287-3] [Medline: 26412086]
13. James DCS, Harville C. Smartphone usage, social media engagement, and willingness to participate in mHealth weight management research among African American women. Health Educ Behav 2018 Jun;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
14. Perrin A, Turner E. Smartphones help blacks, Hispanics bridge some – but not all – digital gaps with whites. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2019 Aug 20. URL: https://www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2019-11-13]
15. Smith A. African Americans and technology use: Detailed demographic tables. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2014 Jan 06. URL: https://www.pewresearch.org/internet/2014/01/06/detailed-demographic-tables/ [accessed 2020-01-20]
16. Mobile fact sheet. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-01-20]
17. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: A review. J Med Internet Res 2013 Jan 04;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
18. 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]
19. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: A pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
20. Gonzalez Gladstein SK. Acceptability Study and Pilot RCT of a Guide to Understanding Reproductive Health for Ladeez (GURHL) Code: An HIV Risk Reduction App Intervention for Black and Latina Young Women in New York City [doctoral dissertation]. New York, NY: City University of New York (CUNY); 2018 Jan 1. URL: https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1018&context=sph_etds [accessed 2020-01-20]
21. Browne FA, Wechsberg WM, Kizakevich PN, Zule WA, Bonner CP, Madison AN, et al. mHealth versus face-to-face: Study protocol for a randomized trial to test a gender-focused intervention for young African American women at risk for HIV in North Carolina. BMC Public Health 2018 Aug 06;18(1):982 [FREE Full text] [doi: 10.1186/s12889-018-5796-8] [Medline: 30081868]
22. Billings DW, Leaf SL, Spencer J, Crenshaw T, Brockington S, Dalal RS. A randomized trial to evaluate the efficacy of a web-based HIV behavioral intervention for high-risk African American women. AIDS Behav 2015 Jul;19(7):1263-1274 [FREE Full text] [doi: 10.1007/s10461-015-0999-9] [Medline: 25616838]
23. Zhang A, Reynolds NR, Farley JE, Wang X, Tan S, Yan J. Preferences for an HIV prevention mobile phone app: A qualitative study among men who have sex with men in China. BMC Public Health 2019 Mar 12;19(1):297 [FREE Full text] [doi: 10.1186/s12889-019-6617-4] [Medline: 30866880]

24. Bailey JV, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Computer-based interventions for sexual health promotion: Systematic review and meta-analyses. Int J STD AIDS 2012 Jun;23(6):408-413. [doi: 10.1258/ijsa.2011.011221] [Medline: 22807534]
25. James DC, Harville C, Whitehead N, Stellefson M, Dodani S, Sears C. Willingness of African American women to participate in e-Health/m-Health research. Telemed J E Health 2016 Mar;22(3):191-197 [FREE Full text] [doi: 10.1089/tmj.2015.0071] [Medline: 26313323]
26. James DCS, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research Among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Nov 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
27. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: Systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
28. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001 Aug;3(3):265-299. [doi: 10.1207/s1532785xmep0303_03]
29. Ending the HIV epidemic: A plan for America. Centers for Disease Control and Prevention. 2020 Feb. URL: https://www.cdc.gov/EndHIV/ [accessed 2020-01-20]
30. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field methods 2016 Jul 24;29(1):3-22. [doi: 10.1177/1525822x16639015]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
33. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. Ann Intern Med 2003 Dec 02;139(11):907-915. [doi: 10.7326/0003-4819-139-11-200312020-00009] [Medline: 14644893]
34. Cooper-Patrick L, Gallo JJ, Gonzales JJ, Vu HT, Powe NR, Nelson C, et al. Race, gender, and partnership in the patient-physician relationship. JAMA 1999 Aug 11;282(6):583-589. [doi: 10.1001/jama.282.6.583] [Medline: 10450723]
35. Darbes L, Crepaz N, Lyles C, Kennedy G, Rutherford G. The efficacy of behavioral interventions in reducing HIV risk behaviors and incident sexually transmitted diseases in heterosexual African Americans. AIDS 2008 Jun 19;22(10):1177-1194 [FREE Full text] [doi: 10.1097/QAD.0b013e3282ff624e] [Medline: 18525264]
36. Rikard RV, Thompson MS, Head R, McNeil C, White C. Problem posing and cultural tailoring: Developing an HIV/AIDS health literacy toolkit with the African American community. Health Promot Pract 2012 Sep;13(5):626-636. [doi: 10.1177/1524839911416649] [Medline: 22102601]
37. Burton WM, White AN, Knowlden AP. A systematic review of culturally tailored obesity interventions among African American adults. Am J Health Educ 2017 Apr 12;48(3):185-197. [doi: 10.1080/19325037.2017.1292876]
38. Peek ME. Can mHealth interventions reduce health disparities among vulnerable populations? Divers Equal Health Care 2017;14(2):44-45 [FREE Full text] [doi: 10.21767/2049-5471.100091]
39. Tufts K, Johnson K, Shepherd J, Lee J, Bait Ajzoon MS, Mahan L, et al. Novel interventions for HIV self-management in African American women: A systematic review of mHealth interventions. J Assoc Nurses AIDS Care 2015;26(2):139-150. [doi: 10.1016/j.jana.2014.08.002] [Medline: 25283352]
40. Simoni JM, Kutner BA, Horvath KJ. Opportunities and challenges of digital technology for HIV treatment and prevention. Curr HIV/AIDS Rep 2015 Dec;12(4):437-440 [FREE Full text] [doi: 10.1007/s11904-015-0289-1] [Medline: 26412082]
41. James D, Harville C, Sears C, Efunbumi O, Bondoc I. Participation of African Americans in e-Health and m-Health studies: A systematic review. Telemed J E Health 2017 May;23(5):351-364. [doi: 10.1089/tmj.2016.0067] [Medline: 27792475]
42. Maven. URL: https://www.mavenclinic.com [accessed 2020-07-04]

Abbreviations

FGD: focus group discussion
mHealth: mobile health
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
SiHLE: Sisters, Informing, Healing, Living, Empowering
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by J Jones, D Levine; comments to author 04.04.20; revised version received 09.04.20; accepted 10.04.20; published 24.07.20.

Please cite as:
Chandler R, Hernandez N, Guillaume D, Grandoit S, Branch-Ellis D, Lightfoot M
A Community-Engaged Approach to Creating a Mobile HIV Prevention App for Black Women: Focus Group Study to Determine Preferences via Prototype Demos
JMIR Mhealth Uhealth 2020;8(7):e18437
URL: http://mhealth.jmir.org/2020/7/e18437/
doi:10.2196/18437
PMID:32706723

Review

Privacy Assessment in Mobile Health Apps: Scoping Review

Jaime Benjumea[1], MSc; Jorge Ropero[1], PhD; Octavio Rivera-Romero[1], PhD; Enrique Dorronzoro-Zubiete[1], PhD; Alejandro Carrasco[1], PhD

Department of Electronic Technology, Universidad de Sevilla, Seville, Spain

Corresponding Author:
Jaime Benjumea, MSc
Department of Electronic Technology
Universidad de Sevilla
ETS Ing Informática
Avda Reina Mercedes s/n
Seville, 41012
Spain
Phone: 34 954559960
Email: jaimebm@us.es

Abstract

Background: Privacy has always been a concern, especially in the health domain. The proliferation of mobile health (mHealth) apps has led to a large amount of sensitive data being generated. Some authors have performed privacy assessments of mHealth apps. They have evaluated diverse privacy components; however, different authors have used different criteria for their assessments.

Objective: This scoping review aims to understand how privacy is assessed for mHealth apps, focusing on the components, scales, criteria, and scoring methods used. A simple taxonomy to categorize the privacy assessments of mHealth apps based on component evaluation is also proposed.

Methods: We followed the methodology defined by Arksey and O'Malley to conduct a scoping review. Included studies were categorized based on the privacy component, which was assessed using the proposed taxonomy.

Results: The database searches retrieved a total of 710 citations—24 of them met the defined selection criteria, and data were extracted from them. Even though the inclusion criteria considered articles published since 2009, all the studies that were ultimately included were published from 2014 onward. Although 12 papers out of 24 (50%) analyzed only privacy, 8 (33%) analyzed both privacy and security. Moreover, 4 papers (17%) analyzed full apps, with privacy being just part of the assessment. The evaluation criteria used by authors were heterogeneous and were based on their experience, the literature, and/or existing legal frameworks. Regarding the set of items used for the assessments, each article defined a different one. Items included app permissions, analysis of the destination, analysis of the content of communications, study of the privacy policy, use of remote storage, and existence of a password to access the app, among many others. Most of the included studies provided a scoring method that enables the comparison of privacy among apps.

Conclusions: The privacy assessment of mHealth apps is a complex task, as the criteria used by different authors for their evaluations are very heterogeneous. Although some studies about privacy assessment have been conducted, a very large set of items to evaluate privacy has been used up until now. In-app information and privacy policies are primarily utilized by the scientific community to extract privacy information from mHealth apps. The creation of a scale based on more objective criteria is a desirable step forward for privacy assessment in the future.

(JMIR Mhealth Uhealth 2020;8(7):e18868) doi:10.2196/18868

KEYWORDS

privacy; mHealth; apps; privacy assessment; data privacy; review; security; mobile phone

Introduction

Although data privacy has always been a concern of the utmost interest, there has been some neglect for years, as changes have taken shape faster than regulations. Only recently have developers and customers really begun to worry about data privacy. The enormous amount of data handled by companies and the exposure of users' sensitive information have led governments to design frameworks to care for the privacy of citizens [1,2]. Likewise, the large amount of data handled by

the Internet of Things through big data techniques has raised concerns about privacy [3,4].

The health domain, however, was probably the first to have privacy regulation. In 1996, the Health Insurance Portability and Accountability Act (HIPAA) required the United States Department of Health and Human Services to safeguard protected health information according to national standards. Some of the requirements deal with data privacy [5].

In Europe, concerns have not been limited to the health domain, and regulations are strict. In 2018, the General Data Protection Regulation (GDPR) replaced the existing 1995 Data Protection Directive, and it became directly applicable to all European Union member states [6]. The GDPR introduced an important and modern change of approach toward a reinforced principle of accountability [7].

These concerns also apply to mobile health (mHealth) apps. mHealth technology has been widely adopted in many countries worldwide, as the number of smartphones and mHealth apps has increased dramatically. In 2018 in the United States, 77% of the population owned a smartphone [8], and in 2017, there were more than 300,000 mHealth apps [9]. The proliferation of this kind of app has allowed individuals to generate significant quantities of data about their lifestyles [10]. This situation has not escaped the attention of scientific researchers, and data privacy is a recurrent topic reported on in qualitative studies focused the needs and preferences of people with chronic conditions regarding mHealth solutions [11].

Although mHealth apps hold promise as self-management, monitoring, and behavior-change tools, among others, many smartphone users do not download mHealth apps because of lack of interest, cost, and concern about apps collecting their data [12]. Some studies have proven that there is cause for users' concerns about both the privacy and security of these apps [13] and some assess only the lack of privacy of several of these apps [14,15]. It is, therefore, important to have the right tools to evaluate privacy and security levels by identifying different methods of assessing mHealth apps.

Despite privacy assessment currently being a relevant topic, there is a lack of objective protocols, methods, and procedures in place to define the necessary metrics and steps for a privacy assessment of an mHealth app. Different methods may be used to analyze privacy, such as assessment of privacy policies, evaluation of app communications, and studying app behavior. Extracting the information used to evaluate the privacy of mHealth apps, and even creating a taxonomy of the privacy components used for the assessment, should be important goals for researchers.

Further, different metrics and items have also been proposed to assess privacy. The types of measurements and items used should be based on laws, recommendations, and best practices. Discovering the different criteria that can be used for privacy assessment and the methods of defining them is imperative. Therefore, our literature review fills this research gap, focusing on describing and comparing how privacy is assessed by researchers in the mHealth domain.

Finally, we consider whether there is any measure of how good the privacy is in an mHealth app and how it would be possible to develop a scale for a privacy score. As such, we must search for any available way of assessing privacy in mHealth apps as well as the information that could potentially be used, and how it has been used, in these evaluations. To the best of our knowledge, no other review regarding the privacy of mHealth apps has been published.

Methods

Overview

This review aims to summarize how privacy is assessed in the literature including any type of study design. For this purpose, we conducted a scoping review using Arksey and O'Malley's proposed framework [16]. We used Tricco et al's PRISMA ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [17] as a guide for reporting the procedure (see Multimedia Appendix 1). The authors of the framework include "summarize and disseminate research findings" and "identify research gaps in the existing literature" in the rationale for conducting a scoping review. Also, Arksey and O'Malley list "addressing a broad topic where many different study designs might be applicable" as a characteristic of scoping studies.

Search Strategy

A systematic search strategy was used to identify relevant papers about the assessment of mHealth app privacy. The search was conducted in July 2019 in English, using terms regarding privacy, mHealth, and assessment; the following electronic databases were used: Scopus, PubMed, IEEE (Institute of Electrical and Electronics Engineers) Xplore, and ACM (Association for Computing Machinery) Digital Library. The search string used was as follows: privacy AND ("health app" OR "health apps" OR "mobile health" OR mhealth) AND (test OR testing OR tested OR framework OR review OR reviewing OR reviewed OR evaluate OR evaluation OR evaluating OR evaluated OR assess OR assessing OR assessment OR assessed OR "comparative analysis" OR "regulation compliance" OR taxonomy). The search terms and strategies for each database are detailed in Multimedia Appendix 2.

The database results were imported into the Mendeley application to further scrutinize the papers.

Selection Criteria

The inclusion criteria for studies were as follows:

1. Papers that assessed the privacy of mHealth apps, regardless of the subject of the assessment, as well as papers that assessed several aspects of mHealth apps, including privacy.
2. Papers published with a title and abstract in English from 2009 onward in research journals, conference proceedings, or book chapters.

Papers that did not propose a method to evaluate privacy were excluded—even if they analyzed privacy—if they focused only on general aspects, such as users' concerns, threat analysis, or challenges identified. Papers that did not evaluate any app were also excluded.

Study Selection

After completing the search process and removing duplicates, the remaining 480 papers were screened. Initially, two authors (JB and JR) independently reviewed 10.0% (48/480) of the titles and abstracts to assess the level of agreement; the Cohen κ statistic, a measure of interrater reliability, was 0.73, which denotes an acceptable level of agreement [18]. Then, each author analyzed half of the remaining titles and abstracts to determine if they were potentially suitable for our objective. As a result, 77 articles were selected. Each author subsequently conducted a full-text review of those papers and 24 articles were ultimately included for data extraction. During this process, any doubt or discrepancy was resolved by consensus.

Charting

The authors followed a collaborative and iterative process to define a charting table for collecting the data from the included studies. Information was gathered into four main groups: general information, evaluation procedure, evaluation criteria, and scoring method. Multimedia Appendix 3 shows the charting table that was used.

The *general information* group includes the year of publication, source title and type, app area, as well as the number of analyzed mHealth apps.

The *evaluation procedure* group comprises all the information related to the way the apps were assessed, according to the assessment design and the object of assessment. The assessment design deals with the type of evaluation that was done. Some papers analyze only privacy, while others assess security and privacy, and some even evaluate privacy as part of the whole functioning of the app. Information regarding whether the study assessed only app privacy, or whether app privacy was a component of a multidimensional evaluation, is included in this category. Additionally, information regarding what privacy components were assessed is also part of this group. After reviewing the full text of the included studies, a taxonomy of privacy components was defined by consensus. The categories, based on our review, that were used for the assessment of privacy were as follows:

1. App properties and behavior: this category refers to the app functionality. An article falls into this category if the app was actively used and some user information was provided to the app. Examples of this category are the type of log in used by the app, such as email or connecting via an external provider like Facebook, or if user registration and/or a password are needed to use the app.
2. In-app information: as with the previous category, the app was analyzed from within to look for information related to privacy, such as security measures or data sharing. Privacy policies were assessed in a separate item because some articles assess this in that fashion.
3. Personal information types: to fall into this category, the article must explicitly analyze the type of personal data collected by the app.
4. App communications: some articles analyze whether the app communications are private by intercepting traffic. Therefore, it is possible not only to know if traffic is encrypted but also, in some cases, to check the content of the traffic. Some authors were also able to find out the traffic destination of app communications, such as third parties and ad sites.
5. Static and dynamic analyses: the use of static and/or dynamic analysis is very common when evaluating the security of an app; however, these analyses can also be used to analyze certain aspects of privacy, such as whether privacy measures are properly implemented in app communications and the types of permissions used by an app.
6. Existence of a privacy policy: articles that check for the existence of a privacy policy are included in this category.
7. Analysis of the content of the privacy policy and/or the Type of Service: the authors of the article have read the privacy policy and searched for the presence or absence of certain information, such as how the data are stored, the use of encryption, and whether the data are shared with third parties, among others. Legibility (see the next category) is excluded from this category because the metrics used to evaluate legibility do not depend on the type of document being assessed.
8. Privacy policy legibility: transparency is one of the pillars of GDPR, and some articles analyze certain metrics regarding the readability of an app's privacy policy, including the length of the document, number of phrases, and use of readability algorithms available in the literature.

The *evaluation criteria* group includes the items used for the assessment and what the assessment criteria are based on. Very heterogeneous information was extracted from each article, and the assessment criteria were decided on in varied ways. Evaluators chose a set of criteria based on the literature, the authors' experience, an existing legal framework, and/or certain privacy recommendations and principles. It is difficult to categorize the criteria that were used to assess privacy, as they were not selected in a purely objective way. Different privacy items are defined according to the categories previously described. After extracting all the data regarding privacy assessment criteria from the studies that met the inclusion criteria, we defined, by consensus, a classification system consisting of 21 elements, listed hereafter.

A privacy policy is important when assessing privacy. The following items can be defined according to the content of a privacy policy: the existence of a data controller, details about the provision of a data protection officer, stating the purpose of data processing, establishing the legal basis, identifying the recipients of personal data, disclosing the occurrence of international data transfers, establishing the subject's data rights (including the right to withdraw consent), whether it is an obligation to provide data, disclosing the occurrence of data profiling, detailing the nature of the collected information, stating the risks of data collection, disclosing the location of the collected information, and using anonymization.

Some of these items may also be defined by in-app information. Details regarding the purpose of data processing, the legal basis, the recipients of personal data, the existence of the subject's data rights, the risks of data collection, and the protection of

minors were extracted from the in-app information for this review.

Personal information types were used to define the nature and location of the collected information. App properties and behavior define whether user registration is necessary and the minimum amount of data collection that must be collected for an app to function correctly. App communications as well as static and dynamic analyses were used to check traffic and whether security measures were implemented; for these last cases, the distinction between security and privacy was not obvious.

Last, the *scoring method* group deals with the existence or nonexistence of a final score in each article. If there was a score, the weighting of assessed items was also considered.

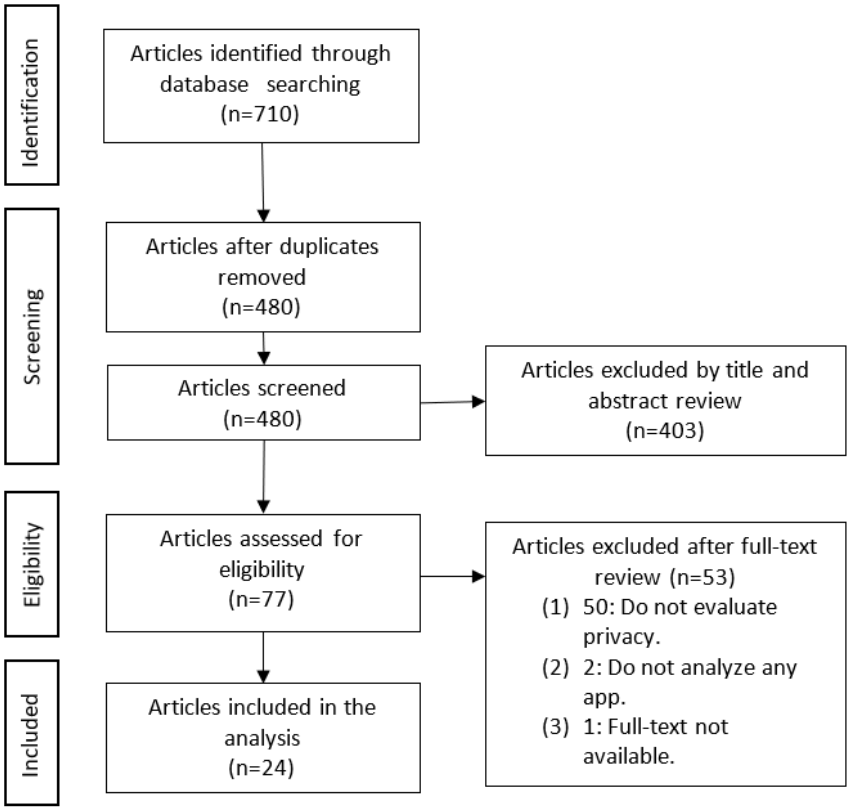
The charting table containing all the data to be extracted was implemented using Microsoft Excel. Two authors (JB and JR) independently extracted data from the 24 selected articles. Discrepancies were resolved by consensus.

Results

Search Results

The database search retrieved a total of 710 citations; 230 duplicates were removed. After an initial screening of the abstracts and titles, 403 articles that did not meet the eligibility criteria were excluded and 77 were selected for full-text screening. After the full-text review, 24 studies [6,14,15,19-39] remained that fulfilled the inclusion criteria for this scoping review (see Figure 1). A full list of the included studies can be found in Multimedia Appendix 4.

Figure 1. Flow diagram of the search strategy.



General Information

The general information contained in each study is summarized in Table 1. The year of publication, source title and type, app areas, and number of analyzed apps comprise the general information from each article. The source type is categorized as either a journal article, conference paper, or book chapter. The app areas were determined according to what the original article stated about the subject matter.

Table 1. General information from each article.

Reference	Source[a]	App areas	Number of analyzed apps
Papageorgiou et al, 2018 [6]	IEEE (Institute of Electrical and Electronics Engineers) Access (J)	Pregnancy and baby growth Family members and assistants Blood pressure and diabetes	20
Minen et al, 2018 [14]	Headache (J)	Headache	14
Huckvale et al, 2019 [15]	JAMA (Journal of the American Medical Association) Network Open (J)	Depression Smoking cessation	36
Scott et al, 2015 [19]	Australasian Journal of Information Systems (J)	General (top 20 mobile health [mHealth] apps)	20
Brüggemann et al, 2016 [20]	Annual Privacy Forum (J)	Medical Health and fitness	298
Mense et al, 2016 [21]	Studies in Health Technology and Informatics (BC)	Health and fitness	20
Hutton et al, 2018 [22]	JMIR mHealth and uHealth (J)	Self-tracking	64
Zapata et al, 2014 [23]	Annual International Conference of the IEEE Engineering in Medicine and Biology Society (C)	Personal health record	24
Sunyaev et al, 2015 [24]	Journal of the American Informatics Association (J)	Medical Health and fitness	600
Leigh et al, 2017 [25]	Evidence-Based Mental Health (J)	Chronic insomnia	19
Baumel et al, 2017 [26]	Journal of Medical Internet Research (J)	Health-related behaviors Mental health	84
Bachiri et al, 2018 [27]	Journal of Medical Systems (J)	Pregnancy	19
de las Aguas Robustillo Cortés et al, 2014 [28]	Telemedicine and e-Health (J)	HIV/AIDS	41
Quevedo-Rodríguez and Wagner, 2019 [29]	Endocrinología, Diabetes y Nutrición (J)	Diabetes	42
Knorr et al, 2015 [30]	IFIP (International Federation for Information Processing) Advances in Information and Communication Technology (J)	Diabetes Blood pressure	154
Zapata et al, 2014 [31]	RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) (J)	Personal health record	24
Bondaronek et al, 2018 [32]	JMIR mHealth and uHealth (J)	Physical activity	65
O'Laughlin et al, 2019 [33]	Internet Interventions (J)	Depression	116
Adhikari et al, 2014 [34]	Australasian Conference on Information Systems (C)	General (top 20 mHealth apps)	20
Aliasgari et al, 2018 [35]	IEEE Conference on Application, Information and Network Security (C)	General (top 25 mHealth apps)	25
Mense et al, 2016 [36]	Modeling and Simulation in Medicine Symposium (C)	Health and fitness	10
Powell el al, 2018 [37]	JMIR mHealth and uHealth (J)	Diabetes Mental health	70
Huckvale et al, 2015 [38]	BMC (BioMed Central) Medicine (J)	General	79
Robillard et al, 2019 [39]	Internet Interventions (J)	Mental health	369

[a]Sources include journal articles (J), conference papers (C), or book chapters (BC).

According to the type of source, 19 out of the 24 articles (79%) were published in journals [6,14,15,19,20,22,24-33,37-39], whereas 4 (17%) were published in conference proceedings [23,34-36] and 1 (4%) was a book chapter [21]. The publication fields were quite heterogeneous, with 12 out of 24 articles (50%) pertaining to the area of *medical informatics* [21-24,26-28,32,33,36,37,39], 5 (21%) to *medicine* [14,15,25,29,38], 4 (17%) to *information technology*

[19,30,31,34], 2 (8%) to *security and privacy* [20,35], and 1 (4%) to a multidisciplinary source [6].

Based on our defined inclusion criteria, we analyzed articles published between January 2009 and July 2019. None of the selected articles was published between 2009 and 2013. Out of the 24 papers, 4 (17%) were published in 2014 [23,28,31,34], 4 (17%) in 2015 [19,24,30,38], 3 (13%) in 2016 [20,21,36], 2 (8%) in 2017 [25,26], 7 (29%) in 2018 [6,14,22,27,32,35,37], and 4 (17%) in the first half of 2019 [15,29,33,39].

A wide range of app types was analyzed in the included studies, and some articles analyzed apps in different areas. For instance, in Knorr et al [30], both diabetes and blood pressure apps were analyzed. Fitness apps, including self-tracking and physical activity apps, were the most analyzed, appearing in 6 articles (25%) [20-22,24,32,36]. Mental health apps, including apps for depression monitoring, were assessed in 5 articles (21%) [15,26,33,37,39], and diabetes-related apps appeared in 4 articles (17%) [6,29,30,37]. Other app areas were HIV/AIDS (1/24, 4%) [28], headache (1/24, 4%) [14], pregnancy and baby growth (2/24, 8%) [6,27], personal health record management (2/24, 8%) [23,31], chronic insomnia (1/24, 4%) [25], and smoking cessation (1/24, 4%) [15]. Top mHealth apps were assessed in 4 articles (17%) [19,34,35,38].

Only 2 articles out of 24 (8%) analyzed certified apps. Huckvale et al [38] analyzed 79 apps certified by the United Kingdom's National Health Service (NHS) and concluded that there were gaps in compliance with data protection principles in these accredited apps. By contrast, Leigh et al analyzed 18 apps for Android and 1 NHS-certified app for iOS [25], and the authors found that the NHS-approved app outscored the others when using their evaluation criteria.

Finally, the number of apps analyzed in each article is disparate, ranging from 10 to 600 apps, with 20 apps being the mode (3/24, 13%) [19,34,35]. The average number of apps assessed was 92.6 (SD 136.9). Most of the articles (13/24, 54%), however, assessed less than 51 apps [6,14,15,19,21,25,27-29,31,34-36].

Evaluation Procedure

A summary of the collected information is shown in Table 2. The objects of the assessments and the basis of the assessment criteria are described in the Methods section.

Table 2. Procedure for evaluation of the apps.

Reference	Area of assessment	Object of the assessment	Basis of the assessment criteria (includes legal framework)
Papageorgiou et al, 2018 [6]	Privacy and security	In-app information Static and dynamic analyses App communications Existence of a privacy policy Content of the privacy policy	Authors Legal
Minen et al, 2018 [14]	Privacy	Static and dynamic analyses App communications Existence of a privacy policy Content of the privacy policy	Authors
Huckvale et al, 2019 [15]	Privacy	In-app information App communications Existence of a privacy policy Content of the privacy policy	Literature
Scott et al, 2015 [19]	Privacy and security	App properties and behavior Existence of a privacy policy	Literature
Brüggemann et al, 2016 [20]	Privacy	App properties and behavior Personal information types App communications	Authors Literature
Mense et al, 2016 [21]	Privacy	App communications	Author
Hutton et al, 2018 [22]	Privacy	App properties and behavior In-app information Existence of a privacy policy Content of the privacy policy	Literature Legal Recommendations or principles
Zapata et al, 2014 [23]	Privacy	App properties and behavior Existence of a privacy policy	Literature Legal
Sunyaev et al, 2015 [24]	Privacy	Existence of a privacy policy Content of the privacy policy Legibility of the privacy policy	Authors
Leigh et al, 2017 [25]	Multidimensional	In-app information Existence of a privacy policy Content of the privacy policy	Legal Recommendations or principles
Baumel et al, 2017 [26]	Multidimensional	Existence of a privacy policy Content of the privacy policy	Literature
Bachiri et al, 2018 [27]	Privacy	App properties and behavior Existence of a privacy policy	Literature Legal Recommendations or principles
de las Aguas Robustillo Cortés et al, 2014 [28]	Multidimensional	App properties and behavior In-app information	Recommendations or principles
Quevedo-Rodríguez and Wagner, 2019 [29]	Multidimensional	App properties and behavior In-app information Existence of a privacy policy Content of the privacy policy	Recommendations or principles
Knorr et al, 2015 [30]	Privacy and security	Legibility of the privacy policy	Recommendations or principles
Zapata et al, 2014 [31]	Privacy	App properties and behavior Existence of a privacy policy Content of the privacy policy	Authors Recommendations or principles

Reference	Area of assessment	Object of the assessment	Basis of the assessment criteria (includes legal framework)
Bondaronek et al, 2018 [32]	Privacy and security	Existence of a privacy policy Content of the privacy policy	Recommendations or principles
O'Laughlin et al, 2019 [33]	Privacy	Existence of a privacy policy Content of the privacy policy	Authors
Adhikari et al, 2014 [34]	Privacy and security	App properties and behavior In-app information Existence of a privacy policy	Literature
Aliasgari et al, 2018 [35]	Privacy and security	App communications	Legal
Mense et al, 2016 [36]	Privacy and security	App communications	Authors Recommendations or principles
Powell el al, 2018 [37]	Privacy	Existence of a privacy policy Legibility of the privacy policy	Authors
Huckvale et al, 2015 [38]	Privacy and security	App properties and behavior In-app information Static and dynamic analyses App communications Existence of a privacy policy Content of the privacy policy	Legal
Robillard et al, 2019 [39]	Privacy	Existence of a privacy policy Content of the privacy policy Legibility of the privacy policy	Authors Literature

Of the 24 articles assessed, 12 (50%) [14,15,20-24,27,31,33,37,39] evaluated only privacy; 8 (33%) evaluated security features, together with privacy [6,19,30,32,34-36,38]; and 4 (17%) conducted a multidimensional assessment [25,26,28,29], with privacy being only part of the evaluation.

When considering the object of the assessment, 19 out of the 24 articles (79%) [6,14,15,19,22-27,29-34,37-39] used the privacy policy as part of the assessment or solely evaluated the privacy policy. App properties and behavior were used for assessment by 10 articles (42%) [14,19,20,22,23,27-29,34,38], and 8 papers (33%) used in-app information [6,15,22,25,28,29,34,38] or app communications [6,15,20,21,30,35,36,38] for privacy evaluation. Finally, only 2 articles (8%) each used personal information types [20,38] and static and dynamic analyses [6,30].

The selected articles used different bases to define criteria to assess privacy of mobile apps. Most of the papers combined some sources to determine the items for assessment. Out of 24 papers, 10 (42%) [15,19,20,22,23,26,27,34,37,39] used the literature to determine the items, while 9 (38%) [6,14,20,21,24,33,36,37,39] were based on the authors' criteria. Not many of the papers used legal frameworks or regulations—only 3 out of 24 papers (13%) [23,27,35] used the HIPAA and just 2 (8%) [6,22] explicitly mentioned the GDPR as a basis for determining the assessment criteria, although none of them checked the GDPR compliance. However, out of 24 articles, 2 (8%) [25,38] did use the previous European privacy regulation (ie, the 1995 Data Protection Directive). A total of 12 other privacy frameworks, recommendations from certification organizations and standard associations, and privacy principles were used. Multimedia Appendix 5 shows a further analysis regarding the object of the privacy assessment in mHealth apps.

Evaluation Criteria

The evaluation criteria are heterogeneous, as were the methods for defining them. Though a very brief summary of the criteria is shown in Table 3, they are described in more detail in Multimedia Appendix 6. The classification items proposed by the different articles to be used for evaluating app privacy are shown in Table 4.

Table 3. Criteria for evaluation of the apps.

Reference	Criteria	Assessment of criteria
Papageorgiou et al, 2018 [6]	Privacy policy: consent, user rights (ie, withdraw and portability), data protection officer, data collection, purpose, and transfer Permission and static analysis Data transmission: https, SSL (Secure Sockets Layer), and secure transmission	Number of apps that meet the different criteria
Minen et al, 2018 [14]	Account functionality Data storage Privacy policy: type of information collected, data sharing, protection of minors, data access, and user rights	Number of apps that meet the different criteria
Huckvale et al, 2019 [15]	Privacy policy availability Uses of data, data transfer, and data collection Mechanisms for security, how long data will be retained, cookies, user rights (ie, opt-out, consequences of not providing data, deletion, editing, and complaints), and protection of minors Identity of data controller Adherence to privacy policy	Percentage of apps that meet the different criteria
Scott et al, 2015 [19]	User registration and authentication Data storing and sharing Enable users to update, correct, and delete their data Data privacy and security measures and existence of privacy policy	Items 1-3: risk score (1 point if there is a risk); Items 4-9: safety score (1 point if it is safe)
Brüggemann et al, 2016 [20]	Information-sharing targets (S), information transfer (T), and information collection (U) Personal information types (P) and log-in (L) Connection security (R)	PrivacyRiskScoreApp = TApp × w(T) + PApp × w(P) + LApp × w(L)+ SApp × w(S) + UApp × w(U) + RApp × w(R) w = weight
Mense et al, 2016 [21]	Use of SSL and certificate pinning Information sent and identification of third parties	Number of apps that meet the different criteria
Hutton et al, 2018 [22]	Notice and awareness: data sharing, nature of data, and explanation of security measures Choice or consent: user-consent control Access or participation: user access to data Social disclosure: privacy control	Most heuristics are valued as 0-2 (0, 1, or 2), though some have slightly different values (ie, 0/1, 0-3, or 0-4)
Zapata et al, 2014 [23]	Privacy policy access and updates Authentication, encryption, and security standards Access can be granted and revoked	All six items are valued as 0, 0.5, or 1
Sunyaev et al, 2015 [24]	Privacy policy availability Privacy policy features: length, readability, scope, and transparency (ie, sharing, collection, and user controls)	Number of apps that meet the different criteria
Leigh et al, 2017 [25]	Data sharing Confidentiality mechanisms Privacy policy availability and content (ie, data collection, use of data, and data encryption)	App privacy features (1-2) and privacy policy (3-8), with 1 point per question
Baumel et al, 2017 [26]	Data communications, storage, and sharing Notification of how personal information is kept confidential Protection of minors Anonymization	Eight items: 1 point if the app does not include the item
Bachiri et al, 2018 [27]	Privacy policy location and updates Access management: permissions, audit, criteria, and authentication Security measures Consideration of the Health Insurance Portability and Accountability Act (HIPAA)	Number of criteria that are met (35 items)

Reference	Criteria	Assessment of criteria
de las Aguas Robustillo Cortés et al, 2014 [28]	Data transmission and confidentiality Registration, purpose of use, information disclosure, and social disclosure Protection of minors and mechanisms to avoid unauthorized access Information storage	–1 (does not meet the criterion), 0 (not applicable), or 1 (meets the criterion)
Quevedo-Rodríguez and Wagner, 2019 [29]	Nature and purpose of the information and data storing Information about privacy, consent, and security measures User access Protection of minors	Compliance with items: 2 (complies), 1 (partially complies), or 0 (does not comply)
Knorr et al, 2015 [30]	Static and dynamic analyses and web connection Inspection of privacy policies	General compliance with the items
Zapata et al, 2014 [31]	Notification: privacy policy access and updates, cookies, and use of safety standards Security: authentication, encryption, server protection, and backup copies Election and access: access can be granted and revoked and access in case of emergency	Compliance with items: 2 (complies), 1 (partially complies), or 0 (does not comply)
Bondaronek et al, 2018 [32]	Privacy information: availability, accessibility, data collecting, data sharing, and data security	Number of apps that meet the different criteria
O'Laughlin et al, 2019 [33]	Privacy policy availability, existence of a log-in process, and identification Data storage and sharing User access: editing and deletion	Some of the items received a white, light-grey, or dark-grey score; other items received a white or light-grey score; 1 item received a white, light-grey, or black score
Adhikari et al, 2014 [34]	User registration and authentication Data storing and sharing Enable users to update, correct, and delete their data Data privacy and security measures and existence of privacy policy	Items 1-3: risk score (1 point if there is a risk); Items 4-8: safety score (1 point if it is safe)
Aliasgari et al, 2018 [35]	SSL configuration Data transfer and collection Compliance with the HIPAA	HIPAA compliance or not: the authors checked if the terms and conditions indicated HIPAA compliance, or they asked the app's support team
Mense et al, 2016 [36]	Encryption Data transmission	Number of apps that meet the different criteria
Powell el al, 2018 [37]	Privacy policy readability: word count, sentences per paragraph, words per sentence, characters per word, average number of sentences per 100 words, average words with 6 or more characters, Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Score, SMOG (Simple Measure of Gobbledygook) Index, Coleman Liau Index, Automated Readability Index, Fry Grade Level, and Raygor Estimate Graph Grade Level	Average score, median, or range for every item comparing diabetes apps vs mental health apps
Huckvale et al, 2015 [38]	Privacy policy: availability and features Concordance of privacy policies and data-handling practices Coverage of privacy policy: data collection, data transfer, anonymization, how long data are retained, use of cookies, user rights (ie, opt-out, consequences of not providing data, data access, and complaints), identification of data controller, and updates	Percentage of apps that meet the different criteria
Robillard et al, 2019 [39]	Collected information (ie, nature and types), use of information, and data sharing Reasons for disclosing information User rights: consent, opt-out, and deletion	Percentage of apps that meet the different criteria

Table 4. Items present in the assessment of criteria for each article.

Item	Reference																							
	[6]	[14]	[15]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]
Existence of a data controller	X		X				X																X	
DPO[a] details are given	X																							
Purposes of the processing are stated		X	X				X			X	X		X	X	X		X	X			X		X	X
Legal basis exists	X		X				X	X		X		X		X		X							X	X
Recipients of personal data are identified		X	X	X		X	X		X	X			X	X	X		X	X	X		X		X	X
International data transfers are disclosed	X					X																		
Data storage period is stated			X																				X	
Existence of users' data rights		X	X	X			X	X	X			X	X	X		X		X	X					X
Existence of the right to withdraw consent	X		X				X																	X
Existence of the right to complain to a supervisory authority			X																				X	
Obligation to provide data			X																				X	
Existence of data processing and profiling														X										
Nature of the collected information is disclosed		X			X	X	X				X		X	X										X

Item	Reference																								
	[6]	[14]	[15]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]
Risks of data collection and management of confidentiality breaches are stated										X	X													
Location of the collected information is disclosed		X		X	X		X												X					
User registration is required				X	X								X					X	X					
Existence of a privacy policy	X	X	X	X					X	X	X	X			X		X	X	X				X	
Privacy policy good practices	X		X					X	X			X			X	X	X					X	X	
Minimum data needed for app functioning are collected										X														
Protection of minors and age of verification exists		X	X								X		X	X										
Anonymization takes place		X	X								X	X					X							X

[a]DPO: data protection officer.

As seen in Table 4, many different items were considered as criteria to assess privacy. We have defined 21 items, but only four of them were taken into account by more than half the selected articles. The identification of the recipients of personal data was used as an evaluation criterion in 16 out of the 24 papers (67%) [14,15,19,21,22,24,25,28-30,32-34,36,38,39]. The existence of a privacy policy was determined by 13 out of 24 articles (54%) [6,14,15,19,24-27,30,32-34,38]. The stating of the purposes of the data processing was also examined by 13 papers (54%) [14,15,22,25,26,28-30,32,33,36,38,39]. Additionally, 13 articles (54%) [14,15,19,22-24,27-29,31,33,35,39] determined the existence of subjects' data rights, though only partially—most of them only considered access and/or data control by the user.

Table 4 also shows two different ways of assessing privacy. Out of 24 papers, 10 (42%) [6,14,15,21,24,30,32,36,38,39] checked whether the analyzed apps met the criteria described in the Evaluation Procedure section. Meanwhile, 14 articles out of 24 (58%) [19,20,22,23,25-29,31,33-35,37] evaluated the different apps according to several criteria.

Scoring Method

Of the 14 articles that assessed apps according to several items, 13 (93%) of them provided a scoring method that enables a comparison of privacy among apps. Only 1 paper (7%) [22] did not give a final score, although every item had an associated score; thus, a scoring method could easily be developed. The items were assessed in a binary manner in 6 out of the 14 papers (43%) [19,25-27,34,35], which produced a score. Out of 14 articles, 7 (50%) [14,20,22,23,28,29,33] used a binary assessment with intermediate values: 0, 0.5, or 1; 0, 1, or 2; or –1, 0, or 1 were used. Hutton et al utilized different discrete values depending on the assessed items [22]. Bondaronek et al used discrete values—white, light grey, dark grey, and black—to

obtain a final score of acceptable, unacceptable, or questionable [32].

Focusing on the articles that developed a scoring method, we have also analyzed whether the scoring was weighted. In that case, all the items would have different weights according to their importance when calculating the final score. Only 2 articles out of 24 (8%) [20,28] proposed a weighted score and 1 article (4%) [6] distinguished between "major issues" and "minor issues" but did not produce a final score. A summary is shown in Table 5.

Table 5. Scoring methods used to assess apps.

Reference	Score	Weighted score
Papageorgiou et al, 2018 [6]	No	No, though there are "major issues" and "minor issues"
Minen et al, 2018 [14]	No	N/A[a]
Huckvale et al, 2019 [15]	No	N/A
Scott et al, 2015 [19]	Yes. Risk score: 0-3; safety score: 0-6	No
Brüggemann et al, 2016 [20]	Yes. Connection security (S), information-sharing targets (T), unspecific information transfer (U), information collection (R), and log-in (L) are binary. Personal information type (P) is more elaborated: 13 types are considered and a correction factor is applied.	Yes, it can be configured by the user
Mense et al, 2016 [21]	No	N/A
Hutton et al, 2018 [22]	The paper does not give a score but, rather, explains how different heuristics are implemented. However, it is easy to assign a score to every app with the available information.	N/A, although it can be calculated (see Scoring Method section above)
Zapata et al, 2014 [23]	Yes: 0-6	No
Sunyaev et al, 2015 [24]	No	N/A
Leigh et al, 2017 [25]	Yes: 0-8	No
Baumel et al, 2017 [26]	Yes: 0-8, with 0 points being maximum privacy	No
Bachiri et al, 2018 [27]	Yes: 0-35	No
de las Aguas Robustillo Cortés et al, 2014 [28]	Yes, but it is a general app score, not only for privacy	Yes, weighted by experts
Quevedo-Rodríguez and Wagner, 2019 [29]	Yes, but as part of the global app quality	No
Knorr et al, 2015 [30]	No	N/A
Zapata et al, 2014 [31]	Yes	No
Bondaronek et al, 2018 [32]	No, at least for the privacy items	N/A
O'Laughlin et al, 2019 [33]	Yes: acceptable, unacceptable, or questionable	No
Adhikari et al, 2014 [34]	Yes. Risk score: 0-3; safety score: 0-5	No
Aliasgari et al, 2018 [35]	Yes. Although there is no global score, there are certain scores pertaining to Transport Layer Security (TLS) and Health Insurance Portability and Accountability Act (HIPAA) compliance.	No
Mense et al, 2016 [36]	No	N/A
Powell el al, 2018 [37]	Average score, median, and range for every item	No
Huckvale et al, 2015 [38]	No	N/A
Robillard et al, 2019 [39]	No	N/A

[a]N/A: not applicable.

Discussion

General Information

This review deals with the privacy assessment for mHealth apps. Finding information about the assessment of privacy of mHealth apps is not a trivial task, as the sources are very heterogeneous, including many areas of application. What is obvious is that the interest in privacy has been growing in the scientific community, with special significance in recent years. Despite studying the period from 2009 to 2019, the 24 selected articles were published in 2014 or later.

Privacy is essential in the health domain, and the app areas are very diverse. Fitness, mental health, and diabetes apps were common in the assessments, but such varied fields as HIV/AIDS, pregnancy, and headaches were considered. Some papers, such as Powell et al, evaluated seemingly unrelated areas, such as mental health and diabetes, at the same time [37]. The number of analyzed apps per paper also varied widely, from 10 apps [36] to 600 apps [24].

Evaluation Procedure

The articles presented in this scoping review evaluated privacy in different ways. Some of them analyzed only privacy, whereas others evaluated it together with security or other app functions.

Several of the articles used the privacy policy to determine information about the app privacy, but researchers should report more detailed information regarding how they assess the privacy of apps to ensure the reliability of their studies. As an example, it is not clear how so much information was obtained by analyzing only the app privacy policies in 3 papers (13%) [27,31,33]—perhaps an in-app information assessment was also performed. None of the articles explained how they evaluated privacy policies when considering certain items, such as informing the user about the secondary uses of their data. Some authors even noted that there were difficulties in evaluating privacy policies due to the complexity of the language used in them (eg, “Disagreements between the raters arose primarily from confusion over the apps’ privacy policies, which were often unclear in terms of language and intent” [22]), but none of them specified the exact criteria used to evaluate the content of the privacy policies. This could lead to inconsistent results if their assessment framework were to be used by others. Specifying the particular criteria used in the assessment could make the evaluations reproducible.

The legal framework is another important issue with privacy assessment. The number of mHealth apps has increased considerably [8], and important privacy regulations have emerged—not only in the mHealth domain—such as the GDPR [6,7]. However, only 7 out of the 24 articles (29%) used law as a direct source for establishing the assessment criteria—4 of them [6,22,25,38] used the European legislation (ie, the GDPR or the 1995 Data Protection Directive) as a source and 3 [23,27,35] were based on the HIPAA. Although some authors were skeptical about the applicability of the HIPAA to mHealth apps [6,40], others suggested that the HIPAA might be applicable [35]. If articles that used recommendations directly from private and/or public bodies, such as the US Federal Trade Commission or the UK Information Commissioner’s Office, are considered in this category, then the number of articles that contemplated laws goes up to 11 (46%). Additionally, data minimization is one of the main principles regarding processing personal data in the GDPR, meaning that data collection should be limited to processing purposes only. However, only 2 papers (8%) [20,38] analyzed the types of data collected by an app.

Several articles in our review also analyzed whether communications were secured, and 8 articles (33%) [6,15,20,21,30,35,36,38] actually checked if they were. Moreover, 1 article (4%) [15] brought to our attention that discrepancies between what the privacy policy states about app data transmission and the real data transmissions are not uncommon. By contrast, Huckvale et al did not observe any discrepancy [38]. Nevertheless, future analyses of privacy policies could verify whether developers properly disclose the nature of app communications.

Although the privacy policy is a common source of data to assess the privacy of apps, there are many challenges to address. The evaluation procedure needs to be straightforward by removing subjective and unclear assessments of privacy. It should also be supported by a legal framework, although that is not the current trend.

Evaluation Criteria

The criteria that have been used to assess the privacy of mHealth apps are very diverse. We have identified 21 items but, within each item, there are particularities that depend on the authors’ criteria. Moreover, as previously mentioned, in many cases, the criteria used to assess the items are not explained clearly enough, or they are not easily reproducible. Therefore, the list of different items and how they are evaluated never ends, and it is extremely subjective. Although the evaluations in this review are useful, we suggest a more objective privacy assessment.

As an example, some articles searched for specific information in the privacy policy, such as whether the user is informed about other uses of their data, whereas other papers looked for this information in the app. We consider that it is possible to miss important information by searching in the wrong place. For instance, 2 articles (8%) [22,38] checked both elements—the privacy policy and the app—while 6 papers (25%) [14,24,28,30,32,33] only checked the app, with no reference to the privacy policy. In 4 papers (17%) [25,26,29,34] it was not clear whether it was the app or the privacy policy that was examined. Finally, 3 papers (13%) [6,15,39] used only the app privacy policy and the terms and conditions.

One of the main issues created by the subjectivity of the evaluation criteria involves the nature of the items used. Sometimes the criteria are not clear enough. This issue may lead to different results when other users and/or developers assess privacy. New evaluation approaches should put special emphasis on defining clear and objective items to evaluate.

Scoring Method

A scoring method or scale to assess app privacy could be a key tool for systematically comparing apps. Many scoring methods were used in the included studies. Most of them are quite simple,

with a methodology that consists of assigning a binary value to some defined items, but they have, nonetheless, proven to be effective in assessing privacy by providing a simple approach to comparing apps. A weighted score, which highlights the importance of some items over others, was also explored in 2 papers (8%).

Despite the promising results derived from the use of a weighted score, further research must be conducted to identify the subjective relevance and importance of the different items perceived by consumers, patients, and experts, in order to assess the privacy of the apps. Further research must also be conducted aimed at defining common legal-based criteria to better assess the privacy of mHealth apps.

Review Limitations

This study has several limitations. Relevant studies may have been missed if they were published with a title or abstract in a language other than English, outside of the specified time frame, or in different databases than those that were used. Some studies may not be included due to the keywords chosen for the search string.

Specifically, for this review, the absence of an existing taxonomy of the privacy components used for the assessment is also a limitation. Although we attempt to compensate for this limitation with our level of expertise and detailed knowledge, charting is still subjective.

Finally, the different requirements implied by different types of apps shows that not all apps are equally sensitive to privacy risks, which suggests the possibility of analyzing how crucial privacy is according to the type of app. As we did not find any such existing classification system, we set this as a point for future research.

Conclusions

Privacy in mHealth apps has been determined based on an analysis of the app user interface, communications privacy, and privacy policy. Checking privacy in communications is usually very straightforward, with objective criteria for its assessment. When analyzing user interfaces and privacy policies, however, the criteria are very heterogeneous and less objective; this is especially true when analyzing privacy policies, which can lead to irreproducible results. In our opinion, it is very important to develop a more detailed assessment of privacy policies, so that the assessment frameworks may be utilized by subsequent users and lead to coherent results.

Another important conclusion from this study is that there is a lack of analyses pertaining to the types of personal information collected by the apps. Minimization is one of the principles of the GDPR, so a greater effort should be made to analyze whether apps gather more personal information than is necessary.

In short, despite great progress made through the scientific community's awareness about the importance of privacy assessment of mHealth apps, there is still a long way to go. A positive step forward would be the creation of a scale or scoring system based on objective criteria, which would, therefore, be less open to interpretation. Another good development would be the use of a certain legal basis for such a scale and explaining in detail how to apply the evaluation criteria.

Acknowledgments

This work was funded by the Cátedra Indra Sociedad Digital de la Universidad de Sevilla. EDZ receives funding and is supported by the V Plan Propio de Investigación de la Universidad de Sevilla, Spain.

Authors' Contributions

JB and JR led the scoping review, contributed to all phases of the study, and wrote the main body of the manuscript. ORR and AC participated in the definition of the review plan, monitored all phases of the review, participated in the decision making, and reviewed the manuscript. EDZ contributed to the data analysis and the discussion of the results and supported the manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews).
[DOCX File , 49 KB - mhealth_v8i7e18868_app1.docx]

Multimedia Appendix 2
Search terms and strategies for each database.
[DOCX File , 13 KB - mhealth_v8i7e18868_app2.docx]

Multimedia Appendix 3
Charting table used for collecting data from the included studies.
[DOCX File , 20 KB - mhealth_v8i7e18868_app3.docx]

Multimedia Appendix 4
Included studies.
[DOCX File , 16 KB - mhealth_v8i7e18868_app4.docx]

Multimedia Appendix 5
Objects of assessment of the apps.
[DOCX File , 24 KB - mhealth_v8i7e18868_app5.docx]

Multimedia Appendix 6
App data extraction details.
[DOCX File , 89 KB - mhealth_v8i7e18868_app6.docx]

References

1. Sokolovska A, Kocarev L. Integrating technical and legal concepts of privacy. IEEE Access 2018;6:26543-26557. [doi: 10.1109/access.2018.2836184]
2. Thorarensen B. The processing of health information: Protecting the individual right to privacy through effective legal remedies. Health Technol 2017 Mar 7;7(4):401-413. [doi: 10.1007/s12553-017-0184-4]
3. Abed Y, Chavan M. The challenges of institutional distance: Data privacy issues in cloud computing. Sci Technol Soc 2018 Dec 16;24(1):161-181. [doi: 10.1177/0971721818806088]
4. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating health data sharing: A solution based on the Internet of Things and distributed ledger technologies. J Med Internet Res 2019 Jun 06;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]
5. Regola N, Chawla NV. Storing and using health data in a virtual private cloud. J Med Internet Res 2013 Mar 13;15(3):e63 [FREE Full text] [doi: 10.2196/jmir.2076] [Medline: 23485880]
6. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: The alarming state of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
7. Demetzou K. Data Protection Impact Assessment: A tool for accountability and the unclarified concept of 'high risk' in the General Data Protection Regulation. Comput Law Secur Rev 2019 Nov;35(6):105342. [doi: 10.1016/j.clsr.2019.105342]
8. Kao C, Liebovitz DM. Consumer mobile health apps: Current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
9. mHealth App Economics 2017: Current Status and Future Trends in Mobile Health. 7th edition. Berlin, Germany: Research2Guidance; 2017 Nov. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2020-03-24]
10. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: Information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
11. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: Mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
12. Krebs P, Duncan DT. Health app use among US mobile phone owners: A national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
13. 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]
14. Minen MT, Stieglitz EJ, Sciortino R, Torous J. Privacy issues in smartphone applications: An analysis of headache/migraine applications. Headache 2018 Jul;58(7):1014-1027 [FREE Full text] [doi: 10.1111/head.13341] [Medline: 29974470]
15. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
16. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
18. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/BM.2012.031]
19. Scott K, Richards D, Adhikari R. A review and comparative analysis of security risks and safety measures of mobile health apps. Australas J Inf Syst 2015 Nov 22;19:1-18 [FREE Full text] [doi: 10.3127/ajis.v19i0.1210]

20. Brüggemann T, Hansen J, Dehling T, Sunyaev A. An information privacy risk index for mHealth apps. In: Schiffner S, Serna J, Ikonomou D, Rannenberg K, editors. Privacy Technologies and Policy. APF 2016. Lecture Notes in Computer Science, vol 9857. Cham, Switzerland: Springer; Sep 3, 2016.
21. Mense A, Steger S, Sulek M, Jukic-Sunaric D, Mészáros A. Analyzing privacy risks of mHealth applications. Stud Health Technol Inform 2016;221:41-45. [Medline: 27071873]
22. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: Heuristic evaluation approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
23. Zapata B, Hernández Niñirola A, Fernández-Alemán JL, Toval A. Assessing the privacy policies in mobile personal health records. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2014). New York, NY: IEEE; 2014 Presented at: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2014); August 26-30, 2014; Chicago, IL p. 4956-4959. [doi: 10.1109/EMBC.2014.6944736]
24. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
25. Leigh S, Ouyang J, Mimnagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov;20(4):e20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]
26. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
27. Bachiri M, Idri A, Fernández-Alemán JL, Toval A. Evaluating the privacy policies of mobile personal health records for pregnancy monitoring. J Med Syst 2018 Jun 29;42(8):144. [doi: 10.1007/s10916-018-1002-x] [Medline: 29959535]
28. de las Aguas Robustillo Cortés M, Cantudo Cuenca MR, Morillo Verdugo R, Calvo Cidoncha E. High quantity but limited quality in healthcare applications intended for HIV-infected patients. Telemed J E Health 2014 Aug;20(8):729-735. [doi: 10.1089/tmj.2013.0262] [Medline: 24849001]
29. Quevedo Rodríguez A, Wägner AM. Mobile phone applications for diabetes management: A systematic review. Endocrinol Diabetes Nutr 2019 May;66(5):330-337. [doi: 10.1016/j.endinu.2018.11.005] [Medline: 30745121]
30. Knorr K, Aspinall D, Wolters M. On the privacy, security and safety of blood pressure and diabetes apps. In: Proceedings of the IFIP (International Federation for Information Processing) International Conference on ICT (Information and Communications Technology) Systems Security and Privacy Protection. Cham, Switzerland: Springer; 2015 May 9 Presented at: IFIP (International Federation for Information Processing) International Conference on ICT (Information and Communications Technology) Systems Security and Privacy Protection; May 26-28, 2015; Hamburg, Germany p. 571-584. [doi: 10.1007/978-3-319-18467-8_38]
31. Cruz Zapata B, Hernández Niñirola A, Fernández-Alemán JL, Toval A. Privacy and security in mobile personal health records for Android and iOS [Article in Spanish]. RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação). 2014 Jun 1. URL: http://www.scielo.mec.pt/pdf/rist/n13/n13a04.pdf [accessed 2020-06-16]
32. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: Review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
33. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
34. Adhikari R, Richards D, Scott K. Security and privacy issues related to the use of mobile health apps. In: Proceedings of the 25th Australasian Conference on Information Systems. 2014 Dec Presented at: 25th Australasian Conference on Information Systems; December 8-10, 2014; Auckland, New Zealand p. 1-11 URL: https://openrepository.aut.ac.nz/bitstream/handle/10292/8117/acis20140_submission_12.pdf?sequence=1&isAllowed=y
35. Aliasgari M, Black M, Yadav N. Security vulnerabilities in mobile health applications. In: Proceedings of the IEEE Conference on Applications, Information & Network Security. 2018 Nov Presented at: IEEE Conference on Applications, Information and Network Security; November 21-22, 2018; Langkawi, Malaysia. [doi: 10.1109/ains.2018.8631464]
36. Mense A, Urbauer P, Sauermann S, Wahl H. Simulation environment for testing security and privacy of mobile health apps. In: Proceedings of the Symposium on Modeling and Simulation in Medicine. 2016 Apr 3 Presented at: Symposium on Modeling and Simulation in Medicine; April 3-6, 2016; Pasadena, CA p. 1-5. [doi: 10.22360/springsim.2016.msm.008]
37. Powell AC, Singh P, Torous J. The complexity of mental health app privacy policies: A potential barrier to privacy. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e158 [FREE Full text] [doi: 10.2196/mhealth.9871] [Medline: 30061090]
38. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: A cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]

39. Robillard JM, Feng TL, Sporn AB, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Sep;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
40. He D, Naveed M, Gunter CA, Nahrstedt K. Security concerns in Android mHealth apps. In: Proceedings of the American Medical Informatics Association 2014 Annual Symposium (AMIA 2014). 2014 Presented at: American Medical Informatics Association 2014 Annual Symposium (AMIA 2014); November 15-19, 2014; Washington, DC p. 645-654 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419898/pdf/1969977.pdf

Abbreviations

ACM: Association for Computing Machinery
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
IEEE: Institute of Electrical and Electronics Engineers
mHealth: mobile health
NHS: National Health Service
PRISMA ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by G Eysenbach; submitted 24.03.20; peer-reviewed by D Szinay, J Suleder; comments to author 06.04.20; revised version received 22.04.20; accepted 13.05.20; published 02.07.20.

Please cite as:
Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A
Privacy Assessment in Mobile Health Apps: Scoping Review
JMIR Mhealth Uhealth 2020;8(7):e18868
URL: https://mhealth.jmir.org/2020/7/e18868
doi:10.2196/18868
PMID:32459640

Original Paper

Assessment of the Fairness of Privacy Policies of Mobile Health Apps: Scale Development and Evaluation in Cancer Apps

Jaime Benjumea[1], MSc; Jorge Ropero[1], PhD; Octavio Rivera-Romero[1], PhD; Enrique Dorronzoro-Zubiete[1], PhD; Alejandro Carrasco[1], PhD

Department of Electronic Technology, Universidad de Sevilla, Sevilla, Spain

Corresponding Author:
Jaime Benjumea, MSc
Department of Electronic Technology
Universidad de Sevilla
Escuela Técnica Superior de Ingeniería Informática
Avda Reina Mercedes s/n
Sevilla
Spain
Phone: 34 6306 29719
Email: jaimebm@us.es

Abstract

Background: Cancer patients are increasingly using mobile health (mHealth) apps to take control of their health. Many studies have explored their efficiency, content, usability, and adherence; however, these apps have created a new set of privacy challenges, as they store personal and sensitive data.

Objective: The purpose of this study was to refine and evaluate a scale based on the General Data Protection Regulation and assess the fairness of privacy policies of mHealth apps.

Methods: Based on the experience gained from our previous work, we redefined some of the items and scores of our privacy scale. Using the new version of our scale, we conducted a case study in which we analyzed the privacy policies of cancer Android apps. A systematic search of cancer mobile apps was performed in the Spanish version of the Google Play website.

Results: The redefinition of certain items reduced discrepancies between reviewers. Thus, use of the scale was made easier, not only for the reviewers but also for any other potential users of our scale. Assessment of the privacy policies revealed that 29% (9/31) of the apps included in the study did not have a privacy policy, 32% (10/31) had a score over 50 out of a maximum of 100 points, and 39% (12/31) scored fewer than 50 points.

Conclusions: In this paper, we present a scale for the assessment of mHealth apps that is an improved version of our previous scale with adjusted scores. The results showed a lack of fairness in the mHealth app privacy policies that we examined, and the scale provides developers with a tool to evaluate their privacy policies.

(JMIR Mhealth Uhealth 2020;8(7):e17134) doi:10.2196/17134

KEYWORDS

privacy; mhealth apps; fairness assessment scale; cancer apps; GDPR

Introduction

Privacy in Mobile Health Apps

Health care systems are putting a great emphasis on the role of the patient and encouraging people to take control of their health [1]. Mobile health (mHealth) apps are one of the technological breakthroughs that make this possible. There are more than 3 billion smartphone users worldwide, and this number is predicted to grow by several 100 million in the next few years [2]. This proliferation of smartphones has led to an increase in the availability and abundance of mHealth apps. In 2017, there were more than 300,000 mHealth apps, and this number tends to grow by 25% every year. In 2018, 52% of smartphone users collected health-related information on their smartphones, and 60% of smartphone users downloaded health-related apps [3].

Among other uses, mHealth apps can provide disease and treatment information; practical tools for avoiding some diseases (prevention and healthy behavior promotion); tools to assist in the identification of symptoms (early detection); practical tools to deal with the medical, behavioral, or emotional aspects of a

specific disease (disease management); and access to peer or professional assistance (support) [4,5].

Despite the potential impact of mHealth apps on patient health, there is a lack of specific regulations and standards regarding the development of mHealth apps [6], which may result in potential risks and poor mHealth app quality. For example, some studies have reported problems with existing mHealth apps such as failure to meet the needs of persons with chronic conditions [7], a lack of cited source material or references [5], and insufficient testing of mHealth apps with respect to usability and validity [8]. Such setbacks reduce health care professional and patient confidence in these apps [9,10].

Criteria have been proposed to assess mHealth apps. Stoyanov et al [11] developed the Mobile Application Rating Scale (MARS) scale to classify and rate the quality of mHealth apps based on a literature review of app evaluations containing explicit quality rating criteria. Llorens-Vernet and Miró [6] recently proposed criteria to be integrated into a general standard for mHealth app development based on a systematic review, searches on professional organization websites, and standards governing the development of software for medical devices.

Privacy is a major concern for mHealth app users [12], as some mHealth apps require the collection, storage, and sharing of personal and sensitive patient data. Guidelines and recommendations for health-related apps—such as those developed by the Andalusian Agency for Healthcare Quality (Spain), Tecnologies de la Informació i la Comunicació Salut Social Foundation (Spain), National Health Service (United Kingdom), and European Commission—include several privacy items that highlight its importance in the context of mHealth. Also, privacy is one of the components included in the criteria proposed by Stoyanov et al [11] and by Llorens-Vernet and Miró [6].

Privacy assessment is a multifaceted issue that, among other things, bears upon privacy policies. In the context of mHealth apps, privacy policies contain privacy-related information that users can review prior to installation in order to get a clear idea of what personal data the app will access and the purposes for their processing. These concerns should be taken seriously, but 70% of the 600 most-used health-related apps do not include a privacy policy [13]. Although the previously mentioned guidelines and recommendations, MARS scale, and criteria include items regarding privacy, developers and evaluators need more specialized tools when it comes to the development and assessment of mHealth app privacy policies.

In this paper, we introduce a novel scale based on the General Data Protection Regulation (GDPR) to assess the fairness of mHealth app privacy policies. This scale provides detailed information regarding items to be included in privacy policies in order to comply with the GDPR. Consequently, we offer an objective and reproducible method for assessing the fairness of mHealth app privacy policies or developing the privacy policy of an mHealth app. This paper presents the final version of the scale (the culmination of an iterative development process) and the results obtained after applying it in a case study.

Legal Background

The GDPR is a regulation (2016/79) passed by the European Parliament and the Council of the European Union (EU). It was published in the Official Journal of the European Union [14] in 2016 and has been applicable since May 25, 2018. The GDPR applies to all EU member countries plus Iceland, Luxemburg, and Norway. Being a regulation (and not a directive), the GDPR applies directly to all these countries.

The GDPR introduces some important changes that replace previous legislation (Directive 95/46/EC). The first one can be found in Article 3 (territorial scope), as the GDPR applies to any controller or processor in the EU, even if processing does not take place in the EU. Also, the GDPR applies to any controller or processor (regardless the country of origin) if it is related to "the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union or the monitoring of their behavior as far as their behavior takes place within the Union" (Article 3.2). The controller must be able to demonstrate compliance with the GDPR (Article 5.2) and is subject to higher fines than before (Articles 66 and 83). Article 4 of the GDPR includes definitions that clarify relevant concepts. These concepts are summarized in Table 1.

Like many legal texts, the GDPR is difficult to understand and comply with, especially when it comes to app developers or users who are not legal experts; reading, interpreting, and understanding 99 articles in 88 pages of legal language is not easy. Our paper helps ordinary people become more familiar with the regulations so they can comply with the laws. We have developed a tool that makes compliance as easy as following simple guidelines such that even small app developers (eg, freelancers) can easily use it.

For example, there are two items in our scale (items 4 and 5) regarding the information to be provided to the data subject, described in Article 13 with a simple sentence: "the purposes of the processing for which the personal data are intended as well as the legal basis for the processing." This sentence is translated in the definition of our scale into approximately 250 words because this article is connected with several parts of the GDPR: recitals 39, 58, 60, 61, and 63, and articles 4, 5, and 6. These recitals and articles must be read and understood to be clear on the intentions of the GDPR.

Table 1. Definition of General Data Protection Regulation concepts.

Concept	Definition
Data subject	A natural person whose personal data are being processed; the GDPR[a] defines personal data not only as the data related to an identified person, but also as the data that can be used to identify, directly or indirectly, a natural person.
Data controller	"The natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data" [14].
Data processor	"The natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller" [14].
Recipient	"The natural or legal person, public authority, agency, or another body, to which the personal data are disclosed" [14].
Representative	A natural or legal person established in the EU[b]; a representative must be designated by data controllers or processors not in the EU (Article 27).
DPO[c]	A person who must be designated by the controller or processor in certain circumstances (see Article 37 for more details); the duties of the DPO[c] are defined in Article 39; they include, among others, advising the controller or processor about their duties related to the GDPR and monitoring compliance with GDPR

[a]GDPR: General Data Protection Regulation.

[b]EU: European Union.

[c]DPO: data protection officer.

Research Background

There are several studies in the literature that have addressed the availability of privacy policies in mobile apps and the assessment of some aspects of their quality [9,15-37].

Heuristics were proposed by Hutton et al [15] to assess privacy in mHealth apps for self-tracking. In their study, Hutton et al [15] used recommendations from the US Federal Trade Commission and GDPR to define the items used in the heuristics. The evaluation of an app consists of checking its behavior and the contents of its privacy policy. They also proposed a scoring method. However, their system is not exclusively based on the GDPR and evaluates other issues that are not related to the content of the privacy policy. For example, some items in the heuristics are closely related to usability.

In 2013, Sunyaev et al [16] assessed the availability and quality of the privacy policies of the 600 most commonly used mHealth apps. They found that only 183 apps out of 600 had a privacy policy. They also determined the lengths of privacy policies, their readability, and whether privacy policies were focused on the app. Moreover, they checked if the contents of privacy policies addressed aspects users considered to be the most important. They did not develop a method or use any legal framework or regulation to design their assessment scheme.

A framework for assessing apps related to chronic insomnia disorder was defined by Leigh et al [17]. The framework was based on 24 criteria, and 6 of them dealt with privacy policies. The framework defines 6 questions about the content of the privacy policy that must be answered yes or no based on the UK Data Protection Act 1998 (and thus, Directive 95/46/EC), the UK Information Commissioner Office, and the Charter of Fundamental Rights of the EU. The authors used the answers to these 6 questions to obtain a privacy score. Blood pressure and diabetes apps were assessed by Knorr et al [18] in a study that rated 154 apps using static and dynamic analysis and evaluating web server connections and privacy policies. They assessed 12 aspects of the privacy policies including recommendations from the Organization for Economic Cooperation and Development on privacy. They discovered that 67% of the apps that stored data on the web had a privacy policy. However, they neither developed a score to evaluate privacy policies nor did they compare them.

A scoring method to evaluate the quality of 116 apps for depression was defined by O'Loughlin et al [19] using 7 questions about the privacy policy and app behavior, such as whether the app requires a personal identification number for access or the privacy policy states if data are encrypted or stored locally. Based on the results of this questionnaire, they classified privacy policies as acceptable, questionable, or unacceptable, but they did not assign a numerical score to the app, and it is not GDPR-based. They also found that only 57 out of the 116 apps they evaluated had a privacy policy.

Huckvale et al [9] assessed 79 mHealth apps, certified as clinically safe and trustworthy by the United Kingdom National Health Service Health Apps Library. They evaluated the app behavior, considering aspects like data transmission, storage, and privacy policies. When assessing privacy policies, they used a coding method based on the UK Information Commissioner Office recommendations and the UK Data Protection Act. Their proposed scheme used four domains (uses of data, technical concerns, user rights, and administrative details) to classify the topics analyzed in the privacy policy. Each of the 24 topics was classified as addressed or absent in the privacy policy. Although they discussed the percentage of apps complying with each topic, they did not develop a scoring method.

Papageorgiou et al [20] conducted a privacy and security analysis of 20 mHealth apps. They had a broad scope of analysis including static and dynamic analysis of the apps, permission analysis, and security in communications. They also investigated whether privacy policies complied with the GDPR, focusing on the right to withdraw consent, the right to portability, data protection officer (DPO) contact information, profiling, and transfers of personal data to countries not within the EU. They

discussed how many apps complied with these items, but they also did not develop a scoring method.

A total of 29 apps were analyzed by Minen et al [21], focusing on data storage and privacy policies in headache apps. When analyzing privacy policies, they searched for the presence or absence of information regarding data collection, data sharing, use by children, and certain user rights.

The scientific community has been searching for a way to assess privacy in mHealth apps for the last few years. Many studies have assessed privacy in apps, usually focusing on their user interfaces, privacy in communications, and privacy policies. However, the established criteria used to analyze privacy policies are heterogeneous and subjective. These solutions are based on the researchers' own experience, the literature, and/or an existing legal framework. The items that are considered for the assessments are very diverse, and the evaluation of these items are, on many occasions, very subjective to the evaluators' criteria.

It is necessary, therefore, to create tools to evaluate privacy policies and establish privacy scales according to objective criteria that are less open to interpretation. Although some papers considered the GDPR, none of them proposed a set of items that enable GDPR compliance. Our aim is to fill that research gap by proposing of a GDPR-based scale to assess privacy policies in mHealth apps.

Methods

Privacy Scale Design

Article 13 of the GDPR summarizes the information that must be given to the user (known as data subject in the GDPR) when the information is collected from them (Table 1). This information is usually delivered to the user via a document called a privacy policy, and in order to be compliant with the GDPR, the privacy policy must meet certain requirements.

In our previous work [38], we developed a scale to assess the fairness of the privacy policies of mHealth apps. The objective of our scale was to analyze privacy policies in a systematic way and design a GDPR-based system to assess and improve such policies. Based on Article 13 of the GDPR and the recommendations of the National Data Protection Authority in Spain [39,40], we identified and summarized the information that should be provided in privacy policies (see Table 2 for a list of the items). We also defined a scoring method to assess each item.

Our scale is not intended to check strict compliance with the GDPR but to assess the fairness of privacy policies. This means that a privacy policy with fewer than 100 points (the maximum score) may be compliant, in a strict way, with the GDPR. This is because some items mentioned in Article 13 (for example, the identity of the data controller) must be analyzed carefully. It is not always easy to classify an item as yes or no in terms of compliance with the GDPR. Also, the GDPR allows some items to be omitted in certain cases, such as transfer to non-EU countries if the personal data are stored in the EU. However, our proposed score penalizes the absence of such information even if the data remain in the EU. In fact, we are just applying one of the principles of the GDPR (Article 5)—"lawfulness, fairness, and transparency"—to privacy policies, as they should go slightly further in their contents.

Indeed, the items defined in our scale may be used by developers as a checklist to design privacy policies that comply with the GDPR. They also could be used by data controllers to check if their apps are GDPR-compliant. Furthermore, a privacy policy scoring 100 points would be fully compliant with the GDPR. Using the proposed scoring method, we consider privacy policies with scores from 75.0 to 100 points as very fair (category 1 [Cat1]). A score from 50.0 to 74.9 is somewhat fair (category 2 [Cat2]). A score from 25.0 to 49.9 is somewhat unfair (category 3 [Cat3]). Finally, we consider a score from 0 to 24.9 as very unfair (category 4 [Cat4]).

Some discrepancies in the interpretation of how to assign a score to some items were found in the first iteration of the scale design process. In the first iteration, the privacy policies of 9 apps were analyzed, and we obtained Kappa-Cohen indexes for each item. Possible scores for each item and Kappa-Cohen indexes are shown in Table 3.

Table 2. Items in the privacy policy (Article 13).

Item	Item number
Identity of data controller	1
Identity of the representative	2
Data protection officer details	3
Purposes for the processing	4
Legal basis for the processing	5
Legitimate interests from controller	6
Recipients (or categories) of the personal data	7
Transfers to non–European Union countries	8
Period for which data will be stored	9
Existence of data subject's rights	10
Existence of right to withdraw consent	11
Right to lodge a complaint with a supervisory authority	12
Obligation to provide personal data	13
Existence of automated decision making or profiling	14

Table 3. Kappa-Cohen indexes for privacy policy items for the 9 apps evaluated in the first iteration.

Item	Item number	Score	Kappa-Cohen index (n=9)
Identity of data controller	1	0: no info; 0.5: partial; 1: full	0.77
Identity of the representative	2	0: no info; 1: info provided; N/A: not applicable	1
Data protection officer details	3	0: no info; 1: info provided	0.61
Purposes for the processing	4	0: no info; 0.5: generic; 1: specific	0.77
Legal basis for the processing	5	0: no info; 1: info provided	0.77
Legitimate interests from controller	6	0: no info; 1: info provided; N/A: not applicable	0.8
Recipients (or categories) of the personal data	7	0: no info; 1: info provided	–0.13
International transfers of data	8	0: no info; 0.5: generic; 1: full details or no international transfers	0.53
Period for which data will be stored	9	0: no info; 0.5: generic; 1: specific	0.66
Existence of data subject's rights	10	0: no info; 0.5: generic; 1: full	0.49
Existence of right to withdraw consent	11	0: no info; 1: info provided; N/A: not applicable	0.08
Right to lodge a complaint with a supervisory authority	12	0: no info; 0.5: generic; 1: specific	0.77
Obligation to provide personal data	13	0: no info; 1: info provided	–0.17
Existence of automated decision making or profiling	14	0: no info; 0.5: generic; 1: specific or no profiling or automated decision making done	0.17

A refinement of the criteria used to assign those scores was performed to resolve the discrepancies. Also, an error in the definition and description of one of the items was corrected. Items and their possible scores are (re)defined as follows:

- Identity of data controller: 1 point if full information is given. Full information means name, postal address, and electronic address (both email and a contact form are considered valid) of the data controller; 0.5 points if some information is missing; 0 points if the information is omitted. If only an electronic address is provided, the score is 0 points. Also, if the street address is not mentioned, the score is 0 points.
- Identity of the representative: The representative is a natural or legal person, established in the EU, who must be designated by the data controller if they are not in the EU. In this case, 1 point is given if full information is given (in the same way as with the data controller) and 0 points otherwise.
- DPO details: The GDPR states that a DPO must be designated if the controller processes a large quantity of data in some special categories, such as health data. We

assume that a DPO must exist in any given mHealth app. At least an email address must be given to get 1 point. In order to be consistent with the definition of DPO in the GDPR, the DPO must be a different person from the data controller, so the email address should also be different. Otherwise, 0 points are given.

- Purposes for the processing: The purposes for the processing must be stated explicitly in the privacy policy. Sometimes the information given is too general. For example, "We collect this information for the purpose of providing our service" does not give any detail about why the data controller needs the personal data. In this case, the score for this item is 0.5 points. If purposes are provided explicitly, 1 point is given. If purposes are not mentioned, 0 points.
- Legal basis for the processing: There are six legal bases for the processing (Article 6 of GDPR): consent, need to perform a contract, legal obligation, protect vital interest of somebody, public interest or exercise official authority, and legitimate interest. This information must be given in the privacy policy in order to get 1 point. However, we found that, in some cases, the legal basis in which the processing was founded was not explicitly stated in a separate paragraph. This is because, sometimes, this information is embedded in the declaration of the purposes for the processing. As the information is, after all, given to the data subject, we decided to give 1 point in these cases. Furthermore, if 0 points are given to this indicator, indicators based on the legal basis (ie, items 6 and 11) are considered N/A (not applicable).
- Legitimate interests from controller: 1 point if this information is given and 0 points otherwise. N/A if legitimate interest is not stated as a legal basis for the processing or if item 5 is 0 points.
- Recipients (or categories of recipients) of the personal data: 1 point if this information is given and 0 points otherwise. We must note that if there are no recipients, this must be explicitly stated. Also, we must keep in mind that in accordance with Article 4 of the GDPR a data processor is considered a recipient.
- Transfers to non-EU countries: This item refers to the fact that personal data may be transferred to a country not in the EU. In this case, the data controller must give enough information about the measures that are in place to achieve a similar level of protection. We give 1 point if privacy policy indicates that this transfer is in place, and there is a reference to the measures taken. We consider enough information to contain a mention to the compliance with Privacy Shield [14] or similar frameworks. We also give 1 point if the transfer is based on an adequacy decision from the Commission. If there is a mention to a transfer to non-EU countries without further information, this item is 0.5 points. The GDPR states that this information must be given if there are transfers to non-EU countries and says nothing if the data are stored within the EU. We consider that this information must be given even where there are no transfers to non-EU countries. Thus, if there is no information about transfers outside the EU, this item is 0 points. When the data controller is in the EU, the fact of not transferring data outside the EU must be explicitly stated. If not, the item is 0 points. Otherwise, the item is N/A.
- Period for which data will be stored: To obtain 1 point, the privacy policy must point out a specific time in the future when the data will be erased. We consider the following time references as valid: a period of inactivity in the data subject's account or a specific reference to a user request to erase the data. The latter is independent of the specification of data subject's rights in the privacy policy.
- Existence of data subject's rights: 1 point is given if the specific user's rights mentioned in Article 13 (right to access, rectification, erasure, restriction of processing, object of processing, and data portability) are enumerated in the privacy policy with a way to exercise these rights. This information may also be provided using a link. We award 1 point if 5 or more rights are mentioned and 0.5 points if partial information is given (for example, some rights are omitted or there is no indication on how to exercise the rights); 0 points if there is no reference to data subject's rights.
- Existence of the right to withdraw consent: This is an additional right that only exists if the legal basis for the processing is consent. As such, the score is 1 point if this right is mentioned (along with the way to exercise it) and 0 points if the right is omitted. This item is N/A if consent is not one of the legal bases for the processing or if item 5 is 0 points.
- Right to lodge a complaint with a supervisory authority: According to the GDPR, there is an obligation to inform users that they have the right to lodge a complaint with a supervisory authority if they believe that their rights have been violated. To score 1 point, the privacy policy must not only identify the appropriate supervisory authority but also provide at least a link to it. Moreover, 1 point is given if the policy links to the list of all supervisory authorities within the EU. Simply naming the supervisory authority is considered insufficient and earns a score of 0.5 points and 0 points otherwise.
- Obligation to provide personal data: The privacy policy must explicitly state what happens if the user does not provide certain personal data. Examples of good practices are the following: "if you choose not to provide data, we may not be able to provide you those services" and "In order to join [...] you must provide [...]." If this information is given, 1 point. Otherwise, 0 points.
- Existence of automated decision making or profiling: We consider that there must be a reference to the existence or absence of automated decision making or profiling based on personal data. This item may be tricky to interpret because sometimes apps make automated decisions such as defining user interface language based on personal data. This example probably does not fit within the reasoning of the GDPR. Thus, we consider as automated decision making or profiling a behavior that goes beyond simple decisions made by the app. This item scores 1 point if this information is shown in the privacy policy and enough information about the logic around this decision or profiling is given. We consider a link to the information to be valid. Therefore,

0.5 points are scored if there is a reference to this item with no additional information, and the score is 0 points if no information is given. Simply using Google Analytics is not sufficient to consider that the app to be profiling. However, the use of cookies, if they modify the app behavior, is considered profiling. We are also aware that, in accordance with the National Data Protection Authority in Spain on mobile apps [41], some information must be given if the app includes targeted advertisements.

As stated in the introduction, an important objective of our research is to provide a reproducible scale for the assessment of privacy policies in mHealth apps, so schematics and further descriptions of the items with explanatory examples for the scoring system are defined in Multimedia Appendix 1.

The proposed scale consists of 14 items to check when assessing a privacy policy, as shown in Table 1. Each item is assigned a score of 0 points or 1 point, though some of them may score 0.5 points. Thus, every item has the same weight within the total score. Also, as some items may be N/A, the final score is expressed as a percentual score. Thus, if an app achieves 7 points when privacy items are assessed but only 12 items are applicable, its final score is 58.3 points. Like the scales previously defined in the literature, our scale must be simple and easy to apply. Our scoring method is in concordance with other mHealth app privacy scales [15,27,32], which are also composed of yes/no questions. Most of the privacy scales that defined a score were based on items that could have 2 or 3 values.

Items 1, 4, 8, 9, 10, 12, and 14 can score 0.5 points, as some items are more complex than others. For example, the item purposes for the processing requires that the data controller be very explicit when defining the app's purposes for the processing. We found that some data controllers state the purposes for the processing but are not as explicit as they should be. In these cases, the item scores 0.5 points. Other items, such as item 3 (DPO details) are so simple that they can only score 1 point (yes) or 0 points (no). Additional details and examples can be found in Multimedia Appendix 1.

We have also added new indicators that did not appear in our previous work. They do not directly influence scores, but they add some information that is relevant to the study of privacy policies:

- Date of last update of the privacy policy: We look for this information in the text of the privacy policy itself. We do not consider alternative ways of obtaining the last update date such as analyzing http last-modified headers [42].
- Data controller's country: We collect this from the identity of the data controller. We do not consider additional information such as any obtained by the WHOIS tool [43]. This tool provides information about the domain name, but it might be misleading.
- GDPR awareness: This indicator only gets a yes/no value indicating whether GDPR is explicitly mentioned in the privacy policy.

Case Study: Cancer Apps

Study Design

A systematic search strategy was followed to identify all relevant mHealth apps for the most common types of cancer (breast, prostate, colorectal, and lung cancer) and for cancer in general. We focused on Android apps due to its market dominance, being the most installed operating system among the new smartphones shipped worldwide from 2017 to 2019 [44]. Two researchers (ORR and EDZ) searched the Spanish version of the Google Play website, taking steps to ensure that no previous searches or cookies influenced the results. Five searches were completed on July 25, 2019, using "cancer mama," "cancer prostata," "cancer," "cancer colon recto," and "cancer pulmon" as search strings.

Selection Criteria

Apps were included in the screening stage if their title or description contained one of the search strings defined. After duplicates were removed, two researchers (ORR and EDZ) reviewed and assessed the title and description of the resulting mHealth apps for eligibility against the selection criteria. Apps whose titles and/or descriptions met the selection criteria were downloaded and installed. A researcher (ORR) checked that they worked properly and met the selection criteria. Disagreements were resolved by consensus.

The following inclusion criteria were used: the title or description referred to at least one of the search strings, it was intended exclusively for cancer patients or survivors, and the app collected user data or allowed users to share their opinions or data.

Apps were excluded if they met at least one of the following conditions: the title or description was written neither in English nor Spanish, the user interface was available neither in English nor Spanish, the privacy policy was written neither in English nor Spanish, it was not focused on cancer, it was intended for people other than cancer patients or survivors, or it was not free.

Data Extraction

The following descriptive characteristics of apps meeting the selection criteria were collected from the Google Play website when available: developer, category, number of ratings, user rating, last update, and number of downloads. Additionally, using the information included in the description, two researchers (ORR and EDZ) independently classified the apps according to main purpose and type of cancer. Discrepancies were resolved by consensus.

URLs linking to the privacy policy of the included apps were also collected when available. If no link was provided, we tried to find the privacy policy on the developer's website. A researcher (JR) reviewed all installed apps and checked if they contained any additional privacy policy information. If the privacy policy contained in the app was different from the one linked in the Google Play website, the former was considered for the assessment. When user registration was required, the privacy policy had to be available before registering for the service. Otherwise, we considered the app to not have a privacy policy, as it was not accessible before using the app. Two

researchers (JB and JR) reviewed and independently assessed all privacy policies of the included apps using the proposed scale. Finally, a score for each privacy policy was assigned according to the scoring scale defined earlier. Discrepancies were resolved by consensus.

Classification

Included apps were classified according to main purpose and type of cancer. We used the classification scheme for an app's main purpose proposed in Giunti et al [7]. We removed the awareness-raising option because the apps designed for this purpose did not meet the selection criteria and were excluded from our study. Therefore, we coded an app's main purpose as disease and treatment information (DTI), disease management (DM), or support (S).

Finally, type of cancer was coded as general, colorectal, breast, prostate, lung, or other. General was used to code included apps that pertained to cancer in general without identifying any specific type. Other was assigned to apps that pertained to a specific type of cancer other than colorectal, breast, prostate, or lung cancer.

Results

App Selection and Extracted Features

Google Play searches resulted in 1249 mHealth apps. After duplicates were removed, 831 mHealth apps were assessed for eligibility; 41 of those apps met the selection criteria and were downloaded and installed on an Android smartphone (Moto G7, Motorola Mobility, LLC) to check if they worked properly. Finally, 31 mHealth apps met the selection criteria and were included in the analysis. Figure 1 shows the flow diagram of the described procedure, while Table 4 shows the selected apps considering the selection criteria. For convenience when analyzing the results, apps were tagged from App1 to App31. For the statistical analysis below, five more features were added to Table 4: Google Play app ratings, number of reviews, number of downloads, app type, and cancer type. Multimedia Appendix 2 contains a list of the apps that were found in Google Play search and those included in the case study.

We applied the second iteration of our privacy policy assessment scale (described in the Methods section) to the 31 selected apps. In the case of App15, a privacy policy link led to a generic privacy policy for the company, so we considered the one in the app. Table 5 shows the privacy scores obtained after assessing all the privacy policies. We also added more information about the apps to the table in order to achieve a more complete analysis of the results. In particular, we added the data controller's location, the last app update, and whether the privacy policy was GDPR-aware (ie, it mentioned the GDPR). If the app did not have a privacy policy, it was assigned a score of 0 points.

Figure 1. Flow diagram.

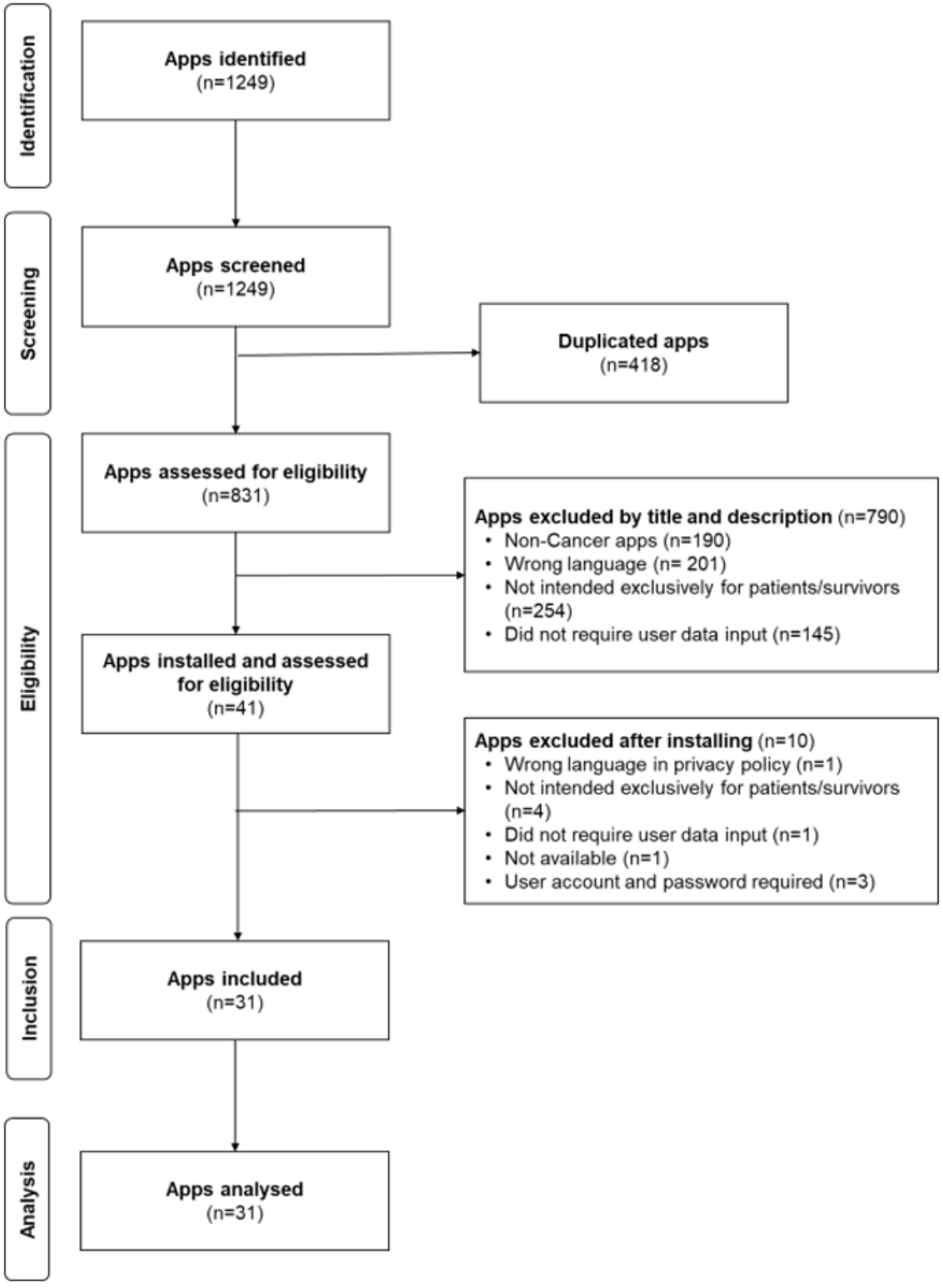


Table 4. Selected apps.

App name	Developer	Rating (stars)	# Ratings	# Downloads	App type	Cancer type	Label
BECCA: Breast Cancer Support	Breast Cancer Care	4.5	63	10,000+	S[a]	Breast	App1
EmotionSpace cáncer de mama	Pfizer Inc	2.5	2	100+	S	Breast	App2
ChemoWave: For Cancer Patients	Treatment Technologies & Insights	4.4	20	1000+	DM[b]	General	App3
OWise Breast Cancer	Px HealthCare BV	4.4	10	1000+	DM	Breast	App4
My Cancer Coach	Genomic Health Inc	4.5	86	10,000+	DM	General	App5
Breast Advocate	Toliman Health	5	1	100+	DTI[c]	Breast	App6
Breast Cancer Support	MyHealthTeams	4.1	47	1000+	S	Breast	App7
KMBCN	Kepharge	5	1	10+	DTI	Breast	App8
Triple Negative Breast Cancer	Kognito	5	2	100+	DTI	Breast	App9
Breast Cancer: Others Like Me	Eli Malki	0	0	5+	S	Breast	App10
Outcomes4Me	Outcomes4Me Inc	5	5	100+	DTI	Breast	App11
Boobytrapp: The Breast Cancer App	Boobytrapp	3.7	3	100+	S	Breast	App12
The BAPS App Wales	The Orchard Media & Events Group Ltd	0	0	100+	DM	Breast	App13
BELONG Beating Cancer Together	BelongTail	4.7	1,151	100,000+	DM	General	App14
Diana	F Hoffmann–La Roche	5	7	1000+	DM	Breast	App15
Got Boobs?	Got Boobs	0	0	100+	S	Breast	App16
inKind Space	PixelEdge	0	0	10+	S	Breast	App17
Cancer Surveillance	GoMLV	3.7	21	1000+	DM	General	App18
Focalyx	Lyx Health	4.8	6	50+	DM	Prostate	App19
Adrenal Cancer: Others Like Me	Eli Malki	5	6	1000+	S	Other	App20
How Are You Today? PC	Intelesant	0	0	100+	DM	Prostate	App21
Cancer.Net Mobile	American Society of Clinical Oncology	4.2	227	10,000+	DM	General	App22
TNM Cancer Staging	International Atomic Energy Agency	4.6	323	10,000+	DTI	General	App23
Untire: Beating cancer fatigue	Tired of Cancer BV	4.5	60	5000+	DM	General	App24
Self-Care During Cancer	NearSpace Inc	4.7	6	1000+	S	General	App25
CanDi: Cancer Diet App	Faculty of Health Sciences UniSZA	4.7	60	500+	DM	General	App26
CancerAid	CancerAid PTY LTD	3.7	25	1000+	DM	General	App27
GRYT Health Cancer Community	GRYT Health	3.9	7	100+	S	General	App28
Target Ovarian Cancer Symptoms Diary	Brandwave Marketing	3.6	8	1000+	DM	Other	App29
Pancreatic Cancer Action: Symptom Tracker	Healthbit Ltd	5	3	100+	DM	Other	App30
My Care Plan (cancer survivor)	NearSpace Inc	4	4	1000+	DM	General	App31

[a]S: support.

[b]DM: disease management.

[c]DTI: disease and treatment information.

Table 5. Privacy scores.

App name	Label	Data controller's location	Last update	GDPR[a] aware	Score
BECCA: Breast Cancer Support	App1	UK[b]	03/2019	No	76.9
EmotionSpace cáncer de mama	App2	Germany	05/2018	No	75
ChemoWave: For Cancer Patients	App3	US[c]	10/2018	No	53.6
OWise Breast Cancer	App4	UK	N/A[d]	Yes	31.8
My Cancer Coach	App5	US	02/2015	No	23.1
Breast Advocate	App6	Unknown	No privacy policy	N/A	0
Breast Cancer Support	App7	US	09/2019	Yes	78.6
KMBCN	App8	Unknown	No privacy policy	N/A	0
Triple Negative Breast Cancer	App9	US	02/2019	No	34.6
Breast Cancer: Others Like Me	App10	Unknown	No privacy policy	N/A	0
Outcomes4Me	App11	Unknown	11/2018	No	34.6
Boobytrapp: The Breast Cancer App	App12	Singapore	06/2018	No	29.2
The BAPS App Wales	App13	UK	N/A	Yes	69.2
BELONG Beating Cancer Together	App14	Israel	09/2018	Yes	75
Diana	App15	Spain	10/2018	No	40.9
Got Boobs?	App16	US	10/2018	No	26.9
inKind Space	App17	US	N/A	No	25
Cancer Surveillance	App18	Unknown	N/A	No	15
Focalyx	App19	Unknown	No privacy policy	N/A	0
Adrenal Cancer: Others Like Me	App20	Unknown	No privacy policy	N/A	0
How Are You Today? PC	App21	Unknown	No privacy policy	N/A	0
Cancer.Net Mobile	App22	US	07/2019	Yes	50
TNM Cancer Staging	App23	Unknown	No privacy policy	N/A	0
Untire: Beating Cancer Fatigue	App24	Netherlands	N/A	Yes	66.7
Self-Care During Cancer	App25	US	03/2014	No	29.2
CanDi: Cancer Diet App	App26	Unknown	No privacy policy	N/A	0
CancerAid	App27	Australia	N/A	No	42.9
GRYT Health Cancer Community	App28	US	12/2018	No	46.2
Target Ovarian Cancer Symptoms Diary	App29	UK	04/2018	Yes	80.8
Pancreatic Cancer Action: Symptom Tracker	App30	UK	06/2018	Yes	75
My Care Plan (cancer survivor)	App31	Unknown	No privacy policy	N/A	0

[a]GDPR: General Data Protection Regulation.

[b]UK: United Kingdom.

[c]US: United States.

[d]N/A: not applicable.

Assessment of Privacy Policies

We assessed the fairness of the privacy policies of 31 cancer apps. Surprisingly, as seen in Figure 2A, 29% (9/31) of the included apps did not have a privacy policy. Thus, we first determined the presence/absence of a privacy policy according to different features in the apps. We considered the app type, the type of cancer, and the number of downloads. A summary of the analysis is shown in Figure 2. Second, we analyzed privacy policies according to the obtained score. As specified in the Methods section, Cat1 (75.0 to 100 points), Cat2 (50.0 to 74.9 points), Cat3 (25.0 to 49.9 points), and Cat4 (0 to 24.9 points). Only 19% (6/31) of apps had a Cat1 privacy policy, while 4 apps belonged to Cat2. Thus, only 32% (10/31) of apps scored above 50.0 points in our GDPR-based privacy policy assessment. The results are shown in Figure 3A, where NPP means no privacy policy. We also analyzed the fairness of

privacy policies according to different features in the apps, considering the app type, type of cancer, and number of downloads. When assessing fairness, we also considered the data controller's country, last privacy policy update, and if the privacy policy was GDPR-aware. A graphic summary of privacy scores can be seen in Figure 3 and Figure 4.

Privacy policy fairness of 4 included apps had been previously analyzed in Benjumea et al [38]: App1, App4, App7, and App12. Despite the minor modifications in the scale for the second iteration, App4, App7, and App12 did not change their scores (even though App7 had recently updated its privacy policy). App7's score was one of the highest, with 78.6 points. More remarkable is the case of App1: with a recent update, its score went from 57.7 to 76.9 points. This indicates a significant effort to follow GDPR guidelines.

Figure 2. Analysis of privacy policy presence.

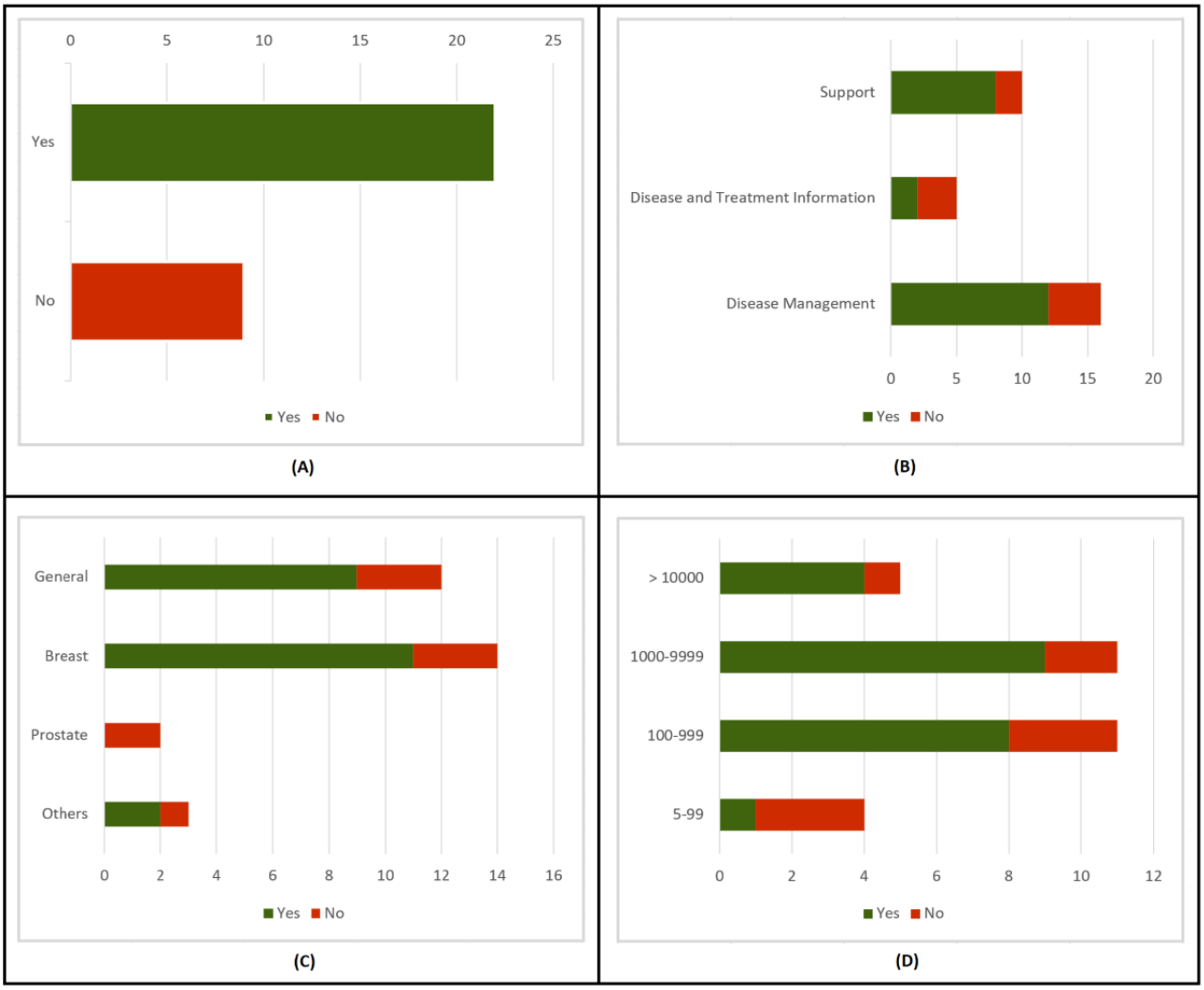


Figure 3. Privacy score summary (part1).

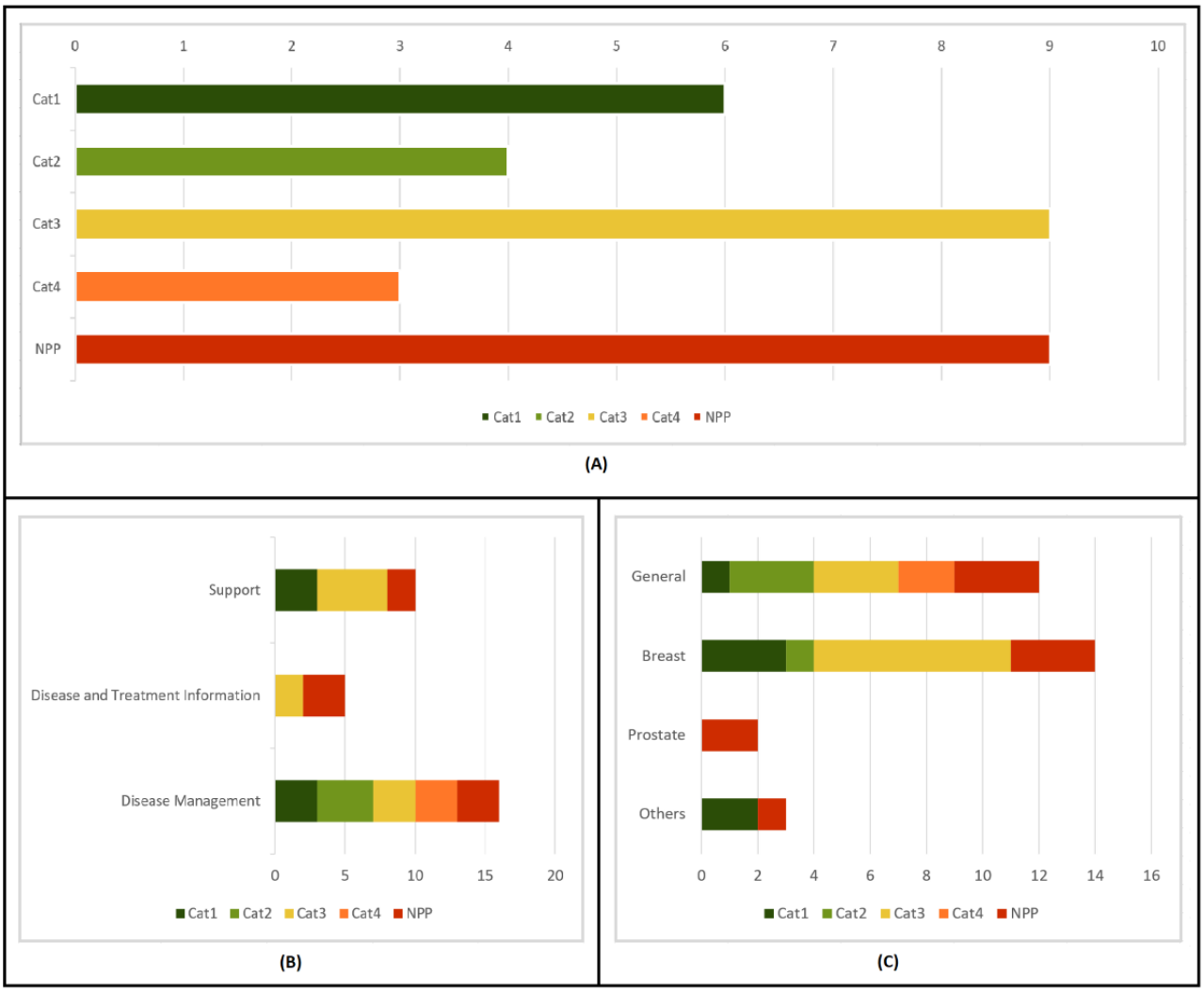
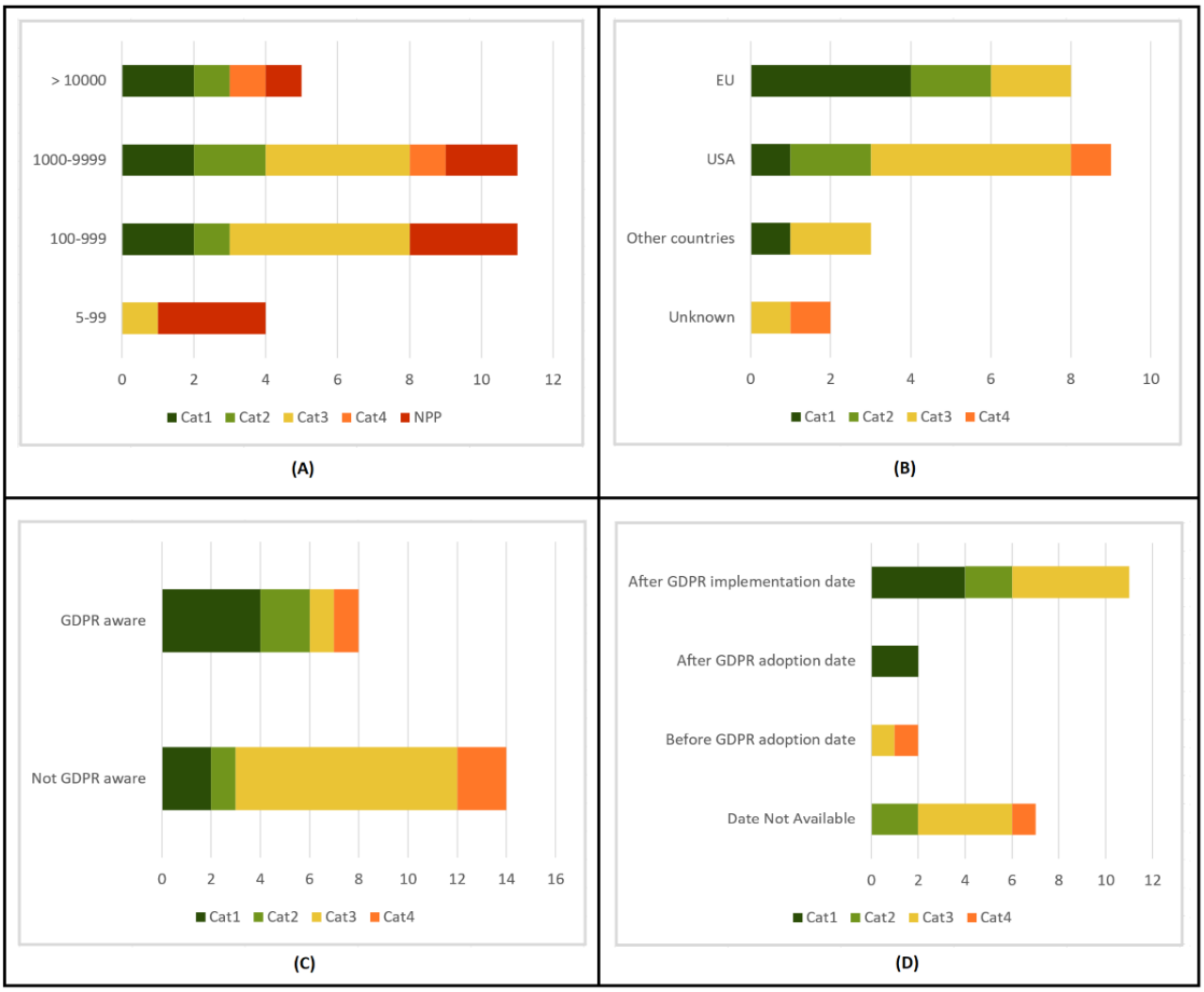


Figure 4. Privacy score summary (part2).



Assessment of Privacy Policies by App Type

We analyzed the presence of privacy policies according to app type and found that 75% (12/16) of disease management apps and 80% (8/10) of support apps had a privacy policy. As seen in Figure 2B, only 40% (2/5) of disease treatment and information apps had a privacy policy. Regarding the score, 44% (7/16) of disease management apps were above 50 points. Only 30% (3/10) of support apps were above 50 points, but all of them were in Cat1. No data treatment and information app reached a score of 50. These results are presented in Figure 3B.

Assessment of Privacy Policies by Type of Cancer

Regarding the presence of a privacy policy according to cancer type, both general cancer apps (9/12, 75%) and breast cancer apps (11/14, 79%) had better results than prostate cancer (0/2). However, we consider that only 2 prostate cancer apps are not representative enough for a further analysis. As for other cancer apps, 67% (2/3) had privacy policies. Results are displayed in Figure 2C. In all, 3 of the breast cancer apps and 1 general cancer app were in Cat1, while the two other type apps that had a privacy policy were also above 75 points. These results do not seem to permit the drawing of any definite conclusions about the relationship between privacy policies and the type of cancer. Results can be seen in Figure 3C.

Assessment of Privacy Policies by Number of Downloads

Next, we evaluated the relationship between number of downloads and the presence of a privacy policy. We gathered apps into 4 groups: 5 to 99 downloads, 100 to 999 downloads, 1000 to 9999 downloads, and more than 10,000 downloads. The last 3 groups yielded similar results: 80% (4/5), 82% (9/11), and 73% (8/11), respectively, had a privacy policy. The only significant difference was observed in apps with 5 to 99 downloads: only 25% (1/4) had a privacy policy. These results can be seen in Figure 2D. There was, however, a difference in scores. None of the apps with 5 to 99 downloads reached 50 points. As for apps with 100 to 999 and 1000 to 9999 downloads, only 27% (3/11) and 36% (4/11) of apps, respectively, were in the first two categories. A total of 60% (3/5) of apps with more than 10,000 downloads were above 50 points. Thus, we found a clear relationship between the number of downloads and the average fairness of privacy policies, which can be observed in Figure 4A.

Assessment of Privacy Policies by Data Controller's Country

Three more features were analyzed but only in cases where a privacy policy was present. The first was the data controller's country. We identified 4 groups: EU apps (8), US apps (9), apps

from other countries (3), and apps from unknown countries (2). According to our analysis, the fairness of EU privacy policies was much better than those from the United States, with 75% (6/8) of EU apps above 50 points, and 4 of them in the first category. Meanwhile, only 33% (3/9) of US apps were in the first two categories. Only 33% (1/3) from other countries were above 50 points. As for the two apps for which the data controller's country was unknown, neither reached 50 points. These results can be seen in Figure 4B.

Assessment of Privacy Policies by General Data Protection Regulation Awareness

The second feature that was analyzed when a privacy policy was present was GDPR awareness. When the GDPR was explicitly mentioned in the app's privacy policy, 75% (6/8) of those apps reached the first two categories. Moreover, half of them reached the first category. When the GDPR was not mentioned, only 21% (3/14) reached 50 points. Figure 4C demonstrates this.

Assessment of Privacy Policies by Last Update

Finally, last privacy policy update was evaluated. We identified 4 groups: apps updated after GDPR implementation (May 25, 2018), apps updated after GDPR adoption (April 14, 2016), apps updated before GDPR adoption, and unknown update date. In fact, half of the 22 apps with a privacy policy were updated after GDPR implementation. Two more were updated after GDPR adoption. We found that 62% (8/13) of the apps updated after GDPR adoption were in the first two categories. Meanwhile, only 22% (2/9) of the apps not updated after GDPR adoption reached 50 points. Thus, update recency seems to be related to app privacy policy fairness. Figure 4D shows the results.

Assessment of Privacy Policies by Popularity

Last, we analyzed how popularity affects fairness in privacy policies. Popularity in an app is often measured by the number of downloads and the number of stars [45]. Consistent with previous literature, apps with fewer than 10 ratings were excluded to avoid unfair ratings [46]. Table 6 shows the scores for the apps that have been rated more than 10 times, ordered by their number of stars. Evidently, it is difficult to find any relationship between popularity and privacy score. The apps ranked second and third did not even have a privacy policy, even though App23 had more than 300 ratings. Moreover, App5 was ranked fourth and had a Cat4 privacy score.

Table 6. Assessment of privacy policies by app popularity.

App label	Stars	Ratings	Downloads	Privacy score
App14	4.7	1151	100,000+	75
App26	4.7	60	500+	0
App23	4.6	323	10,000+	0
App5	4.5	86	10,000+	23.1
App1	4.5	63	10,000+	76.9
App24	4.5	60	5000+	66.7
App3	4.4	20	1000+	53.6
App4	4.4	10	1000+	31.8
App22	4.2	227	10,000+	50
App7	4.1	47	1000+	78.6
App27	3.7	25	1000+	42.9
App18	3.7	21	1000+	15

Analysis of Item Compliance

Finally, we analyzed item compliance for the 22 apps with a privacy policy. Table 7 summarizes the results. We see a heterogeneous compliance of the different items that we checked. Only a few of the items mostly complied. Item 7, showing the recipients of personal data, was satisfied by 95% (21/22) of apps, while item 4, regarding the purposes of processing, was satisfied by 91% (20/22) of apps, with the other two giving partial information. Another item with positive results was item 1: 77% (17/22) of apps provided the identity of the data controller, with 3 more apps giving partial information. The last positive item was item 5. A total of 68% (15/22) of apps determined a legal basis for the processing.

Three items showed varied behavior. For item 10, 45% (10/22) of apps showed the existence of data subject's rights, with 2 more apps giving partial information. For both items 8 and 9, 36% (8/22) of apps disclosed transfers to other countries and about the period of personal data storage. Some apps gave partial information about them.

A negative behavior was observed for the rest of the items. Only 27% (6/22) of apps satisfied items 3 and 13. Item 3 regarded the DPO's contact details, while item 13 dealt with the obligation of providing personal data and the possible consequences of not providing such data. A total of 6 apps did not comply with item 11, which regarded the right of withdrawing consent at any time. However, item 11 was not applicable in 8 apps, as consent was not a legal basis for data processing. We determined that 23% (5/22) of apps satisfied

item 12, the right to lodge a complaint with a supervisory authority, with 3 more apps giving partial information. Item 6 is quite particular, as it is only applicable when the legitimate interests of the data controller constitute the legal basis for the processing. Only 33% (3/9) of apps complied with it, while it was not applicable in 13 apps. Item 14 was satisfied by only 9% (2/22) of apps. Information about profiling was not available in most of the cases. Last, none of the 13 apps outside the EU complied with item 2. The apps outside the EU should provide the identity of a representative inside the EU. This item was not applicable to the 9 apps in the EU.

Table 7. Summary of compliance with General Data Protection Regulation items.

Item number	Full information	Partial information	No information	Not applicable
1	17	3	2	0
2	0	0	13	9
3	6	0	16	0
4	20	2	0	0
5	15	0	7	0
6	3	0	6	13
7	21	0	1	0
8	8	7	6	13
9	8	5	9	0
10	10	2	10	0
11	6	0	8	8
12	5	3	14	0
13	6	0	16	0
14	2	2	18	0

Discussion

Principal Findings

This paper proposes a GDPR-based scale for assessing the fairness of privacy policies. We defined 14 items that provide developers with a tool to comply with the GDPR, while data controllers and users can use the scale to obtain a score that defines the fairness of privacy policies. Countries are really starting to be concerned about privacy and its implications. This has led the EU to develop laws that help protect user privacy. As a result, the GDPR was adopted in April 2016 and implemented in May 2018. In this study, we developed the second iteration of our GDPR-based method to assess the fairness of privacy policies. In this iteration, we refined the scores and criteria to obtain such scores with the aim of assessing not only compliance with the GDPR but fairness of the privacy policies in an objective way. Discrepancies between researchers have been critically reduced. A percentual score was defined, with a maximum of 100 points. A low score indicated not necessarily that an app did not comply with the GDPR but could indicate a low fairness of its privacy policy.

As privacy is crucial in mHealth and particularly for cancer patients, we assessed 31 Android cancer apps from Google Play. In order to foster straightforward interpretation of the results, we classified scores into four categories according to their fairness (from most to least fair): Cat1, Cat2, Cat3, and Cat4. We analyzed the fairness of privacy policies by app type, cancer type, number of downloads, data controller country, GDPR awareness, and last privacy policy update.

The first disappointing result was the absence of a privacy policy in 9 of 31 apps. This means that only 71% of the apps had a privacy policy. According to the literature, when top mHealth apps were analyzed, the percentage of apps with a privacy policy was about 90% [22,23]. However, when the type of app was selected, results were similar to ours: according to Bondaronek et al [32], 75% of physical activity apps had a privacy policy, while in Adhikari et al [33], they found that 69% of depression and smoking cessation apps had a privacy policy. When the selection is smaller, results are even worse. O'Loughlin et al [19] found privacy policies in only 49% of depression apps, while Sunyaev et al [16] showed that, surprisingly, 31% of medical or health and fitness apps had privacy policies. Moreover, when we analyzed the scores according to the type of app, it was noticeable that disease management apps and support apps obtained better scores than disease treatment and information apps. We believe that this is a positive fact, as disease management apps and support apps handle more sensitive information about patients.

In the literature, there are two ways to assess privacy. Some articles evaluated the different apps according to several items, eventually obtaining a score [27,28,36], while others checked if the analyzed apps met the criteria they had defined [20,32,37].

Regarding scores, only 45% of apps with a privacy policy that we assessed had a score greater than or equal to 50%, with an average score of 50.5 points. Only Hutton et al [15] had a comparable scoring system. They built a 26-item heuristic to assess privacy in mHealth apps, although only the first 7 items dealt with privacy policies. They applied their heuristic to 64

self-tracking mHealth apps and found an average score of 46.2%, with a high dispersion. Thus, their results are quite in line with ours.

The most complied-with items were the following: item 1 (identity of data controller), item 4 (purposes for the processing) and item 7 (recipients or categories of the personal data). Still, only 45% (5/22) of apps fully informed users about their rights (item 10) and only 5 fully informed users about their right to lodge a complaint with a supervisory authority. Finally, it is interesting that none of the 13 apps whose data controller was not within the EU informed users of the identity of their representative in the EU.

It is difficult to find such a complete analysis in the literature, but some of the items were assessed by different articles. Item 1 was evaluated in Hutton et al [15] and Papageorgiou et al [20], which were complied with by 75% and 63% of apps, respectively. Results were similar to our study, where 77% of apps satisfied this item. Item 1 was also analyzed in Huckvale et al [22], but results were very different. Only 25% of apps identified the data controller. Item 3 was also evaluated in Papageorgiou et al [20], with none of the apps having a DPO. Our study showed that 27% of apps had a DPO. Item 4 was assessed in Hutton et al [15] and Minen et al [21]: 61% and 64% of apps complied with this item, respectively, whereas a better result (91%) was obtained in this paper. Item 7 was assessed in Hutton et al [15], with 61% of apps stating the recipients of personal data; 96% of apps makes item 7 the most complied-with item in our study. Item 9 was evaluated in Huckvale et al [22]: 32% of apps stated the period for which personal data will be stored, compared with 36%. In Minen et al [21], item 10 was analyzed: 36% of apps informed users about their rights, whereas we obtained a result of 46%. Item 11 was assessed in Hutton et al [15] and Papageorgiou et al [20]: 55% and 37% of apps complied with this item, respectively. In our study, 43% of apps informed users about the right to withdraw consent. Items 12 and 13 were assessed in Huckvale et al [22]: 32% of apps complied with item 12, and 36% of apps satisfied item 13; 23% and 27% complied with these items in our study. Finally, Papageorgiou et al [20] evaluated item 14: 58% of apps informed users about profiling. This result was quite different from ours: 9% satisfied item 14.

Like other privacy scales [18,19,27], our scale considers each item to be equally important. In further research, we will work on the next iteration of the scale, wherein this approach will be reconsidered. We will evaluate whether using weighted scores provides a better assessment of the privacy policies of mHealth apps or only makes the scale more complex without any additional benefit.

Limitations

This study has some limitations. Some relevant apps may have been missed during our searches due to limitations of the Google Play search algorithm. Also, it is possible that developers may not have included some relevant information in the app description. As the eligibility assessment was based on app descriptions in the first search, this lack of information might have resulted in app exclusion. Only the Spanish version of the Google Play website was used during the search, and potentially relevant apps published on other versions of Google Play might have been excluded. Our study focused on Android apps, and this restriction also could have introduced a selection bias.

Conclusions

In this paper, we presented an improved version of our GDPR-based scale for the assessment of the fairness of privacy policies of mHealth apps. This new version has been successfully applied in a case study where the privacy policies of 31 cancer apps were analyzed, yielding results in line with similar studies. This analysis uncovered a surprising lack of fairness in these policies. The nature of the data and the concerns that patients have regarding privacy suggest that it should be a major concern for developers, users, and data controllers. Thus, the proposed scale seems to be suitable for evaluating the fairness of mHealth app privacy policies and for use by developers to ensure compliance with the GDPR.

Acknowledgments

This work was partially funded by the Cátedra de Telefónica Inteligencia en la red of the Universidad de Sevilla. This work was partially funded by the Cátedra Indra Sociedad Digital of the Universidad de Sevilla. ED-Z receives funding from and is supported by the V Plan Propio de Investigación de la Universidad de Sevilla, Spain.

Authors' Contributions

JB directed the study and took the lead in defining the scale as well as applying it in the case study. JB also supported the data analysis and interpretation of the data. JR took the lead in drafting the manuscript, supported by JB, OR-R, ED-Z, and AC. JR also participated in study direction, contributed to the scale definition and its application in the case study, and supported the data analysis and interpretation of the data. OR-R took the lead in data collection and participated in study direction, data analysis, and interpretation of the data. ED-Z participated in study direction, data collection, analysis, and interpretation of the data and reviewed the final version of the manuscript. AC resolved discrepancies, acquired funding through a research project, participated in the interpretation of data, and reviewed the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
User's guide.
[DOCX File , 71 KB - mhealth_v8i7e17134_app1.docx]

Multimedia Appendix 2
List of mHealth apps.
[DOCX File , 108 KB - mhealth_v8i7e17134_app2.docx]

References

1. Ham C, Dixon A, Brooke B. Transforming the delivery of health and social care: the case for fundamental change. London: The King's Fund URL: https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/transforming-the-delivery-of-health-and-social-care-the-kings-fund-sep-2012.pdf [accessed 2020-07-06]
2. Wyatt K, Finley A, Uribe R, Pallagi P, Willaert B, Ommen S, et al. Patients' experiences and attitudes of using a secure mobile phone app for medical photography: qualitative survey study. J Med Internet Res 2020 May 12;22(5):e14412 [FREE Full text] [doi: 10.2196/14412] [Medline: 32396127]
3. Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006 [FREE Full text] [doi: 10.2196/11006] [Medline: 30355554]
4. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
5. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of breast cancer mobile applications: state of the practice review. Int J Med Inform 2018 Dec;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
6. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide (preprint). JMIR mHealth uHealth 2019. [doi: 10.2196/preprints.13057]
7. Giunti G, Guisado Fernández E, Dorronzoro Zubiete E, Rivera Romero O. Supply and demand in mhealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
8. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul—a systematic review of pain-related apps. PLoS One 2014 Jul;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
9. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
10. Germanakos GP, Mourlas C, Samaras G. A mobile agent approach for ubiquitous and personalized ehealth information systems. 2005. URL: https://cgi.csc.liv.ac.uk/~floriana/UM05-eHealth/Germanakos.pdf [accessed 2020-07-06]
11. 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]
12. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
13. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
14. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN [accessed 2020-07-06]
15. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mhealth apps for self-tracking: heuristic evaluation approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
16. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2014 Aug 21:E28-E33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
17. Leigh S, Ouyang J, Mimnagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov;20(4):e20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]

18. Knorr K, Aspinall D, Wolters M. On the privacy, security and safety of blood pressure and diabetes apps. Proc Int Conf ICT Syst Security Privacy Protection 2015. [doi: 10.1007/978-3-319-18467-8_38]
19. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
20. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
21. Minen MT, Stieglitz EJ, Sciortino R, Torous J. Privacy issues in smartphone applications: an analysis of headache/migraine applications. Headache 2018 Jul;58(7):1014-1027 [FREE Full text] [doi: 10.1111/head.13341] [Medline: 29974470]
22. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
23. Scott K, Richards D, Adhikari R. A review and comparative analysis of security risks and safety measures of mobile health apps. AJIS 2015 Nov 22;19. [doi: 10.3127/ajis.v19i0.1210]
24. Brüggemann BT, Hansen J, Dehling T, Sunyaev A. Privacy technologies and policy. In: Schiffner S, Serna J, editors. An Information Privacy Risk Index for mHealth Apps. Cham: Springer; 2016.
25. Mense A, Steger S, Sulek M, Jukic-Sunaric D, Mészáros A. Analyzing privacy risks of mhealth applications. Stud Health Technol Inform 2016;221:41-45. [Medline: 27071873]
26. Zapata B, Hernández Niñirola A, Fernández-Alemán J, Toval A. Assessing the privacy policies in mobile personal health records. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4956-4959. [doi: 10.1109/EMBC.2014.6944736] [Medline: 25571104]
27. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
28. Bachiri M, Idri A, Fernández-Alemán JL, Toval A. Evaluating the privacy policies of mobile personal health records for pregnancy monitoring. J Med Syst 2018 Jun 29;42(8):1-14. [doi: 10.1007/s10916-018-1002-x] [Medline: 29959535]
29. Robustillo Cortés MDLA, Cantudo-Cuenca MR, Morillo-Verdugo R, Calvo-Cidoncha E. High quantity but limited quality in healthcare applications intended for HIV-infected patients. Telemed J E Health 2014 Aug;20(8):729-735. [doi: 10.1089/tmj.2013.0262] [Medline: 24849001]
30. Quevedo Rodríguez A, Wägner AM. Mobile phone applications for diabetes management: a systematic review. Endocrinol Diabetes Nutr 2019 May;66(5):330-337. [doi: 10.1016/j.endinu.2018.11.005] [Medline: 30745121]
31. Zapata B, Niñirola A, Fernández-Alemán J, Toval A. [Privacy and security in mobile personal health records for Android and iOS]. RISTI 2014 Jun 01(13):35-50. [doi: 10.4304/risti.13.35-50]
32. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
33. Adhikari AR, Richards D, Scott K. Security and privacy issues related to the use of mobile health apps. 2014 Presented at: Proceedings of the 25th Australasian Conference on Information Systems; 2014; Auckland p. 8-10.
34. Aliasgari AM, Black M, Yadav N. Security vulnerabilities in mobile health applications. 2018 Presented at: Proceedings of the IEEE Conference on Applications, Information & Network Security Nov ; Langkawi, Malaysia; AINS; 2018; Langkawi p. 21-22. [doi: 10.1109/ains.2018.8631464]
35. Mense MA, Urbauer P, Sauermann S, Wahl H. Simulation environment for testing security and privacy of mobile health apps. 2016 Presented at: Modeling and Simulation in Medicine Symposium; 2016; Pasadena p. 3-6. [doi: 10.22360/springsim.2016.msm.008]
36. Powell AC, Singh P, Torous J. The complexity of mental health app privacy policies: a potential barrier to privacy. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e158 [FREE Full text] [doi: 10.2196/mhealth.9871] [Medline: 30061090]
37. Robillard JM, Feng TL, Sporn AB, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Sep;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
38. Benjumea J, Dorronzoro-Zubiete E, Ropero J, Rivera-Romero O, Carrasco A. Privacy in mobile health applications for breast cancer patients. 2019 Presented at: Proceedings of the IEEE 32nd International Symposium on Computer-Based Medical Systems; 2019; Córdoba. [doi: 10.1109/cbms.2019.00131]
39. Informe sobre políticas de privacidad en internet. URL: https://www.aepd.es/media/estudios/informe-politicas-de-privacidad-adaptacion-RGPD.pdf [accessed 2019-11-01]
40. Decálogo para la adaptación al RGPD de las políticas de privacidad en Internet. URL: https://www.aepd.es/media/estudios/decalogo-politicas-de-privacidad-adaptacion-RGPD.pdf [accessed 2020-07-06]
41. El deber de informar y otras medidas de responsabilidad proactiva en apps para dispositivos móviles. URL: https://www.aepd.es/media/notas-tecnicas/nota-tecnica-apps-moviles.pdf [accessed 2019-11-01]
42. Fielding R, Reschke J. IETF request for comments. Hypertext transfer protocol (HTTP/1.1): conditional requests. 2014. URL: http://www.ietf.org/rfc/rfc7232.txt [accessed 2020-07-09]

43. About WHOIS. URL: https://whois.icann.org/en/about-whois [accessed 2019-11-01]
44. Smartphone market share. IDC Corporate. 2020 Jun 22. URL: https://www.idc.com/promo/smartphone-market-share/os [accessed 2019-11-01]
45. Ubaid Ur Rehman M, Aleem M, Islam MA, Ahmed S. Smart applications for diabetes management: a comprehensive survey and ranking. Health Informatics J 2019 Sep 30:1460458219869159. [doi: 10.1177/1460458219869159] [Medline: 31566463]
46. Mojica Ruiz IJ, Nagappan M, Adams B, Berger T, Dienst S, Hassan AE. An examination of the current rating system used in mobile-app stores. IEEE Softw 2016 Nov;33(6):86-92. [doi: 10.1109/ms.2015.56]

Abbreviations

Cat1: category 1
Cat2: category 2
Cat3: category 3
Cat4: category 4
DPO: data protection officer
EU: European Union
GDPR: General Data Protection Regulation
MARS: Mobile Application Rating Scale
mHealth: mobile health
N/A: not applicable
NPP: no privacy policy

Edited by G Eysenbach; submitted 20.11.19; peer-reviewed by B Richardson, B Brumen; comments to author 20.01.20; revised version received 27.05.20; accepted 22.06.20; published 28.07.20.

Please cite as:
Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A
Assessment of the Fairness of Privacy Policies of Mobile Health Apps: Scale Development and Evaluation in Cancer Apps
JMIR Mhealth Uhealth 2020;8(7):e17134
URL: http://mhealth.jmir.org/2020/7/e17134/
doi:10.2196/17134
PMID:32720913

Original Paper

Need for the Development of a Specific Regulatory Framework for Evaluation of Mobile Health Apps in Peru: Systematic Search on App Stores and Content Analysis

Leonardo Rojas Mezarina[1*], MD; Javier Silva-Valencia[1*], MD, MSc; Stefan Escobar-Agreda[1*], MD; Daniel Hector Espinoza Herrera[1*], MD, MS; Miguel S Egoavil[1*], MD; Mirko Maceda Kuljich[2*], MD, MS; Fiorella Inga-Berrospi[3], MD, MSc; Sergio Ronceros[4*], MD, MS, PhD

[1]Unidad de Telesalud, School of Medicine, Universidad Nacional Mayor de San Marcos, Lima, Peru

[2]School of Medicine, Universidad San Martin de Porres, Lima, Peru

[3]Universidad Continental, Lima, Peru

[4]Deanery of the School of Medicine, Universidad Nacional Mayor de San Marcos, Lima, Peru

[*]these authors contributed equally

Corresponding Author:
Javier Silva-Valencia, MD, MSc
Unidad de Telesalud, School of Medicine
Universidad Nacional Mayor de San Marcos
Av Grau 755, Cercado de Lima
Lima
Peru
Phone: 51 978978368
Email: javier.silva@unmsm.edu.pe

Abstract

Background: In Peru, there is an increase in the creation of mobile health (mHealth) apps; however, this situation could present problems related to the quality of information these apps share, data security and privacy, usability, and effectiveness, as there is no specific local regulation about their creation and use.

Objective: The objective of this study was to review mHealth apps created, uploaded, or used in Peru, and perform an analysis of the national regulatory framework that could be applied to evaluate whether there is a need to develop and implement a specific regulation to these apps.

Methods: A total of 3 reviews were performed. First, we reviewed information about Peruvian mHealth apps created up to May 2019 from scientific publications, news, government communications, and virtual stores, and evaluated their purpose, creator, and the available evidence of their usability and effectiveness. The second review was carried out by taking a sample of the 10 most commonly used mHealth apps in Peru (regardless of the country of creation), to evaluate the information they collect and classify them according to the possible risks that they could present in terms of security and privacy. In addition, we evaluated whether they refer to or endorse the information they provided. Finally, in the third review, we searched for Peruvian standards related to electronic health (eHealth) that involve information technology that can be applied to regulate these apps.

Results: A total of 66 apps meeting our inclusion criteria were identified; of these, 47% (n=31) belonged to government agencies and 47% (n=31) were designed for administrative purposes (private and government agencies). There was no evidence about the usability or effectiveness of any of these apps. Concerning the 10 most commonly used mHealth apps in Peru, about the half of them gathered user information that could be leaked, changed, or lost, thus posing a great harm to their users or to their related patients. In addition, 6/10 (60%) of these apps did not mention the source of the information they provided. Among the Peruvian norms, the Law on the Protection of Personal Data, Law on Medical Devices, and administrative directives on standards and criteria for health information systems have some regulations that could be applied to these apps; however, these do not fully cover all aspects concerning the evaluation of security and privacy of data, quality of provided information, and evidence of an app's usability and effectiveness.

Conclusions: Because many Peruvian mHealth apps have issues related to security and privacy of data, quality of information provided, and lack of available evidence of their usability and effectiveness, there is an urgent need to develop a regulatory

framework based on existing medical device and health information system norms in order to promote the evaluation and regulation of all the aforesaid aspects, including the creation of a national repository for these apps that describes all these characteristics.

(JMIR Mhealth Uhealth 2020;8(7):e16753) doi:10.2196/16753

KEYWORDS

mhealth apps; mHealth; regulatory framework; Peru, eHealth

Introduction

Mobile apps are any software programs that can be downloaded and executed on a mobile device such as cell phones or tablets [1]. Since their large-scale introduction approximately 10 years ago, apps have been greatly developed and diversified, taking on a leading role in practically all the daily activities that we carry out [2]. When apps are related to aspects of health, these are called *mobile medical apps* [3], *mobile apps in health*, or simply *apps in health*. Such apps are within the technological branch known as mobile health (mHealth), which is defined as the practice of public health using mobile devices to improve health and medical care through new forms of interactive services that promote wellness, care, prevention, and disease management [4,5].

mHealth apps are classified into 2 large groups: apps for health personnel and apps for general public. Those directed at health personnel can be used for training, consultations, communication between professionals, referrals, monitoring, or patient management [6], whereas those aimed at the general public are related to medication management; symptoms of chronic diseases; or monitoring of aspects such as weight, obesity, nutrition, fertility, lactation, and pregnancy [6]. Therefore, the use of mHealth apps could aid in achieving better disease control, changing habits, improving the quality of care, allowing greater access to information [2], reducing costs, avoiding unnecessary medical consultations, and bringing health services closer to people.

Nevertheless, the possible risks of the indiscriminate creation and use of mHealth apps are worrisome [7]. Many studies have reported limited or inconclusive results regarding their effectiveness, which is mainly because it is impossible to formulate a successful and adequate design strategy for medical apps [8]; or in some cases, the effect achieved cannot be attributed solely to the use of these apps [9]. In addition, doubts have been raised about the quality of the information they provide that would be used for promotion of health, prevention of diseases, or in some cases, even management and treatment of diseases. Finally, there has been an increasing interest in evaluating the data privacy and security policies of those apps, which when left unmonitored pose a huge risk to their users [10]. To manage these situations, some international institutions such as the US Food and Drug Administration (FDA) and the European Union (EU) have developed standards and norms to regulate the development and use of mHealth apps [3].

In Peru, given the steady increase in the rate of access to mobile phones in the last few years, a need arose to promote the development of mHealth apps, as in other developing countries [11], through different strategies such as hackathons (ie, meetings for collaborative development) [12] and funding calls to support information and communications technology projects by public and private institutions [13]. However, it remains unknown whether the current Peruvian norms adequately regulate aspects related to the creation and use of mHealth apps or whether there is a need to implement a specific regulatory framework for this purpose. Therefore, in this study, we review all mHealth apps created and published in Peru and perform an analysis of the current Peruvian norms that could be applied for their regulation, in order to evaluate whether there is a need to implement a specific mHealth regulatory framework.

Methods

Overview

A total of 3 reviews were performed up to May 2019. The first was a review of mHealth apps created in Peru, the second was a review on mHealth apps most used, and the third was a study on the Peruvian regulatory aspects that could be applied to mHealth.

Review of Mobile Apps Created in Peru

Data Sources

A search for information about Peruvian mHealth apps was performed on scientific publications, gray literature, and app stores. For the scientific literature review, the following databases were searched: PubMed, SciELO, Scopus, IEEE Xplore, and Web of Knowledge (Web of Science); for the gray literature search, we reviewed newspapers, reports of research institutes and nongovernmental organizations, thesis, conference proceedings, and websites (identified through Google Search); and concerning medical apps published in virtual app stores, we searched only Google Play and Apple Store, as most smartphones used in Peru run on either Android or iOS (Table 1).

Table 1. Search strategy.

Search strategy/Review employed	Data sources
Scientific literature review	• Bibliographic database and search engines: PubMed, SciELO, Scopus, IEEE Xplore, Web of Knowledge (Web of Science), Google Scholar, and Google Search.
Gray literature	• Peruvian National Digital Repository of Science, Technology and Innovation, Library of Theses and Dissertations (ALICIA) • Newspapers: Correo, El Comercio, El Peruano, Expreso, Primera, Razón, La Republica, Gestión, Ojo, and Perú21 • Websites: gob.pe/minsa, dge.gob.pe, web.ins.gob.pe, nongovernmental organizations • Proceedings from related conferences • Not-for-profit organizations websites • Blogs and other networks
Virtual mobile app stores	• Google Play: medical apps • Apple App Store: medical apps

Search Methods

To search mHealth apps on app stores, the following search terms in Spanish were included: health, Perú, Peruvian, disease, patient care, drugs, pharmacy, insurance, self-care, Clinics, Hospitals, MINSA, DIGESA, INS, SuSalud, EsSalud, DIRESA, and SIS.

For the search in scientific and gray literature, we used 30 keywords in Spanish and English to identify apps published up to May 9, 2019. These keywords included the following:

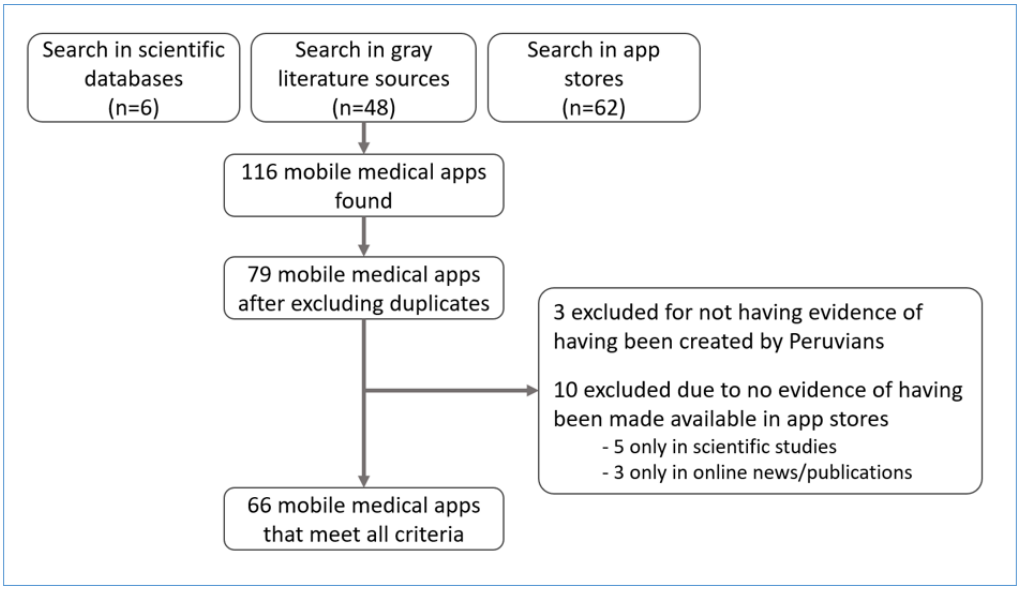
- Peru-related keywords: related to Peru or Peruvian institutions. For example, Peru, Ministry of Health (or its acronym in Spanish MINSA), National Institute of Health (or its acronym in Spanish INS).
- Health-related keywords: Related to apps associated with providing health services for education, prevention, diagnosis, or treatment purposes. For example, patient care, self-care, health services, diagnosis.
- Mobile technology–related keywords: Related to the use of mHealth technology. For example, mobile, mHealth, app, mobile software.

Additional details about the search algorithms used are presented in Multimedia Appendix 1.

Eligibility Criteria

The eligibility criteria were as follows: only apps published in Spanish and with evidence that they were created in Peru were studied. For this purpose, we reviewed information about each app available on the information section of app stores and verified if these were related to Peruvian Institutions (eg, INS, MINSA) or to Peruvian companies or clinics (eg, MAPFRE Perú, Inkafarma, Clinica San Pablo) or if their developers were Peruvian. Apps that were not published openly or that were only used for internal studies and those that could not be accessed to assess their characteristics were discarded (Figure 1).

Figure 1. Flowchart of Peruvian mobile medical app research.



Screening Process

The screening process for identifying apps meeting our inclusion criteria on mobile app stores (Google Play and Apple Store), scientific databases (for scientific literature) and Google Search (for gray literature) was performed by 2 independent authors (LRM and DHEH). To describe the main characteristics of these apps, the authors documented each app's name, entity that created it, year of last update, description of its functionalities, type of use, target audience, and if there was any evidence of efficacy. A third author compiled a database based on the information provided by the 2 authors, eliminating duplicated information and evaluating if the identified apps and related literature meet our eligibility criteria.

Review of the Most Used mHealth Apps in Peru

Considering that the Android operating system is used by approximately 89.5% of the population in Peru, we searched for the 10 most downloaded mHealth apps in Peru on the Google Play Store. Apps were chosen based on app market statistics from analytical platforms such as App Annie [14] and SimilarWeb [15], which were reviewed in May 2019. For each app, security and privacy aspects were evaluated through the tool proposed by Dehling et al [16], which has shown good reliability. This tool allows one to evaluate the kind of information collected by mHealth apps, defining whether they are medical or nonmedical apps, and whether they leak, modify, or provide user information to third parties. Thus, we evaluated whether these apps represent a high, low, or no risk to the users, irrespective of whether they were intended to be used by the general public or the health care personnel who might collect patient data. In this study, this tool was applied by 2 independent researchers (JS-V and SE-A) and if there was a doubt it was solved by either consensus or the participation of a third researcher. Additional details about the structure and results obtained using this tool are described in Multimedia Appendix 2.

We also evaluated if these apps had a disclaimer, and whether they declared the terms of use and privacy policies. The quality of the information for each app was assessed by evaluating if it had references, mentioned where the source of information came from, or presented any evidence that the information was accurate. To evaluate all these aspects, we first installed the app, verified all kinds of information required by the app, and then reviewed their terms and conditions of use if these were available.

Review of the Legal and Regulatory Aspects Related to mHealth in Peru

Peruvian laws and regulations in the area of digital health were reviewed to understand which current legal standards apply or could apply to medical apps. The search terms used in this review were the same as those used in the previous one. The search was performed on *Portal de Estrategia Digital*, which is a web platform designed by the Peruvian Ministry of Health that displays all the contents related to the health information and communications technology norms of Peru [17]. Specifically, we looked for all norms that regulate aspects related to the quality of information provided by health systems, information security, and protection of personal data, as we considered these to be the main aspects that could represent an adequate scenario for the use of mHealth technologies in a useful and safe manner in our country.

Data Analysis

All the statistical analyses were performed through the Statistical Package Stata version 14.0 for Windows (StataCorp). We performed a descriptive analysis of the characteristics of mHealth apps created in Peru, and evaluated the security and privacy aspects of the 10 most commonly used mHealth apps in Peru. All the quantitative results were presented as categorical data through frequencies and proportions.

Results

Mobile Medical Apps Created in Peru

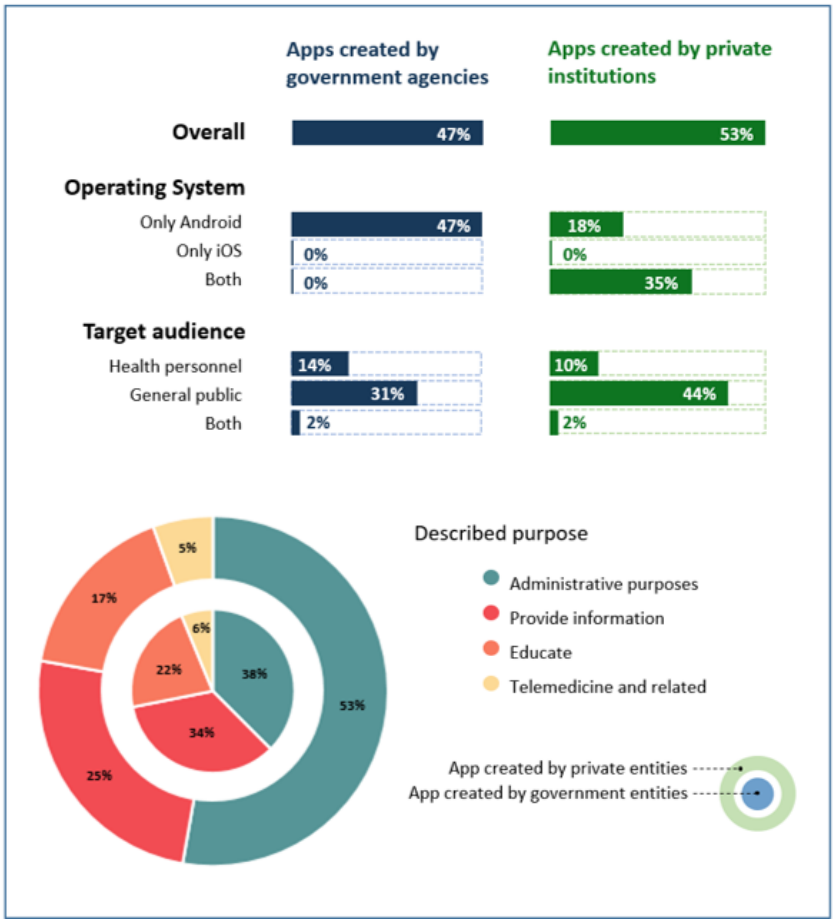
We analyzed mHealth apps created in Peru and updated within the last 5 years. A total of 66 apps were found, of which almost half (31/66, 47%) were created by government agencies. The general characteristics are presented in Figure 2. Only 35% (23/66) of these apps could be installed and run on both Android and iOS. In particular, none of the apps created by government agencies (such as Ministry of Health, National Institute of Health; National Health Superintendence) could be run on Apple iOS. Most apps were aimed at the general public for administrative purposes or to provide information on health-related issues.

Concerning the purpose, we identified 4 categories of apps: (1) those created for administrative purposes related, for example, to schedule and manage appointments, pay for health services, field work management, digital health cards; (2) those created to provide information, for example, about the location of health facilities, insurance status, pharmaceutical products, prices, or food composition; (3) those created with the aim of educating, for example, to promote psychological education, or share virtualized books, care guides, and health advices; and finally, (4) those related to telehealth functions to facilitate the offering of virtual and distant communication services among health personnel. The full list of apps is summarized in Multimedia Appendix 3.

We did not find any publications, reports, or articles that provide information on how well the apps work, based on their purpose (effectiveness), how they are perceived by the users (usability), the source of the information they provided, updating frequency, or other mechanisms to ensure the quality of information these apps provide.

Besides the 66 apps we found, it is important to mention that in the scientific literature, there were also other mobile medical apps that were created in Peru, but these were used only for research purposes. Ruiz et al [18] discussed 4 mHealth interventions in Peru, but their apps were not published openly.

Figure 2. Characteristics of the mobile medical apps created in Peru in the last 5 years (n=66).



Most Used mHealth Apps in Peru

Table 2 presents data on the 10 most downloaded medical apps in the Peruvian territory. All of them were from private institutions or independent creators. The most downloaded apps were for telehealth services for either symptoms assessment or communication with health professionals. As much as 6 of the 10 apps were aimed at the general public to answer health questions, inform about illnesses, and to monitor conditions such as menstruation and pregnancy. Among the apps aimed at health professionals were those used to consult information about pharmaceutical products, disease codes, and medical concepts.

Concerning the privacy and security aspects of the 10 most downloaded mHealth apps in Peru, which were evaluated using the tool proposed by Dehling et al [16], we found that 5 of 10 (50%) collected medical information such as symptoms, history, doctor, allergies, menstrual cycle, or pregnancy; and 4 of 10 (40%) collected information about sexual history, pregnancy status, or the presence of current illnesses; however, these 4 apps leaked user data and thus could cause embarrassment or affect the employment prospects of their users. In addition, 40% (4/10) of these apps collect information whose loss could cause potential harm to their users (mainly apps that provide telehealth services or estimate conditions/diseases). Finally, 4 of the apps ask for information about the date of birth and contact information (eg, ID in social network, email, or phone number) that are invaluable to third parties, for example, to sell products. Half of these apps did not mention whether security protocols are followed to protect the information they collect (Table 3).

Regarding the quality of information provided by these apps, only 4 mentioned the source of the information provided. The others neither had any bibliographical references nor mentioned if a health professional participates within the operating team.

Table 2. Top 10 most downloaded medical apps in the Peruvian territory as of May 2019.

Rank	App	App rating (stars)	Description	Target audience	Purpose
1	Ada – Tu guía de salud	4.7	Ada, your health guide. Ada can help you get accurate health reports. Ada will evaluate your symptoms and, in an instant, will give you relevant information and the next steps to take.	General Public	Telehealth service
2	Chat Médico mediQuo	4.5	Make your consultations to specialized doctors in an easy and simple way and through your smartphone. Chat with our doctors and specialists who will answer questions instantly.	General public	Telehealth service
3	Vademecum Perú	3.8	The app allows access to information on medicines, laboratories, pathologies, interactions, pharmacological index, and therapeutic index. It gathers all the relevant information related to pharmaceutical products.	Health personnel	Provide information
4	CIE10 Español	4.3	This app contains the entire International Classification of Diseases (ICD-10) in one place.	Health personnel	Provide information
5	Calendario Menstrual	4.7	An app that helps and assists women in monitoring their periods, cycles, ovulation, and the possibility of pregnancy (fertile days).	General public	Management purposes
6	Mi Calendario de Embarazo	4.8	Get detailed information about your baby's development in each week of pregnancy. It includes daily information about pregnancy, feeding advice, control of weight gain per month and belly growth, weight gain calculator, and much more.	General public	Information/management purposes
7	Anatomy Learning - 3D Atlas	4.6	Real-time 3D Atlas. Dissect and learn anatomy.	General public	Education
8	Disorder & Diseases Dictionary	4.7	Medical dictionary that provides more than 1000 definitions of diseases.	General public	Provide information
9	Medical Terminology Dictionary	4.4	This dictionary of medical terms includes more than 10,000 entries with anatomical, pathological, diagnostic definitions, etc.	Health personnel	Provide information
10	Diagnósticos de Enfermería	4.4 stars	Nanda Nursing Diagnostics 2018-2020	Health personnel	Provide information

Table 3. Aspects of security, privacy, and quality of the information in the 10 most downloaded medical apps in Peru (data as of May 2019).

Aspects	Apps, n (%)
Security and privacy[a]	
Specificity: Health specificity of information available to apps	
Standard	4 (40)
Nonstandard	1 (10)
Medical	5 (50)
Leaks: Potential damage through information leaks	
None	6 (60)
Low	0 (0)
High	4 (40)
Change: Potential damage through manipulation (change) of information	
None	1 (10)
Low	1 (10)
High	8 (80)
Loss: Potential damage through loss of information	
None	4 (40)
Low	2 (20)
High	4 (40)
Value: Value of information to third parties	
None	4 (40)
Low	2 (20)
High	4 (40)
Quality of information	
Include a disclaimer to report limitations	5 (50)
Mention the source of the provided information	4 (40)
Mention how often the information is updated	0 (0)

[a]Analyzed using the security and privacy assessment tool for mHealth apps proposed by Dehling et al [16].

Legal and Regulatory Aspects Related to mHealth in Peru

Among the Peruvian norms related to electronic health (eHealth) and health information technology services that could be applied to Peruvian mHealth apps, we found the following laws.

Law No. 29733: Law on the Protection of Personal Data

It establishes that "one of the rights that all people have is the right to have their personal data protected, so that such information would not be used in a way that harms them, threatening the dignity and integrity of the people" [19]. Therefore, the necessary guarantees must be given to control the use of personal data, regardless of how they are used. mHealth apps developed in Peru have the obligation to comply with such provisions.

Law No. 29459: Law of Pharmaceutical Products, Medical Devices, and Health Products

The law defines a medical device as any instrument, device, implement, machine, reagent, in vitro calibrator, computer application, material or another similar article provided to be used either alone or in combination with human beings for functions of [20]:

- Diagnosis, prevention, monitoring, treatment, or relief of a disease.
- Diagnosis, monitoring, treatment, relief, or compensation of an injury.
- Research, replacement, or modification of the anatomy or a physiological process.
- Support or maintenance of life.
- Control of conception.
- Disinfection of medical devices.

Although this law does not specifically mention mobile app, it does mention that any computer application that declares any of these functions as intended use will be considered a medical device. It is requested that in order to approve a medical device, it must have an approval from a foreign entity (such as the FDA) or follow a national approval process.

Administrative Directive No. 230: Administrative Directive That Establishes the Standards and Technical Criteria for the Development of Health Information Systems

This administrative directive is only for establishments and dependencies of the Ministry of Health [21]. It aims to establish standards for the development of information systems and mentions that health sector information systems could be of 2 types:

1. Administrative information systems that integrate data from administrative processes and procedures such as planning, budgeting, logistics, finance, human resources, and
2. Health information systems that integrate data on the processes and procedures involved in people's health; health insurance; epidemics and health emergencies; environmental health and food safety; health intelligence; pharmaceutical and medical products, medical devices, and pharmaceutical establishments; human resources in health; infrastructure and sanitary equipment; and health research and technologies.

Law No. 30421: Telehealth Framework Law

Telehealth has been regulated in Peru since 2016 and its definition was recently modified according to Legislative Decree 1303 [22]. It is defined as the provision of distance health services by trained health personnel. The Law defines 4 areas of Telehealth app: Telemedicine, Telemanagement, Teletraining, and TeleECI (education, communication, and information to the general population). Telehealth regulations so far do not specifically mention the use of mobile apps as a tool or component necessary to achieve the objectives.

Discussion

Summary of Evidence

According to our results, there are many issues on mHealth apps created in Peru with regard to security and privacy of data, quality of information provided, and evidence of their effectiveness that need to be evaluated and regulated. Because the current Peruvian norms do not fully cover all these aspects, there is a need to create and implement a specific Peruvian regulatory framework to mHealth apps.

Characteristics of Peruvian mHealth Apps and Need for Their Regulation

Security and Privacy Aspects

Concerning the security and privacy aspects, we found that 40% (4/10) of the most used mHealth apps in Peru included data whose manipulation, leak, or loss could induce potential harm or damage to their users, for example, affecting their reputation and employment prospects or misorient aspects related to their personal health care. Despite this issue, we found that many of these apps do not clearly mention how the data they collect will be used in their terms and conditions policies. In addition, the information related to the security and privacy aspects of these apps is usually presented in long texts with complex structures, so most users would likely accept these terms and conditions without reviewing them [23,24]. In this context, there is a need to ensure that all apps include a plan or strategy to avoid or prevent the change, leak, or loss of information (either provided or asked) as well as implement other security and privacy features of personal data management according to the Peruvian norms.

Quality of Information

Concerning the quality of information presented, we found that among the most used mHealth apps in Peru, only about half report the source of information they provide; notably, none of them reported whether this information is regularly updated. This could represent a potential risk to users of these apps, as most apps are oriented toward general public that do not have enough background to evaluate the quality of information they receive. Indeed, one study showed that about 30%-40% of Peruvian adult population do not have adequate knowledge and competence to access, understand, and apply information received about health topics [25]. This situation suggests the need to ensure that all of these apps include the references used to elaborate the information they provide, or at least indicate that this information is provided by health professionals or subject matter experts.

Evidence of Usability and Effectiveness

We found that none of the Peruvian mHealth apps available on app stores have been evaluated in published scientific studies about their usability and effectiveness. Usually the quality of these apps is evaluated in the app stores by a 5-star rating system and reviews from their users that give an indication about their usability, which is then used as a feedback to their developers to improve the characteristics of these apps. However, this information does not necessarily reflect the actual effectiveness of these apps in relation to their mentioned purpose of use. In this context, there is a need to consolidate a list of mHealth apps that provide information not only about the user's satisfaction level (star rating) but also about the studies that support their effectiveness in relation to their mentioned purpose.

In addition, we found that none of Peruvian mHealth apps that have been evaluated in scientific studies or developed in technology innovation events was currently available on app stores. This is likely because these apps were developed for research purposes and their developers may not have the intention or capabilities to continue their use for commercial purposes which could sustain their development and use in the future. In this context, it is important for business incubators, founders, or interested companies to identify these apps and support their developers until these apps could be made sustainable.

Proposal of a Legal Framework About Peruvian mHealth Apps

The Experience of the FDA

In the United States, the US FDA published in February 2015 a guide that details how it regulates mHealth apps [3]. According to this, the FDA only regulates medical apps that meet the following definition of a medical device: (1) Those that are intended to be used as accessories for a medical device (eg,

extension of a vital signs monitor or an app that controls an infusion pump); and (2) Those that transform a mobile platform into a medical device by including accessories (eg, app that connects a glucose strip so that the mobile phone functions as a glucometer) and by providing a specific diagnosis or treatment (eg, apps that provide a diagnosis when monitoring clinical parameters).

It is important to mention that this control is determined by the intended use that the app declares. For this, it uses a risk-based approach (from level I [low risk] to level III [high risk]). Among the examples of major risks (level III), it is worth mentioning apps such as arrhythmia detector or electrocardiogram or blood pressure measurement system. Examples of low-risk apps are those linked to electronic medical records, patient self-care, and patient communication with their health care provider.

The Experience of the European Union

In Europe, the EU, like the FDA, regulates the use of mHealth apps if these are considered medical devices. According to the EU regulation, an mHealth app can be a medical device if it contemplates the intention of being used by people with at least one of the following purposes: (1) diagnosis, prevention, control, treatment, or relief of a disease; (2) the diagnosis, control, treatment, relief, or compensation of an injury or deficiency; (3) examination, replacement, or alteration of the anatomical structure or a physiological process; and (4) contraception. In this situation, a same app with different purposes may or may not considered a medical device according to its use; however, its design, use, and data collection policies should adhere to the proper medical device regulations of EU.

EU regulations propose that the use of medical devices in Europe do not need state approval; however, they must comply with the so-called Essential Safety and Performance Requirements of the European Medical Device Directive, which varies according to the potential risk of the product [26]. Therefore, they need technical documentation that contains a report on risk analysis and a clinical evaluation. In addition, the EU states are in charge of monitoring the product continuously after commercialization. According to these regulations, medical devices are classified into 4 categories based on risk level: class I risk (low-risk potential), class IIa (medium risk potential), class IIb (higher risk), and class III (particularly high-risk potential) [27].

Proposed Regulatory Framework

Considering all aforementioned aspects, a regulatory framework for mHealth apps in Peru could be based on the current Peruvian regulations with certain considerations according to international experiences.

As proposed by international regulatory agencies, the main step to regulate mHealth apps would be to determine whether they meet the definition of a medical device [28,29]. For this, we could use the Law of Medical Devices (Law No. 29459) which points out that all medical devices in Peru must obtain the approval of a foreign entity (eg, the FDA) or follow a national approval process (which since its implementation in 2017, despite multiple requests, has only managed to register a few). This is an important issue because it must be decided whether every mobile app needs a comprehensive approval, or as international agencies do, whether each kind of risk requires its own form of approval.

In addition, as the environment of the mHealth apps experiences constant innovation and re-design, the legal framework should include a not-so-exhaustive list of the characteristics of Peruvian mHealth apps considered as medical devices and also those that are not in order to provide clarity and assistance in decision making. This list should include characteristics such as security and privacy aspects, the quality of information provided, and the existence of available evidence (preferably, published in scientific journals) about the usability and effectiveness of these apps [30].

Apps that do not meet the definition of a medical device must also be regulated. The FDA stresses the importance of long-term monitoring of apps that are not considered medical devices, such as those designed to help patients to control their disease, those that help organize health information, and those that help to access information. In Peru, mHealth apps that are not considered medical devices could be considered as an information system and thus could be regulated by Administrative Directive 230, which establishes the technical standards and criteria for health information system–based apps created by public institutions.

Unlike Law No. 29459, whose scope applies to the entire national territory, Directive 230 only applies to establishments of the Ministry of Health, so it would be necessary to create a regulation whose scope applies to the entire national level, irrespective of the health subsector the app belongs to.

Conclusions and Recommendations

In Peru, the rapid increase of the use of mobile technology in health has led to a widespread creation of mHealth apps. However, their use presents a risk to users, as there are lack of strategies or norms to manage security and privacy aspects, quality of information provided, and available evidence about their usability and effectiveness. Given that in Peru the existing norms do not exhaustively cover all these aspects, we recommend the formulation and promulgation of a legal framework specific to mHealth apps based on international models and current national legislation. It should include the creation of a national repository of mHealth apps that provides information related to the aspects discussed in this work. In the same way, we suggest performing studies that evaluate the use and impact of these apps on patients and health professionals, to establish their associated benefits and risks.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search algorithms.
[DOCX File , 14 KB - mhealth_v8i7e16753_app1.docx]

Multimedia Appendix 2
Security and privacy tool to mhealth apps.
[XLSX File (Microsoft Excel File), 19 KB - mhealth_v8i7e16753_app2.xlsx]

Multimedia Appendix 3
Database of Mobile Apps created in Perú in the last 5 years.
[XLSX File (Microsoft Excel File), 24 KB - mhealth_v8i7e16753_app3.xlsx]

References

1. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
2. Milošević M, Shrove MT, Jovanov E. Applications of Smartphones for Ubiquitous Health Monitoring and Wellbeing Management. JITA 2011 Jun 15;1(1):1. [doi: 10.7251/JIT1101007M]
3. Food and Drug Administration. Policy for Device Software Functions and Mobile Medical Applications. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications [accessed 2020-03-17]
4. World Health Organization. New Horizons for Health through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011.
5. PAHO eHealth/WHO. 51.° Consejo Directivo. Estrategia y Plan de Acción Sobre ESalud. 2011. URL: https://www.paho.org/hq/dmdocuments/2011/CD51-13-s.pdf [accessed 2020-06-15]
6. Aitken M. The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2020-06-01]
7. 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]
8. 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 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
9. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile Apps to Support the Self-Management of Hypertension: Systematic Review of Effectiveness, Usability, and User Satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]
10. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33 [FREE Full text] [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
11. Kahn JG, Yang JS, Kahn JS. 'Mobile' Health Needs And Opportunities In Developing Countries. Health Affairs 2010 Feb;29(2):252-258. [doi: 10.1377/hlthaff.2009.0965]
12. Olson KR, Walsh M, Garg P, Steel A, Mehta S, Data S, et al. Health hackathons: theatre or substance? A survey assessment of outcomes from healthcare-focused hackathons in three countries. BMJ Innov 2017 Jan 04;3(1):37-44. [doi: 10.1136/bmjinnov-2016-000147]
13. Ministerio de la Producción - Perú. Innóvate Perú. 2020. URL: https://www.innovateperu.gob.pe/ [accessed 2020-03-19]
14. App Annie Intelligence. App Annie. URL: https://www.appannie.com/en/product/intelligence/ [accessed 2020-03-18]
15. SimilarWeb: Website Traffic Statistics & Market Intelligence. URL: https://www.similarweb.com/ [accessed 2020-03-18]
16. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
17. Presidencia del Consejo de Ministros. Portal de Estrategia Digital. 2018. URL: http://digital.minsa.gob.pe/ [accessed 2020-03-18]
18. Ruiz EF, Proaño Á, Ponce OJ, Curioso WH. Tecnologías móviles para la salud pública en el Perú: lecciones aprendidas. Rev Peru Med Exp Salud Publica 2015 Jun 19;32(2):364-372. [doi: 10.17843/rpmesp.2015.322.1634]
19. Congreso de la República. Ley N° 29733. In: Ley de Protección de Datos Personales. Lima, Peru: Congreso de la República; 2011:1-8.
20. Congreso de la República. Ley de Los Productos Farmacéuticos, Dispositivos Médicos y Productos Sanitarios. Lima, Peru: Congreso de la República; 2009:1-11.
21. Ministerio de Salud. Directiva Administrativa Que Establece Los Estándares y Criterios Técnicos Para El Desarrollo de Los Sistemas de Información de Salud. Lima, Peru: Ministerio de Salud; 2017:1-59.
22. Congreso de la República. Ley Marco de Telesalud. Lima, Peru: Congreso de la República; 2016:1-2.

23. Martínez-Pérez B, de LTI, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
24. 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. [doi: 10.1186/s12911-017-0535-0]
25. Macha Quillama LF, Samanamu Leandro AE, Rodriguez Hurtado D. Alfabetización sanitaria y factores relacionados en pacientes que acuden a consulta externa de Medicina Interna en un Hospital Nacional en Lima,Perú. Rev. salud pública 2017 Sep 01;19(5):679-685. [doi: 10.15446/rsap.v19n5.53008]
26. Official Journal of the European Communities. Council Directive 93/42/EEC of 14 June 1993 Concerning Medical Devices. European Union: European Union; 1993.
27. Chai J. Medical device regulation in the United Statesthe European Union: a comparative study. Food Drug Law J 2000;55(1):57-80.
28. Kramer DB, Xu S, Kesselheim AS. How Does Medical Device Regulation Perform in the United States and the European Union? A Systematic Review. PLoS Med 2012 Jul 31;9(7):e1001276. [doi: 10.1371/journal.pmed.1001276]
29. Crico C, Renzi C, Graf N, Buyx A, Kondylakis H, Koumakis L, et al. mHealth and telemedicine apps: in search of a common regulation. Ecancermedicalscience 2018;12:853 [FREE Full text] [doi: 10.3332/ecancer.2018.853] [Medline: 30079115]
30. Abroms LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]

Abbreviations

EU: European Union
FDA: Food and Drug Administration

Edited by G Eysenbach; submitted 23.10.19; peer-reviewed by J Vargas Herrera, T Alessa; comments to author 19.01.20; revised version received 04.04.20; accepted 26.04.20; published 10.07.20.

Please cite as:
Rojas Mezarina L, Silva-Valencia J, Escobar-Agreda S, Espinoza Herrera DH, Egoavil MS, Maceda Kuljich M, Inga-Berrospi F, Ronceros S
Need for the Development of a Specific Regulatory Framework for Evaluation of Mobile Health Apps in Peru: Systematic Search on App Stores and Content Analysis
JMIR Mhealth Uhealth 2020;8(7):e16753
URL: https://mhealth.jmir.org/2020/7/e16753
doi:10.2196/16753
PMID:32352926

Original Paper

Smartphone Apps Targeting Physical Activity in People With Rheumatoid Arthritis: Systematic Quality Appraisal and Content Analysis

Lindsay M Bearne[1*], MCSP, MSc, PhD; Mandeep Sekhon[1*], BSc, MSc, PhD; Rebecca Grainger[2*], BMedSci, MBChB, FRACP, PhD; Anthony La[1*], BSc, MSc; Mehrdad Shamali[1*], BSc, MSc; Aliya Amirova[1*], BSc, MSc; Emma L Godfrey[1*], BA, MA, PhD; Claire M White[1*], BSc, MSc, PhD

[1]Department of Population Health Sciences, School of Population Health and Environmental Sciences, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom

[2]University of Otago, Wellington, New Zealand

[*]all authors contributed equally

Corresponding Author:
Lindsay M Bearne, MCSP, MSc, PhD
Department of Population Health Sciences, School of Population Health and Environmental Sciences
Faculty of Life Sciences & Medicine
King's College London
2nd Floor, Addison House, Guy's Campus
London, SE1 1UL
United Kingdom
Phone: 44 02078486283
Email: lindsay.bearne@kcl.ac.uk

Abstract

Background: Rheumatoid arthritis (RA) is a disabling, inflammatory joint condition affecting 0.5%-1% of the global population. Physical activity (PA) and exercise are recommended for people with RA, but uptake and adherence tend to be low. Smartphone apps could assist people with RA to achieve PA recommendations. However, it is not known whether high quality, evidence-informed PA apps that include behavior change techniques (BCTs) previously identified as effective for PA adherence are available for people with RA.

Objective: This study aims to systematically identify apps that include goals to facilitate PA for adults with RA and assess app quality and content for the inclusion of relevant BCTs against recommendations for cardiorespiratory, resistance, flexibility, and neuromotor PA and exercise.

Methods: A systematic search of the Apple App Store and Google Play Store in the United Kingdom was conducted to identify English language apps that promote PA for adults with RA. Two researchers independently assessed app quality (mobile app rating scale [MARS]; range 0-5) and content (BCT Taxonomy version 1, World Health Organization, the American College of Sports Medicine, and the European League against Rheumatism recommendations for PA). The completeness of reporting of PA prescription was evaluated using a modified version of the Consensus on Exercise Reporting Template (CERT; range 0-14).

Results: A total of 14,047 apps were identified. Following deduplication, 2737 apps were screened for eligibility; 6 apps were downloaded (2 on the Apple App Store and 4 on the Google Play Store), yielding 4 unique apps. App quality varied (MARS score 2.25-4.17). Only 1 app was congruent with all aspects of the PA recommendations. All apps completely or partially recommended flexibility and resistance exercises, 3 apps completely or partially advised some form of neuromotor exercise, but only 2 offered full or partial guidance on cardiorespiratory exercise. Completeness of exercise reporting was mixed (CERT scores 7-14 points) and 3-7 BCTs were identified. Two BCTs were common to all apps (information about health consequences and instruction on how to perform behavior). Higher quality apps included a greater number of BCTs and were more closely aligned to PA guidance. No published trials evaluating the effect of the included apps were identified.

Conclusions: This review identifies 4 PA apps of mixed quality and content for use by people with RA. Higher quality apps were more closely aligned to PA guidance and included a greater number of BCTs. One high-quality app (Rheumatoid Arthritis

Information Support and Education) included 7 BCTs and was fully aligned with PA and exercise guidance. The effect of apps on PA adherence should be established before implementation.

(JMIR Mhealth Uhealth 2020;8(7):e18495) doi:10.2196/18495

KEYWORDS

rheumatoid arthritis; physical activity; exercise; mobile applications; behavior change techniques; mobile phone

Introduction

Background

Rheumatoid arthritis (RA) is a disabling, autoimmune inflammatory condition that affects 0.5% to 1% of the global population [1,2]. Evidence-informed guidelines recommend physical activity (PA) and exercise for people with RA tailored to an individual's baseline PA level, disease activity, and symptoms. PA prescriptions comprise exercise type (ie, cardiorespiratory, resistance, flexibility, or neuromotor training), the number of sets/repetitions, load and/or intensity, recovery time/method of progression, and frequency and duration of exercise sessions [3-7].

However, people with RA tend not to meet the recommended levels of PA [8,9], and there are several barriers that can make changing PA behavior challenging for people with RA without appropriate support and guidance [8-11].

Interventions that target the factors that influence adherence to PA using behavior change techniques (BCTs; ie, strategies that help an individual change their behavior) can improve PA levels and health outcomes [12-14]. Many people with RA would like help to increase their PA from health care professionals [9], but interventions can be difficult to implement due to lack of time, resources, and/or the limited number of appropriately trained health care professionals [15-17]. Consequently, novel methods of delivering interventions that can be tailored for people with RA are needed to support adherence to PA.

With the rapid increase in the availability of mobile apps [18], the development and use of high-quality apps may be a promising approach to support people with RA to reach evidence-informed PA recommendations. However, appraisals of apps for RA symptom monitoring and self-management, including PA, suggested that app quality and content were heterogeneous and they did not consistently provide evidence-based management strategies or include validated symptom measures [19-22]. Apps were seldom developed in collaboration with people with RA or clinicians, and older adults often found them difficult to use [21-23].

The effects of digital interventions on PA adherence in people with RA are unclear. A systematic review of randomized controlled trials (4 trials; n=492 participants) found limited evidence of an effect of interactive digital interventions (ie, interactive information and communication technologies to support behavior change, such as online fora) on PA adherence in people with RA or juvenile idiopathic arthritis [24]. Trials included 3 to 9 BCTs [24]. No quality rating of the interactive digital interventions using standardized measures was conducted. Thus, the systematic identification and evaluation of apps that can support adherence to PA and evidence to support the effectiveness of these apps are required.

Recommendations suggest that the safety, quality, and content of self-management apps (including PA apps) for people with RA should be considered during all stages of development, evaluation, and implementation [25]. Features such as engagement, esthetics, functionality, and information quality should be assessed using reliable tools, such as the mobile app rating scale (MARS) [26]. Content should be evaluated for (1) BCTs using a recognized framework (eg, BCT Taxonomy version 1, BCTT v1) [27], (2) congruence with evidence-informed recommendations on PA (eg, World Health Organization and the European League against Rheumatism) [3,5] and exercise (eg, American College of Sports Medicine) [4,6,28], and (3) described using standardized reporting formats such as the Consensus on Exercise Reporting Template (CERT) [7,29].

Objectives

This study aims to systematically identify and evaluate the quality and content of publicly available mobile apps aiming to support the uptake of and adherence to PA in people with RA.

Methods

Protocol and Registration

The review protocol was developed a priori by a team consisting of physiotherapists, rheumatologists, and health psychologists with experience in conducting systematic reviews, evaluating mobile health apps, PA, and behavior change interventions. The protocol was not eligible for registration on the International Prospective Register of Systematic Reviews (PROSPERO) as PROSPERO does not register reviews of apps.

Data Sources

Where possible, this review followed the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines for reporting systematic reviews [30].

Systematic individual searches of the Google Play (Samsung Galaxy s8 operating G950FXXS4DSD3/ G950FOXM4DSBA/G950FXXS4DSD3 software with Android version 9) and Apple App Store (iOS 12.3.1 software operating on iPhone 7) were conducted on June 19 and 20, 2019.

Key search terms for RA, PA, and exercise were identified from the literature [22,24] and refined by all coauthors. Search terms were used in isolation and combination to search for all relevant apps (Table 1).

To ensure that all potentially relevant apps were captured, searches for *rheumatoid arthritis* and *arthritis* were also conducted in the United Kingdom National Health Service app library and the Apple App Store in New Zealand, Australia, Canada, and United States using the fnd website [31].

Table 1. Key search terms used for identifying mobile apps targeting physical activity in people with rheumatoid arthritis.

Search terms	Rheumatic	Arthritis	RA[a]	Rheumatoid arthritis	Inflammatory arthritis
Physical activity	Rheumatic physical activity	Arthritis physical activity	RA physical activity	Rheumatoid arthritis physical activity	Inflammatory arthritis physical activity
Exercise	Rheumatic exercise	Arthritis exercise	RA exercise	Rheumatoid arthritis exercise	Inflammatory arthritis exercise
Walking	Rheumatic walking	Arthritis walking	RA walking	Rheumatoid arthritis walking	Inflammatory arthritis walking
Strength	Rheumatic strength	Arthritis strength	RA strength	Rheumatoid arthritis strength	Inflammatory arthritis strength
Fitness	Rheumatic fitness	Arthritis fitness	RA fitness	Rheumatoid arthritis fitness	Inflammatory arthritis fitness
Training	Rheumatic training	Arthritis training	RA training	Rheumatoid arthritis training	Inflammatory arthritis training
Running	Rheumatic running	Arthritis running	RA running	Rheumatoid arthritis running	Inflammatory arthritis running

[a]RA: rheumatoid arthritis.

App Selection

Eligibility Criteria

Deduplication based on app store title and description (provided in the *read more* section in the app) was conducted by 2 independent reviewers (R1 and R2), and potentially eligible apps were downloaded on one or both devices, where possible. The information section for each potentially suitable app was reviewed against the eligibility criteria. Inclusion criteria comprised (1) a smartphone-based app available in at least one app store; (2) targeted at adults (≥18 years) with RA as users specifically, (3) focused on promoting uptake and adherence to PA or exercise, and (4) available in English. Apps were excluded if they (1) targeted people with a condition other than RA, (2) were solely for use by health care practitioners, and (3) were specific clinic, congress/conference, or product apps. No cost restrictions were applied and full app content was purchased, if required.

Data Items and Extraction

Data extraction was conducted by 2 independent reviewers (R1 and R2) using a data extraction tool developed a priori. Each app was used for at least 10 min. The following app characteristics were extracted: app name, platform, version, developer, stakeholder involvement in app development, size, star rating, number of installs, privacy policy statements, and medical product status. The availability of published trials evaluating app efficacy or effectiveness was checked on developer websites and by searching electronic databases (last search in June 2019) using the search strategy described in our systematic review to synthesize the evidence for the effectiveness of mobile apps designed to enhance adherence to PA for people with inflammatory arthritis (PROSPEROCRD42019129341) [32]. In addition, searches in Google, Google Scholar, and PubMed were undertaken using the app name as a search term. The review of search results was stopped after the first 50 irrelevant results.

Quality Appraisal of Individual Apps

The quality of the included apps was assessed using the simple and reliable MARS [26]. A total of 23 items were rated on a 5-point Likert scale (1=inadequate to 5=excellent) and summarized into 4 categories: engagement (5 items), functionality (4 items), esthetics (3 items), information quality (7 items), and a subjective quality scale (4 items). The mean score for each category and the MARS total score (the mean of the 4 category scores excluding subjective quality scale) was calculated (maximum score=5). Item 19 of the MARS, *evidence base*, was excluded from all calculations because no apps had been studied in clinical trials, as specified by Stoyanov et al [26].

Content Analysis of Individual Apps

PA and exercise recommendations for people with RA were evaluated (yes, partial, or no) for congruence with evidence-informed guidance for cardiorespiratory, resistance, flexibility, and neuromotor PA [3-6]. This three-point scale reflected the format of the checklists used to assess the fidelity of rehabilitation and exercise intervention delivery [33-35]. Apps were considered to be fully aligned with the guidance if they included details of exercise type, intensity (eg, load, sets, repetitions), frequency, and time/duration of exercise/sessions (*yes*), partially aligned with guidance if they included at least one of these parameters (*partial*) and not aligned with the guidance if they did not include any of these parameters (*no*). Content was assessed against the PA guidance that was most appropriate for the age and exercise experience of the target user for each app.

Guidance indicates that adults (aged 18 to 64 years) should perform ≥150 min of cardiorespiratory exercise at moderate

intensity or 75 min of vigorous/high-intensity activity or an equivalent combination per week in bouts ≥10 min. Adults who are novice (ie, unaccustomed to exercise) or intermediate (ie, some experience of exercise) exercisers should also perform between 2 and 4 sets of resistance exercises (8 and 12 repetitions per set) at a moderate/hard intensity (ie, 60% to 70% of one repetition maximum) for each major muscle group on at least two days per week. However, experienced exercisers (ie, engaged in habitual exercise) should work at a hard or very hard intensity (ie, ≥80% one repetition maximum). Exercises to increase or maintain flexibility are advised at least twice per week for a minimum of 10 min. Each stretch should be held to the point of tightness or slight discomfort for 10 to 30 seconds up to a total of 60 seconds of stretching time per exercise. Adults with poor mobility and balance are advised to perform 20 to 30 min of neuromotor exercises (eg, balance, agility, coordination, proprioceptive exercise training, or multifaceted activities such as yoga) on at least two days per week to enhance physical function, balance, and prevent falls [3-5].

Guidelines for older adults (≥65 years) and adults between 50 and 64 years with long-term conditions are similar, although activities should be tailored to health and disease status, baseline fitness, and initially comprise very light or light intensity resistance exercises (ie, 40% to 50% of one repetition maximum) and stretches of a 30- to 60-second duration [3,5,6].

The CERT is a reliable, 16-item (7 domain) specification to evaluate the reporting of exercise interventions (Table 2) [7,29,36]. Each item was rated (yes=1, no=0, or not applicable) and summed to produce a total CERT score. It was not possible to score items 11 or 16; therefore, the maximum possible CERT score was 14 [7,29].

Table 2. Abbreviated item description for the Consensus of Exercise Reporting Template.

Item category and item number	Abbreviated item description
What	
1	Description of type of exercise equipment
Who	
2	Description of qualifications/expertise/training of instructor
How	
3	Description of whether exercises are performed individually or in a group
4	Description of whether exercises are supervised/unsupervised
5	Description of the measurement/reporting of adherence to exercise
6	Description of motivational strategies
7	Decision rules for determining exercise progression and how exercise was progressed
8	Description of each exercise to enable replication (eg, illustrations, photos)
9	Description of any home programme component
10	Description of any nonexercise component
11	Description of the type/number of adverse events that occurred during exercise
Where	
12	Description of exercise setting
When, how much	
13	Description of exercise intervention and dosage
Tailoring	
14	Description of whether exercises are generic or tailored to the individual
15	Decision rule for starting level of exercise
How well	
16	Description of whether the exercise is delivered and performed as planned

Two postgraduate physiotherapy students (R1 and R2) independently assessed app quality using the MARS and PA and exercise content for congruence with evidence-informed recommendations and reporting using the CERT. Discrepancies were resolved by discussion and another assessor acted as an arbiter if required. As recommended by the MARS developers, assessors received training on applying the MARS by viewing the MARS training video [26,37] and bespoke training to assess app content for congruence with PA recommendations and completeness of reporting by a member of the CERT development group [7].

Before rating the included apps, the assessors evaluated up to 5 randomly selected apps (identified in the search but previously excluded from the analysis) and discussed their results to ensure

an understanding of the MARS, PA and exercise recommendations, and CERT items and processes.

The BCTT v1 is a comprehensive and reliable tool that consists of 93 distinct BCTs that can be used to identify specific *active components* of behavioral interventions [27]. It is used to design and retrospectively evaluate behavioral health interventions, such as PA [38].

Two postgraduate health psychologists (R3 and R4) completed the training for applying the BCTT v1 before independently coding BCTs in the included apps [27]. Discrepancies were resolved by discussion, and another rater acted as an arbiter if required.

Statistical Analysis

Interrater reliability using an intraclass correlation coefficient (ICC) was calculated using IBM SPSS version 25.0 (two-way random-effects model of absolute agreement between single ratings). Scores <0.5 were considered poor agreement, moderate reliability (0.5-0.75), good reliability (0.75-0.9), and excellent reliability (>0.9) [39].

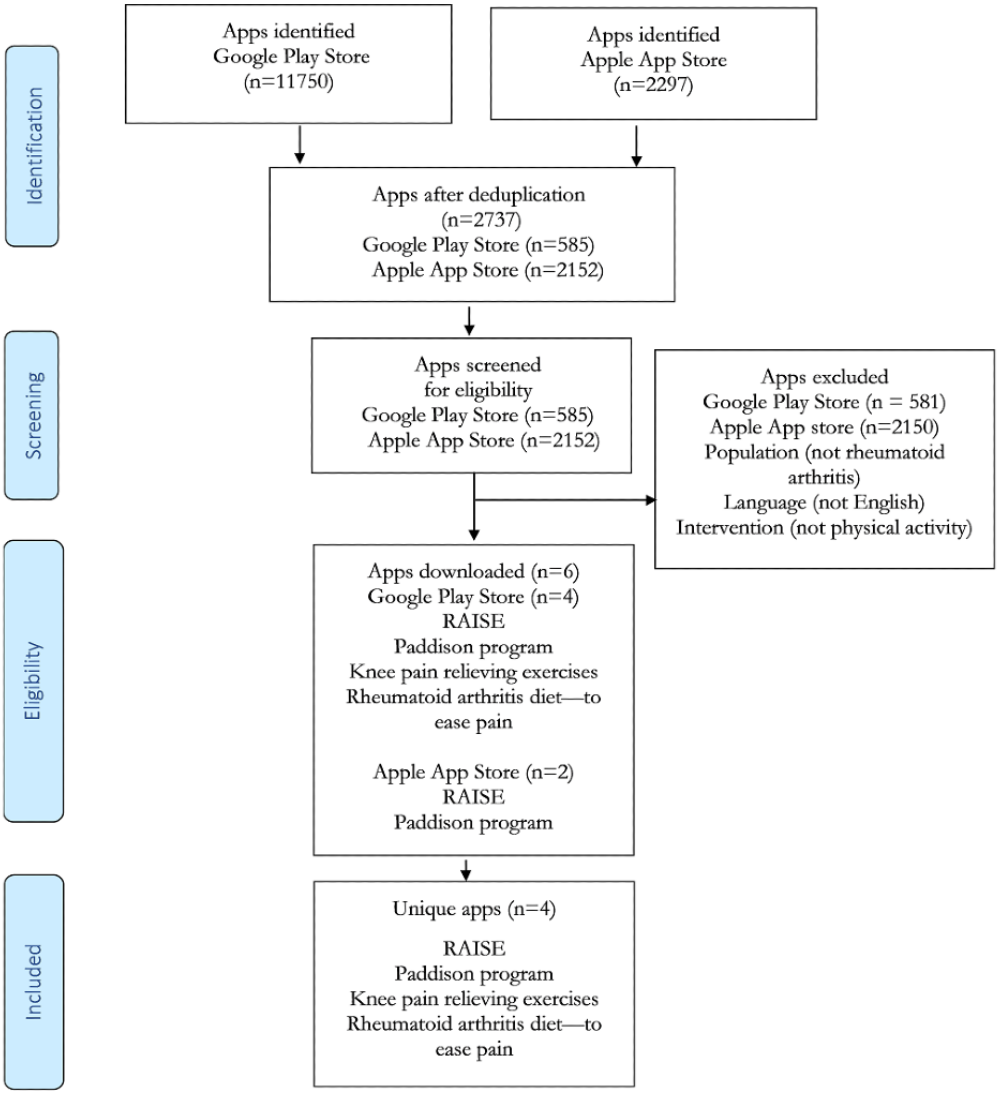
Results

App Selection

The systematic search identified 14,047 apps (UK Google Play Store n=11,750 and Apple App Store n=2297). No further apps were identified from the United Kingdom National Health Service app library and the Apple App Store in the United States, New Zealand, Australia, or Canada.

Following deduplication, 2737 app titles and descriptions were screened and 6 apps (Google Play Store n=4 and Apple App Store n=2) met the eligibility criteria, yielding 4 unique apps (Rheumatoid Arthritis Information Support and Education [RAISE], Paddison program, Knee pain relieving exercises, Rheumatoid arthritis diet—to ease pain; Figure 1). Two apps (Knee pain relieving exercises and Rheumatoid arthritis diet—to ease pain) were exclusively available on the Google Play Store, and 2 apps were available on both platforms (RAISE and Paddison program; Figure 1).

Figure 1. Flow diagram illustrating the process of mobile app selection.



Characteristics of the Included Apps

The characteristics of the included apps are detailed in Table 3. A total of 3 apps were developed by or in conjunction with people with RA (RAISE and Paddison program) or clinicians (RAISE and Knee pain relieving exercises). Apple App store star ratings of included apps ranged from 2 (RAISE on Android) to 5 stars (RAISE on iOS), although the number of ratings was generally low (Knee pain relieving exercises n=140 and RAISE-iOS n=1 ratings). All mobile apps were free to download. The Paddison program app required a payment of US $69 to access the Paddison program resources and daily videos for 12 days (*essential package*). A fee of US $99 provided access to all content, that is, Paddison program resources, 12 daily videos, and additional updated video/website content (*advanced healing package*).

The apps had been installed 500 (RAISE on Android) to 50,000 (Knee pain relieving exercises) times. None of the included apps had a medical product status. No published trials evaluating the effect of the included apps were identified.

Table 3. Characteristics and quality rating of mobile apps targeting physical activity in people with rheumatoid arthritis.

App	Knee pain relieving exercises	Rheumatoid arthritis diet—to ease pain	RAISE[a]		Paddison program	
A platform	Android[b]	Android[b]	Android	iOS	Android	iOS
Developer	Dr Kavin Khatri	TP Topics	Publicis Dublin	Publicis Dublin	Paddison program	Paddison program
Stakeholder involvement in app development	Clinician	N/A[c]	Patient organization and clinicians	Patient organization and clinicians	Patient representative	Patient representative
Version	1.0.0	1.1.2	1.1.2	1.0.0	1.2.0	2.1.0
Size, MB[d]	4.84	16.27	10.18	18.90	36.31	48.80
Star rating (number of ratings)	4.0 (140)	3.0 (3)	2.0 (2)	5.0 (1)	4.4 (8)	—[e]
Number of installations	>50,000	>1000	>500	—	>1000	—
Privacy policy statement (yes/no)	Yes	Yes	Yes	Yes	Yes	Yes
Medical product status (yes/no)	No	No	No	No	No	No
Peer-reviewed publications (yes/no)	No	No	No	No	No	No
MARS[f]-engagement	2.00	2.40	3.90	3.90	3.70	3.70
MARS-functionality	4.50	3.12	4.38	4.38	3.88	3.88
MARS-esthetics	2.33	3.00	4.33	4.33	3.66	4.00
MARS-information	2.83	2.83	4.08	4.08	3.33	3.33
MARS-subjective	1.88	2.84	3.25	3.25	2.00	2.00
MARS-total	2.92	2.83	4.17	4.17	3.64	3.73

[a]RAISE: Rheumatoid Arthritis Information Support and Education.

[b]Apps not available on the iOS platform.

[c]N/A: not applicable.

[d]MB: megabytes.

[e]Data not available.

[e]MARS: mobile app rating scale.

Quality Appraisal of the Included Apps

The MARS total score ranged from 2.83 (Rheumatoid arthritis diet—to ease pain) to 4.17 (RAISE), indicating variation in app quality (Table 3). The esthetics category showed the greatest variability (2.30 for Knee pain relieving exercises to 4.33 for RAISE). A total of 3 of the 4 apps (Knee pain relieving exercises, Rheumatoid arthritis diet—to ease pain, and RAISE) scored their highest MARS score for the functionality category and their lowest MARS score for the engagement category. The Paddison program achieved its highest score for esthetics on the iOS platform but not the Android platform. This is because some items (eg, logos) were distorted or not visible on the Android platform. The Paddison program achieved its lowest score for the information quality category on both platforms.

Only 1 app (RAISE) was rated >4 out of 5 for the MARS total score and achieved the highest score in all categories (ie, engagement, esthetics, information, subjective) except

functionality. The Knee pain relieving exercise app scored highest for functionality.

The interrater reliability for the MARS indicated excellent agreement for all subscales: engagement (ICC 0.96; 95% CI 0.71-0.99), functionality (ICC 0.92; 95% CI 0.01-0.99), esthetics (ICC 1.00 absolute agreement), information (ICC 0.95; 95% CI 0.64-0.99), subjective (ICC 0.98; 95% CI 0.83-0.99), and the MARS total score (ICC 0.99; 95% CI 0.92-0.99).

Content Analysis of the Included Apps

The content of the included apps is summarized in Tables 4 to 6. All apps completely or partially recommended flexibility and resistance exercises. Three apps completely or partially advised some form of neuromotor exercise (Rheumatoid arthritis diet—to ease pain, RAISE, and Paddison program). Two apps offered full or partial guidance on cardiorespiratory exercise (Rheumatoid arthritis diet—to ease pain and RAISE; Table 4).

Only 1 app (RAISE) was congruent with all aspects of the evidence-informed PA recommendations. However, the RAISE app recommended 30 min of cardiorespiratory exercise daily (equivalent to 210 min per week), which is in excess of the minimum weekly PA recommendations.

The Rheumatoid arthritis diet—to ease pain app was primarily focused on providing dietary advice but also recommended all types of PA. It was rated as partially adhering to the guidance because PA dosages were not specified.

Table 4. Congruence with evidence-informed recommendations in mobile apps targeting physical activity in people with rheumatoid arthritis.

Physical activity and exercise recommendations	Knee pain relieving exercises	Rheumatoid arthritis diet—to ease pain	RAISE[a]	Paddison program
Cardiorespiratory exercise	No	Partial	Yes	No
Resistance exercise	Partial	Partial	Yes	Partial
Flexibility exercise	Partial	Partial	Yes	No
Neuromotor exercise	No	Partial	Yes	Partial

[a]RAISE: Rheumatoid Arthritis Information Support and Education.

Completeness of exercise reporting in apps using CERT is summarized in Table 5. All apps offered information on the (1) exercise format (individual, unsupervised, or home-based), (2) the potential positive benefits of exercise as a motivational strategy, and (3) nonexercise advice (eg, lifestyle or medication). All apps specified some form of guidance for exercise progression, although the parameters for exercise progression did not always align with guidance. All apps, except Rheumatoid arthritis diet—to ease pain, provided details of (1) the exercise equipment required; (2) the qualification/training of the exercise instructor; and (3) descriptions, adaptations, and dosage of the exercises. Two apps offered decision rules to help users determine an initial exercise dose (RAISE and Paddison program). One app (RAISE) offered the option to document exercise adherence.

The RAISE app reported all types of PA completely and achieved the maximum possible score (14 out of 14). The Rheumatoid arthritis diet—to ease the pain app achieved the lowest CERT score (7 out of 14), predominantly because it did not explicitly report PA dosages. Interrater reliability for CERT scoring was good (ICC 0.796; 95% CI 0.806-0.933).

Table 5. Physical activity and exercise reporting in accordance with the Consensus of Exercise Reporting Template in mobile apps targeting physical activity in people with rheumatoid arthritis.

Items and item number	Knee pain relieving exercises	Rheumatoid arthritis diet—to ease pain	RAISE[a]	Paddison program
What				
1	Yes	No	Yes	Yes
Who				
2	Yes	No	Yes	Yes
How				
3	Yes	Yes	Yes	Yes
4	Yes	Yes	Yes	Yes
5	No	No	Yes	No
6	Yes	Yes	Yes	Yes
7	Yes	Yes	Yes	Yes
8	Yes	No	Yes	Yes
9	Yes	No	Yes	Yes
10	Yes	Yes	Yes	Yes
11	N/A[b]	N/A	N/A	N/A
Where				
12	Yes	Yes	Yes	Yes
When and how much				
13	Yes	No	Yes	No
Tailoring				
14	Yes; tailored	Yes; generic	Yes; tailored	Yes; tailored
15	No	No	Yes	Yes
How well				
16	N/A	N/A	N/A	N/A
Total Score	12	7	14	12

[a]RAISE: Rheumatoid Arthritis Information Support and Education.

[b]N/A: not applicable.

The apps included between 3 (Rheumatoid arthritis diet—to ease pain) and 7 (RAISE and Paddison program) BCTs. Two BCTs were identified in all apps (instructions on how to perform behavior and information about health consequences; Table 6). Credible source was present in all apps but the Paddison program. All apps included demonstration of behavior apart from the Rheumatoid arthritis diet—to ease pain app. The RAISE and Paddison program apps both included 7 BCTs. Five of these BCTs were common to both apps (instruction on how to perform behavior, information about health consequences, demonstration of behavior, goal setting, and social comparison). Only the RAISE app included self-monitoring of behavior. The interrater reliability for BCTs was good (ICC 0.874; 95% CI 0.799-0.921).

Table 6. Behavior change techniques included in mobile apps targeting physical activity in people with rheumatoid arthritis.

Behavior change techniques	Knee pain relieving exercises	Rheumatoid arthritis diet—to ease pain	RAISE[a]	Paddison program
Instruction on how to perform behavior	Yes	Yes	Yes	Yes
Information about health consequences	Yes	Yes	Yes	Yes
Demonstration of behavior	Yes	No	Yes	Yes
Credible source	Yes	Yes	Yes	No
Goal setting behavior	No	No	Yes	Yes
Social comparison	No	No	Yes	Yes
Graded task	Yes	No	No	Yes
Self-monitoring behavior	No	No	Yes	No
Generalization of target behavior	No	No	No	Yes
Framing/reframing	No	No	No	Yes
Total number	5	3	7	7

[a]RAISE: Rheumatoid Arthritis Information Support and Education.

Discussion

Principal Findings

This is the first systematic identification, quality appraisal, and content analysis of widely available PA and exercise apps for people with RA. Up to June 20, 2019, there were only 4 unique apps that met our inclusion criteria available on iOS and/or Android platforms. The quality and content of the included apps varied considerably and did not always align with PA recommendations. Notably, higher-quality apps tended to include a greater number of BCTs and most closely aligned to PA recommendations. The highest quality app (RAISE) was the only app to explicitly report PA prescriptions aligned to evidence-informed recommendations for people with RA and embedded the highest number of BCTs.

Despite guidance for the development and evaluation of mobile apps [25,40], the quality ratings of the apps included in our review were mixed. This finding is consistent with reviews of publicly available rheumatology self-management apps [20-22] and apps targeting PA and exercise in the general population [41,42]. For example, Simoes et al [43] identified 51 moderate-quality PA apps for use by the general population (MARS total score 3.16 to 4.41) with the functionality and esthetics domains scoring most highly. This is broadly similar to our findings and suggests that the included apps were intuitive, logical to follow, easy to learn and navigate, which is particularly important for people with RA who may have fatigue or hand and wrist disability [44,45]. However, the apps achieved lower ratings for information quality and engagement and to optimize utility, app content should be high quality, interesting, simple to understand, and have the option to be tailored with user data [25].

The RAISE app was the only app that fully and explicitly reported PA prescription that aligned to all evidence-informed recommendations for people with RA. This may be because the RAISE app was the only app to be developed in conjunction with both clinicians and people with RA; thus, concordance with evidence-based guidelines and the acceptability and user experience of people with RA were likely to be considered from inception.

Even though the RAISE app was rated as congruent with all PA guidance, it recommended a weekly duration of cardiorespiratory PA in excess of the minimum dosage stated in the guidelines, which may be inappropriate for novice exercisers with RA. Interestingly, the RAISE app had the lowest number of installs and number of user ratings, suggesting that it is not widely used by people with RA. This may be because people with RA were unaware of the app or because they found the PA recommendations to be too ambitious and unacceptable.

The other apps did not completely align with PA recommendations, similar to other research [41,42]. Common reasons for apps not completely adhering to the PA guidance were the lack of or incorrect specific PA dosages (eg, sets, repetitions). For example, the Rheumatoid arthritis diet—to ease pain app recommended all PA types but did not provide specific PA dosages so was only partially congruent with the guidance. This may be because the primary focus of this app was dietary advice.

PA prescription may need to be modified for people with RA [3,4,6] and some apps offered a reduced starting dose or suggested ways to tailor exercises or exercise progression. However, these recommendations did not always align with guidance. Incomplete or unclear PA prescriptions maybe confusing to users who are novice exercisers, and inappropriate prescriptions may impact users' engagement and adherence with PA and compromise PA effectiveness or safety [7].

All apps included some BCTs that may promote adherence to PA, and the higher-quality apps included a greater number of BCTs. No app contained more than 7 BCTs, which is similar to the findings of reviews of apps targeting PA in the general

population [41,43,46,47]. Although the optimum number of BCTs needed to support PA adherence is not known, a recent systematic review including 8 randomized controlled trials (1018 participants) found that interventions with less than 7 BCTs were most effective at enhancing adherence to exercise in people with persistent musculoskeletal pain [48]. In addition, there was a moderate level of evidence that 5 BCTs (social support [unspecified], goal setting [behavior], instruction of behavior, demonstration of behavior, and behavior practice/rehearsal) supported PA adherence [48]. Two of the higher-quality apps in our review included 3 of these 5 BCTs adding evidence-based integrity, although the effectiveness of these apps has not been investigated.

Methodological Considerations

The strengths of this review include the comprehensive search of both the UK Google Play Store and Apple App Store. This was complemented by a search of the United Kingdom National Health Service app library and the Apple App Store in the United States, Canada, New Zealand, and Australia. No new apps were identified, suggesting that all English language apps were captured. However, this review only focuses on publicly available apps for use by people with RA, so we may not have captured apps primarily designed for research.

Where possible, we followed rigorous processes that were aligned to PRISMA guidelines [30]. Two reviewers independently screened identified apps for eligibility, extracted data, and rated the quality and content of the apps using standardized tools. Interrater reliability was good or excellent, which lends confidence to our findings. App quality was assessed with the widely used MARS [21,22,26]. However, the MARS rating is subjective, and people with RA may have different perceptions of the app features to our assessors, who did not have RA.

Notably, it is possible to reset star ratings in the iOS app store when new app versions are released. It is not known if the star ratings extracted at the time of our appraisal refer to overall or current app versions ratings. However, all included apps have limited versions, so the impact of this on our findings is likely to be minimal. Finally, no evaluation of the content of the privacy policies of included apps was completed, so we do not know whether the policy adequately protects users' rights.

This comprehensive review of PA apps for people with RA identified 4 apps of mixed quality and content. Higher quality apps more closely aligned to PA guidance and included a greater number of BCTs previously shown to promote PA. The RAISE app was the highest quality app. Future apps should be rigorously developed with key stakeholders, and include evidence-based PA guidance and BCTs, to optimize their acceptability and impact on PA. Robustly designed research into the effect of apps on PA adherence is crucial before implementation.

Authors' Contributions

LMB wrote the draft manuscript. All authors discussed the draft and provided comments and suggestions for change. All authors approved the final report.

Conflicts of Interest

None declared.

References

1. Abhishek A, Doherty M, Kuo C, Mallen CD, Zhang W, Grainge MJ. Rheumatoid arthritis is getting less frequent-results of a nationwide population-based cohort study. Rheumatology (Oxford) 2017 May 01;56(5):736-744 [FREE Full text] [doi: 10.1093/rheumatology/kew468] [Medline: 28064207]
2. Hunter TM, Boytsov NN, Zhang X, Schroeder K, Michaud K, Araujo AB. Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004-2014. Rheumatol Int 2017 Sep;37(9):1551-1557. [doi: 10.1007/s00296-017-3726-1] [Medline: 28455559]
3. Osthoff AR, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018 Sep;77(9):1251-1260. [doi: 10.1136/annrheumdis-2018-213585] [Medline: 29997112]
4. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, American College of Sports Medicine. American college of sports medicine position stand quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
5. World Health Organization. Global recommendations on physical activity for health. In: World Health Organization Report. Switzerland: World Health Organization Press; 2010:1-58.
6. Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical activity and public health in older adults: recommendation from the American college of sports medicine and the American heart association. Med Sci Sports Exerc 2007 Aug;39(8):1435-1445. [doi: 10.1249/mss.0b013e3180616aa2] [Medline: 17762378]
7. Slade S, Dionne CE, Underwood M, Buchbinder R, Beck B, Bennell K, et al. Consensus on exercise reporting template (CERT): modified Delphi study. Phys Ther 2016 Oct;96(10):1514-1524. [doi: 10.2522/ptj.20150668] [Medline: 27149962]

8. Sokka T, Häkkinen A, Kautiainen H, Maillefert JF, Toloza S, Hansen TM, QUEST-RA Group. Physical inactivity in patients with rheumatoid arthritis: data from twenty-one countries in a cross-sectional, international study. Arthritis Rheum 2008 Jan 15;59(1):42-50 [FREE Full text] [doi: 10.1002/art.23255] [Medline: 18163412]
9. Manning VL, Hurley MV, Scott DL, Bearne LM. Are patients meeting the updated physical activity guidelines? Physical activity participation, recommendation, and preferences among inner-city adults with rheumatic diseases. J Clin Rheumatol 2012 Dec;18(8):399-404. [doi: 10.1097/RHU.0b013e3182779cb6] [Medline: 23188205]
10. van Zanten JJ, Rouse PC, Hale ED, Ntoumanis N, Metsios GS, Duda JL, et al. Perceived barriers, facilitators and benefits for regular physical activity and exercise in patients with rheumatoid arthritis: a review of the literature. Sports Med 2015 Oct;45(10):1401-1412 [FREE Full text] [doi: 10.1007/s40279-015-0363-2] [Medline: 26219268]
11. Bearne LM, Manning VL, Choy E, Scott DL, Hurley MV. Participants' experiences of an education, self-management and upper extremity exercise training for people with rheumatoid arthritis programme (EXTRA). Physiotherapy 2017 Dec;103(4):430-438. [doi: 10.1016/j.physio.2016.12.002] [Medline: 28823567]
12. Thomsen T, Aadahl M, Beyer N, Hetland ML, Løppenthin K, Midtgaard J, et al. The efficacy of motivational counselling and SMS reminders on daily sitting time in patients with rheumatoid arthritis: a randomised controlled trial. Ann Rheum Dis 2017 Sep;76(9):1603-1606 [FREE Full text] [doi: 10.1136/annrheumdis-2016-210953] [Medline: 28584189]
13. Knittle K, de Gucht V, Hurkmans E, Peeters A, Ronday K, Maes S, et al. Targeting motivation and self-regulation to increase physical activity among patients with rheumatoid arthritis: a randomised controlled trial. Clin Rheumatol 2015 Feb;34(2):231-238. [doi: 10.1007/s10067-013-2425-x] [Medline: 24213780]
14. Manning VL, Hurley MV, Scott DL, Coker B, Choy E, Bearne LM. Education, self-management, and upper extremity exercise training in people with rheumatoid arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2014 Feb;66(2):217-227 [FREE Full text] [doi: 10.1002/acr.22102] [Medline: 23925924]
15. Ndosi M, Ferguson R, Backhouse MR, Bearne L, Ainsworth P, Roach A, et al. National variation in the composition of rheumatology multidisciplinary teams: a cross-sectional study. Rheumatol Int 2017 Sep;37(9):1453-1459 [FREE Full text] [doi: 10.1007/s00296-017-3751-0] [Medline: 28551723]
16. British SOR. National Early Inflammatory Arthritis Audit (NEIAA) 1st Annual Report. In: British Society of Rheumatology. -: British Society of Rheumatology; Oct 2019:1-51.
17. McCuish WJ, Bearne LM. Do inpatient multidisciplinary rehabilitation programmes improve health status in people with long-term musculoskeletal conditions? A service evaluation. Musculoskeletal Care 2014 Dec;12(4):244-250. [doi: 10.1002/msc.1072] [Medline: 24840767]
18. Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
19. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile health apps for self-management of rheumatic and musculoskeletal diseases: systematic literature review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730 [FREE Full text] [doi: 10.2196/14730] [Medline: 31769758]
20. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. J Clin Rheumatol 2019 Apr;25(3):133-141. [doi: 10.1097/RHU.0000000000000800] [Medline: 29933327]
21. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: review and analysis using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
22. 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]
23. Bhattarai P, Newton-John TR, Phillips JL. Quality and usability of arthritic pain self-management apps for older adults: a systematic review. Pain Med 2018 Mar 01;19(3):471-484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]
24. Griffiths AJ, White CM, Thain PK, Bearne LM. The effect of interactive digital interventions on physical activity in people with inflammatory arthritis: a systematic review. Rheumatol Int 2018 Sep;38(9):1623-1634 [FREE Full text] [doi: 10.1007/s00296-018-4010-8] [Medline: 29556750]
25. Najm A, Nikiphorou E, Kostine M, Richez C, Pauling JD, Finckh A, et al. EULAR points to consider for the development, evaluation and implementation of mobile health applications aiding self-management in people living with rheumatic and musculoskeletal diseases. RMD Open 2019;5(2):e001014 [FREE Full text] [doi: 10.1136/rmdopen-2019-001014] [Medline: 31565245]
26. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
27. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
28. American College Of Sports Medicine. ACSMs Guidelines for Exercise Testing and Prescription. Netherlands: Wolters Kluwer; Feb 10, 2017.

29. Slade SC, Dionne CE, Underwood M, Buchbinder R. Consensus on exercise reporting template (CERT): explanation and elaboration statement. Br J Sports Med 2016 Dec;50(23):1428-1437. [doi: 10.1136/bjsports-2016-096651] [Medline: 27707738]
30. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Br Med J 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
31. Mack J, Mack R. fnd. 2014. URL: https://fnd.io/ [accessed 2019-06-19]
32. Sekhon M, White C, Amirova A, Godfrey E, Revenas A, Bearne L. Effectiveness of web-based and mobile health interventions designed to support adherence to physical activity for people with inflammatory arthritis: a systematic review. PROSPERO International prospective register of systematic reviews. United Kingdom: University of York; 2019. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019129341 [accessed 2020-05-08]
33. Breitenstein SM, Gross D, Garvey CA, Hill C, Fogg L, Resnick B. Implementation fidelity in community-based interventions. Res Nurs Health 2010 Apr;33(2):164-173 [FREE Full text] [doi: 10.1002/nur.20373] [Medline: 20198637]
34. Godfrey E, Wileman V, Galea Holmes M, McCracken LM, Norton S, Moss-Morris R, et al. Physical therapy informed by acceptance and commitment therapy (PACT) versus usual care physical therapy for adults with chronic low back pain: a randomized controlled trial. J Pain 2019 Jun 05 [FREE Full text] [doi: 10.1016/j.jpain.2019.05.012] [Medline: 31173921]
35. Bearne L, Galea Holmes M, Bieles J, Eddy S, Fisher G, Modarai B, et al. Motivating structured walking activity in people with intermittent claudication (MOSAIC): protocol for a randomised controlled trial of a physiotherapist-led, behavioural change intervention versus usual care in adults with intermittent claudication. BMJ Open 2019 Aug 24;9(8):e030002 [FREE Full text] [doi: 10.1136/bmjopen-2019-030002] [Medline: 31446416]
36. Slade SC, Finnegan S, Dionne CE, Underwood M, Buchbinder R. The consensus on exercise reporting template (CERT) applied to exercise interventions in musculoskeletal trials demonstrated good rater agreement and incomplete reporting. J Clin Epidemiol 2018 Nov;103:120-130. [doi: 10.1016/j.jclinepi.2018.07.009] [Medline: 30055247]
37. Huston N. Mobile Application Rating Scale Training Module 2.: Daily Motion; 2015. URL: https://www.dailymotion.com/video/x3443jz [accessed 2019-06-20]
38. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
40. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. United Kingdom; 2019 Mar. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2020-06-18]
41. Kebede M, Steenbock B, Helmer SM, Sill J, Möllers T, Pischke CR. Identifying evidence-informed physical activity apps: content analysis. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e10314 [FREE Full text] [doi: 10.2196/10314] [Medline: 30563810]
42. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
43. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, behavioral change techniques, and quality of the most popular mobile apps to measure physical activity: systematic search in app stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281 [FREE Full text] [doi: 10.2196/11281] [Medline: 30368438]
44. Bearne LM, Coomer AF, Hurley MV. Upper limb sensorimotor function and functional performance in patients with rheumatoid arthritis. Disabil Rehabil 2007 Jul 15;29(13):1035-1039. [doi: 10.1080/09638280600929128] [Medline: 17612988]
45. Primdahl J, Hegelund A, Lorenzen AG, Loeppenthin K, Dures E, Appel Esbensen B. The experience of people with rheumatoid arthritis living with fatigue: a qualitative metasynthesis. BMJ Open 2019 Mar 20;9(3):e024338 [FREE Full text] [doi: 10.1136/bmjopen-2018-024338] [Medline: 30898808]
46. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
47. Middelweerd A, Mollee JS, van der Wal CN, Brug J, te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
48. Meade LB, Bearne LM, Sweeney LH, Alageel SH, Godfrey EL. Behaviour change techniques associated with adherence to prescribed exercise in patients with persistent musculoskeletal pain: systematic review. Br J Health Psychol 2019 Feb;24(1):10-30 [FREE Full text] [doi: 10.1111/bjhp.12324] [Medline: 29911311]

Abbreviations

BCT: behavior change technique
BCTT: Behavior Change Technique Taxonomy
CERT: Consensus on Exercise Reporting Template
ICC: intraclass correlation coefficient
MARS: mobile app rating scale
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RA: rheumatoid arthritis
RAISE: Rheumatoid Arthritis Information Support and Education

Edited by G Eysenbach; submitted 01.03.20; peer-reviewed by P Clarkson, M Fajardo; comments to author 04.04.20; revised version received 18.04.20; accepted 14.05.20; published 21.07.20.

Please cite as:
Bearne LM, Sekhon M, Grainger R, La A, Shamali M, Amirova A, Godfrey EL, White CM
Smartphone Apps Targeting Physical Activity in People With Rheumatoid Arthritis: Systematic Quality Appraisal and Content Analysis
JMIR Mhealth Uhealth 2020;8(7):e18495
URL: http://mhealth.jmir.org/2020/7/e18495/
doi:10.2196/18495
PMID:32706727

Original Paper

Assessing Apps for Patients with Genitourinary Tumors Using the Mobile Application Rating Scale (MARS): Systematic Search in App Stores and Content Analysis

Miguel Ángel Amor-García[1], PharmD; Roberto Collado-Borrell[1], PhD, PharmD; Vicente Escudero-Vilaplana[1], PhD, PharmD; Alejandra Melgarejo-Ortuño[1], PharmD; Ana Herranz-Alonso[1], PharmD, PhD; José Ángel Arranz Arija[1], MD, PhD; María Sanjurjo-Sáez[1], PharmD

Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

Corresponding Author:
Vicente Escudero-Vilaplana, PhD, PharmD
Hospital General Universitario Gregorio Marañón
Instituto de Investigación Sanitaria Gregorio Marañón
Doctor Esquerdo, 46
Madrid, 28007
Spain
Phone: 34 915867714
Fax: 34 915866621
Email: vicente.escudero@salud.madrid.org

Abstract

Background: The large number of available cancer apps and their impact on the population necessitates a transparent, objective, and comprehensive evaluation by app experts, health care professionals, and users. To date, there have been no analyses or classifications of apps for patients with genitourinary cancers, which are among the most prevalent types of cancer.

Objective: The objective of our study was to analyze the quality of apps for patients diagnosed with genitourinary cancers using the Mobile Application Rating Scale (MARS) and identify high-quality apps.

Methods: We performed an observational cross-sectional descriptive study of all smartphone apps for patients diagnosed with genitourinary cancers available on iOS and Android platforms. In July 2019, we searched for all available apps for patients with genitourinary cancers (bladder, prostate, cervical, uterine, endometrial, kidney, testicular, and vulvar) or their caregivers. Apps were downloaded and evaluated, and the general characteristics were entered into a database. The evaluation was performed by 2 independent researchers using the MARS questionnaire, which rates 23 evaluation criteria clustered in 5 domains (Engagement, Functionality, Esthetics, Information, and Subjective Quality) on a scale from 1 to 5.

Results: In total, 46 apps were analyzed. Of these, 31 (67%) were available on Android, 6 (13%) on iOS, and 9 (20%) on both platforms. The apps were free in 89% of cases (41/46), and 61% (28/46) had been updated in the previous year. The apps were intended for prostate cancer in 30% of cases (14/46) and cervical cancer in 17% (8/46). The apps were mainly informative (63%, 29/46), preventive (24%, 11/46), and diagnostic (13%, 6/46). Only 7/46 apps (15%) were developed by health care organizations. The mean MARS score for the overall quality of the 46 apps was 2.98 (SD 0.77), with a maximum of 4.63 and a minimum of 1.95. Functionality scores were quite similar for most of the apps, with the greatest differences in Engagement and Esthetics, which showed acceptable scores in one-third of the apps. The 5 apps with the highest MARS score were the following: “Bladder cancer manager,” “Kidney cancer manager,” “My prostate cancer manager,” “Target Ovarian Cancer Symptoms Diary,” and “My Cancer Coach.” We observed statistically significant differences in the MARS score between the operating systems and the developer types (P<.001 and P=.01, respectively), but not for cost (P=.62).

Conclusions: MARS is a helpful methodology to decide which apps can be prescribed to patients and to identify which features should be addressed to improve these tools. Most of the apps designed for patients with genitourinary cancers only try to provide data about the disease, without coherent interactivity. The participation of health professionals in the development of these apps is low; nevertheless, we observed that both the participation of health professionals and regular updates were correlated with quality.

(JMIR Mhealth Uhealth 2020;8(7):e17609) doi:10.2196/17609

KEYWORDS

genitourinary cancer; mobile apps; eHealth; mHealth; rating tool

Introduction

Genitourinary cancers represent 25% of all types of cancer [1]. For men, the most commonly diagnosed cancers are prostate cancer and bladder cancer, while for women, cervical cancer and uterine cancer are the most common [2]. Although these cancers are leading causes of death, recent advances in diagnosis and therapy have led to significant improvements in the overall survival of these patients [1].

The chronification of many genitourinary cancers and the special features of newer treatments, such as oral anticancer agents and immunotherapy, have also changed the profile of patients with this type of cancer [3-5]. As a result of the more exhaustive amount of information that patients and relatives need to know about the characteristics of the disease and treatment, there are new challenges for improving communication between patients and caregivers, and greater opportunities for contact with health care professionals. In this regard, information and communications technologies, especially mobile apps and remote assistance services [6-8], could help to improve the autonomy and communication options of these patients [9].

Currently, more than 200 health apps are released daily, and in the last 2 years, the number of available apps has doubled to reach more than 300,000 [10]. This development, which has not been specifically regulated, has led to the diffusion of some poor-quality apps [11,12]. Apps are downloaded from one of the operating system stores ("Play Store" for Android and "App Store" for iOS), where they are valued based on only 2 criteria (ie, the number of downloads and the user ratings) [13,14]. They are nonspecific search engines that do not enable the user to apply filters to assess their disease, the purpose of the app, or the quality of the app [6]. Consequently, searching for high-quality information is becoming even more difficult, with the result that the user downloads apps of uncertain reliability that are likely not the most appropriate option for his/her needs [15-17]. This aspect is particularly important for patients with cancer, where receiving poor-quality information may have a negative impact on prognosis [18].

The large number of available health care apps and their impact on the population necessitates a transparent, objective, and comprehensive evaluation by app experts, health care professionals, and users [15,19,20]. There are several methods to evaluate the quality of health apps. The most appropriate for use in online stores where patients can search and contrast health care apps is the Mobile Application Rating Scale (MARS). This tool provides a simple, quantitative, and validated system that enables rapid evaluation with little variation [12,13,21]. The number of apps for patients with cancer is continuously increasing owing to the availability of new information, requirements for communication, and the empowerment of patients who wish to participate in their care [5,9]. However, there is no standardized methodology for the classification, assessment, and validation of apps for patients with genitourinary cancers, although MARS is the most widely recommended.

The objective of our study was to analyze the quality of apps for patients diagnosed with genitourinary cancers using the MARS scale to identify high-quality apps.

Methods

Study Design

We performed an observational cross-sectional descriptive study of all smartphone apps for patients diagnosed with genitourinary cancers available on the iOS and Android platforms.

Our study followed a methodology to select the apps and adhered to the PRISMA (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) guidelines [22]. In July 2019, a search was conducted in the App Store (iOS) and Play Store (Android) within the categories "medicine" and "health and fitness." The terms used in this search were the following: "bladder cancer," "kidney cancer," "testicular cancer," and "prostate cancer" for urological tumors and "reproductive cancer," "endometrial cancer," "cervical cancer," "uterine cancer," "ovarian cancer," and "vulvar cancer" for gynecological tumors.

Once the search was completed, all available information on the platform was analyzed, and only apps that were in English or Spanish and intended for patients and caregivers were selected. For this study, we excluded apps aimed specifically at health care professionals, those for charitable purposes or without scientific content, and those not specific to genitourinary cancers. Apps that met the indicated criteria were downloaded and evaluated, regardless of cost. The iOS apps were downloaded to an iPhone 8 (version 12.3.2) and the Android apps were downloaded to a Xiaomi Mi A1 (version 9.0).

Characteristics and Content of the Apps

The general characteristics of the applications were entered into a database. Recorded characteristics included the name, platform (Android or iOS), cost (€), category (medicine and health and fitness), date of the last update, language, and target type of cancer. The content of the applications was classified into 1 of 3 categories according to its purpose: informative, preventive, and diagnostic. Furthermore, any information about the participation of health professionals in the app design or development was included. Qualified professionals were considered to have contributed to the app contents if the app had been developed by health care organizations such as local health authorities, universities, scientific societies and foundations, and hospitals.

MARS Evaluation

The quality of the apps was then assessed using MARS. This methodology includes 23 evaluation criteria, clustered within 5 domains: (1) "Engagement," which assesses the entertainment, customization, and interactivity of the app (feedback, reminders, and notifications); (2) "Functionality," which examines the

functionality of the app, ease of use, transition between screens, and intuitive design; (3) "Esthetics," which assesses graphic design, visual appeal, and stylistic consistency; (4) "Information," which evaluates the quality of the content (text, measures, and references), determined by the credibility of the source; and (5) "Subjective quality," which determines whether the app could be recommended to people who might benefit from it, if they would be prepared to pay for it, how many times it would be used, and what overall star rating it would be given. Each evaluation criterion was rated from 1 to 5 (1=Inadequate, 2=Poor, 3=Acceptable, 4=Good, 5=Excellent) [12,13,22]. The MARS evaluation was carried out by 2 independent researchers with experience in app design and development and familiarity with genitourinary cancers. After all the evaluation criteria were scored, the mean score of the domains was calculated to obtain the total mean MARS score, which describes the overall quality of the app.

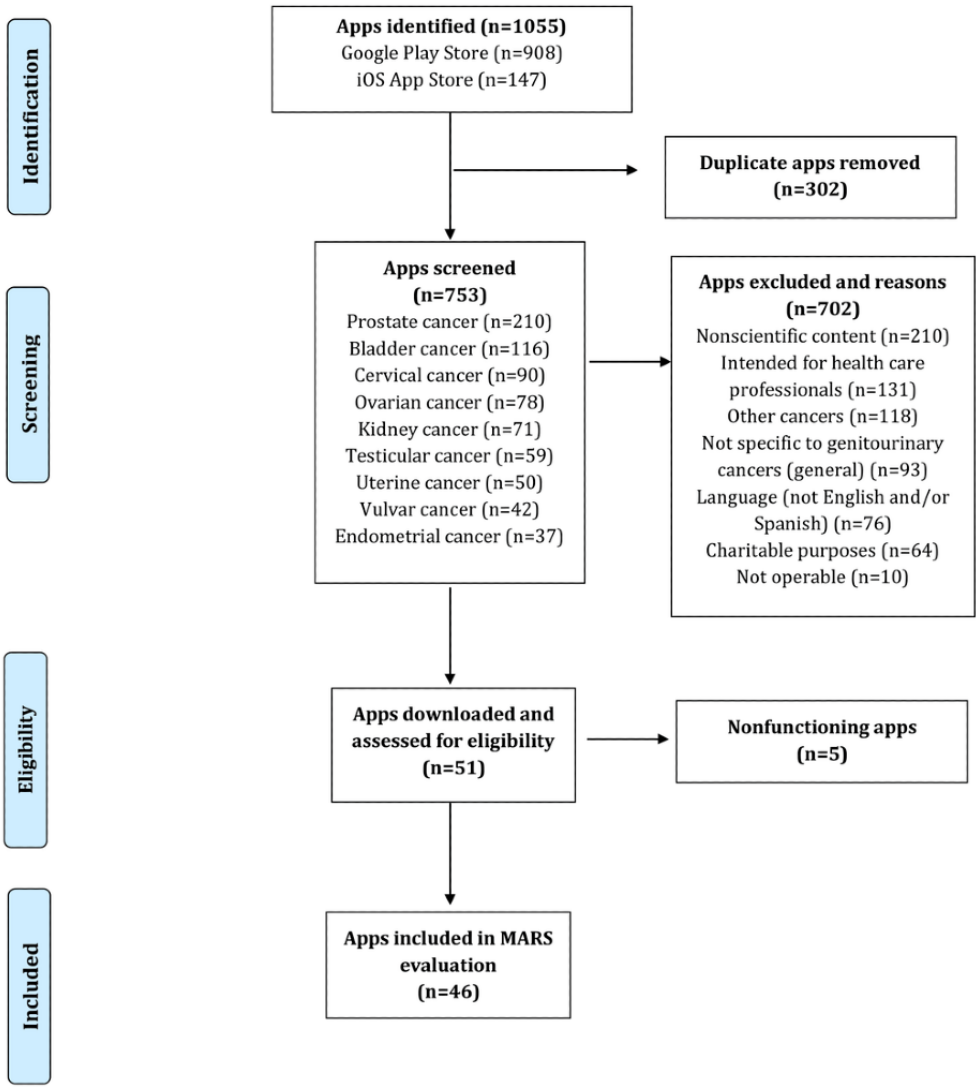
The quantitative variables were described using mean and standard deviation. The categorical variables were described using frequencies and percentages. The numerical variables were compared using the *t* test. A weighted Cohen κ test was performed to guarantee the reliability of those data analyzed by 2 independent observers, and the joint probability agreement (estimated as the percentage of times the raters agreed on an item) was measured. Results with a *P* value <.05 were considered statistically significant. Data were analyzed using Stata (version IC-15, StataCorp LLC).

Results

Study Design

The app search provided a total of 1055 apps, of which 51 were finally downloaded (Figure 1).

Figure 1. Study flowchart. MARS: Mobile Application Rating Scale.



Characteristics and Content of the Apps

At the time of the MARS evaluation, 5 of those 51 applications had been removed from the store or could not be opened, leading to a final total of 46. Of these, 31 (67%) were available on Android, 6 (13%) on iOS, and 9 (20%) on both platforms. Most of the apps (89%, 41/46) were free, with only 5 apps (11%, 5/46) requiring payment (mean cost of €3.51 [US $3.98], SD

1.21). The general characteristics of the apps are shown in Table 1. Table 2 describes the characteristics of the apps.

Most of the apps were informative (63%, 29/46), followed by preventive (24%, 11/46) and diagnostic (13%, 6/46). Only 7 of the 46 apps (15%) were developed by health care organizations, which included 3 scientific societies (43%), 3 universities (43%), and 1 hospital (14%). Regarding the type of cancer, 27/46 (59%) were intended for patients with urological cancers and 21/46 (47%) for those with gynecologic cancers. Apps for urological cancers were intended for prostate cancer in 14/46 apps (30%), testicular cancer in 7/46 (15%), bladder cancer in 4/46 (9%), and kidney cancer in 2/46 (4%). Gynecologic cancers were represented by cervical cancer in 9/46 apps (20%), ovarian cancer in 8/46 (15%), uterine cancer in 2/46 (4%), endometrial cancer in 2/46 (4%), and vulvar cancer in 1/46 (2%). One app contained information on cervical, testicular, and ovarian cancer.

Table 1. General characteristics of the apps.

Characteristics	Apps, n (%)
Platform	
Android	31 (67)
Android/iOS	9 (20)
iOS	6 (13)
Cost	
No	41 (89)
Yes	5 (11)
Category	
Medicine	26 (57)
Health and fitness	20 (43)
Year of the last update	
2014	1 (2)
2015	3 (7)
2016	2 (4)
2017	12 (26)
2018	28 (61)
Language	
English	43 (94)
Spanish	2 (4)
English/Spanish	1 (2)

Table 2. Characteristics of the apps analyzed[a].

Name of the app	Type of cancer	Purpose[b]			Platform		Free	Updated in the last year	Developed by a health organization	Language[c]	
		I	P	D	iOS	Android				E	S
Ball Checker	Testicular			✓	✓	✓	✓	✓	✓		✓
Best Prostate Cancer Treatment	Prostate	✓			✓		✓	✓		✓	
Bladder cancer (Bedieman)	Bladder	✓				✓	✓	✓		✓	
Bladder cancer manager	Bladder		✓		✓		✓	✓		✓	
Cancer cervix fact	Cervical	✓				✓	✓	✓		✓	
Cáncer de cuello uterino	Cervical	✓				✓	✓			✓	
Cáncer de ovarios	Ovarian	✓				✓	✓			✓	
Cancer de prostata (Anastore)	Prostate	✓				✓	✓			✓	
Cancer de prostata (Pen Drouzi)	Prostate	✓				✓	✓	✓		✓	
Cancer de RIÑON	Kidney	✓				✓	✓			✓	
Cáncer de vejiga	Bladder	✓				✓	✓			✓	
Cáncer testicular (Anass apps)	Testicular	✓				✓	✓			✓	
Cancer testicular (Expert Health Studio)	Testicular	✓				✓	✓	✓		✓	
Cancer testicular (Health Advice Ideas)	Testicular	✓				✓	✓	✓		✓	
Cancer testicular (Pen Drouzi)	Testicular	✓				✓	✓	✓		✓	
Cáncer uterino	Uterine	✓				✓	✓			✓	
Cervical cancer (Bedieman)	Cervical	✓				✓	✓	✓		✓	
Cervical cancer (Natural health care)	Cervical	✓				✓	✓	✓		✓	
Cervical cancer (Nougat spring)	Cervical	✓				✓	✓	✓		✓	
Cervical cancer (Personal Remedies LLC)	Cervical		✓		✓	✓		✓		✓	
Common causes of cervical cancer	Cervical		✓			✓	✓			✓	
El cancer de vulva	Vulvar	✓				✓	✓			✓	
Endometrial cancer (Bedieman)	Endometrial	✓				✓	✓	✓		✓	
Endometrial cancer (online Global Groups)	Endometrial	✓				✓	✓	✓		✓	
Global Pap App	Cervical		✓		✓		✓	✓		✓	
How to prevent ovarian cancer	Ovarian	✓				✓	✓	✓		✓	
IPCRC (Prostate Ca Calculator)	Prostate			✓		✓	✓			✓	
itsaMANTHING-Prostate cancer	Prostate	✓			✓	✓	✓		✓	✓	

Name of the app	Type of cancer	Purpose[b]			Platform		Free	Updated in the last year	Developed by a health organization	Language[c]	
		I	P	D	iOS	Android				E	S
Kidney cancer manager	Kidney		✓		✓		✓	✓		✓	
My Cancer Coach	Prostate	✓				✓	✓	✓	✓	✓	
My prostate cancer manager	Prostate		✓		✓		✓	✓		✓	
OddBalls-Check Yourself	Testicular		✓			✓	✓			✓	
Ovarian cancer (Personal Remedies LLC)	Ovarian		✓		✓	✓		✓		✓	
Ovarian Cancer Awareness	Ovarian	✓				✓	✓			✓	
Ovarian cancer symptoms diary	Ovarian		✓		✓	✓	✓		✓	✓	
Prostate cancer (Dinatale)	Prostate	✓				✓	✓	✓		✓	
Prostate cancer (Personal Remedies LLC)	Prostate		✓		✓	✓		✓		✓	
Prostate Cancer Calculator	Prostate			✓		✓	✓			✓	
Prostate PRO-Tracker	Prostate			✓	✓		✓		✓	✓	
ProstateCheck	Prostate			✓	✓	✓			✓		✓
Reproductive cancers	Ovarian, cervical, testicular	✓				✓	✓	✓		✓	
Rotterdam Prostate Cancer Risk	Prostate			✓	✓	✓		✓	✓	✓	✓
Target Ovarian Cancer Symptoms Diary	Ovarian		✓		✓	✓	✓	✓		✓	
Treat prostate cancer	Prostate	✓				✓	✓			✓	
Treating bladder cancer	Bladder	✓				✓	✓	✓		✓	
Ways to treat uterine cancer	Uterine	✓				✓	✓	✓		✓	

[a]In apps with the same name, the developer is indicated in parentheses.

[b]I: informative; P: preventive; D: diagnostic.

[c]E: English; S: Spanish.

MARS Evaluation

The mean MARS score for the overall quality of the 46 apps was 2.98 (SD 0.77), with a maximum of 4.63 and a minimum of 1.95 (Table 3).

The Functionality scores were similar for most of the apps. The apps provided adequate and rapid movement between the screens and menus. The greatest differences were found in the Engagement domain because of deficiencies in areas such as customization and interactivity. Similar differences were found in the Information domain because of the visual explanation and evidence base. The Esthetics domain showed acceptable scores in one-third of the apps; this was associated with a greater update rate and visual appeal.

Table 3. Mobile Application Rating Scale scores of the evaluated apps out of 5[a].

Name of app	Engagement	Functionality	Esthetics	Information	Subjective quality	Overall
Bladder cancer manager	4.60	4.63	4.83	4.57	4.50	4.63
Kidney cancer manager	4.60	4.63	4.83	4.57	4.50	4.63
My prostate cancer manager	4.60	4.63	4.83	4.57	4.50	4.63
Target Ovarian Cancer Symptoms Diary	4.10	4.75	4.83	3.93	3.63	4.25
My Cancer Coach	4.10	4.63	3.67	4.50	4.25	4.23
Rotterdam Prostate Cancer Risk	3.20	4.63	3.67	4.43	3.25	3.83
ProstateCheck	3.40	4.88	3.50	3.36	3.25	3.68
Ball Checker	3.50	4.25	3.67	4.21	2.75	3.68
Prostate PRO-Tracker	3.40	4.75	3.67	4.21	2.88	3.64
Global Pap App	2.90	4.88	3.67	3.71	3.00	3.63
OddBalls - Check yourself	3.30	3.88	3.50	4.00	3.25	3.59
Prostate Cancer Calculator	3.80	4.63	2.00	3.71	3.5	3.53
itsaMANTHING-Prostate Cancer	2.60	4.88	3.17	3.71	3.13	3.50
Endometrial cancer (online Global Groups)	2.60	4.63	3.50	3.29	3.25	3.45
Prostate cancer (Dinatale)	2.50	4.63	3.67	3.64	2.38	3.36
IPCRC (Prostate Ca Calculator)	2.50	4.88	3.50	2.71	3.00	3.32
Treating bladder cancer	2.50	4.50	3.67	3.29	2.63	3.32
Ovarian Cancer Symptoms Diary	2.90	4.38	3.17	3.14	2.88	3.29
Cáncer de prostata (Pen Drouzi)	1.90	4.50	3.50	3.36	2.25	3.10
Cáncer testicular (Pen Drouzi)	1.90	4.50	3.50	3.36	3.25	3.10
Cervical cancer (Nougat spring)	2.20	4.75	2.50	2.93	2.75	3.03
Best Prostate Cancer Treatment	2.40	4.00	3.33	3.21	2.13	3.02
Bladder cancer	2.10	4.63	3.00	2.86	2.38	2.99
Cervical cancer (Personal Remedies LLC)	3.20	2.25	3.50	2.86	1.88	2.74
Ovarian cancer (Personal Remedies LLC)	3.20	2.25	3.50	2.86	1.88	2.74
Prostate cancer (Personal Remedies LLC)	3.20	2.25	3.50	2.86	1.88	2.74
Cancer testicular (Expert Health Studio)	1.80	3.50	3.33	3.14	1.63	2.68
Cancer testicular (Health Advice Ideas)	1.80	3.38	3.33	3.14	1.50	2.63
Reproductive cancers	1.60	4.50	3.00	2.43	1.50	2.61
Cervical cancer (Natural health care)	1.70	4.38	2.33	2.86	1.75	2.60
Cervical cancer (Bedieman)	1.40	3.75	2.50	2.29	1.75	2.34
Endometrial cancer (Bedieman)	1.40	3.75	2.50	2.29	1.75	2.34
How to prevent ovarian cancer	1.70	3.88	2.33	2.57	1.13	2.32
Ovarian Cancer Awareness	1.80	4.38	2.00	1.93	1.50	2.32
Treat prostate cancer	1.60	4.00	1.67	2.57	1.75	2.32
Cancer de próstata (Anastore)	1.60	3.50	2.17	2.71	1.50	2.32
Cancer de RIÑON	1.70	3.00	1.67	2.64	1.63	2.13
Cáncer de vejiga	1.70	3.00	1.67	2.64	1.63	2.13
Cancer testicular (Anass apps)	1.70	3.00	1.67	2.64	1.63	2.13
Cancer de cuello uterino	1.70	3.00	1.67	2.64	1.63	2.13

Name of app	Engagement	Functionality	Esthetics	Information	Subjective quality	Overall
Cáncer de ovarios	1.70	3.00	2.00	2.57	1.25	2.10
Cáncer uterino	1.70	3.00	1.67	2.64	1.50	2.10
El cáncer de vulva	1.70	3.00	1.67	2.64	1.50	2.10
Common causes of cervical cancer	1.80	3.13	1.50	2.50	1.50	2.09
Ways to treat uterine cancer	1.60	3.50	1.33	2.00	1.63	2.01
Cancer cervix fact	1.40	3.63	1.33	2.14	1.25	1.95

[a]For apps with the same name, the developer is indicated in parentheses.

Comparison by the operating system (iOS and Android) revealed an overall MARS score of 3.64 for apps available in the App Store (n=15) and 2.19 for those available in the Play Store (n=40); the difference was statistically significant (*P*<.001). However, when the overall MARS scores were analyzed considering whether the apps were free (n=41) or required payment (n=5), the only significant differences were in the Functionality domain. Comparison by developer type revealed statistically significant differences between the apps that had been supported by a health organization (n=7), with a score of 3.70, and those that had not (n=39), with a score of 2.85 (Table 4).

Interrater agreement was substantial across all evaluation criteria, except for the Entertainment and Interest criteria of the Engagement domain and the Subjective Quality criterion "What is your overall star rating for the app?" The joint probability of agreement was >85% in all the items and >90% in 11 of the 23 evaluation criteria and 4 of the 5 domains analyzed (Table 5). The mean κ score for the 5 domains was 0.748, indicating that substantial agreement was observed between the 2 evaluators.

Table 4. Results of the Mobile Application Rating Scale evaluation: comparison by different characteristics.

Category	Operating system			Developer			Cost		
	Android (n=40)	iOS (n=15)	*P* value	Non–health organization (n=39)	Health organization (n=7)	*P* value	Free (n=41)	Paid (n=5)	*P* value
Engagement	2.30	3.45	<.001	2.34	3.30	.01	2.40	3.24	.06
Functionality	3.87	4.13	.32	3.85	4.63	.02	4.05	3.25	.03
Esthetics	2.75	3.84	<.001	2.83	3.50	.10	2.86	3.53	.16
Information	3.00	3.72	<.001	3.00	3.88	.002	3.12	3.27	.67
Subjective quality	2.18	3.05	.001	2.21	3.18	.01	2.35	2.43	.88
Overall	2.82	3.64	<.001	2.85	3.70	.01	2.96	3.14	.62

Table 5. Interrater agreement for the Mobile Application Rating Scale domains and evaluation criteria.

Domains and evaluation criteria	Weighted Cohen κ	Agreement (%)
Engagement	0.76	92.2
Entertainment	0.55	89.4
Interest	0.56	87.8
Customization	0.86	96.8
Interactivity	0.77	91.5
Target group	0.64	88.8
Functionality	0.71	90.0
Performance	0.68	89.9
Ease of use	0.62	89.4
Navigation	0.69	87.9
Gestural design	0.64	87.9
Esthetics	0.80	93.6
Layout	0.75	90.8
Graphics	0.76	93.1
Visual appeal	0.91	96.8
Information	0.79	93.6
Accuracy of the app in the description (App/Play Store)	0.63	89.4
Goals	0.72	93.6
Quality of information	0.67	88.7
Quantity of information	0.65	87.2
Visual information	0.86	94.1
Evidence base	0.66	94.1
Credibility	0.61	89.4
Subjective quality	0.68	89.9
Would you recommend this app to people who might benefit from it?	0.64	90.4
Would you pay for this app?	0.74	90.8
How many times do you think you would use this app in the next 12 months if it was relevant to you?	0.72	90.4
What is your overall star rating of the app?	0.48	86.7

Discussion

Principal Findings and Comparison With Previous Work

Based on a systematic and validated questionnaire (MARS), our study provided an objective ranking of 46 apps for patients with genitourinary cancers available in the Apple and Android stores. Apps for patients with prostate and cervical cancer accounted for almost half of all the apps evaluated (30% and 17%, respectively). This frequency is consistent with the findings of the Globocan 2018 report [2], according to which genitourinary cancers have the highest global incidence (second in men and fourth in women). Therefore, the number of apps analyzed correlated with the incidence of the specific genitourinary cancers, unlike other types of cancer, such as lung and colorectal cancer, which are extremely prevalent, yet for which few apps have been released [15].

More than half of the apps (61%) had been updated in the last year and therefore provided better quality information, which is increasingly necessary given the advantages of apps in diagnosis and therapy in this area. Our result is comparable to that obtained in a review of 166 apps for patients with cancer, where it was observed that 52.4% had been updated in the previous year [15]. However, the proportion of apps that had been developed or promoted by health care organizations was very low (15.2%), thus potentially reducing the quality and reliability of the apps. This result was consistent with the conclusions of authors such as Giunti et al [14], who showed an evident absence of health professionals in the development of health care apps. Apps are mostly developed by non–health

professionals who are creative and skilled in design but lack scientific knowledge.

Most of the apps included in our study were informative, with generic data on pathophysiology, treatments, and symptoms of individual cancers, as reported elsewhere [14]. An important component of some of these apps, such as Treating Bladder Cancer, was including information about the most common symptoms and signs of bladder cancer (hematuria, pain or burning sensation, and increased frequency in urination) that could alert patients to talk to the doctor. These features are important to reduce the risk of having a more advanced stage of cancer. We found that most of the informative apps, like those for cervical cancer, included information about treatment and prognosis. Nonetheless, this information should be given by physicians, and informative apps provide doctors with a tool to improve early diagnosis through successful screening. Additionally, as indicated by Bender et al [23], most of the apps try to increase awareness of cancer in the population. Very few apps focused on how to handle the disease after a diagnosis, correct administration of the treatment prescribed (eg, dosing, management of adverse effects, possible interactions with other long-term medications), and adequate monitoring of symptoms. Only 6 diagnostic apps (13.0%) were evaluated; of these, 5 were intended for patients with prostate cancer and 1 for patients with testicular cancer. Diagnostic apps are more frequent in other types of cancer, such as melanoma and breast cancer [16].

Several methods for the evaluation of mHealth apps have been developed, although in most cases, the absence of a systematic methodology and the fact that they were not developed by scientific professionals made their routine use impossible [12]. The MARS methodology, as reported by Stoyanov et al [21], is an easy-to-use, simple and logical tool that is considered highly reliable because it is promoted by expert technicians and health care professionals. This evaluation proposes a multidisciplinary analysis designed for all health care apps, with 5 domains covering the main aspects for correct evaluation. Stoyanov et al [13] reported that the MARS questionnaire showed high levels of internal consistency (Cronbach α=.9) and interrater reliability (two-way mixed intraclass correlation coefficient, 0.79; 95% CI, 0.75-0.83) when it was applied to rate 50 mental health apps.

As for the MARS score, our study showed a mean score of 2.98 for the overall quality of the apps, with a score of 3.13 for quality content. These scores were significantly higher than those found by Böhme et al [6] for mobile cancer apps for prostate cancer, breast cancer, and colorectal cancer (1.96). On the other hand, our results were similar to the scores of apps for other diseases. For example, the mean score found by Richardson et al [19] for mobile apps targeted to parents of infants in the neonatal intensive care unit was 3.37, without considering the subjective quality. Salazar et al [24] showed a mean score of 3.92 for mobile apps for the management of pain, while Siddique et al [25] found a median score of 3.70 for apps targeted to the care management of chronic kidney and end-stage renal disease. To note, 22 apps (48%) scored a value equal to or greater than 3 points (ie, “Acceptable”). Only 5 apps (11%) exceeded 4 points in the overall quality score (ie, “Good”). Jupp et al [3] found that the apps evaluated stood out in the Functionality domain, with high scores in most of them because they were developed to be highly efficient and easy to use. Consistent with our findings, these authors found that the 3 strategies necessary for optimal use of the smartphone among patients with cancer were the management of symptoms and medications, quality of the information resources, and ability to export data. In contrast, the Esthetics domain had a better correlation with the overall MARS score. It is assumed that this is due to the attempt to develop a more engaging appearance which is directly related to other features, such as a higher frequency of use.

In reports based on MARS for assessment of apps aimed at patients with cancer or other diseases, the domain that scored the lowest was Engagement. The main reason is that the apps were unable to make patients feel that they were participating in the management of their disease. We drew the same conclusion, with Engagement being the domain with the poorest mean score compared to the others. According to the literature, the fundamental aspects that can be improved in this section are customization, user interactivity, and entertainment, which leads to a higher score [3,20]. The participation of patients in the development, design, and validation of apps through focus groups considerably improves the score for this domain. Furthermore, the participation of patients in the management of their symptoms by registering and sending messages and reminders significantly enhances the health outcomes [26]. Collado et al [18] found that more than 40% of patients would be interested in communicating with their physician or pharmacist using an app.

The apps that scored best in the MARS evaluation were “Bladder cancer manager,“ ”Kidney cancer manager,“ and ”My prostate cancer manager.“ These apps were available in the App Store and stood out because of their high scores in the Engagement and Esthetics domains, as did the next 2 apps in the ranking, “Target Ovarian Cancer Symptoms Diary” and “My Cancer Coach.” The “Top 5” apps contained reminders and schedules and offered the possibility of registering analytical information and treatments prescribed, thus enabling a greater score in the Subjective Quality domain because they achieve the main goal. Likewise, the 4 best apps had an explicit preventive purpose, in contrast to the informative purpose that was more frequent in the global analysis.

Of note, 9 of the 10 best apps had been updated during 2018, and 5 of the 10 best apps had been developed by a health care organization. This is an important observation because neither of these 2 aspects is specifically evaluated in MARS. The analysis by the domain of the apps developed by health care organizations and those that were not revealed statistically significant differences for each of the items evaluated. Therefore, the quality of health apps is based on the frequency of updating and the participation of health care organizations in their development, thus confirming the hypothesis proposed elsewhere [15,27]. The first three apps (“Bladder cancer manager,” “Kidney cancer manager.” and “My prostate cancer manager”) had the same developer (Point of Care) which is a platform of health apps intended to patients and clinicians. However, clinicians were not involved in their development. Also, “Target Ovarian Cancer Symptoms Diary” had a clinical

advisory panel made up of many oncology specialists who answer patients´ questions but they were not involved in promoting the app. Otherwise, "My Cancer Coach" was developed by some health organizations and the coaches formed a multidisciplinary team (nurses, physicians…) to provide reliable information to patients. However, the lowest scores were obtained for apps that are merely informative and provide links to other web pages but not their information.

The analysis by operating system revealed a statistically significant difference that was more favorable to the apps from the App Store than those from Play Store, probably because verification requirements for publishing and application are stricter for iOS than for Android.

Apps are beginning to show a significant impact on users´ health. Because of that, regulatory authorities are responsible for evaluating these technologies to control their availability in the stores. In 2013 and regarding this issue, the Food and Drug Administration (FDA) published a guide containing recommendations to assess the quality of these apps. However, due to the rise of health apps, an objective, comprehensive and clear evaluation of apps is necessary. This evaluation should allow users and health care professionals (who recommend them to patients) which apps meet minimum standards of quality and safety in their content.

Patients are exposed to unreliable information related to their health so quality certifications are needed to identify those apps that offer the best content for users. For example, App Saludable is a free and open-access certification given to some apps which were developed using strict guidelines in Spain [28]. This is one of the first certifications in Europe that evaluate the quality and safety of health apps.

Limitations

MARS is limited by its subjectivity [13,21]. However, the high interrater reliability obtained between the 2 evaluators, both of whom had experience in the development and validation of health apps and were familiar with genitourinary cancers, highlights the considerable coherence of our results.

Conclusions

The quality of health apps should be evaluated using approaches such as MARS to decide which apps could be prescribed to patients and to identify which features should be addressed to improve these tools. Most of the apps designed for patients with genitourinary cancers only try to increase awareness and provide data about the disease, without ensuring coherent interactivity. Although the participation of health professionals in the development of these apps is low, we observed that their participation was associated with the app quality and the recency of updates. Greater scores in quality were observed in iOS apps, although no correlation between quality and price was found.

Conflicts of Interest

None declared.

References

1. Lopez CM. Avances en supervivencia en tumores genitourinarios gracias a las nuevas técnicas diagnósticas. Gaceta Medica. 2018. URL: https://www.gacetamedica.com/especializada/avances-en-supervivencia-en-tumores-genitourinarios-gracias-a-las-nuevas-tecnicas-diagnosticas-EG1833350 [accessed 2019-09-19]
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
3. Jupp JCY, Sultani H, Cooper CA, Peterson KA, Truong TH. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. Pediatr Blood Cancer 2018 Nov;65(11):e27278. [doi: 10.1002/pbc.27278] [Medline: 29943893]
4. Ali EE, Chan SSL, Leow JL, Chew L, Yap KY. User acceptance of an app-based adherence intervention: Perspectives from patients taking oral anticancer medications. J Oncol Pharm Pract 2019 Mar 23;25(2):390-397 [FREE Full text] [doi: 10.1177/1078155218778106] [Medline: 29792123]
5. Jiang Y, West BT, Barton DL, Harris MR. Acceptance and Use of eHealth/mHealth Applications for Self-Management Among Cancer Survivors. Stud Health Technol Inform 2017;245:131-135 [FREE Full text] [Medline: 29295067]
6. Böhme C, von Osthoff MB, Frey K, Hübner J. Qualitative evaluation of mobile cancer apps with particular attention to the target group, content, and advertising. J Cancer Res Clin Oncol 2018 Jan 19;144(1):173-181. [doi: 10.1007/s00432-017-2533-0] [Medline: 29052039]
7. Owens OL, Beer JM, Reyes LI, Thomas TL. Systematic Review of Commercially Available Mobile Phone Applications for Prostate Cancer Education. Am J Mens Health 2019 Dec 07;13(1):1557988318816912 [FREE Full text] [doi: 10.1177/1557988318816912] [Medline: 30526243]
8. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr 8;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]

9. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The Impact of Information Technology on Patient Engagement and Health Behavior Change: A Systematic Review of the Literature. JMIR Med Inform 2016 Jan 21;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
10. The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System Institute Report. IQVIA. 2017. URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2019-09-03]
11. Boyle JA, Xu R, Gilbert E, Kuczynska-Burggraf M, Tan B, Teede H, et al. Ask PCOS: Identifying Need to Inform Evidence-Based App Development for Polycystic Ovary Syndrome. Semin Reprod Med 2018 Dec;36(1):59-65. [doi: 10.1055/s-0038-1667187] [Medline: 30189452]
12. Stec MA, Arbour MW, Hines HF. Client-Centered Mobile Health Care Applications: Using the Mobile Application Rating Scale Instrument for Evidence-Based Evaluation. J Midwifery Womens Health 2019 May 18;64(3):324-329. [doi: 10.1111/jmwh.12941] [Medline: 30887711]
13. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
14. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of Breast Cancer mobile applications: state of the practice review. Int J Med Inform 2018 Dec;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
15. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-Tested Mobile Apps for Breast Cancer Care: Systematic Review. JMIR mHealth uHealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
16. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
17. Crico C, Renzi C, Graf N, Buyx A, Kondylakis H, Koumakis L, et al. mHealth and telemedicine apps: in search of a common regulation. Ecancermedicalscience 2018;12:853 [FREE Full text] [doi: 10.3332/ecancer.2018.853] [Medline: 30079115]
18. Berkowitz CM, Zullig LL, Koontz BF, Smith SK. Prescribing an App? Oncology Providers' Views on Mobile Health Apps for Cancer Care. JCO Clinical Cancer Informatics 2017 Nov(1):1-7. [doi: 10.1200/CCI.17.00107] [Medline: 29113731]
19. Richardson B, Dol J, Rutledge K, Monaghan J, Orovec A, Howie K, et al. Evaluation of Mobile Apps Targeted to Parents of Infants in the Neonatal Intensive Care Unit: Systematic App Review. JMIR mHealth uHealth 2019 Apr 15;7(4):e11620 [FREE Full text] [doi: 10.2196/11620] [Medline: 30985282]
20. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile Health in Oncology: A Patient Survey About App-Assisted Cancer Care. JMIR mHealth uHealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]
21. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
22. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
23. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
24. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the Quality of Mobile Apps for the Management of Pain: Systematic Search and Evaluation Using the Mobile App Rating Scale. JMIR mHealth uHealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
25. Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, et al. Mobile Apps for the Care Management of Chronic Kidney and End-Stage Renal Diseases: Systematic Search in App Stores and Evaluation. JMIR mHealth uHealth 2019 Sep 04;7(9):e12604 [FREE Full text] [doi: 10.2196/12604] [Medline: 31486408]
26. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile Apps for Increasing Treatment Adherence: Systematic Review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
27. Cantudo-Cuenca M, Robustillo-Cortés M, Cantudo-Cuenca M, Morillo-Verdugo R. A better regulation is required in viral hepatitis smartphone applications. Farm Hosp 2014 Apr 01;38(2):112-117 [FREE Full text] [Medline: 24669895]
28. Apps i projectes de mobilitat. TicSalut. URL: https://ticsalutsocial.cat/es/ [accessed 2020-02-01]

Abbreviations

FDA: Food and Drug Administration
MARS: Mobile Application Rating Scale

Edited by G Eysenbach; submitted 27.12.19; peer-reviewed by T Muto, M Gaman; comments to author 27.01.20; revised version received 06.02.20; accepted 23.03.20; published 23.07.20.

Please cite as:
Amor-García MÁ, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija JÁ, Sanjurjo-Sáez M
Assessing Apps for Patients with Genitourinary Tumors Using the Mobile Application Rating Scale (MARS): Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2020;8(7):e17609
URL: http://mhealth.jmir.org/2020/7/e17609/
doi:10.2196/17609
PMID:32706737

Original Paper

Explaining Differences in the Acceptability of 99DOTS, a Cell Phone–Based Strategy for Monitoring Adherence to Tuberculosis Medications: Qualitative Study of Patients and Health Care Providers

Beena E Thomas[1*], PhD, MSW; J Vignesh Kumar[1], PhD; Chidiebere Onongaya[2], MD, MS; Spurthi N Bhatt[2], MD, MPH; Amith Galivanche[2]; Murugesan Periyasamy[1], MSW; M Chiranjeevi[1], MSc; Amit Subhash Khandewale[1], MSW; Geetha Ramachandran[1], PhD; Daksha Shah[3], MBBS, MPH; Jessica E Haberer[4], MD, MS; Kenneth H Mayer[5], MD; Ramnath Subbaraman[2*], MD, MSc, FACP

[1]Department of Social and Behavioural Research, National Institute for Research in Tuberculosis, Chennai, India

[2]Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States

[3]Public Health Department, Municipal Corporation of Greater Mumbai, Mumbai, India

[4]Center for Global Health, Massachusetts General Hospital, Boston, MA, United States

[5]The Fenway Institute, Fenway Health, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Ramnath Subbaraman, MD, MSc, FACP
Department of Public Health and Community Medicine
Tufts University School of Medicine
136 Harrison Ave
MV237
Boston, MA, 02111
United States
Phone: 1 6176360985
Email: ramnath.subbaraman@tufts.edu

Abstract

Background: 99DOTS is a cell phone–based strategy for monitoring tuberculosis (TB) medication adherence that has been rolled out to more than 150,000 patients in India's public health sector. A considerable proportion of patients stop using 99DOTS during therapy.

Objective: This study aims to understand reasons for variability in the acceptance and use of 99DOTS by TB patients and health care providers (HCPs).

Methods: We conducted qualitative interviews with individuals taking TB therapy in the government program in Chennai and Vellore (HIV-coinfected patients) and Mumbai (HIV-uninfected patients) across intensive and continuation treatment phases. We conducted interviews with HCPs who provide TB care, all of whom were involved in implementing 99DOTS. Interviews were transcribed, coded using a deductive approach, and analyzed with Dedoose 8.0.35 software (SocioCultural Research Consultants, LLC). The findings of the study were interpreted using the unified theory of acceptance and use of technology, which highlights 4 constructs associated with technology acceptance: performance expectancy, effort expectancy, social influences, and facilitating conditions.

Results: We conducted 62 interviews with patients with TB, of whom 30 (48%) were HIV coinfected, and 31 interviews with HCPs. Acceptance of 99DOTS by patients was variable. Greater patient acceptance was related to perceptions of improved patient-HCP relationships from increased phone communication, TB pill-taking habit formation due to SMS text messaging reminders, and reduced need to visit health facilities (performance expectancy); improved family involvement in TB care (social influences); and from 99DOTS leading HCPs to engage positively in patients' care through increased outreach (facilitating conditions). Lower patient acceptance was related to perceptions of reduced face-to-face contact with HCPs (performance expectancy); problems with cell phone access, literacy, cellular signal, or technology fatigue (effort expectancy); high TB- and

HIV-related stigma within the family (social influences); and poor counseling in 99DOTS by HCPs or perceptions that HCPs were not acting upon adherence data (facilitating conditions). Acceptance of 99DOTS by HCPs was generally high and related to perceptions that the 99DOTS adherence dashboard and patient-related SMS text messaging alerts improve quality of care, the efficiency of care, and the patient-HCP relationship (performance expectancy); that the dashboard is easy to use (effort expectancy); and that 99DOTS leads to better coordination among HCPs (social influences). However, HCPs described suboptimal facilitating conditions, including inadequate training of HCPs in 99DOTS, unequal changes in workload, and shortages of 99DOTS medication envelopes.

Conclusions: In India's government TB program, 99DOTS had high acceptance by HCPs but variable acceptance by patients. Although some factors contributing to suboptimal patient acceptance are modifiable, other factors such as TB- and HIV-related stigma and poor cell phone accessibility, cellular signal, and literacy are more difficult to address. Screening for these barriers may facilitate targeting of 99DOTS to patients more likely to use this technology.

(JMIR Mhealth Uhealth 2020;8(7):e16634) doi:10.2196/16634

KEYWORDS

tuberculosis; medication adherence; mobile phone; mHealth; implementation science; qualitative research

Introduction

Background

Tuberculosis (TB) is a leading infectious cause of mortality globally [1]. Poor adherence to medications may contribute to suboptimal TB care delivery outcomes by leading to higher mortality, treatment failure, posttreatment disease relapse, and development of drug-resistant TB strains [2,3]. In recent years, interest has increased in the use of digital adherence technologies (DATs), which include cell phone–based strategies, digital pillboxes, and ingestible sensors—to monitor and improve adherence to TB medications [4], especially given the growing concerns regarding more restrictive directly observed therapy (DOT) monitoring approaches [5-7]. Despite growing interest in DATs, data regarding these technologies' acceptability, accuracy for measuring adherence, and effectiveness in improving treatment outcomes are lacking, especially in resource-constrained contexts [4,8].

99DOTS is a feature (ie, nonsmart) phone–based DAT that has been implemented widely in India's National TB Elimination Program (NTEP) to monitor more than 150,000 patients taking therapy for drug-susceptible TB since 2015 [9]. The rapid expansion of 99DOTS in the NTEP was partly driven by the introduction in 2014 of daily TB medication dosing for all patients, replacing India's prior intermittent (ie, thrice weekly) dosing regimen. Under India's prior facility-based monitoring approach, most patients traveled to health centers for DOT. With the switch to daily dosing, program managers were concerned that traveling to health centers on a daily basis—rather than 3 times a week, as was previously the case—might be too challenging for patients [9]. In this context, remote monitoring of medication adherence using 99DOTS appeared to be a promising alternative [9].

Under the 99DOTS strategy, dispensing pills for each daily medication dose results in revealing a hidden phone number that patients call for free. A computer program records these phone call-reported doses, allowing remote visualization of each patient's adherence record by health care providers (HCPs). 99DOTS' potential advantages include its relatively low cost compared with other DATs, monitoring of individual adherence data in near real time, and automated notification of HCPs regarding potentially nonadherent patients, which has the potential to facilitate individualized care based on a patient's risk for poor outcomes [9].

Despite these promising features, routine data and research studies have revealed suboptimal patient engagement with 99DOTS, that is, many patients do not call 99DOTS throughout the entire treatment course. For example, in a cohort study in Mumbai and Chennai conducted in parallel with this qualitative study, in which unannounced home visits were conducted on patients with TB across all months of therapy, only approximately two-thirds of patients had called 99DOTS in the 48 hours preceding the home visit [10]. Programmatic data from the rollout of 99DOTS to more than 20,000 patients in Mumbai similarly suggest that fewer than 40% of patients called 99DOTS to report at least 80% of their doses throughout the full treatment duration [9]. Suboptimal patient engagement may undermine 99DOTS' benefits for measuring and promoting medication adherence and preventing loss to follow-up [10].

To our knowledge, only one study has been published evaluating the impact of 99DOTS on TB treatment outcomes. This study, conducted in four districts in Karnataka, compared TB treatment outcomes at HIV centers where 99DOTS was implemented with those at HIV centers where the technology was not rolled out [11]. Notably, implementation of 99DOTS at these centers occurred concurrently with the transition to daily (rather than thrice weekly) TB therapy. In addition, responsibility for dispensing TB medications and managing TB treatment was switched to HCPs at these HIV centers, whereas these tasks had previously been the responsibility of HCPs in the NTEP [11]. After adjusting for potential confounding factors, implementation of 99DOTS was significantly associated with 1.3 times increased adjusted relative risk of unfavorable outcomes, including death, lost to follow-up, and treatment failure [11]. Problems identified in that study, including increased workload for HIV center staff and reduced patient-provider interaction, have also been described in a study of 99DOTS conducted in Delhi [12]. In addition, most studies of other nonsmartphone-based technologies for promoting adherence to TB medications—in particular, one- and two-way SMS text messaging strategies—have not been shown to

improve treatment completion [4,8]. These prior studies highlight the importance of understanding patient- and HCP-related barriers to the acceptance and the use of 99DOTS.

Objectives

In this manuscript, we present findings from a qualitative study of patients and HCPs that aimed to understand the acceptability of 99DOTS in India's NTEP. We interviewed drug-susceptible TB patients with and without HIV coinfection in three major Indian cities across both the intensive (early) and continuation (later) phases of therapy. Given the known problem of variable use of 99DOTS by TB patients, one goal of this analysis was to understand why some patients engage with 99DOTS, whereas others do not. In addition, we aimed to understand the acceptance of 99DOTS by HCPs with various roles in TB care provision, as the 99DOTS strategy represents a transformation in the approach by which HCPs monitor patients in the NTEP, which could have implications for work efficiency and quality of care. We analyzed patient and HCP findings using the unified theory of acceptance and use of technology (UTAUT), which synthesizes constructs that have previously been shown to predict the acceptance and use of novel technologies [13,14].

Methods

Ethics Approvals

The study protocol was approved by the Institutional Ethics Committee of the National Institute for Research in TB (NIRT; FWA00005104) on October 7, 2016; the Institutional Review Board of Brigham and Women's Hospital (Partners Healthcare; FWA00000484) on January 31, 2017; and the Tufts Health Sciences Institutional Review Board (FWA00004517) on June 6, 2018. Written informed consent was obtained from all patients and HCPs who participated in the interviews.

Study Setting

We recruited patients in Mumbai (population of about 18.4 million in the greater metropolitan area) and 2 cities in the south Indian state of Tamil Nadu—Chennai (population of approximately 7.1 million in the greater metropolitan area) and Vellore (population of about 480,000). All these cities have a relatively high TB burden in the general population [15,16].

99DOTS Implementation

With 99DOTS, medications are dispensed in blister packs wrapped in a custom envelope. 99DOTS envelopes were produced separately from the medication blister packs. As a result, TB personnel at the district or city level were responsible for wrapping the envelopes over medication blister packs before they were given to patients.

When a patient dispenses pills for a daily medication dose, this action breaks a perforated flap on the envelope, revealing a hidden phone number that the patient is requested to call for free. A computer program then records that the pill was *in-hand* and likely ingested. The phone number associated with each dose is unpredictable to the patient. When dispensing medications, pharmacists or other HCPs are responsible for explaining to patients how to call 99DOTS and its potential relevance for their care. Patients did not receive any monetary or nonmonetary incentives to call.

99DOTS allows HCPs to remotely *observe* a patient's adherence history by logging in to a website on a computer or using an app on a mobile device, such as a smartphone. In this app, the 99DOTS dashboard presents a color-coded adherence calendar for each patient, with green showing that a call was made on a given day (suggesting that the dose may have been taken) or red showing that a call was not made (suggesting that the dose may not have been taken). These data are designed to help HCPs provide individualized feedback to patients regarding their adherence. A series of possible missed doses (ie, doses not *called in*) triggers automated SMS text message notifications to HCPs to help them identify patients who may be at higher risk for poor outcomes. HCPs are also provided with prioritized task lists recommending certain actions, such as calling to check-in with, or conducting a home visit for, a patient who has missed reporting a few doses in a row. As home visits are not normally conducted for patients taking unmonitored self-administered therapy, unless patients are lost to follow-up, we considered patient reports of multiple home visits by HCPs to be evidence that 99DOTS was potentially triggering such outreach.

Patient and HCP perceptions regarding whether 99DOTS facilitates more or less HCP-patient interaction were shaped by the alternative monitoring approach with which these individuals implicitly compared 99DOTS. For example, HCPs and some patients who had previously been treated for TB had experiences with facility-based DOT, which required patients to come to health facilities 3 times a week, therefore entailing a greater baseline level of HCP-patient interaction than 99DOTS. Other patients did not have experience with facility-based DOT and therefore implicitly compared 99DOTS with unmonitored self-administered therapy for other diseases, in which there is relatively minimal HCP-patient interaction apart from patient visits to health facilities for medication refills.

During the rollout of 99DOTS, in-person training took place at state-level offices, where nominated staff—from HIV treatment centers (commonly called antiretroviral therapy [ART] centers) or district TB programs—would gather at a single location. In addition, for ART centers, 2 virtual training sessions—each attended by more than 1000 HCPs—were conducted for individuals who could not attend centralized trainings. Subsequently, local trainings were conducted if requested by individual district TB programs. A team at Everwell Health Solutions, the organization managing 99DOTS' implementation, monitored the rollout process to identify implementation gaps and provide monthly reports on patient engagement with 99DOTS. The HCPs reported problems via a WhatsApp group. These challenges were resolved remotely via WhatsApp or in person during regular field visits by personnel from Everwell Health Solutions.

HIV-coinfected patients underwent registration into 99DOTS and picked up TB medication refills at ART centers; however, home visits for these patients were conducted by HCPs in district TB programs. The National AIDS Control Organization (which manages ART centers) and the NTEP (which manages TB

programs) were under different government bodies at the time of this study. As such, challenges in coordination between ART centers and district TB programs have the potential to impact linkage to care, and follow-up of HIV-coinfected patients started on TB therapy via 99DOTS.

Recruitment of Study Participants and Collection of Qualitative Data

99DOTS was first rolled out to monitor HIV-coinfected patients diagnosed with TB at ART centers across India. In Chennai and Vellore, all TB patients in this study were HIV-coinfected individuals recruited from 5 different ART centers where 99DOTS had already been rolled out. Mumbai was the first city where 99DOTS was rolled out to facilitate monitoring of the broader (ie, mostly HIV uninfected) TB patient population under the care of the NTEP. We recruited mostly HIV-uninfected TB patients from 11 DOT centers in Mumbai. In India, approximately 95% of TB patients are HIV uninfected; however, given the unique challenges faced by HIV-coinfected patients, including having to concurrently take HIV and TB medications, and their generally poorer TB treatment outcomes [16], our goal in this study was to ensure that perspectives from HIV-coinfected and HIV-uninfected patients were well represented.

We had a goal of recruiting 60 TB patients using purposeful sampling, stratifying across the following criteria: (1) roughly equal representation of HIV-coinfected and uninfected patients; (2) roughly equal representation of patients in the intensive (early) and continuation (later) phases of therapy, as patients' engagement with 99DOTS wanes throughout the course of TB therapy [9]; and (3) at least one-fifth of participants being women. Owing to recruitment in parallel across different study sites, we exceeded the target sample slightly, for both patient and HCP interviews, before this was recognized during routine biweekly virtual meetings among site personnel. Our anticipated sample size was somewhat large for a qualitative study, as we wanted to generate robust data for each of these stratified subgroups. One of the patient interviews was not transcribed and analyzed because the audio recording was muffled.

Patient interview guides included questions and follow-up probes to assess various aspects of patients' experiences with 99DOTS, including counseling regarding this monitoring strategy, the process of calling 99DOTS, and HCP feedback and actions taken in response to patients' adherence records. For HIV-coinfected patients, the interview guide included questions aimed at understanding the indirect impact of 99DOTS on HIV medication adherence. For each aspect of patients' 99DOTS experience, questions assessed constructs for technology acceptance included in the UTAUT model.

We aimed to conduct interviews with 30 HCPs recruited using purposeful sampling, with the goal of including individuals with diverse roles in caring for HIV-coinfected and HIV-uninfected TB patients. We recruited HCPs who interact routinely with patients, including health visitors (individuals with at least a high school level of education who monitor TB therapy), senior treatment supervisors (individuals with at least a high school level of education who supervise health visitors), pharmacists, HIV counselors (who also counsel TB patients in HIV centers), TB counselors (who counsel TB patients in DOT centers), staff nurses, and medical officers (doctors with an MBBS or higher degree). We also recruited personnel with administrative roles, including data managers, TB officers (doctors who supervise TB care at multiple DOT centers), district TB officers (doctors who supervise TB care across a district), city TB officers (doctors who supervise TB care across a city), and district executive health officers (doctors who manage public health services across a district). In the representative quotations, we identify district TB officers, city TB officers, and district executive health officers as *higher-level administrative officers* to ensure that individuals are not identifiable.

We asked questions and follow-up probes to assess various aspects of HCPs' experiences with 99DOTS, including using the 99DOTS dashboard to visualize patient data, notifications and task lists regarding high-risk patients and using these data to guide patient interactions. Questions for HCPs also assessed constructs in the UTAUT.

Interviews were conducted between February 2017 and August 2018. Interviews lasted 30 to 45 min and were conducted by 10 field researchers with a master's degree in social work or another social science field. Field researchers at the Chennai, Vellore, and Mumbai sites underwent a uniform 2-day training at the NIRT in Chennai before starting data collection. All field researchers were taught to follow a uniform approach using a common interview guide. Patient interviews were conducted at DOT or ART centers where patients were receiving treatment. Patients were given nonmonetary compensation for their time equivalent to Indian rupees 100-150 (US $1.30-2.00), which consisted of helpful items such as dhal (lentils) or hygiene products.

HCP interviews were conducted in private spaces in DOT or ART centers or in other private locations of the study participant's preference. Interviews were conducted in Tamil, Hindi, Marathi, or English using appropriately translated interview guides and were audio recorded, transcribed, and translated to produce English-language transcripts. Quantitative and qualitative data were deidentified before analysis, and care was taken to ensure that specific patients or health facilities cannot be identified based on the narrative excerpts included in the manuscript.

Analytical Framework: UTAUT

The UTAUT integrates findings from prior models that were used to evaluate the acceptance and use of technologies [13,14]. By synthesizing diverse constructs used in prior models, the UTAUT identifies 4 broader constructs that explain technology acceptance: performance expectancy, effort expectancy, social influences, and facilitating conditions. Performance expectancy, or perceived usefulness, refers to the degree to which an individual believes that the technology will help with their medical care and daily life (for patients) or with work efficiency or the quality of care they are able to deliver (for HCPs). Effort expectancy, or ease of use, refers to how easy the technology is to use, for example, calling 99DOTS (for patients) or using the digital adherence dashboard (for HCPs). Social influences refer to the influence that other individuals—for example, family or community members (for patients) or other HCPs (for

HCPs)—have on someone's ability to accept or use the technology. Facilitating conditions refer to the quality of the organizational infrastructure that exists to support individuals using the technology; in this case, we assume that this refers to the training and infrastructure provided by the health system to ensure that 99DOTS functions appropriately for patients and HCPs.

In the UTAUT, 3 constructs—performance expectancy, effort expectancy, and social influences—influence behavioral intention to use a technology, which subsequently influences the actual use of a technology [13,14]. In contrast, facilitating conditions directly affect the actual use of a technology. Although all these constructs play a crucial role in shaping technology acceptance, in prior research, performance expectancy was found to be the strongest predictor of intention to use a technology [13,14].

Analysis of Qualitative Data

We used a deductive approach for this thematic analysis, in which our coding was guided by the constructs in the UTAUT (Multimedia Appendices 1 and 2). An initial coding scheme was created based on collective discussion within the research team. Interview transcripts were independently coded by 3 researchers using Dedoose software (version 8.0.35; SocioCultural Research Consultants, LLC). Researchers met frequently to reconcile differences in code application and identify new themes emerging from the data. These new themes were incorporated into the coding scheme, and all transcripts were coded a second time using the revised coding scheme. After coding was complete, we identified key themes that could influence the acceptability and use of 99DOTS based on the general representativeness or salience of particular themes. We assembled these themes within the broader constructs of the UTAUT and identified representative quotations for each selected theme.

Although we could have quantified codes applied across interviews using Dedoose software, we did not do so for a few reasons. First, quantification implies that our findings are representative of those in a larger population; however, we used purposeful sampling of study participants. In addition, rather than just reporting common themes (ie, those reported most frequently), we also reported salient themes (ie, ones that seem important even if reported by a minority of participants). Second, quantification often implies that the same questions were asked to all participants in a systematic manner, as is done in structured surveys. Although field researchers underwent rigorous training and followed a uniform interview guide, the questions asked were open ended or semistructured, with a goal of eliciting narrative data that would be followed up with probe questions that could vary depending on the participant's response. This malleable approach to interviewing is a strength of qualitative research, as it often yields unexpected findings; however, quantifying such findings could be misleading by implying a structured interviewing strategy.

In our analysis, we did not classify patients into those with *high acceptance* or *low acceptance* of 99DOTS. Although patients did have differences in the use of 99DOTS, individual patients often found different components of 99DOTS to be both acceptable and unacceptable depending on their individual circumstances. For example, the same patient might appreciate that 99DOTS averts the time spent traveling to health facilities (when hypothetically compared with facility-based DOT) but simultaneously find it difficult to use 99DOTS at home because of concerns about the disclosure of their TB diagnosis to family members. As such, we coded both positive and negative perceptions of 99DOTS that were reported by a given patient to provide a broader understanding of the factors that might increase or decrease the technology's acceptability and use.

Results

Characteristics of the Study Participants

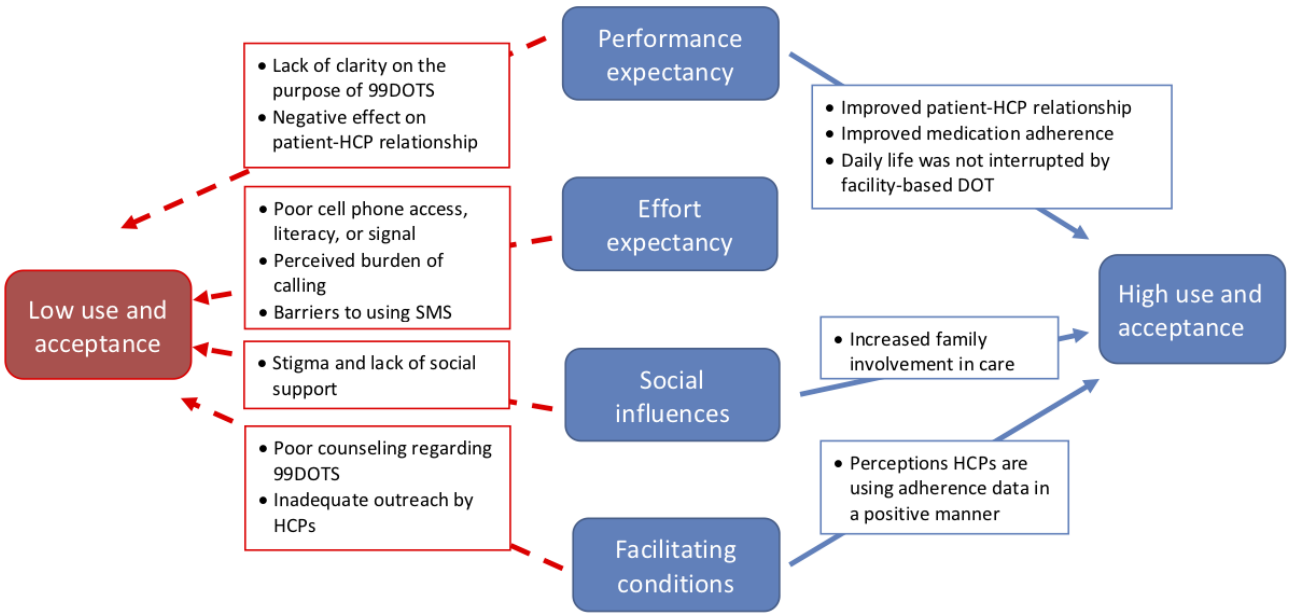
Excluding one nontranscribed interview, 62 patients with TB were interviewed, of whom 36 (58%) were men, 30 (48%) were HIV coinfected, and 34 (55%) were in the continuation phase of therapy. Patients' ages ranged from 18 to 65 years (median 35 years). The median household monthly income was Indian rupees 8500 (IQR Indian rupees 7000-12,000), which is approximately US $111 (IQR US $92-157). With regard to educational attainment, 10 (16%) patients were illiterate, 21 (34%) had completed some schooling, 20 (32%) had a secondary school certificate, and 11 (18%) had pursued or completed postsecondary education (ie, college education or higher). With regard to occupation, 14 (23%) patients were unemployed, 13 (21%) were homemakers, 16 (26%) worked in the informal sector or as private low-wage employees (eg, waste collector or housekeeper), 18 (29%) worked in semiskilled or skilled trades or self-employment (eg, electrician or auto-rickshaw driver), and 1 (2%) worked in a high-wage profession (ie, engineer).

In total, 31 HCPs were interviewed, of whom 13 (42%) were men. HCPs' ages ranged from 25 to 56 years (median 38 years). Of the 31 HCPs, 4 (13%) were health visitors, 4 (13%) were senior treatment supervisors, 2 (6%) were pharmacists, 2 (6%) were ART counselors, 3 (10%) TB were counselors, 1 (3%) was a nurse, 6 (19%) medical were officers, 2 (6%) were data managers, 2 (6%) were TB officers, 4 (13%) were district or city TB officers, and 1 (3%) was a district executive health officer.

Findings From Patients With TB

Patient interviews revealed evidence of both high and low acceptability of 99DOTS (Figure 1).

Figure 1. Key findings regarding the determinants of high and low acceptance and use of 99DOTS by patients with tuberculosis, based on the framework of the Unified Theory of Acceptance and Use of Technology. Note that there is no arrow between effort expectancy and high use and acceptance, because the findings did not reveal a meaningful relationship. DOT: directly observed therapy; HCP: health care provider; SMS: short messaging service.



Determinants of High Patient Acceptability and Use of 99DOTS

With regard to performance expectancy (perceived usefulness), some TB patients perceived that 99DOTS contributed to improved patient-HCP relationships, improved medication adherence, and fewer interruptions to their daily life activities (Table 1 and Figure 1). Some patients reported that 99DOTS facilitated increased phone contact by HCPs. Such episodes of phone outreach as well as the perception that their adherence was being remotely observed facilitated a perception that the government was taking a personal interest in their health (Table 1, Q1) and was more *caring* about patients:

> *I am thankful that the government is taking good care of my health. It feels good that the government is calling me... they are caring.* [50-year-old woman, HIV uninfected, continuation phase of therapy]

Some patients also reported benefits of the SMS text messaging reminders in facilitating habit formation with regard to daily pill taking; they reported that these reminders "trained [their] minds" and created a new "awareness" about adherence (Q2 and Q3). Some HIV-coinfected patients reported that 99DOTS provided collateral benefit by serving as a reminder to take their antiretroviral medications:

> *It's actually also reminding me to take ART tablets.* [49-year-old man, HIV coinfected, intensive phase of therapy]

A few patients who had also experienced the previous facility-based DOT model, which required thrice-weekly visits to health care facilities, reported benefits to their daily life, including saving money and time (Q4), from not having to visit health facilities as frequently in the 99DOTS model:

> *The previous approach [facility-based DOT] was time consuming and expensive too. This one [99DOTS] is good because you just have to take the pills and make the calls; it saves your time.* [19-year-old woman, HIV uninfected, intensive phase of therapy]

With regard to social influences, some patients said that 99DOTS increased social support by involving family members in their TB care (Q5); for example, family members routinely reminded or helped patients to call 99DOTS:

> *I have never forgotten [to call]. If I forget on one day, my daughter reminds me to make the call.* [50-year-old woman, HIV uninfected, continuation phase of therapy]

With regard to facilitating conditions, some patients noted that HCPs regularly used 99DOTS to guide their interactions, engaging in frequent phone outreach to patients or multiple home visits for counseling (Q6 and Q7). None of these patients reported having previous experiences with facility-based DOT. As such, their perception of increased HCP outreach may have been shaped by an implicit comparison of 99DOTS with unmonitored self-administered therapy, which is the standard of care for most other diseases. The perception that HCPs and the health system were using the adherence data from 99DOTS enhanced the motivation of these patients to continue using the technology. For example, the following patient appreciated how an HCP checked on his adherence and then wished him well:

> *One sir [HCP] called me and asked, "Are you taking tablets correctly?" I replied, "Sir I am taking them." He would then say to me, "OK, take them regularly. Take care of your health. Be safe."* [35-year-old man, HIV coinfected, continuation phase of therapy]

Table 1. Representative quotations on determinants of high patient acceptability and use of 99DOTS.

Determinant of acceptability	Quotations
Performance expectancy	
99DOTS improved the patient–health care provider relationship	• Q1. "[Being monitored by 99DOTS] doesn't mean she [the health visitor] will forget about us; I feel that madam still remembers me... if I don't call, she makes me understand what happens if I don't take the pills." (19-year-old woman, HIV uninfected, continuation phase of therapy)
SMS text messaging alerts improved medication adherence	• Q2. "Now I am remembering that I have to call at 11 O'clock. I have a new awareness that I should take tablets at 11'O clock." (36-year-old man, HIV coinfected, continuation phase of therapy) • Q3. "If we forget to call, the message alerts us to take tablets. So it trains our mind to take pills on time, so we will not forget to call. It is useful." (52-year-old woman, HIV coinfected, continuation phase of therapy)
Daily life was interrupted less when compared with facility-based directly observed therapy	• Q4. "I prefer [99DOTS] because it saves our time. We can take the pills at home, and we can also do our domestic work." (18-year-old woman, HIV uninfected, continuation phase of therapy)
Social influences	
Increased family involvement in the patient's care	• Q5. "[My son] taught me, 'You have to take 2 tablets per day and follow the arrow mark from the starting point'... He gave me my medicines from the beginning [of therapy] and reminds me to take tablets...he also dials the toll free number for me." (43-year-old woman, HIV coinfected, intensive phase of therapy)
Facilitating conditions	
Perception that health care providers are using the 99DOTS adherence data in a positive manner	• Q6. "They have called me four or five times. They tell me that I missed taking my tablets and that I should call them. They will know [if] we don't call them back." (49-year-old man, HIV coinfected, intensive phase of therapy) • Q7. Interviewer: "Did anybody come to your home from the hospital to see if you have taken your pills?"Respondent: "Yes... Madam came last week and another person also." Interviewer: "How many times?"Respondent: "4 to 5 times." (40-year-old man, HIV uninfected, continuation phase of therapy)

Determinants of Low Patient Acceptability and the Use of 99DOTS

Determinants of low patient acceptability and the use of 99DOTS are presented in Table 2 and Figure 1. With regard to performance expectancy (perceived usefulness), many patients who were supposed to use the technology had trouble understanding what 99DOTS was until it was explained to them by the field researchers. This reflected a suboptimal understanding of the purpose of 99DOTS in their TB care. Some patients also felt that 99DOTS resulted in a poorer patient-HCP relationship by reducing the frequency of face-to-face contact. This led to some patients feeling *isolated* (Table 2, Q8), whereas others described difficulties in reporting potential adverse reactions to medications because of the reduced frequency of visits:

> *I developed leg numbness during treatment... Maybe if I came to the center [more frequently] and discussed my problem with the doctor they could have given some suggestion. But I am able to share my problems only once a month with the doctor.* [32-year-old woman, HIV coinfected, continuation phase of therapy]

Other patients did not appreciate receiving negative feedback from HCPs for not calling 99DOTS (Q9), although they reported regularly taking their pills:

> *The HCP scolded me for not calling... If the patient doesn't give the missed call, then the government thinks the patient has not taken the tablet.* [65-year-old woman, HIV coinfected, continuation phase of therapy]

Several concerns emerged regarding effort expectancy (ease of use). A few patients reported not having any access to a phone either because they live in situations (eg, college dormitory with restrictive policies) in which they were not allowed to keep or use cell phones (Q10) or because they and their family members do not own a phone:

> *I live in an interior village. No one has a phone I can use. Only my sister has a phone. I don't have a phone in my own hand. We don't have many resources, Madam.* [32-year-old man, HIV coinfected, continuation phase of therapy]

A more prominent problem was shared cell phone use (Q11 and Q12), which reflected problems with accessibility (eg, another family member keeps the cell phone), literacy (eg, older patients did not know how to use the phone), or both, as described in the following quotation:

> *I don't know how to use it, so I do not keep the phone with me... My son leaves our hometown for employment. I can give the missed calls when he is at home... what can I do if my son is not available?* [65-year-old woman, HIV coinfected, continuation phase of therapy]

Sharing of cell phones often resulted in patients using 99DOTS in ways that would result in adherence data not being reported or being reported inaccurately in the 99DOTS system. For example, patients reported medication doses using a third-party phone number not registered with 99DOTS (Q12), called the same phone number over several days (Q13), or called all the phone numbers listed on an envelope on the same day (Q14). In addition, patients reported separating envelopes from medication blister packs to have family members call 99DOTS on their behalf, which could result in calls into the 99DOTS system happening independent of actual pill ingestion:

> *I leave that cover [envelope sleeve] with my son and just take the tablets with me [to consume at work]... I ask him at night [if he made the call].* [50-year-old woman, HIV coinfected, continuation phase of therapy]

Patients reported other technical barriers to using 99DOTS, including getting a busy signal when calling the 99DOTS number (Q15), lack of electricity in the home to keep their cell phones charged (Q16), or inability to read the number on the 99DOTS envelope (Q17). The most common technical problems related to challenges with SIM cards (eg, lost or nonfunctional cards, Q11 and Q12) or patients lacking cellular signal in their homes (Q18):

> *There is not enough [cellular] signal in the area where I live, so I borrow a phone from someone else and call the toll-free number.* [55-year-old man, HIV coinfected, continuation phase of therapy]

A few patients also described a sense of *fatigue* with calling 99DOTS, especially in the continuation phase of therapy (Q19). The following patient describes this feeling of decreasing motivation as treatment progressed:

> *Yes, my interest in calling has decreased compared to the initial phase... Now I don't [call after taking medication].* [35-year-old man, HIV uninfected, intensive phase of therapy]

Other patients articulated a vaguer feeling of tiredness or sleepiness after taking medication that prevented them from calling (Q20).

Several patients also reported barriers in ease of use for SMS text messaging reminders, including language barriers (SMS text messages were often sent in English rather than local Indian languages, Q21) and not noticing 99DOTS SMS text messaging reminders (Q22), in light of the high volume of *spam* SMS text messages sent by advertisers.

With regard to social influences, TB- and HIV-related stigma were major barriers to using 99DOTS. Some patients had not revealed their TB or HIV diagnoses to their family members because of stigma and, as a result, wanted to take their pills in the dark (when family members were sleeping) or in a private place outside the home, for fear of these diagnoses becoming disclosed (Q23 and Q24). One patient described having to step away from others to make the 99DOTS phone calls:

> *I am very worried about my children coming to know [about the TB diagnosis]... so I am making calls while hiding from others.* [51-year-old man, HIV coinfected, intensive phase of therapy]

The 99DOTS envelope served as a barrier to being able to cut up medication blister packs to facilitate taking TB medications discreetly. Patients also had concerns when HCPs visited their homes—in response to a lack of phone calls registered in the 99DOTS dashboard, as these visits sometimes resulted in disclosure of the patient's TB and/or HIV diagnoses to family members (Q25).

In terms of facilitating conditions, many patients reported poor counseling by HCPs regarding how to use 99DOTS and what its potential benefits might be (Q26 and Q27). For example, one patient did not realize that the SMS text messaging reminders he was receiving were from 99DOTS; he had assumed that they were from his son's employer:

> *I didn't know why [the SMS reminders from 99DOTS] came... I thought that I was getting messages from construction sites that my son works at.* [55-year-old man, HIV coinfected, continuation phase of therapy]

Some patients expressed disappointment in the lack of outreach by HCPs via phone calls or home visits. The reasons that patients felt disappointed due to lack of outreach varied. A few patients had prior experiences with facility-based DOT and implicitly compared this prior model with 99DOTS, which involves a lower frequency of HCP-patient interaction. Other patients felt disappointed because they received no outreach at all from HCPs by phone or home visit (Q28) or because HCPs had called to say they would conduct a home visit but did not actually do so, as described in the following quotation:

> *[HCPs] told me that they will come to my home, but they never came.* [28-year-old man, HIV uninfected, continuation phase of therapy]

Table 2. Representative quotations on determinants of low patient acceptability and use of 99DOTS.

Determinant of low acceptability	Quotations
Performance expectancy	
Technology negatively affects patient-provider relationship	• Q8. "One person came to see me since I started regularly taking TB tablets. They came and saw me once and that was it... I told them no one is visiting me and that I feel isolated. But nobody is interested in my worries." (32-year-old man, HIV coinfected, continuation phase of therapy) • Q9. "We wake up at 4 in the morning and we are fasting all day [for Ramadan], so in the morning there is no time to call... madam told me that I had not called and scolded me." (50-year-old woman, HIV uninfected, continuation phase)
Effort expectancy	
Inability to call because of lack of phone access or restricted phone access	• Q10. "I have a phone in my office. But students are not allowed to use mobile phones until we go home." (19-year-old man, HIV coinfected, continuation phase of therapy)
Inability to call because of shared phone use	• Q11. Interviewer: "Has it ever happened that you had not made the call because of your phone problems?" Respondent: "[When f]ather is not at home whole day or when there is a SIM card network issue also." (20-year-old woman, HIV uninfected, continuation phase of therapy) • Q12. "My mother used to take [the cellphone] with her... for those days I couldn't call... There was an alternative phone — my brother-in-law's mobile phone. After taking medication I would tell my brother-in-law... But he ended up losing his SIM card." (34-year-old man, HIV coinfected, intensive phase of therapy)
Inappropriate calling of phone numbers in the envelopes (may lead to inaccurate adherence information)	• Q13. "I saved the toll free number in the first blister [of the envelope] and called that number only. Later they said I should not call like that. They advised to call according to [the corresponding] blister [for each day]." (50-year-old man, HIV coinfected, continuation phase of therapy) • Q14. "My daughter calls...but she does not call daily. She calls all the numbers on the strip at once after all the medication [in the blister pack] has been taken." (55-year-old man, HIV coinfected, continuation phase of therapy)
Other barriers to cell phone use or calling 99DOTS	• Q15. "The first time I call, it gives a busy signal . . . after thinking that I dialed a wrong number, I dial the number again and it works... this has happened two or three times." (43 year-old-woman, HIV coinfected, intensive phase of therapy) • Q16. "I am staying in a hut, so I don't have electricity in my home; we burn [wood] sticks to get light." (49-year-old man, HIV coinfected, intensive phase of therapy) • Q17. "My vision is not clear enough to see the small print [on the envelope]... I get help from my daughter or someone who can see the letters and call." (42-year-old man, HIV coinfected, continuation phase of therapy) • Q18. "Yes, sometimes there are network [cellular signal] issues at my house." (25-year-old woman, HIV uninfected, continuation phase of therapy)
Perceived high burden of calling or "technology fatigue"	• Q19. "[I am] tired of calling daily." (19-year-old man, HIV uninfected, continuation phase of therapy) • Q20. "I forget to call Madam... I fall asleep as soon as I take [the medication]." (42-year-old man, HIV coinfected, continuation phase of therapy)
Barriers to using SMS text messaging reminders	• Q21. "If I receive a message in Tamil, I'll try to read by spelling out the letters... but I don't know how to read in English." (30-year-old man, HIV coinfected, continuation phase of therapy) • Q22. "Honestly I did not notice [the SMS reminders] or I did not check my phone for messages." (25-year-old woman, HIV uninfected, continuation phase of therapy)
Social influences	
Stigma and lack of social support present barriers to patient engagement	• Q23. "We do not take medication in front of others. People think TB is the worst disease and spreads by touching and that even if you talk, the disease will spread... all people think I am dirty; nobody wants to come close to me, so that's why we take our pills behind closed doors." (19-year-old woman, HIV uninfected, continuation phase of therapy) • Q24. "Sometimes I take [medication] when nobody else is at home or... when everyone at home is sleeping... Sometimes I leave my home to take the tablets." (19-year-old man, HIV coinfected, continuation phase of therapy) • Q25. "[The HCPs visiting my house] revealed my status. They told others that I have TB... I feel unworthy to live after others came to know that I have AIDS." (54-year-old man, HIV coinfected, intensive phase of therapy)
Facilitating conditions	

Determinant of low acceptability	Quotations
Poor counseling regarding 99DOTS	• Q26. "One informational paper was given but... I thought it was useless and threw it out. Then when I came again to collect medicine, they asked me why I was not calling and I told them that I never received a number to call." (27-year-old woman, HIV uninfected, intensive phase of therapy) • Q27. "Sometimes I wonder, why do they ask us to call? What is the reason for calling?" (29-year-old man, HIV coinfected, intensive phase of therapy)
Perceptions of inadequate or negative outreach by health care professionals	• Q28. "They [HCPs] didn't call... no one visited my house yet either." (35-year-old man, HIV coinfected, intensive phase of therapy)

Findings From HCPs

The HCP interview findings generally reflected high acceptance of 99DOTS, with a few exceptions (Figure 2 and Table 3). With regard to performance expectancy (perceived usefulness), HCPs generally found 99DOTS to be a helpful tool for monitoring patients and felt that the technology improved work efficiency. Higher-level managers found the ability to view patient adherence and outcomes at an aggregate level to be helpful, as described in the following quotation:

> *I am able to see the adherence of the patient in an analytical way, so it helps me and my staff to directly monitor the patient. I can randomly call patients from Mumbai. It's very easy. Without it, I would have to open my Excel [spreadsheet] to see [patient details].* [A higher-level administrative officer from Mumbai]

HCPs who interacted directly with patients found that automated notifications helped them to prioritize high-risk patients and encourage phone outreach (Table 3, Q29), which strengthened the patient-HCP relationship and resulted in early identification of medication adverse effects:

> *There was one patient who was registered for 99DOTS, but she was not calling. When she was contacted... she said she was experiencing some side effects of the drug... So definitely it is helpful.* [A higher-level administrative officer from Mumbai]

Some HCPs reported feeling that they had more time to perform other tasks (such as documentation), given the reduced time spent on directly observing medication ingestion by patients (Q30). HCPs also perceived that patients had improved adherence because of the feeling that they were being *cared for* by the HCPs on the other side of the monitoring technology (Q31). They also felt that 99DOTS enhanced the frequency and ease of outreach to patients by phone:

> *Actually, due to the 99DOTS program, communication [with the patient] has increased... [99DOTS] is useful...we can call the patient and inform them to take pills, which is beneficial.* [A TB Officer from Mumbai]

However, in contrast, a few HCPs described concerns about the accuracy of the adherence record and the quality of the patient-HCP interaction in the absence of direct observation, as noted in the following excerpt:

> *[Patients] may call us even without taking the tablet too. We cannot directly observe them with 99DOTS... [patients] may dial the phone number for our sake and throw out the tablets.* [TB counselor in Tamil Nadu]

With regard to effort expectancy (ease of use), HCPs generally found the 99DOTS adherence dashboard to be easy to use and preferred electronic data entry into the 99DOTS dashboard, as compared with written records (Q32). They also found it easy to generate reports through the 99DOTS dashboard in a manner that prioritizes high-risk patients, as described in the following quotation:

> *It's easy when it comes to reports. Immediately the report can be pulled out to determine which patients need to be given priority.* [A medical officer from Tamil Nadu]

The HCPs perceived that some patients found 99DOTS to be easy to use, but HCPs also affirmed findings from the patient interviews that many patients, especially those with low literacy or educational level (Q33), had challenges using 99DOTS:

> *[Patients] find it easy, but it goes hand in hand with literacy rate.* [A higher-level administrative officer from Mumbai]

With regard to social influences, HCPs described high uptake and utilization of 99DOTS by their supervisors and peers, such that 99DOTS improved coordination among HCPs, which likely served as a social influence that increased its acceptability (Q34). For example, the following nurse said that she is easily able to contact her colleagues in a timely manner due in part to 99DOTS:

> *We [HCPs] always have contact [with each other] through mobile phones. After registering the patient [in 99DOTS] they will get a message, and then they will call us and ask for a report on the patient's status and we will inform them. I have everyone's mobile number.* [A staff nurse from Tamil Nadu]

Although HCPs described positive findings across most UTAUT constructs, many HCPs described suboptimal facilitating conditions during the rollout of 99DOTS. Technical problems with implementation included stock outs of 99DOTS envelopes for wrapping medication blister packs (Q35) and the dashboard not being updated with adherence information, despite patients calling to report doses (Q36). Some HCPs reported a lack of training in 99DOTS (Q37) and difficulty in understanding how to practically apply the training (Q38) due to lack of demonstrative teaching:

> *Program clarity was there, but how it was going to be implemented in the field was not that clear, because there was no live demonstration of patient*

> *[interaction] and enrollment during the training session.* [A higher-level administrative officer from Mumbai]

Some HCPs also described inadequate preparation of personnel for 99DOTS implementation and unequal changes in workload, without commensurate increases in personnel. For example, HCPs in ART centers were previously not responsible for dispensing TB medications, as HIV-coinfected patients diagnosed with TB had previously been referred to DOT centers close to their homes to start treatment (Q39). However, with implementation of 99DOTS, because HCPs in DOT centers were no longer monitoring TB patients using facility-based DOT, the responsibility for dispensing TB medications was transferred to pharmacists at ART centers, as these pharmacists also dispensed ART to these patients for their HIV. As a result, medical officers (ie, doctors), pharmacists, and counselors at ART centers found that their workload increased substantially (Q39), as they had new responsibility for monitoring and counseling patients regarding their TB treatment and use of 99DOTS, without increased personnel support:

> *For just the ART tablets, we would count out missing pills and document them. Now for TB tablets, we also have to note down how many pills are missing... We have to fill so many pages... So workload is difficult. We did not have staff allocated for that change. We just have to do all the work.* [A counselor from Tamil Nadu]

These concerns by staff at ART centers were in contrast to the more general opinion among TB staff, in centers that did not take care of HIV-coinfected patients, that 99DOTS decreased workload.

Figure 2. Key findings regarding the determinants of acceptance and use of 99DOTS by health care providers, based on the framework of the Unified Theory of Acceptance and Use of Technology. Most determinants suggest high acceptance, except for facilitating conditions, as indicated by the red dotted line, which indicates a negative association. HCP: health care provider.

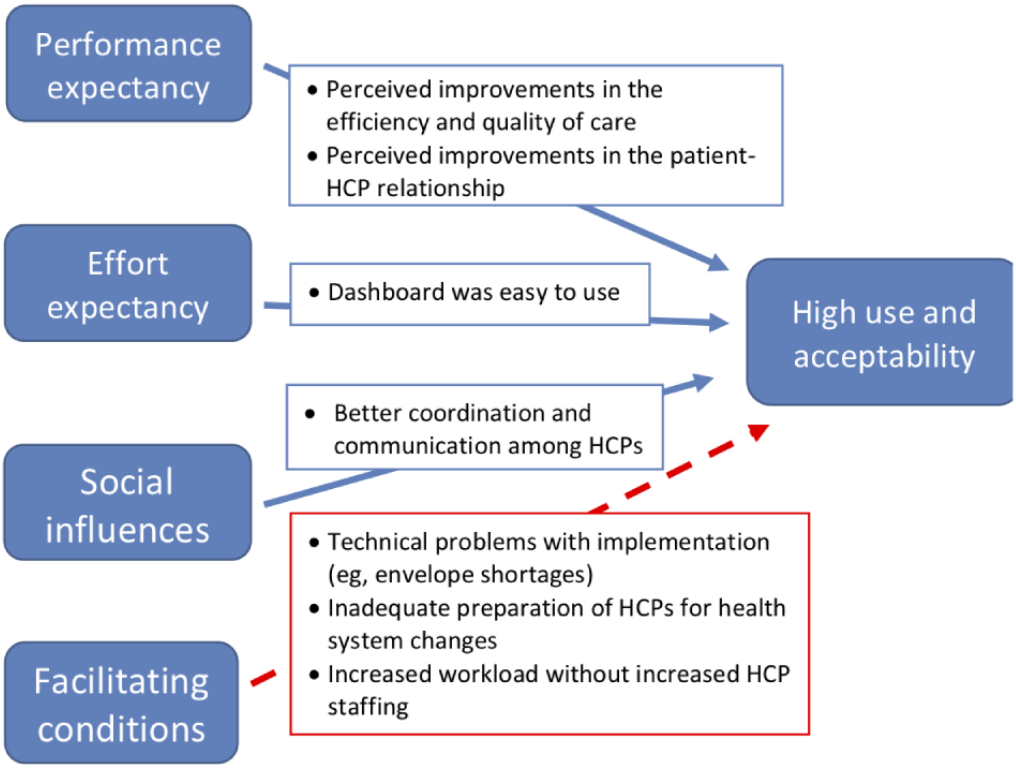


Table 3. Representative quotations on determinants of health care provider acceptability and use of 99DOTS.

Determinant of health care provider acceptability	Quotations
Performance expectancy	
Perceived improvements in the efficiency and quality of care	• Q29. "The health visitors and TB officers were coming to know who had missed doses in a span of one or two successive days. So it became easy for the staff to contact those patients and come to know the reason why patients were missing doses or why they had not called." (A higher-level administrative officer from Mumbai) • Q30. "Because [99DOTS uses remote monitoring] instead of daily supervision [i.e., DOT], I can relax and focus on my work... if I am only meeting a patient once a month, it reduces my work load." (A health visitor from Mumbai)
Perceived improvements in the patient-HCP[a] relationship	• Q31. "The patient mindset has changed...the patient starts to feel that [HCPs] are taking care of me; that is why they are telling me to make a phone call. So he will continue to take tablets correctly." (A counselor from Tamil Nadu)
Effort expectancy	
Dashboard was easy to use	• Q32. "[99DOTS] is easy from all angles. We are entering data manually as well as updating everything on the online portal. [99DOTS] is better to handle compared to manual entry for maintaining the [patient] register." (A senior treatment supervisor from Tamil Nadu)
Patient ease of use was perceived to be variable, depending on educational level	• Q33. "Generally if the patient completed a 10th grade education they know how to use the 99DOTS program." (A senior treatment supervisor from Mumbai)
Social influences	
Better coordination and communication among HCPs	• Q34. "High and medium priority patients are flagged... The patient list is taken and given to the counselor. A message is forwarded to the [senior treatment supervisors]. Every Thursday, a HIV/TB meeting is conducted by the HIV/TB coordinator. Data managers share patient information with the medical officers..." (A medical officer from Tamil Nadu)
Facilitating conditions	
Technical problems with implementation	• Q35. "One of the main problems with 99DOTS is that there was shortage [of envelopes]." (A medical officer from Tamil Nadu) • Q36. "I have a few patients call repeatedly...the number is correct and registered in the 99DOTS system. They have been getting the 'thank you' message also [which confirms the call was received]. But his treatment status does not get updated on the dashboard." (A higher-level administrative officer from Mumbai)
Inadequate training of HCPs for health system changes	• Q37. Interviewer: "Now, before implementation of 99DOTS program did you receive any training?" Respondent: "No, nowhere." (A health visitor from Mumbai) • Q38. "Initially when they were training [staff in 99DOTS], I did not understand anything. Then I took the class again. That time we got a better understanding of the project. I could only get a better understanding after I had seen patients individually." (A pharmacist from Tamil Nadu)
Changes in workload with 99DOTS implementation without changes in HCP staffing	• Q39. "Compared to ART [HIV treatment] this is difficult to handle... We didn't have any connection with DOTS [the national TB program] before. Once we diagnosed TB, we would give a referral form, and patients would get tablets in the DOT centers... Compared to our previous work, ART staff are doing much more work now." (A medical officer from Tamil Nadu)

[a]HCP: health care provider.

Discussion

Implications of Findings From Patients With TB

In this multisite study of the large-scale rollout of a novel DAT, we identified multiple factors that help to explain differences in the acceptability of 99DOTS as a monitoring strategy for TB patients and HCPs. TB patients often had substantially different perspectives regarding the technology, which were shaped by variability in patients' access to resources and social circumstances, including levels of TB- and HIV-related stigma in their families and communities.

With regard to resources, lack of reliable cell phone access—including not having a phone, only having access to a phone shared with family members, lack of electricity for phone charging, and unreliable cellular signal—considerably hindered ease of use (effort expectancy). However, cell phone–related barriers went beyond access alone. Some patients who had reliable cell phone access reported technology fatigue, expressed as *loss of interest* in calling 99DOTS. Technology fatigue has

also been reported in previous studies of nonsmartphone-based TB treatment monitoring strategies, in particular two-way SMS text messaging approaches [17]. This finding may be particularly concerning because prior evidence suggests that performance expectancy, or perceived usefulness of the technology, is the strongest predictor of intention to use a technology. Consistent with these findings, SMS text messaging reminders alone or two-way SMS text message–based strategies (ie, in which patients send SMS text messages to report doses taken) have been found to be ineffective or less effective than strategies that combine SMS text messages with other interventions to promote TB and HIV medication adherence [17-22]. These findings suggest that reminders and remote monitoring may be insufficient and that strategies that use technology to facilitate human interaction, rather than replacing it, may be more effective in improving medication adherence, a principle supported by recent trials of DATs in TB and HIV care that have had more successful outcomes [23-25].

Patients' individual social circumstances also shaped their acceptance of 99DOTS. For example, some patients felt that 99DOTS improved the patient-HCP relationship by facilitating more frequent phone outreach from HCPs, whereas other patients felt isolated by the reduced face-to-face contact with HCPs under 99DOTS. In addition, some patients, who presumably had low TB- and HIV-related stigma within their household, felt 99DOTS involved family members in their TB care in a beneficial manner. In contrast, other patients were so concerned about TB- and HIV-related stigma that they hid their disease (and pill taking) from family members, which made it challenging to regularly call 99DOTS.

For patients concerned about stigma, alternative monitoring strategies may be more or less conspicuous to family members and others in the community, depending on individual circumstances. For example, facility-based DOT, in which patients visit local health centers, where HCPs observe them take their medications, was previously widely used for monitoring in India's TB program. In this approach, individuals who work outside of the home may be able to visit health facilities during working hours without their family members' knowledge; however, for individuals who spend most of their time at home, engagement with facility-based DOT may be more noticeable than taking and reporting medications via the 99DOTS strategy.

In contrast, our findings suggest that unmonitored self-administered therapy—in which patients take their medication blister packs home with them without having to call 99DOTS—is a strategy that would be less conspicuous for most patients with concerns about TB-related stigma. Such an approach makes it easier to take medications discreetly for a few reasons. First, when medication blister packs were not wrapped in 99DOTS envelopes, patients could cut up blister packs into smaller strips that allowed them to be hidden more easily within the home. Second, cutting up blister packs into small strips also allowed patients to easily carry these medications in their pockets, so that medications could be ingested discreetly outside the home (eg, in a local temple in one case). Third, not having to dispense and call the 99DOTS phone number considerably reduced the visibility of the act of pill taking, particularly for patients who might rely on a family member's cell phone to call 99DOTS. For example, one patient described not being able to call 99DOTS because she took her medications discreetly in the dark after other family members went to sleep.

These various factors help to explain suboptimal patient use of 99DOTS [9,10], which contributes to its suboptimal accuracy for measuring adherence [10]. In our concurrent patient cohort study assessing 99DOTS' accuracy for measuring adherence [10], approximately one-third of patients had not called 99DOTS in the 3 days before our study team's unannounced home visit. However, more than three-fourths of these patients were actually taking their TB medications, as measured by a positive urine isoniazid test. Such patterns make sense in light of our qualitative findings, as many patients had barriers to calling 99DOTS (eg, stigma and lack of phone access) but still expressed motivation to take TB medications. That cohort study also found that a small proportion of patients shown as being *adherent* on the dashboard were calling 99DOTS but not actually taking their TB medications [10]. The qualitative findings in this study suggest that some of these inaccurate signals may be attributable to shared cell phone use, which sometimes inadvertently led to inappropriate calling of 99DOTS to report doses, for instance, when envelopes were separated from blister packs and given to family members to call in doses.

Our findings also have implications for the potential strengths and weaknesses of using other DATs to monitor TB medication adherence in this context. For example, some of the barriers to ease of use with 99DOTS could potentially be rectified by providing TB patients with alternative DATs that have a lower patient burden. For example, digital pillboxes record openings and closings of the pillbox as a proxy for medication ingestion. As patients have to open these pillboxes to access their medications, the patient burden is often considered to be lower than two-way SMS text messages or cell phone–based DATs such as 99DOTS, which require patients to take an additional step (SMS text message or phone call) to report a dose [4]. As such, the use of digital pillboxes or other alternative technologies may help to expand the reach of DATs to TB patients who have unreliable cell phone access or who lose interest in calling every day.

In contrast, patients who face high TB- and HIV-associated stigma in their family environment may have challenges not only with 99DOTS but also with alternative DATs, such as digital pillboxes or video DOT, because all these technologies result in pill taking being more noticeable to family members than might be the case with unmonitored self-administered therapy. For example, a digital pillbox would likely be even more difficult to hide from family members than the 99DOTS medication envelopes and, therefore, may have lower acceptability for some patients. As such, it is critical that rigorous research involving diverse patient populations be conducted before novel DATs are rolled out in the Indian setting to better understand their potential strengths and weaknesses in this context.

Implications of Findings From HCPs

The acceptability of 99DOTS for HCPs was generally high. HCPs with a variety of roles in the NTEP found the 99DOTS dashboard to be intuitive and user friendly and most had a general perception that the system had improved efficiency and quality of care, the patient-HCP relationship, and coordination among different HCPs involved in providing care.

However, some HCPs raised concerns about the facilitating conditions for the 99DOTS rollout. In particular, for HCPs at ART centers, the implementation of 99DOTS coincided with a *single window* approach to dispensing and monitoring both HIV and TB medications for HIV coinfected TB patients, partly because 99DOTS allowed centralized monitoring of TB medication adherence. As a result, counselors, pharmacists, and medical officers at ART centers voiced concerns about the increased workload of having to provide longitudinal TB care for the first time, in addition to their usual responsibilities. A recent study conducted at ART centers in Karnataka identified similar problems and found that implementation of 99DOTS with a *single window* approach may have adversely affected TB patient outcomes [11]. These concerns highlight the importance of anticipating changes in workload with the implementation of DATs. It was notable that despite the concerns raised by some HCPs regarding facilitating conditions, most HCPs still seemed to appreciate many features of 99DOTS.

We have previously described how DATs have the potential to transform the delivery of TB care if implemented in a manner that is patient centered [4,26]. For example, DATs have the potential to help HCPs to remotely view adherence data, prioritize their outreach efforts to focus on patients at highest risk, and provide differentiated care to patients based on their individual adherence records. Our findings suggest that HCPs understood the potential benefits of these innovations and were generally open to them. However, a limitation of this qualitative study is that it only assessed HCPs' perceptions of 99DOTS' benefits, rather than the actual accuracy of this technology or its impact on patient outcomes. 99DOTS' acceptability may have been overestimated because of social desirability bias (ie, patients or HCPs telling interviewers what they thought they wanted to hear). As already noted, a cohort study, conducted concordantly with this qualitative study, has revealed concerning limitations in 99DOTS' accuracy for measuring true TB medication adherence [10]. Although HCPs may be enthusiastic about 99DOTS, there may be serious shortcomings in the quality of the adherence data they are receiving, which is particularly concerning when these data are used for public health and clinical decision making.

Another limitation of our study is that we used a deductive approach to analyzing data based on the fact that the constructs in the UTAUT framework have previously been shown to predict technology acceptance and use. A limitation of this analytical approach is that we could have potentially missed some findings from these interviews that did not clearly fit into this predetermined framework.

As our patient sample only included individuals currently being monitored by 99DOTS, we were not able to compare patients' perceptions of 99DOTS with alternative care delivery strategies, such as facility-based DOT or unmonitored self-administered therapy. Although a few patients in our sample compared their experiences with 99DOTS with prior experiences with facility-based DOT, a larger sample would be required to draw definitive conclusions about whether and what types of patients prefer one monitoring approach over another. In addition, patients in our sample with prior experience of facility-based DOT voiced preference for 99DOTS primarily because they did not have to travel to health facilities on a regular basis; however, the same benefit could be achieved with unmonitored self-administered therapy. Future qualitative research should investigate patient preferences for different types of treatment monitoring strategies.

Future research may also evaluate whether the provision of monetary or nonmonetary incentives—or alternatively, disincentives, such as transition to facility-based DOT for patients who do not call 99DOTS—could facilitate greater engagement with the technology. Such studies incentivizing patients to call 99DOTS should be conducted with caution. Existing research comparing 99DOTS' accuracy against adherence measured via urine isoniazid testing suggests that a considerable proportion of nonadherent patients, as confirmed by negative urine isoniazid test results, still call 99DOTS to report taking medications [10]. The provision of incentives to increase patient engagement with 99DOTS has the potential to worsen this problem, which would further decrease the capacity of 99DOTS to identify patients who are truly not taking their TB medications.

Conclusions

In this multisite study of the large-scale implementation of 99DOTS, we found that 99DOTS had high acceptability among HCPs, but patients in India's public sector TB program had variable, sometimes negative, perceptions. Our findings highlight some barriers to the acceptance and use of 99DOTS that could potentially be mitigated through better rollout and implementation of the technology, including improvements in 99DOTS counseling (for patients) or training (for HCPs) and better anticipation of changes in workload for HCPs. In addition, screening TB patients for specific barriers to accepting or using 99DOTS, such as high levels of TB- and HIV-associated stigma in the household, lack of reliable cell phone access, or lack of cell phone literacy—could help target 99DOTS to patients who are more likely to engage with this technology while also identifying patients for whom alternative monitoring approaches (eg, digital pillboxes, video DOT, in-person DOT, or unmonitored self-administered therapy) should be offered.

Acknowledgments

This research was primarily supported by grants from the Bill and Melinda Gates Foundation (OPP1154670 and OPP1154665). RS also received support from a Harvard Catalyst KL2/Catalyst Medical Investigator Training Award (KL2 TR001100), a grant from the Harvard Center for AIDS Research (5P30AI060354-13), and a Doris Duke Clinical Scientist Development Award (grant #2018095). The funding bodies had no role in the study design, data collection, data analysis, data interpretation, or manuscript writing.

Conflicts of Interest

JEH consults for Merck. The remaining authors declare no conflict of interest.

Multimedia Appendix 1
Coding Scheme for the 99DOTS patient interviews.
[PDF File (Adobe PDF File), 38 KB - mhealth_v8i7e16634_app1.pdf]

Multimedia Appendix 2
Coding scheme for the 99DOTS HCP interviews.
[PDF File (Adobe PDF File), 68 KB - mhealth_v8i7e16634_app2.pdf]

References

1. Tuberculosis (TB): Global Tuberculosis Report. World Health Organization. 2019. URL: https://www.who.int/tb/publications/global_report/en/ [accessed 2020-07-15]
2. Imperial MZ, Nahid P, Phillips PP, Davies GR, Fielding K, Hanna D, et al. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. Nat Med 2018 Nov;24(11):1708-1715 [FREE Full text] [doi: 10.1038/s41591-018-0224-2] [Medline: 30397355]
3. Thomas A, Gopi PG, Santha T, Chandrasekaran V, Subramani R, Selvakumar N, et al. Predictors of relapse among pulmonary tuberculosis patients treated in a DOTS programme in South India. Int J Tuberc Lung Dis 2005 May;9(5):556-561. [Medline: 15875929]
4. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
5. Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2015 May 29(5):CD003343. [doi: 10.1002/14651858.CD003343.pub4] [Medline: 26022367]
6. Yellappa V, Lefèvre P, Battaglioli T, Narayanan D, van der Stuyft P. Coping with tuberculosis and directly observed treatment: a qualitative study among patients from South India. BMC Health Serv Res 2016 Jul 19;16:283 [FREE Full text] [doi: 10.1186/s12913-016-1545-9] [Medline: 27430557]
7. Sagbakken M, Frich JC, Bjune GA, Porter JD. Ethical aspects of directly observed treatment for tuberculosis: a cross-cultural comparison. BMC Med Ethics 2013 Jul 2;14:25 [FREE Full text] [doi: 10.1186/1472-6939-14-25] [Medline: 23819555]
8. Ngwatu BK, Nsengiyumva NP, Oxlade O, Mappin-Kasirer B, Nguyen NL, Jaramillo E, Collaborative Group on the Impact of Digital Technologies on TB. The impact of digital health technologies on tuberculosis treatment: a systematic review. Eur Respir J 2018 Jan;51(1):1701596 [FREE Full text] [doi: 10.1183/13993003.01596-2017] [Medline: 29326332]
9. Cross A, Gupta N, Liu B, Nair V, Kumar A, Kuttan R, et al. 99DOTS: A Low-Cost Approach to Monitoring and Improving Medication Adherence. In: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development. 2019 Presented at: ICTD'19; January 4, 2019; Ahmedabad, India URL: https://dl.acm.org/doi/10.1145/3287098.3287102 [doi: 10.1145/3287098.3287102]
10. Thomas BE, Kumar JV, Chiranjeevi M, Shah D, Khandewale A, Thiruvengadam K, et al. Evaluation of the accuracy of 99DOTS, a novel cellphone-based strategy for monitoring adherence to tuberculosis medications: comparison of digital adherence data with urine isoniazid testing. Clin Infect Dis 2020 Mar 28:- epub ahead of print. [doi: 10.1093/cid/ciaa333] [Medline: 32221550]
11. Thekkur P, Kumar AN, Chinnakali P, Selvaraju S, Bairy R, Singh AR, et al. Outcomes and implementation challenges of using daily treatment regimens with an innovative adherence support tool among HIV-infected tuberculosis patients in Karnataka, India: a mixed-methods study. Glob Health Action 2019;12(1):1568826 [FREE Full text] [doi: 10.1080/16549716.2019.1568826] [Medline: 30712507]
12. Prabhu A, Agarwal U, Tripathy JP, Singla N, Sagili K, Thekkur P, et al. '99DOTS'techno-supervision for tuberculosis treatment - a boon or a bane? Exploring challenges in its implementation at a tertiary centre in Delhi, India. Indian J Tuberc 2020 Jan;67(1):46-53. [doi: 10.1016/j.ijtb.2019.08.010] [Medline: 32192617]
13. Venkatesh V, Morris M, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]

14. Venkatesh V, Thong J, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
15. Dhanaraj B, Papanna MK, Adinarayanan S, Vedachalam C, Sundaram V, Shanmugam S, et al. Prevalence and risk factors for adult pulmonary tuberculosis in a metropolitan city of South India. PLoS One 2015;10(4):e0124260 [FREE Full text] [doi: 10.1371/journal.pone.0124260] [Medline: 25905900]
16. TB India 2018: Revised National TB Control Programme Annual Status Report. Central TB Division. 2018. URL: https://tbcindia.gov.in/showfile.php?lid=3314 [accessed 2020-07-20]
17. Mohammed S, Glennerster R, Khan AJ. Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial. PLoS One 2016;11(11):e0162944 [FREE Full text] [doi: 10.1371/journal.pone.0162944] [Medline: 27802283]
18. Liu X, Lewis JJ, Zhang H, Lu W, Zhang S, Zheng G, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med 2015 Sep;12(9):e1001876 [FREE Full text] [doi: 10.1371/journal.pmed.1001876] [Medline: 26372470]
19. Lester RT. Ask, don't tell - mobile phones to improve HIV care. N Engl J Med 2013 Nov 7;369(19):1867-1868. [doi: 10.1056/NEJMc1310509] [Medline: 24195570]
20. Campbell JI, Haberer JE. Cell phone-based and adherence device technologies for HIV care and treatment in resource-limited settings: recent advances. Curr HIV/AIDS Rep 2015 Dec;12(4):523-531. [doi: 10.1007/s11904-015-0282-8] [Medline: 26439917]
21. Linnemayr S, Huang H, Luoto J, Kambugu A, Thirumurthy H, Haberer JE, et al. Text messaging for improving antiretroviral therapy adherence: no effects after 1 year in a randomized controlled trial among adolescents and young adults. Am J Public Health 2017 Dec;107(12):1944-1950. [doi: 10.2105/AJPH.2017.304089] [Medline: 29048966]
22. Kanters S, Park JJ, Chan K, Socias ME, Ford N, Forrest JI, et al. Interventions to improve adherence to antiretroviral therapy: a systematic review and network meta-analysis. Lancet HIV 2017 Jan;4(1):e31-e40. [doi: 10.1016/S2352-3018(16)30206-5] [Medline: 27863996]
23. Yoeli E, Rathauser J, Bhanot SP, Kimenye MK, Mailu E, Masini E, et al. Digital health support in treatment for tuberculosis. N Engl J Med 2019 Sep 5;381(10):986-987. [doi: 10.1056/NEJMc1806550] [Medline: 31483974]
24. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
25. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
26. Haberer JE, Subbaraman R. Digital technology for tuberculosis medication adherence: promise and peril. Ann Am Thorac Soc 2020 Apr;17(4):421-423. [doi: 10.1513/AnnalsATS.202001-027ED] [Medline: 32233860]

Abbreviations

ART: antiretroviral therapy
DAT: digital adherence technology
DOT: directly observed therapy
HCP: health care provider
NIRT: National Institute for Research in TB
NTEP: National Tuberculosis Elimination Program
TB: tuberculosis
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 17.10.19; peer-reviewed by P Shete, WT Ooi, L Rabinovich; comments to author 03.03.20; revised version received 26.05.20; accepted 22.06.20; published 31.07.20.

Please cite as:
Thomas BE, Kumar JV, Onongaya C, Bhatt SN, Galivanche A, Periyasamy M, Chiranjeevi M, Khandewale AS, Ramachandran G, Shah D, Haberer JE, Mayer KH, Subbaraman R
Explaining Differences in the Acceptability of 99DOTS, a Cell Phone–Based Strategy for Monitoring Adherence to Tuberculosis Medications: Qualitative Study of Patients and Health Care Providers
JMIR Mhealth Uhealth 2020;8(7):e16634
URL: http://mhealth.jmir.org/2020/7/e16634/
doi:10.2196/16634
PMID:32735220

Original Paper

A Comprehensive App That Improves Tuberculosis Treatment Management Through Video-Observed Therapy: Usability Study

Xujun Guo[1*], MPH, MD; Yarui Yang[1*], MPH, MD; Howard E Takiff[1,2,3], MPH, MD; Minmin Zhu[1], MD, PhD; Jianping Ma[1], MPH, MD; Tao Zhong[1], MD; Yuzheng Fan[1], MPH, MD; Jian Wang[1*], MD; Shengyuan Liu[1], MD, PhD

[1]Department of Tuberculosis Control and Prevention, Shenzhen Nanshan Centre for Chronic Disease Control, Shenzhen, China

[2]Pathogenomique Mycobacterienne Integree, Institut Pasteur, Paris, France

[3]Laboratorio de Genética Molecular, Instituto Venezolano de Investigaciones Cientificas, Caracas, Venezuela

[*]these authors contributed equally

Corresponding Author:
Shengyuan Liu, MD, PhD
Department of Tuberculosis Control and Prevention
Shenzhen Nanshan Centre for Chronic Disease Control
7 Huaming Road
Nanshan District
Shenzhen, 518054
China
Phone: 86 0755 26407238
Email: liushenglb@126.com

Abstract

Background: Treatment of pulmonary tuberculosis (TB) requires at least six months and is compromised by poor adherence. In the directly observed therapy (DOT) scheme recommended by the World Health Organization, the patient is directly observed taking their medications at a health post. An alternative to DOT is video-observed therapy (VOT), in which the patients take videos of themselves taking the medication and the video is uploaded into the app and reviewed by a health care worker. We developed a comprehensive TB management system by using VOT that is installed as an app on the smartphones of both patients and health care workers. It was implemented into the routine TB control program of the Nanshan District of Shenzhen, China.

Objective: The aim of this study was to compare the effectiveness of VOT with that of DOT in managing the treatment of patients with pulmonary TB and to evaluate the acceptance of VOT for TB management by patients and health care workers.

Methods: Patients beginning treatment between September 2017 and August 2018 were enrolled into the VOT group and their data were compared with the retrospective data of patients who began TB treatment and were managed with routine DOT between January 2016 and August 2017. Sociodemographic characteristics, clinical features, treatment adherence, positive findings of sputum smears, reporting of side effects, time and costs of transportation, and satisfaction were compared between the 2 treatment groups. The attitudes of the health care workers toward the VOT-based system were also analyzed.

Results: This study included 158 patients in the retrospective DOT group and 235 patients in the VOT group. The VOT group showed a significantly higher fraction of doses observed (P<.001), less missed observed doses (P<.001), and fewer treatment discontinuations (P<.05) than the DOT group. Over 79.1% (186/235) of the VOT patients had >85% of their doses observed, while only 16.4% (26/158) of the DOT patients had >85% of their doses observed. All patients were cured without recurrences. The VOT management required significantly (P<.001) less median patient time (300 minutes vs 1240 minutes, respectively) and transportation costs (¥53 [US $7.57] vs ¥276 [US $39.43], respectively; P<.001) than DOT. Significantly more patients (191/235, 81.3%) in the VOT group preferred their treatment method compared to those on DOT (37/131, 28.2%) (P<.001), and 92% (61/66) of the health care workers thought that the VOT method was more convenient than DOT for managing patients with TB.

Conclusions: Implementation of the VOT-based system into the routine program of TB management was simple and it significantly increased patient adherence to their drug regimens. Our study shows that a comprehensive VOT-based TB management represents a viable and improved evolution of DOT.

(JMIR Mhealth Uhealth 2020;8(7):e17658) doi:10.2196/17658

KEYWORDS

tuberculosis; management; video-observed therapy; directly observed therapy; mobile phone

Introduction

Over 10 million people have been reported to develop tuberculosis (TB) in a year globally [1]. Although the incidence of TB in China has been decreasing faster than the rates of global decline, 800,000 TB cases are still being diagnosed annually [2-4], thereby imposing a significant cost on the nation's health care system. To achieve a high probability of cure with the current treatment regimens, patients with TB must take a combination of drugs for at least 6 months if the strain is sensitive to all antibiotics and for even longer periods if the strain is drug-resistant. If patients stop taking drugs when they feel better or if they miss a significant proportion of their doses, their chances of being cured are reduced. To encourage patient compliance with the treatment regimen, the World Health Organization has a strategy known as the directly observed therapy (DOT), which requires the patients to come to a health post daily to receive their pills so that their pill intakes are monitored. DOT has been credited with improved cure rates in some settings [5,6] but it presents challenges for both patients and the health care workers such that its implementation has become limited [7-10]. DOT imposes a burden on patients in terms of the time needed each day to visit the health post and the consequent absence from home or work responsibilities, the possible cost of transportation, particularly in rural areas, as well as the potential stigma. In addition, the daily administration of medication to all patients with TB is challenging for the health care staff.

Using the technical capabilities of smartphones, several studies have explored alternatives that employ video-observed therapy (VOT), which are less disruptive to the patients' work and family life, more cost-effective, and improves the patients' access to doctors [11-13]. The World Health Organization conditionally recommended VOT as an alternative to DOT in 2017, but the evidence was graded as weak because only few randomized controlled trials had been published [14,15]. In the VOT method, the patients either connect via a computer or a smartphone to health care workers who will observe them swallowing the medication in real time [16,17] or alternatively, the patients record videos of themselves taking the pills and then send the videos to the health care workers who can view these videos at their convenience [18].

We developed a comprehensive software for the management of patients with TB by using video monitoring through mobile phones, thereby combining the advantages and the convenience of the VOT method mentioned above. The patients visit their health posts weekly to collect their medications and send videos of themselves taking their pills daily in real time to the health care workers who can review these videos at their convenience. In this study, we evaluate the software for this VOT strategy and its implementation into the routine TB control program of the Nanshan District of Shenzhen, China, and compare the effectiveness of VOT with that of DOT by examining the retrospective data from patients during the year preceding the implementation of the VOT-based program.

Methods

Settings

Shenzhen is a rapidly growing metropolis in the south of China that attracts internal migrant workers from all over the country. The Nanshan District is located in the southwest of Shenzhen, with an area of 187.53 square kilometers and a per capita gross domestic product of US $51,000 in 2018 [19], which is the highest for all districts or counties in China. The penetration rate of mobile phones among the residents and workers in Shenzhen is estimated to be over 90%.

Study Design and Participants

We assessed the effectiveness of the comprehensive TB management system that was implemented in September 2017 as the routine management procedure in the Center for Chronic Disease Control (CCDC) of the Nanshan district of Shenzhen, China. Patients beginning treatment for TB between September 2017 and August 2018 were enrolled into the VOT group, and their data were compared with the retrospective data collected from patients who began TB treatment between January 2016 and August 2017 and were managed with routine DOT. All health care workers involved in the management of patients with TB in the Nanshan District participated in the study, and they had managed patients on DOT before VOT was implemented.

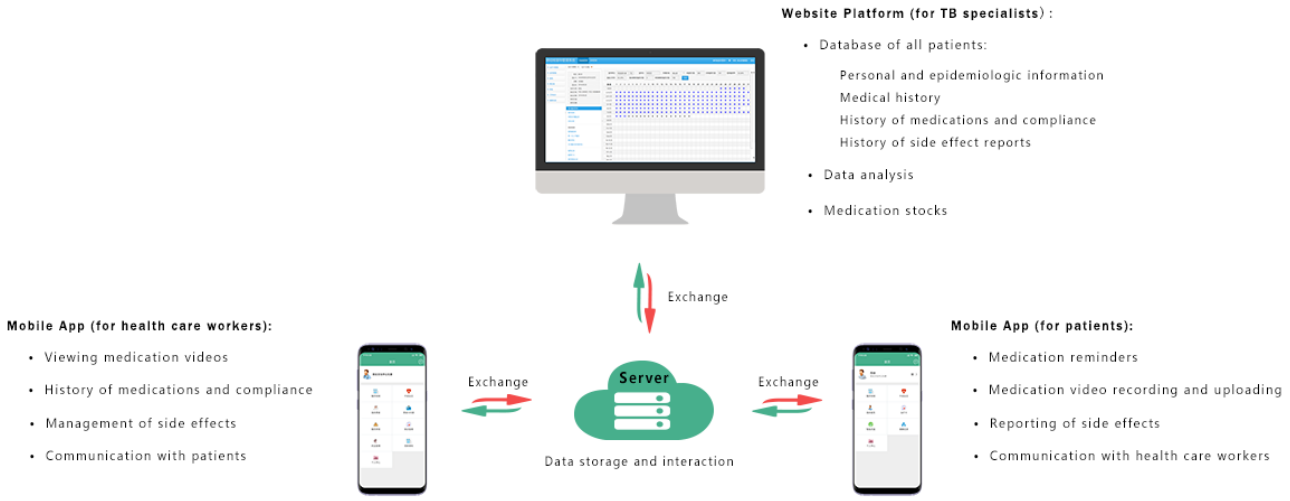
The inclusion criteria were as follows: (1) patients with pulmonary drug-sensitive TB who were treated and managed in the Nanshan CCDC; (2) patients who received a regimen containing only oral medication with fixed-dose combination (FDC) pills at an outpatient facility; and (3) patients who voluntarily provided informed consent to allow their data to be used in research studies. Exclusion criteria were as follows: (1) the presence of other conditions that could compromise the performance of VOT or DOT such as mental disorders, visual, hearing, or speech impairment, or the inability to live on their own; (2) patients who for any reason stopped their anti-TB treatment at the Nanshan CCDC, and (3) patients who did not have a smartphone with an internet connection or a family member with a smartphone and internet connection who could help them.

Comprehensive TB Management System

The software for the comprehensive TB management system was independently developed by the Nanshan CCDC. It consists of an app loaded onto the mobile phones of both patients and the medical personnel and a web-based monitoring platform. This app can be used in smartphones with Android or iPhone operating systems, but it is currently available only in Mandarin (Figure 1). This app has multiple functions, including an easy interface to report the side effects, reminders for scheduled medical appointments during treatment, and quantity management of the medicines for the health posts. The patients in the VOT group received training for the following modules in the app: setting up an account, logging into the account,

setting the time for their medication reminder, recording and uploading a video, reporting side effects, and communicating with a health care worker. When the patient logs in to the app, a daily pop-up message asks if they notice any adverse reactions, with a selection of options concerning their symptoms and duration and an empty space to describe the problem in detail. Patients can also log into the app at any time to report adverse reactions. If there are no adverse reactions, the patient ignores this pop-up message. This app includes a daily alarm set by the patients based on their own preferences, which reminds them to take their medication. The patients takes a video of them swallowing the pills, which is uploaded in real time to the app server. Patients cannot store the video to upload at a later time. The health care workers assigned to monitor each patient can view the videos at their convenience. If a patient does not send a video for a whole day, after midnight, the app shows a pop-up message on the patient's phone reminding him/her to take the medication and upload the video. The health care worker for each patient also receives a message that the patient did not upload the video and the health care worker will contact the patient on the following day through the app or by a phone call to remind him/her to take the medication and send the video.

Figure 1. Schematic representation of the data collection and transmission on the comprehensive tuberculosis management system. TB: tuberculosis.



DOT and VOT procedures

Treatment management with DOT was performed according to the current clinical practice. Most patients with TB in Nanshan are prescribed pills containing FDC of more than one drug per pill, and the patients are instructed to come daily to the community health service center (CHSC) to be observed while taking their medication. The patients requiring nonstandard doses are prescribed pills with the individual drugs that must be taken 2 or 3 times a day. These patients obtain their medicines every 15 days and self-administer these medicines at home. Of the 158 patients in the DOT group who began taking FDC pills once a day at the CHSC, 56 patients had their therapeutic regimen changed to pills with individual drugs, usually because of adverse effects, and the period during which they did not take the FDC pills once daily was excluded from the study.

Patients on VOT management were prescribed FDC drugs and they went to the CHSC every 7 days to receive their medication. Of the 235 patients in the VOT group, 31 patients changed their therapeutic regimen from FDC pills to individual drug pills, and only the period when the patients took the FDC pills once a day was evaluated.

Patients on DOT who failed to visit the health posts daily and patients on VOT who did not send a video were telephoned by health care workers the next day to ask why they missed sending the video and to remind them to take medication. If they could not be telephoned successfully, the health care workers would visit their home within 7 days. If the patient in the VOT group did not upload a video for 3 consecutive days, they were advised to transfer to DOT management. Several patients on DOT adamantly refused to come to the health post daily and were allowed to take their medications at home unobserved.

After successfully finishing the therapy, patients were advised to return to the Nanshan CCDC every 6 months to undergo chest radiography and a sputum smear test for investigation of a recurrence. If they did not return to the CCDC, they would be telephoned and asked to go to the clinic or to report the results of the follow-up tests that they might have done elsewhere. Moreover, periodically, the Nanshan CCDC searched for new diagnostic results of the patients who had finished their TB treatment on the web-based TB information management system set up by the National Center for TB Control and Prevention. All diagnosed TB cases were required to be listed in this registry.

A total of 66 trained health care workers at 66 CHSCs were involved in viewing the videos on the app and briefly seeing the patients when they collected their medications each week. In addition, the patients routinely visited a TB specialist at the Nanshan CCDC every 2 weeks for the first 2 months of treatment and every month for the remainder of the treatment—normally 4 additional months—and blood and urine specimens were obtained for routine tests during these visits. In these visits, the specialists asked the patients whether they missed their doses in order to check the validity of the DOT or VOT records. The TB specialists also periodically reviewed the information provided by the app on the medication stocks. Follow-up sputum smears for microscopy were taken after 2,

3, 5, and 6 months of treatment. Chest radiography images were obtained every 1 or 2 months.

Data Collection

Responding to questionnaires before and after treatment is the standard procedure for all patients with TB. These questionnaires were handwritten by the patients in the DOT group and electronically completed by the patients in the VOT group. The questionnaire consisted of questions on the sociodemographic variables. After the patients completed the treatment, they were asked to assess their acceptance and satisfaction of their treatment management as well as their incurred costs. The data on compliance and clinical results were collected from handwritten records of the patients on DOT and from the comprehensive TB management system records for the patients on VOT. All patient information was made irreversibly anonymous. After the implementation of the project, we asked 66 health care workers to complete an electronic questionnaire on their attitudes toward patient management with DOT and VOT.

Treatment adherence was measured by the number of missed medications, the number of times the patient discontinued their medications for 3 or more continuous days, and the fraction of doses observed, which is the number of the directly observed or video-observed doses divided by the total number of doses expected to be taken during the course of the treatment.

Patients were asked about the time required to go from their point of departure to the CHSC to obtain their medications, the travel costs, and how long the they typically spent in the CHSC to obtain their medicines and see a health care worker. The total number of trips to the CHSC during the course of treatment was collected from the patient records. The time and costs for single visits were then multiplied by the total number of trips to the CHSC to obtain the total time and total costs for the patients to obtain their medications. The median cost during the VOT study period was ¥7 (US $1).

Statistical Analysis

We used Kruskal-Wallis test, chi-square test, and Fisher's exact test within SPSS version 19.0 (IBM Corp) to evaluate the possible differences between the DOT and VOT groups in the sociodemographic characteristics, clinical features, treatment adherence, patient time and costs, positive sputum smear findings, and patient and health care worker satisfaction.

Results

Patient Characteristics

This study initially consisted of 393 cases of rifampicin-sensitive pulmonary TB, with 158 patients in the DOT group and 235 patients in the VOT group. There were no statistical differences between the 2 groups in any of the clinical or sociodemographic parameters recorded (P>.05) (Table 1 and Table 2). Of the 290 patients initially registered for TB treatment with FDC pills at the Nanshan CCDC between September 2017 to August 2018, only 2.1% (6/290) were excluded from the study because they or their family members lacked a smartphone with an internet connection. Another 49 patients were excluded because they stopped their treatment in Nanshan for reasons including moving out of Shenzhen, stopping their treatment on their own, or because it was no longer possible to contact them. The mean time for learning to use the app was 5.7 (SD 6.6) minutes.

Validity and Utility of the TB Management Systems

Medication Adherence

The fraction of the doses observed in the VOT group was significantly higher than that observed in the DOT group (P<.001), with 63.0% (148/235) versus 4.4% (7/158) having ≥95% of their doses observed, respectively (Table 3). The VOT group also had significantly less missed observed doses (P<.001) and discontinuations (P=.002) than the DOT group. In the DOT group, 55.1% (87/158) of the patients had less than 50% of their doses observed at the CHSC, but 74.1% (117/158) reported no missed doses. This discrepancy is because many of the patients in the DOT group often refused to come to the CHSC daily and were given medications to self-administer at home, which they reported to have taken. Similarly, in the VOT group, only 63.0% (148/235) of the patients had ≥95% of their expected doses observed and 91.1% (214/235) of the patients reported no missed doses. The striking difference was that 79.1% (186/235) of the VOT patients had ≥85% of their doses observed, while only 16.4% (26/158) of the DOT patients had ≥85% of their doses observed. Five patients in the VOT group failed to send a video for at least three consecutive days and were advised to transfer to DOT management, but all of them requested to be allowed to remain on VOT (Table 3).

Table 1. Sociodemographic characteristics of the patients in the DOT[a] and VOT[b] groups.

Patient demographics	DOT group (n=158), n (%)	VOT group (n=235), n (%)	*P* value
Gender			.28
Male	102 (64.6)	139 (59.1)	
Female	56 (35.4)	96 (40.9)	
Ethnicity			.06
Han nationality	157 (99.4)	225 (95.7)	
Others	1 (0.6)	10 (4.3)	
Occupation			.59
Food and beverage industry	3 (1.9)	3 (1.3)	
Cadre staff	78 (49.4)	134 (57.0)	
Workers	27 (17.1)	33 (14.0)	
Domestic and unemployed	33 (20.9)	33 (14.0)	
Teachers	1 (0.6)	1 (0.4)	
Retired personnel	2 (1.3)	4 (1.7)	
Students	9 (5.7)	17 (7.2)	
Other	5 (3.2)	10 (4.3)	
Education level			.31
Junior high school and below	29 (18.4)	39 (16.6)	
High school	65 (41.1)	115 (48.9)	
Undergraduate and above	64 (40.5)	81 (34.5)	
Social health insurance status			.91
Shenzhen Medical Insurance 1st	64 (40.5)	87 (37.0)	
Shenzhen Medical Insurance 2nd	38 (24.1)	61 (26.0)	
Shenzhen Medical Insurance 3rd	13 (8.2)	26 (11.0)	
Medical insurance outside Shenzhen	15 (9.5)	19 (8.1)	
No medical insurance	16 (10.1)	22 (9.4)	
Other	12 (7.6)	20 (8.5)	
Was there a stable residence during the period of treatment?			.98
No	10 (6.3)	15 (6.4)	
Yes	148 (93.7)	220 (93.6)	
Annual family total income (¥), (~Max USD)			.47
<10,000 ($1,430)	1 (0.6)	2 (0.9)	
10,000-99,999 ($14,300)	64 (40.5)	104 (44.3)	
100,000-199,999 ($28,600)	49 (31.0)	67 (28.5)	
200,000-299,999 ($42,900)	32 (20.3)	34 (14.5)	
300,000-399,999 ($57,200)	7 (4.4)	13 (5.5)	
≥400,000	5 (3.2)	15 (6.4)	
Age (years)			.22
<25	39 (24.7)	75 (31.9)	
25-44	99 (62.7)	139 (59.1)	
45-64	17 (10.8)	20 (8.5)	
≥65	3 (1.9)	1 (0.4)	

[a]DOT: directly observed therapy.

[b]VOT: video-observed therapy.

Table 2. Clinical features of the patients in the DOT[a] and VOT[b] groups.

Clinical features	DOT group (n=158), n (%)	VOT group (n=235), n (%)	*P* value
Abnormal chest radiographs			N/A[c]
Yes	158 (100.0)	235 (100.0)	
No	0 (0)	0 (0)	
Cavity			.28
Yes	36 (22.8)	43 (18.3)	
No	122 (77.2)	192 (81.7)	
Registration classification			.65
New patients	157 (99.4)	232 (98.7)	
Patients with recurrence	1 (0.6)	3 (1.3)	
Symptoms			.16
Night sweat	1 (0.6)	0 (0)	
Cough for less than 2 weeks	14 (8.9)	22 (9.4)	
Cough for more than 2 weeks	34 (21.5)	57 (24.3)	
Cough sputum for less than 2 weeks	3 (1.9)	0 (0)	
Cough sputum for more than 2 weeks	1 (0.6)	0 (0)	
Hemoptysis or blood sputum	7 (4.4)	17 (7.2)	
Chest tightness	0 (0)	3 (1.3)	
Chest pain	10 (6.3)	19 (8.1)	
Asymptomatic	88 (55.7)	117 (49.8)	
Sputum smears at the time of diagnosis			.91
Negative	125 (79.1)	194 (82.6)	
2-3 bacilli/300 fields of view	0 (0)	1 (0.4)	
4-10 bacilli/300 fields of view	3 (1.9)	2 (0.9)	
3-9 bacilli/100 fields of view	7 (4.4)	11 (4.7)	
1-9 bacilli/10 fields of view	8 (5.1)	9 (4.3)	
1-9 bacilli/field of view	6 (3.8)	7 (3.0)	
≥10 bacilli/field of view	9 (5.7)	11 (4.7)	
Sputum cultures at the time of diagnosis			.37
Positive	72 (45.6)	118 (50.2)	
Negative	86 (54.4)	117 (49.8)	

[a]DOT: directly observed therapy.

[b]VOT: video-observed therapy.

[c]Not applicable: Chi-square test or Fisher's exact test could not be used when 2 groups are compared with all cases representing the same results.

Table 3. Compliance with medication and adverse event management of the patients.

Characteristics	DOT[a] group (n=158), n (%)	VOT[b] group (n=235), n (%)	*P* value
Fraction of the doses observed[c] (%)			<.001
<50	87 (55.1)	7 (3.0)	
50-74	22 (13.9)	19 (8.1)	
75-84	23 (14.9)	23 (9.8)	
85-89	10 (6.3)	17 (7.2)	
90-94	9 (5.7)	21 (8.9)	
≥95	7 (4.4)	148 (63.0)	
Reported missed medication doses[d]			<.001
0	117 (74.1)	214 (91.1)	
1-5	35 (22.2)	19 (8.1)	
6-10	5 (3.2)	0 (0)	
>10	1 (0.6)	1 (0.4)	
Number of times the medication was discontinued for ≥3 days[e]			.002
0	147 (93.0)	230 (97.9)	
1	10 (6.3)	2 (0.9)	
2	0 (0)	2 (0.9)	
4	0 (0)	1 (0.4)	
5	1 (0.6)	0 (0)	
Adverse events reported by patients			.13
Yes	30 (19.0)	60 (25.5)	
No	128 (81.0)	175 (74.5)	
Number of adverse events reported by patients			.22
0	128 (81.0)	175 (74.5)	
1	26 (16.5)	54 (23.0)	
2	3 (1.9)	6 (2.6)	
3	1 (0.6)	0 (0)	
Suspension due to medical advice			.19
Yes	33 (20.9)	37 (15.7)	
No	125 (79.1)	198 (84.3)	
Number of suspensions due to medical advice			.22
0	125 (79.1)	198 (84.3)	
1	33 (20.9)	35 (14.9)	
2	0 (0)	1 (0.4)	
4	0 (0)	1 (0.4)	

[a]DOT: directly observed therapy.

[b]VOT: video-observed therapy.

[c]The number of observed doses divided by the total number of doses expected to be taken. The numerator includes either directly observed or video-observed doses, excluding those obtained by self-report. The denominator includes the self-administered medications; it does not include periods when the treatment was suspended based on medical advice because of side effects.

[d]Missed doses reported by the patients, as the patients on VOT could have taken their medications but they may have forgotten to record a video or they may have self-administered DOT medications at home; therefore, the patients could have taken the reported doses more than the ones observed by the health care workers.

[e]The times when patients missed their doses for at least three consecutive days on their own, not including the occasions when medication was suspended owing to medical advice.

Clinical Results

At the time of the diagnosis, 33 of the 158 patients in the DOT group showed positive sputum smear findings, while 40 showed smear-negative and culture-positive findings. Of the 235 patients in the VOT group, 40 showed positive smear findings and 82 showed negative smears but positive culture findings. Because clinical improvement was followed based upon sputum smear findings, only patients with positive smear findings could be compared to monitor the treatment success by observing for the clearance of the bacilli from the sputum. In most patients, the bacilli were cleared from the sputa after 2 months of therapy, but more patients in the DOT group required 3 months (15%, 5/33 vs 5%, 2/40) or 5 months (3%, 1/33 vs 0%, 0/40) for their sputum smear findings to be negative than those in the VOT group, respectively, although the difference was not statistically significant (*P*=.17). All patients showed negative smears by the end of their 6 months of treatment, and no recurrences were recorded in either group after at least six months of follow-up.

Management of the Side Effects

Adverse events were reported to the health care workers by 19.0% (30/158) of the patients in the DOT group and 25.5% (60/235) patients in the VOT group (Table 3). Medication was suspended in 20.9% (33/158) of the patients in the DOT group and 15.7% (37/235) of the patients in the VOT group, most commonly for abnormal results for liver function tests and for rash, itching, or hives. None of the differences between the 2 groups related to side effects was statistically significant (*P*>.05), but significantly more patients in the DOT group (*P*<.001) than those in the VOT group had their therapeutic regimen changed (35.4%, 56/158 vs 13.2%, 31/235, respectively).

Patient Transportation Time and Expenditures

Time Required

After completing the treatment, the patients were given questionnaires about various aspects of their experiences during the treatment. These questionnaires were completed by 82.9% (131/158) of the patients in the DOT group and 100.0% (235/235) of the patients in the VOT group, although not all the participants answered all the questions. Approximately 64.8% (81/125) of the patients in the DOT group and 59.1% (130/220) of the patients in the VOT group walked from their homes or places of work to the CHSC, and the median distance was 1.0 km. The median time for a one-way trip to the CHSC in the DOT group and VOT group was 10 minutes and 15 minutes, respectively, but the difference was not statistically significant (*P*=.25). If all patients complied with the expected number of visits to the CHSC and CCDC, the median total time spent traveling for their daily visits over 6 months would have been about 3600 minutes or 60 hours in the DOT group compared to 720 minutes or 12 hours in the VOT group. The time calculated from the actual median number of visits was much less in the DOT group, that is, 1240 minutes or 21 hours, because 55.1% (87/158) of the patients skipped more than half of the expected daily CHSC visits. In contrast, the median time calculated from the actual visits to the CHSC in the VOT group was 300 minutes or 5 hours (Table 4).

Table 4. Transportation time to the community health service center and the transportation costs incurred by the patients.

Transportation time and costs	DOT[a] group, median (IQR)	VOT[b] group, median (IQR)	*P* value
One-way distance (km)	1 (0.55, 2)	1 (1, 2)	.01
Time for one-way trip (min)	10 (5, 30)	15 (10, 20)	.25
Cumulative total time for the trips (min)	1240 (540, 3041)	300 (140, 720)	<.001
One-way cost of transportation (¥)[c]	2 (2, 5)	2 (1, 5)	.32
Cumulative total cost of transportation (¥)	276 (115, 552)	53 (25.5, 208)	<.001

[a]DOT: directly observed therapy.

[b]VOT: video-observed therapy.

[c]¥1=US $0.14

Transportation Costs

Of the patients reporting the details of their visits to the CHSC, 35.2% (44/125) in the DOT group and 40.9% (90/220) in the VOT group could not walk to the CHSC and were given funds for public transportation. The median transportation cost of a round trip visit to the CHSC, according to the patients who completed the questionnaire, was about ¥4 (US $0.57), resulting in an estimated total transportation cost of ¥96 (US $13.71) for the VOT patients compared with ¥720 (US $102.86) for the DOT patients. Although the actual costs were less, especially because most DOT patients self-administered many of their doses at home, the VOT patients still spent much less on transportation than the DOT patients (¥53 [US $7.57] vs ¥276 [US $39.43], respectively; *P*<.001) (Table 4).

Acceptance and Satisfaction

Satisfaction with the Management Methods

The patients in the DOT group were evenly divided about the ease of completing their drug regimen with half (49.6%, 65/131) regarding it as easy or very easy and half (50.4%, 66/131) regarding it as difficult or very difficult. In contrast, 93.2% (219/235) of the patients in the VOT group described it as easy or very easy, 6.8% (16/235) as difficult, and none described this method as very difficult. In the VOT group, 80.4% (189/235) thought that the use of the app could reduce the number of missed doses, while only 32.1% (42/131) in the DOT group thought that the daily visits would reduce the missed doses. Of the DOT patients, 40.5% (53/131) were satisfied or very satisfied, while 35.1% (46/131) were not satisfied or were very

dissatisfied with their treatment management, and 71.8% (94/131) referred a self-administered scheme to the daily visits. Of the patients on VOT, 81.3% (191/235) were satisfied or very satisfied, with only 3.4% (8/235) not satisfied, and 81.3% (191/235) indicated they would prefer this method of treatment (Table 5).

The majority of the patients in both the groups (64.9%, 85/131 in DOT and 56.6%, 133/235 in VOT) felt that there was no violation of their privacy, but a significant minority (25.2%, 33/131 in DOT and 37%, 87/235 in VOT) thought that there were occasional violations, and privacy violations were felt to be often or always by 9.9% (13/131) of the patients in the DOT group and 6.4% (15/235) of the patients in the VOT group (Table 5).

Of the 66 health care workers who completed the survey, 24% (16/66) were males with a mean age of 36.20 (SD 7.19) years; 30 were nurses, 24 were physicians, 7 were pharmacists, 2 were public health doctors; and there was a laboratory worker, a social worker, and a cashier. All had worked with both the DOT and VOT methods. Most health care workers (92%, 61/66) thought that the VOT method was convenient to manage patients with TB, as it could improve the management of the patients with TB (86%, 57/66) and they preferred the VOT method over the DOT method (85%, 56/66) (Table 5).

Table 5. Measurements of the satisfaction of the patients and the health care workers with the DOT[a] or VOT[b] methods.

Questions	DOT group (n=131), n (%)	VOT group (n=235), n (%)	*P* value	Health care workers (n=66), n (%)
What do you think of going to the CHSC[c] daily or of using an app to upload a video daily?			<.001	
Very easy	25 (19.1)	122 (51.9)		N/A[d]
Easy	40 (30.5)	97 (41.3)		N/A
Little difficult	52 (39.7)	16 (6.8)		N/A
Very difficult	14 (10.7)	0 (0)		N/A
What is the frequency of the perceived privacy violation going to the CHSC or taking and uploading a video?			.12	
Always	6 (4.6)	7 (3.0)		N/A
Often	7 (5.3)	8 (3.4)		N/A
Occasionally	33 (25.2)	87 (37.0)		N/A
No	85 (64.9)	133 (56.6)		N/A
Does your management method reduce the number of missed medications?			<.001	
Yes	42 (32.1)	189 (80.4)		N/A
No	89 (67.9)	46 (19.6)		N/A
Which management method do you prefer?			<.001	
Their current management method	37 (28.2)	191 (81.3)		N/A
Self-administered medication	94 (71.8)	44 (18.7)		N/A
Are you satisfied with the prescribed management method?			<.001	
Very satisfied	23 (17.6)	98 (41.7)		N/A
Fairly Satisfied	30 (22.9)	93 (39.6)		N/A
Just Satisfied	32 (24.4)	36 (15.3)		N/A
Not satisfied	35 (26.7)	8 (3.4)		N/A
Very dissatisfied	11 (8.4)	0 (0)		N/A
Was the VOT app convenient for managing patients with TB[e]?			N/A	
To a large extent	N/A	N/A		61 (92)
Rarely	N/A	N/A		4 (6)
No	N/A	N/A		1 (2)
Could the VOT app improve the standard management of patients with TB?			N/A	
To a large extent	N/A	N/A		57 (86)
Rarely	N/A	N/A		8 (12)
No	N/A	N/A		1 (2)
Which management method did the health care workers prefer?			N/A	
VOT	N/A	N/A		56 (85)
DOT	N/A	N/A		1 (2)
No special preference	N/A	N/A		9 (14)

[a]DOT: directly observed therapy.

[b]VOT: video-observed therapy.

[c]CHSC: community health service center.

[d]Not applicable.

[e]TB: tuberculosis.

Acceptance of the App

In the VOT group, 96.6% of the patients (227/235) could record the medication administration video by themselves, and the same proportion of patients thought that uploading the video was simple or very simple. Eight patients needed the help of their family to record and upload the videos; 88.9% (209/235) of the patients reported that problems using the app or in uploading the videos occurred less than 10% of the time (Table 6).

The health care workers reported that the duration of the videos was approximately 10-30 seconds each. The mean, median, maximum, and minimum number of patients that each health care worker managed were 3.6, 2, 1, and 19, respectively. The majority (89%, 59/66) of the health care workers were satisfied with the app interface, the time required to work with the app (85%, 56/66), and the quality of the video recordings (88%, 58/66). For the majority of the health care workers (96%, 63/66), problems with the app occurred less than 10% of the time.

Table 6. Acceptance of the patients on VOT[a] and health care workers with the use of the app.

Questions	Patients on VOT (n=235), n (%)	Health care workers (n=66), n (%)
Could you record the medicine administration video by yourself?		
Yes	227 (96.6)	N/A[b]
No	8 (3.4)	N/A
What is the degree of difficulty in using the app to upload the medicine administration video?		
Very simple	143 (60.9)	N/A
Simple	84 (35.7)	N/A
A little difficult	8 (3.4)	N/A
How frequently (%) did the app cause problems during the course of the treatment?		
0	57 (24.3)	18 (27)
<10	152 (64.7)	45 (68)
<50	21 (8.9)	2 (3)
≥50	5 (2.1)	1 (2)
Are you satisfied with the app interface?		
Very satisfied	153 (65.1)	32 (48)
Satisfied	80 (34.0)	27 (41)
Not satisfied	2 (0.9)	7 (11)
Are you satisfied with the time spent on the medicine administration videos?		
Very satisfied	156 (65.1)	37 (56)
Satisfied	77 (32.76)	19 (29)
Not satisfied	2 (0.9)	10 (15)
Are you satisfied with the quality of the video recording?		
Very satisfied	N/A	36 (55)
Satisfied	N/A	22 (33)
Not very satisfied	N/A	8 (12)

[a]VOT: video-observed therapy.

[b]Not applicable.

Discussion

Principal Findings

In this study, we compared the management of patients with TB with a comprehensive VOT management system and the management of patients on DOT by using retrospective data from the year preceding the implementation of VOT. In terms of the number of medication doses observed, missed doses, patient time required and transportation costs, ease of use, and acceptance by patients and medical staff, the VOT management system performed better than the DOT management system.

On the basis of the findings for smears and cultures, 46.2% (73/158) of the DOT patients and 51.9% (122/235) of the VOT patients had a bacteriologically confirmed diagnosis of TB, which is higher than the overall rate of bacteriologically confirmed diagnosis of TB cases of 37% for China [20]. However, the proportion of patients who showed positive smears was low: 20.9 % (33/158) of the patients in the DOT group and 17.4% (41/235) of the patients in the VOT group. Because

therapy was only followed on the basis of the monthly sputum smears, only this small proportion of patients could be followed up for the clinical results. In these patients with positive smears, there was no statistical difference (P=.17) in the clinical results, and all the patients showed negative smears after completing 6 months of treatment. However, 95% (38/40) of the patients on VOT had negative smears at 2 months and 100% (40/40) of the patients on VOT showed negative smears at 3 months, while 82% (27/33) of the patients on DOT showed negative smears at 2 months, 97% (32/33) showed negative smears at 3 months, and 1 patient on DOT showed a positive smear at 5 months.

Even in cities with local health posts that are close to the residence of most patients, daily DOT visits can interfere with the responsibilities at home and at work. Although DOT management requires patients come to the health post daily to be observed while taking their medications, many patients refused to come every day and they were given medications to take home, which they generally reported as faithfully self-administered. While we suspected that compliance with DOT might be a problem in China, and perhaps also in other places [21], we did not expect to find that only 1 out of 6 (16.4%) DOT patients had >85% of their doses observed. In contrast, 4 out of 5 (79.1%) of the VOT patients had >85% of their doses observed. Our study also found that the patient costs for transportation and patient time required for health post visits were considerably decreased with VOT, which likely contributed to the success and acceptance of VOT.

DOT can also be a burden on the health care system [22]. In the health care system of the Nanshan district, each health care worker had few patients with TB, but the health care workers had a heavy work burden because they also routinely cared for patients with 13 other conditions, including diabetes, hypertension, and mental illness, and they are involved in maternal and child health care. It was thus not surprising that they were appreciative of the management app that reduced patient visits, thereby saving them time. Similarly, in the typical settings of countries with a high TB burden, wherein overburdened health care workers are responsible for many patients on TB treatment, the VOT app should produce considerable savings in time and effort for the health care staff.

An additional advantage of the comprehensive TB management system is that it automatically monitors the number of patients on therapy, the medication inventories, the clinical results, the incidence of adverse events, and patient compliance in real time. Noncompliant patients or patients with side effects can be identified without delay and they can be contacted. The system thus avoids the need to collect records from various sources or databases so that both patient care and record keeping are more efficient and require less work by the health care system [22-26].

Limitations

This study had the following limitations. In this study, only patients who took FDC pills once a day were included and patients who took pills more than once a day were excluded. However, if patients need to take a medication more than once a day at home, it should be feasible to send 2 or 3 videos daily, and we plan to implement this possibility in the future. Another limitation is that this study included only patients with drug-sensitive TB, whereas compliance is even more critical for curing patients with drug-resistant TB, who require long treatment regimens with drugs that often have more adverse reactions. VOT would likely be especially useful for these patients [17] with whom DOT has been shown to improve compliance [5].

China is a country wherein VOT could be very successful, especially in large cities like Shenzhen, wherein smartphone penetration is very high and users are accustomed to working with a variety of apps. However, the penetration of mobile phones and facilities with their use may be less in the rural areas of China and in many countries with high TB burden and less resources and infrastructure, and their cost and inadequate internet data connections could be obstacles for VOT management. However, VOT could be the most valuable and cost-effective in these settings because it is in these settings that patients live far from the nearest health posts [27,28]. To function adequately in rural, low-infrastructure settings, the uploading protocol may need to be adapted to work with intermittent internet connectivity while still ensuring that each video has a time stamp and can only be sent once. This was a retrospective study that compared the results of DOT management for TB control in the Nanshan district with the results after the implementation of the comprehensive TB management system with VOT. If the entire study had been carried out in a carefully optimized prospective manner, perhaps the results of all the aspects would have been better than what we observed [29]. However, the results reported here represent what can routinely be expected from VOT implementation in a TB control program because the low proportion of doses that were actually observed with patients on DOT is likely commonplace in many DOT programs in China and perhaps elsewhere [15,30]. We believe that the significant increase in the observed doses with VOT could be a robust finding for programs adopting a comprehensive VOT management system such as that described by us [12,15,18,20].

Conclusions

After the implementation of a VOT-based comprehensive TB management system into the routine management of patients with TB in the Nanshan District of Shenzhen, the treatment compliance of patients was found to be significantly higher than that with the previous model using DOT. The VOT management system for TB required less patient time, low transportation costs, was easily adopted by the patients, and achieved high rates of satisfaction by both patients and health care workers. The findings of our study suggest that the adoption of the VOT system is feasible for TB control programs and will perform better than using the DOT strategy for the treatment of TB.

Acknowledgments

We thank the physicians, the supervisors who helped to collect the data, and all the 66 health care workers who participated in this study. We are grateful for the cooperation and invaluable feedback provided by each of the participating health departments. This study was funded by the National Science and Technology Major Project of China (2018ZX10715004-002), Medical Scientific Research Foundation of the Guangdong Province of China (C2018044), Natural Science Foundation of the Guangdong Province of China (2018A030313123), the Sanming Project of Medicine in Shenzhen (SZSM201603029), and the Scientific and Technology Research of the Bureau of Science and Technology Innovation of Nanshan (2018054, 2018061, 2018076).

Authors' Contributions

SL designed the study. SL and JW provided administrative, technical, and material support. YY, TZ, and YF implemented the study. XG and YY performed the data analysis. JM and MZ were responsible for the quality control and participated in the data interpretation. XG and YY drafted the manuscript, which was critically revised and edited by HET. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. Global Tuberculosis Report 2019. World Health Organization. 2019. URL: https://www.who.int/tb/publications/global_report/en/ [accessed 2019-10-17]
2. Zhu S, Xia L, Yu S, Chen S, Zhang J. The burden and challenges of tuberculosis in China: findings from the Global Burden of Disease Study 2015. Sci Rep 2017 Nov 06;7(1):14601 [FREE Full text] [doi: 10.1038/s41598-017-15024-1] [Medline: 29097809]
3. Glaziou P, Floyd K, Raviglione MC. Global Epidemiology of Tuberculosis. Semin Respir Crit Care Med 2018 Jun;39(3):271-285. [doi: 10.1055/s-0038-1651492] [Medline: 30071543]
4. National Health and Family Planning Commission of China. Diagnosis for pulmonary tuberculosis. http://www.nhc.gov.cn/fzs/s7852d/201711/0819ad84540b4d97a1644bbc6ec4306d.shtml. 2017 Nov 09. URL: http://www.nhc.gov.cn/fzs/s7852d/201711/0819ad84540b4d97a1644bbc6ec4306d.shtml [accessed 2017-11-09]
5. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in New York City--turning the tide. N Engl J Med 1995 Jul 27;333(4):229-233. [doi: 10.1056/NEJM199507273330406] [Medline: 7791840]
6. Weis SE, Slocum PC, Blais FX, King B, Nunn M, Matney GB, et al. The effect of directly observed therapy on the rates of drug resistance and relapse in tuberculosis. N Engl J Med 1994 Apr 28;330(17):1179-1184. [doi: 10.1056/NEJM199404283301702] [Medline: 8139628]
7. Tian J, Lu Z, Bachmann MO, Song F. Effectiveness of directly observed treatment of tuberculosis: a systematic review of controlled studies. Int J Tuberc Lung Dis 2014 Sep;18(9):1092-1098. [doi: 10.5588/ijtld.13.0867] [Medline: 25189558]
8. Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2015 May 29(5):CD003343 [FREE Full text] [doi: 10.1002/14651858.CD003343.pub4] [Medline: 26022367]
9. Sun Q, Meng Q, Yip W, Yin X, Li H. DOT in rural China: experience from a case study in Shandong Province, China. Int J Tuberc Lung Dis 2008 Jun;12(6):625-630. [Medline: 18492328]
10. Zhang H, Ehiri J, Yang H, Tang S, Li Y. Impact of Community-Based DOT on Tuberculosis Treatment Outcomes: A Systematic Review and Meta-Analysis. PLoS One 2016;11(2):e0147744 [FREE Full text] [doi: 10.1371/journal.pone.0147744] [Medline: 26849656]
11. Ingram D. Video directly observed therapy: Enhancing care for patients with active tuberculosis. Nursing 2018 May;48(5):64-66. [doi: 10.1097/01.NURSE.0000531912.91585.85] [Medline: 29697568]
12. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
13. Macaraig M, Lobato MN, McGinnis PK, Wegener D. A National Survey on the Use of Electronic Directly Observed Therapy for Treatment of Tuberculosis. J Public Health Manag Pract 2017 Jul 07;24(6):567-570. [doi: 10.1097/PHH.0000000000000627] [Medline: 28692611]
14. Ngwatu BK, Nsengiyumva NP, Oxlade O, Mappin-Kasirer B, Nguyen NL, Jaramillo E, Collaborative group on the impact of digital technologies on TB. The impact of digital health technologies on tuberculosis treatment: a systematic review. Eur Respir J 2018 Jan 11;51(1):1701596 [FREE Full text] [doi: 10.1183/13993003.01596-2017] [Medline: 29326332]
15. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]

16. Buchman T, Cabello C. A New Method to Directly Observe Tuberculosis Treatment: Skype Observed Therapy, a Patient-Centered Approach. J Public Health Manag Pract 2017;23(2):175-177. [doi: 10.1097/PHH.0000000000000339] [Medline: 27598709]
17. Chuck C, Robinson E, Macaraig M, Alexander M, Burzynski J. Enhancing management of tuberculosis treatment with video directly observed therapy in New York City. Int J Tuberc Lung Dis 2016 May;20(5):588-593. [doi: 10.5588/ijtld.15.0738] [Medline: 27084810]
18. Garfein RS, Liu L, Cuevas-Mota J, Collins K, Muñoz F, Catanzaro DG, et al. Tuberculosis Treatment Monitoring by Video Directly Observed Therapy in 5 Health Districts, California, USA. Emerg Infect Dis 2018 Dec;24(10):1806-1815. [doi: 10.3201/eid2410.180459] [Medline: 30226154]
19. Bureau of Development and Reform of Nanshan District of ShenZhen. Report on the implementation of the national economic and social development plan in 2018 and the draft plan arrangement in 2019 in Nanshan District, Shenzhen. http://www.szns.gov.cn/nsqfgj/gkmlpt/content/6/6402/post_6402756.html#14928. 2019 May 13. URL: http://www.szns.gov.cn/nsqfgj/attachment/0/33/33237/6402756.pdf [accessed 2019-05-13]
20. World Health Organization. WHO Tuberculosis profile: China 2018. https://worldhealthorg.shinyapps.io/tb_profiles. URL: https://worldhealthorg.shinyapps.io/tb_profiles [accessed 2020-05-26]
21. Ai X, Men K, Guo L, Zhang T, Zhao Y, Sun X, et al. Factors associated with low cure rate of tuberculosis in remote poor areas of Shaanxi Province, China: a case control study. BMC Public Health 2010 Mar 07;10:112 [FREE Full text] [doi: 10.1186/1471-2458-10-112] [Medline: 20205941]
22. Lam CK, Fluegge K, Macaraig M, Burzynski J. Cost savings associated with video directly observed therapy for treatment of tuberculosis. Int J Tuberc Lung Dis 2019 Nov 01;23(11):1149-1154. [doi: 10.5588/ijtld.18.0625] [Medline: 31718750]
23. Au-Yeung KY, DiCarlo L. Cost comparison of wirelessly vs. directly observed therapy for adherence confirmation in anti-tuberculosis treatment. Int J Tuberc Lung Dis 2012 Nov;16(11):1498-1504. [doi: 10.5588/ijtld.11.0868] [Medline: 23006834]
24. Holzman SB, Zenilman A, Shah M. Advancing Patient-Centered Care in Tuberculosis Management: A Mixed-Methods Appraisal of Video Directly Observed Therapy. Open Forum Infect Dis 2018 Apr;5(4):ofy046 [FREE Full text] [doi: 10.1093/ofid/ofy046] [Medline: 29732378]
25. Do D, Garfein RS, Cuevas-Mota J, Collins K, Liu L. Change in Patient Comfort Using Mobile Phones Following the Use of an App to Monitor Tuberculosis Treatment Adherence: Longitudinal Study. JMIR Mhealth Uhealth 2019 Feb 01;7(2):e11638 [FREE Full text] [doi: 10.2196/11638] [Medline: 30707103]
26. Wade VA, Karnon J, Eliott JA, Hiller JE. Home videophones improve direct observation in tuberculosis treatment: a mixed methods evaluation. PLoS One 2012;7(11):e50155 [FREE Full text] [doi: 10.1371/journal.pone.0050155] [Medline: 23226243]
27. Krueger K, Ruby D, Cooley P, Montoya B, Exarchos A, Djojonegoro BM, et al. Videophone utilization as an alternative to directly observed therapy for tuberculosis. Int J Tuberc Lung Dis 2010 Jun;14(6):779-781. [Medline: 20487619]
28. Mirsaeidi M, Farshidpour M, Banks-Tripp D, Hashmi S, Kujoth C, Schraufnagel D. Video directly observed therapy for treatment of tuberculosis is patient-oriented and cost-effective. Eur Respir J 2015 Sep;46(3):871-874 [FREE Full text] [doi: 10.1183/09031936.00011015] [Medline: 25792632]
29. Guo P, Qiao W, Sun Y, Liu F, Wang C. Telemedicine Technologies and Tuberculosis Management: A Randomized Controlled Trial. Telemed J E Health 2019 Dec 02. [doi: 10.1089/tmj.2019.0190] [Medline: 31794684]
30. Lam CK, McGinnis Pilote K, Haque A, Burzynski J, Chuck C, Macaraig M. Using Video Technology to Increase Treatment Completion for Patients With Latent Tuberculosis Infection on 3-Month Isoniazid and Rifapentine: An Implementation Study. J Med Internet Res 2018 Nov 20;20(11):e287 [FREE Full text] [doi: 10.2196/jmir.9825] [Medline: 30459146]

Abbreviations

CCDC: Center for Chronic Disease Control
CHSC: community health service center
DOT: directly observed therapy
FDC: fixed-dose combination
TB: tuberculosis
VOT: video-observed therapy

Edited by G Eysenbach; submitted 01.01.20; peer-reviewed by F Zeng, C Lam; comments to author 10.03.20; revised version received 23.04.20; accepted 03.06.20; published 31.07.20.

Please cite as:
Guo X, Yang Y, Takiff HE, Zhu M, Ma J, Zhong T, Fan Y, Wang J, Liu S
A Comprehensive App That Improves Tuberculosis Treatment Management Through Video-Observed Therapy: Usability Study
JMIR Mhealth Uhealth 2020;8(7):e17658
URL: https://mhealth.jmir.org/2020/7/e17658
doi:10.2196/17658
PMID:32735222

Original Paper

Development and Preliminary Evaluation of the Effects of an mHealth Web-Based Platform (HappyAir) on Adherence to a Maintenance Program After Pulmonary Rehabilitation in Patients With Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial

Begoña Jiménez-Reguera[1], MSc; Eva Maroto López[2], MSc; Shane Fitch[2], MSc; Lourdes Juarros[3], MD; Marta Sánchez Cortés[3], PT, MSc; Juan Luis Rodríguez Hermosa[4], MD; Myriam Calle Rubio[4], MD; María Teresa Hernández Criado[5], PT, MSc; Marta López[5], MD; Santiago Angulo-Díaz-Parreño[1], MSc; Aitor Martín-Pintado-Zugasti[1], PT, PhD; Jordi Vilaró[6], PT, PhD

[1]Departamento de Fisioterapia, Facultad de Medicina, Universidad San Pablo-CEU, CEU Universities, Madrid, Spain

[2]Lovexair Foundation, Madrid, Spain

[3]Hospital Universitario 12 de octubre, Madrid, Spain

[4]Hospital Clínico San Carlos, Madrid, Spain

[5]Hospital Universitario de La Princesa, Madrid, Spain

[6]Ramon Llull University, Barcelona, Spain

Corresponding Author:
Jordi Vilaró, PT, PhD
Ramon Llull University
Padilla 326
Barcelona
Spain
Phone: 34 932 533 256
Email: jordi.gestos@gmail.com

Abstract

Background: Pulmonary rehabilitation is one of the main interventions to reduce the use of health resources, and it promotes a reduction in chronic obstructive pulmonary disease (COPD) costs. mHealth systems in COPD aim to improve adherence to maintenance programs after pulmonary rehabilitation by promoting the change in attitude and behavior necessary for patient involvement in the management of the disease.

Objective: This study aimed to assess the effects of an integrated care plan based on an mHealth web-based platform (HappyAir) on adherence to a 1-year maintenance program applied after pulmonary rehabilitation in COPD patients.

Methods: COPD patients from three hospitals were randomized to a control group or an intervention group (HappyAir group). Patients from both groups received an 8-week program of pulmonary rehabilitation and educational sessions about their illness. After completion of the process, only the HappyAir group completed an integrated care plan for 10 months, supervised by an mHealth system and therapeutic educator. The control group only underwent the scheduled check-ups. Adherence to the program was rated using a respiratory physiotherapy adherence self-report (CAP FISIO) questionnaire. Other variables analyzed were adherence to physical activity (Morisky-Green Test), quality of life (Chronic Obstructive Pulmonary Disease Assessment Test, St. George's Respiratory Questionnaire, and EuroQOL-5D), exercise capacity (6-Minute Walk Test), and lung function.

Results: In total, 44 patients were recruited and randomized in the control group (n=24) and HappyAir group (n=20). Eight patients dropped out for various reasons. The CAP FISIO questionnaire results showed an improvement in adherence during follow-up period for the HappyAir group, which was statistically different compared with the control group at 12 months (56.1 [SD 4.0] vs 44.0 [SD 13.6]; *P*=.004) after pulmonary rehabilitation.

Conclusions: mHealth systems designed for COPD patients improve adherence to maintenance programs as long as they are accompanied by disease awareness and patient involvement in management.

Trial Registration: ClinicalTrials.gov NCT04479930; https://clinicaltrials.gov/ct2/show/NCT04479930

(JMIR Mhealth Uhealth 2020;8(7):e18465) doi:10.2196/18465

KEYWORDS

adherence; pulmonary rehabilitation; mHealth; COPD; chronic obstructive pulmonary disease

Introduction

Pulmonary Rehabilitation in Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease (COPD) is a chronic, preventable, and treatable disease process characterized by airflow limitation that is not fully reversible [1]. It is a major public health problem because it represents a high health cost due to the direct and indirect expenses it generates, including the significant consumption of resources and medical and pharmaceutical services, as well as the demand for support and social assistance arising from sickness from work [2-4].

The benefits of pulmonary rehabilitation are such that it has been compared favorably with other strategies, such as drug treatment or telemedicine, in terms of its cost-effectiveness. Thus, it is well known that it can reduce the use of health resources and promotes a reduction in COPD costs [5-7].

Pulmonary rehabilitation is defined as "a comprehensive intervention based on a thorough patient assessment followed by patient tailored therapies that include, but are not limited to, exercise training, education, and behavior change designed to improve the physical and psychological condition of people with chronic respiratory disease and to promote the long-term adherence to health-enhancing behaviors" [8].

However, despite evidence of the benefits provided by pulmonary rehabilitation, these do not last over time, disappearing progressively between 6 and 12 months after the end of rehabilitation, with patients having values even lower than those presented pre–pulmonary rehabilitation [9]. The lack of adherence to maintenance programs seems to be one of the possible causes explaining the loss of health benefits [10,11], so the creation of effective strategies to increase adherence to such programs is the key to maintaining the effects achieved after pulmonary rehabilitation. Aspects such as self-management, patient empowerment, and the acquisition of co-responsibility in therapy are the focus of current research, as they seem to guide determinant behavioral changes to maintain disease control.

Telehealthcare as a Solution

Mobile health (mHealth) systems in COPD, designed according to the needs of the patients, aim to improve adherence to maintenance programs by promoting the change in attitude and behavior necessary for patient involvement in the management of the disease [12,13].

Both patients and professionals recognize the importance of designing individualized mHealth interventions that encompass the different aspects associated with the disease and facilitate self-control through appropriate feedback for each dimension, without replacing or dominating the patient's decision. Health informatics platforms should not replace, at any time, the personal and regular relationship with health care professionals but complement it, as it is very important to maintain continuous and open contact with the multidisciplinary team that provides the necessary support and attention. But, in any case, patients should be responsible for their care, which may lead to an improvement of the therapy's effectiveness [12].

The follow-up programs designed and analyzed so far are not effective, as they do not achieve patient adherence to them, showing a high dropout rate. There is also great controversy regarding the methods used, types of mHealth systems, specific design of maintenance programs, and duration and frequency of follow-up. Most of the studies consulted highlight the influence of the biopsychosocial context of the individual on the involvement with their illness and treatment [14-16].

As stated above, and due to the improvement in adherence to the treatments that digital platforms promote, this study aimed to evaluate whether an mHealth web-based platform (HappyAir) would improve adherence to a 1-year maintenance program applied after pulmonary rehabilitation in COPD patients.

Methods

Study Design and Clinical Trial Protocol

The initial objective of the study was the development of a clinical tool, the HappyAir system, which included the web-based platform and mobile app to allow its use by COPD patients for the management of their long-term pathology. As previous studies have found a possible reluctance in many COPD patients toward new technologies [17], it was decided to generate a new web-based app with an intuitive design and easy operation in order to avoid this possible reticence. The HappyAir system was then integrated into the long-term follow-up program of the intervention group. The completion of the protocol at different stages of the preliminary evaluation study, some not initially planned, meant that the clinical trial was not registered in due course.

A multicenter, longitudinal, prospective, randomized controlled clinical trial was conducted on 44 COPD patients who underwent an integrated care plan monitored with mHealth. The recruitment and follow-up were carried out between December 2015 and May 2017. Patients, who were recruited from participating hospitals, underwent a randomization process, establishing two groups: the intervention group (HappyAir) and control group (see later randomization procedure section).

The study was conducted in two stages, with a total duration of 12 months. The first stage corresponded to the 8-week pulmonary rehabilitation program that was conducted on both groups at the hospital. It included several procedures that followed the guidelines of the Spanish Society of Pulmonology

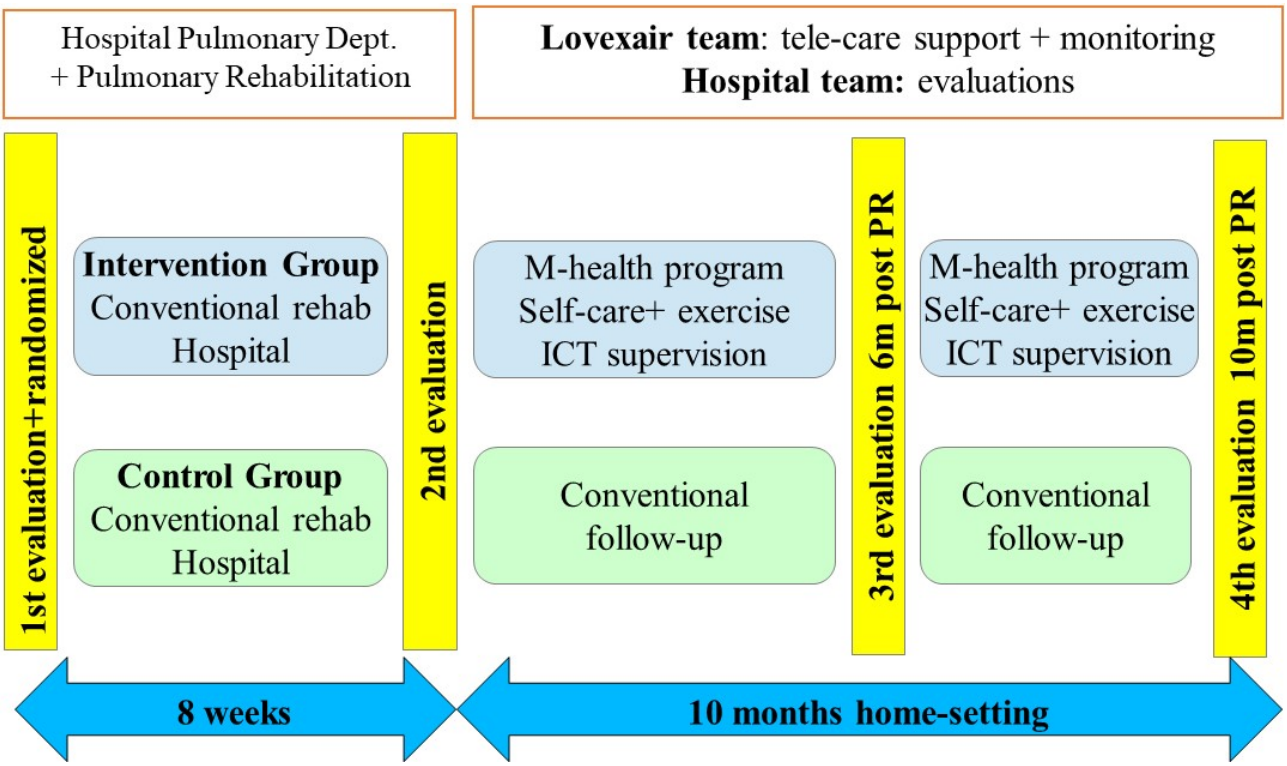
and Thoracic Surgery and are included in hospital protocols, such as muscle training, respiratory physiotherapy, and education on relevant aspects of chronic respiratory disease [18].

The second stage corresponded to a 10-month follow-up period. Patients from both groups underwent a maintenance follow-up community-based program at home and in the neighborhood, in which they were advised to perform physical activity and breathing exercises daily. Patients assigned to the HappyAir group followed an integrated care plan using a mobile device with the pulmonary care web-based app (HappyAir app) and were instructed in its use (Multimedia Appendix 1). The control group only went to the hospital for the scheduled evaluations during the follow-up period, without receiving integrated supervision or using the HappyAir app.

In order to ensure the correct follow-up of the two populations studied, four evaluations were completed by a blinded assessor: at baseline (pre–pulmonary rehabilitation), immediately after pulmonary rehabilitation (post–pulmonary rehabilitation), after 6 months of follow-up, and after 12 months from the beginning of the study (10-month postrehabilitation follow-up; Figure 1).

The study protocol was approved by the ethics committees of 12 de Octubre University Hospital (No.15/308), La Princesa University Hospital, and San Carlos Clinical University Hospital (16/111-E). All patients gave written informed consent to participate in the study.

Figure 1. Intervention and control group process and follow-up.



Study Population and Recruitment

Patients were recruited by convenience sampling through face-to-face interviews at participating hospitals. The recruitment of subjects was performed from patients attending pneumology consultations at the rehabilitation service of the hospitals participating in the study. The participants belonged to the geographical area of Madrid.

COPD patients were selected according to the following inclusion criteria: COPD patient, aged between 55 and 85 years, with degree of severity II, III, or IV of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) scale, in a stable clinical situation (no exacerbations in the last 6 weeks).

Exclusion criteria were patient with unstable cardiovascular disease or muscular, osteoarticular, auditory, visual, or central or peripheral nervous system impairment that prevented the performance of the rehabilitation program or evaluation tests or cognitive impairment that made it difficult to understand the education program and manage the HappyAir system.

After participants were informed about the aim and characteristics of the investigation, they were asked to carefully read and sign an informed consent to be able to participate in the study.

Pulmonary Care Web-Based App (HappyAir App)

The HappyAir app comprises two main parts: an educational program providing patients useful information and advice about their illness and data collection related to physical activity and disease. The HappyAir app reminded the HappyAir group daily to use the app, indicating that they record medication intake, daily exercise time (minutes), level of tiredness after the exercises (good, little tired, very tired, or exhausted), and daily mood (happy, little sad, sad, or very sad). These records allowed for usability evaluation of the HappyAir app. During development, a focus group of patients, health care professionals, and app developers met to detect usability

problems and adapt the app to the characteristics of the target group.

After patients finished pulmonary rehabilitation, we offered an education session that consisted of 3 to 4 hours of practical class demonstration plus an online support aid; the handling of the device and the web-based app were explained so the patients could become familiar with them.

The therapeutic educators had access to the platform to supervise the evolution of the HappyAir group during the follow-up period, collect data, see the results of clinical evaluations, record weekly and monthly goals, and contact the physician responsible for the patients (in case they detected warning signs of possible exacerbations or relapses). In addition, the pulmonologist and physiotherapist of the hospital had access, by login, to the data collection platform to enter the clinical data, record results of the evaluations, see the evolution of patients, and contact the therapeutic educator.

The HappyAir integrated plan was designed as a model of a therapeutic program based on communication that introduced the figure of the therapeutic educator (physiotherapist or respiratory coach) in order to design interventions focused on the patients and their needs, with minimal intervention and presence, making the patients responsible for their self-care and management of their illness. Patient and educator shared responsibility.

Outcome Measures

Treatment Adherence

Adherence to the maintenance program was measured with the respiratory physiotherapy adherence self-report questionnaire (CAP FISIO) questionnaire administered by a blinded assessor at each follow-up session in the hospital. This is a questionnaire created to assess the adherence to physiotherapy treatments [19]. We adapted some terms of the questionnaire to the purpose of this study, because to our knowledge, no questionnaire existed that measured adherence and perception in pulmonary rehabilitation for chronic respiratory disease. It consists of a total of 16 items, with a Likert scale to score each one, with 1 point as totally disagree and 4 points as totally agree. Three different dimensions of results are obtained: total score, perception, and adherence. The final rating scale is set to a range of minimum of 16 points to a maximum of 64 points. A higher final score reflects better adherence to the intervention. The internal consistency was set with a Cronbach alpha [19].

Adherence to physical activity was measured with the therapeutic compliance questionnaire (Morisky-Green Test), administered by a blinded assessor at each follow-up session in the hospitals. This method, which has been validated for various chronic diseases, was originally developed by Morisky et al [20] to assess medication compliance in patients with high blood pressure [21]. Since the test was introduced, it has been used in the evaluation of different diseases. It consists of a series of 4 contrast questions with a yes/no dichotomous answer, reflecting the patient's behavior with respect to compliance. In order to consider the patients compliant with or adherent to treatment, the first and last two answers must be no and the second answer must be yes. Answering at least one of the questions incorrectly indicates poor adherence. Given the scarcity of published questionnaires for the evaluation of adherence to physical exercise of patients with chronic respiratory disease, it has been decided to adapt an already validated test (as described earlier, Morisky-Green Test) to verify adherence in our study. To do this, the word medication was change to physical exercise.

Quality of Life

Quality of life was measured with three different self-reported questionnaires administered by a blinded assessor at each follow-up session in the hospitals.

Chronic Obstructive Pulmonary Disease Assessment Test

The Chronic Obstructive Pulmonary Disease Assessment Test (CAT) is a self-administered questionnaire that measures the impact of COPD on the patient's quality of life. It consists of 8 items, each of which has 5 possible answers from 1 (being absence of symptoms) to 5 (being the worst possible situation). The final score scale is set in a range of 8 to 40 points. A difference of 2 points or more would represent a clinically significant difference in pre- and posttreatment health status. The difference between stable status and exacerbation is a 5-point increase on the 40-point scale. Higher scores mean greater deterioration in COPD-related quality of life [22].

St. George's Respiratory Questionnaire

The St. George's Respiratory Questionnaire (SGRQ) is a validated questionnaire that measures quality of life related to health or perceived health in COPD patients. It consists of 50 items divided into 3 dimensions: symptoms (frequency and severity of symptoms; 8 items), activity (limitation of activity due to dyspnea; 16 items), and impact on daily life (psychological and social functioning disorders; 26 items).

The final score scale is set in a range from 0 (no limitation of the quality of life) to 100 (maximum limitation of the quality of life). A difference of 4 points is considered a clinically significant difference [23,24].

EuroQOL-5D

The EuroQOL-5D is a questionnaire that measures quality of life related to health or perceived health, which, unlike the SGRQ, is simpler to administer. Therefore, it was also decided to use it in case there was not much acceptance of the SGRQ, which, despite being very complete and extended in its application, is a very long questionnaire.

The EuroQOL-5D consists of 5 items divided into 5 dimensions: mobility, self-care, habitual activities, pain or discomfort, and anxiety or depression. It has a minimum number of levels (3) for each dimension (1=no problem, 2=some problem or moderate, and 3=many problems). Preferably, the questionnaire should be self-administered, although administration by personal interview or by mail has been shown to be acceptable. It generates an index that allows the evaluation of health conditions. The state of health of the individual is defined as the combination of the level of problems described in each of the 5 dimensions, using a 5-digit number that reflects the value of each dimension [25,26].

Exercise Capacity

Exercise capacity was measured with the 6-Minute Walk Test (6MWT). This is a simple test in which the subject must walk in a circuit straight and without irregularities, at least 30 meters, for a period of 6 minutes, with the aim to reach the maximum possible distance (walking as quickly as possible, without running). This test was performed following the protocol established by the American Thoracic Society in 2002, according to the 2014 update [27].

Randomization Procedure

We used a computer-generated simple randomization procedure, using the online randomization tool Research Randomizer (Geoffrey C Urbaniak and Scott Plous). Before the beginning of the study, distribution was made in two groups through the Research Randomizer program, and a list of patients designated to each group was drawn up, considering a homogeneous distribution of groups for each hospital. This listing was sequentially numbered and coded to ensure the confidentiality of participants and masking of the professionals who performed the rehabilitation protocol.

Patient data collected during the study were documented anonymously and dissociated and linked to an ID code (patient identification number) so that only the hospital investigator could associate such data to an identified person. The principal researcher of the study, external to the hospital, established the relationship between the ID provided by hospital staff with the code assigned in the tracking platform.

The follow-up assessment of outcome measures of both groups was carried out by a blinded assessor. Due to the characteristics of the intervention, health care professionals and patients could not be blinded to the group assignment.

Sample Size

For the calculation of the sample size to test the difference between the treated and control groups in the total score variable after 12 months of follow-up (the main objective of the study), a pilot study was conducted with 7 individuals in each group, with the following results: in the control group, the mean was 44.2 (SD 12.8), while in the HappyAir group, the mean was 55.6 (SD 6.2). With these results and using the G*Power version 3.1.9.4 program (Heinrich-Heine-Universität Düsseldorf) with a 2-sided test of an alpha level of .05 and a power of 80%, it was determined that 14 individuals would be needed in each group. Assuming the probability of dropouts of 20% because of the long-term intervention, we decided to recruit the maximum number of patients available.

Statistical Analysis

Statistical analyses were performed using SPSS Statistics 25.0 statistical software for Windows (IBM Corporation). Thus, a descriptive study of absolute and relative frequencies and distributions for each of the qualitative variables was completed. The normal distribution of the quantitative variables was demonstrated using the Kolmogorov-Smirnov normality test when the number of data exceeded 50 and the Shapiro-Wilk test when the number of data were fewer than 50.

Subsequently, the existence of statistically significant differences over time or in the different measurements of the quantitative variables was analyzed, and for this purpose, the *t* test was used for repeated measurements and the Wilcoxon test in the nonparametric or qualitative case.

The statistical significance of the intergroup and intragroup comparisons at all levels of segmentation was analyzed using parametric tests (*t* tests [for 2 samples], analyses of variance, and Welch tests) and nonparametric tests (Mann-Whitney *U* and Kruskal-Wallis tests), according to the distribution of the sample.

Results

Patient Characteristics

A total of 44 patients diagnosed with COPD were included in the study, randomly assigned to the control group or HappyAir group; among them, 8 patients dropped out of the study for different reasons. In the end, 36 patients completed the 12-month follow-up process and were included in the final analysis (22 men and 14 women; median age 68.11 [SD 6.74] years). A flowchart of participants in the study is shown in Figure 2. Demographic and baseline characteristics are shown in Table 1. None of the variables showed differences between the groups at baseline.

Figure 2. Distribution of patients Consolidated Standards of Reporting Trials flow diagram.

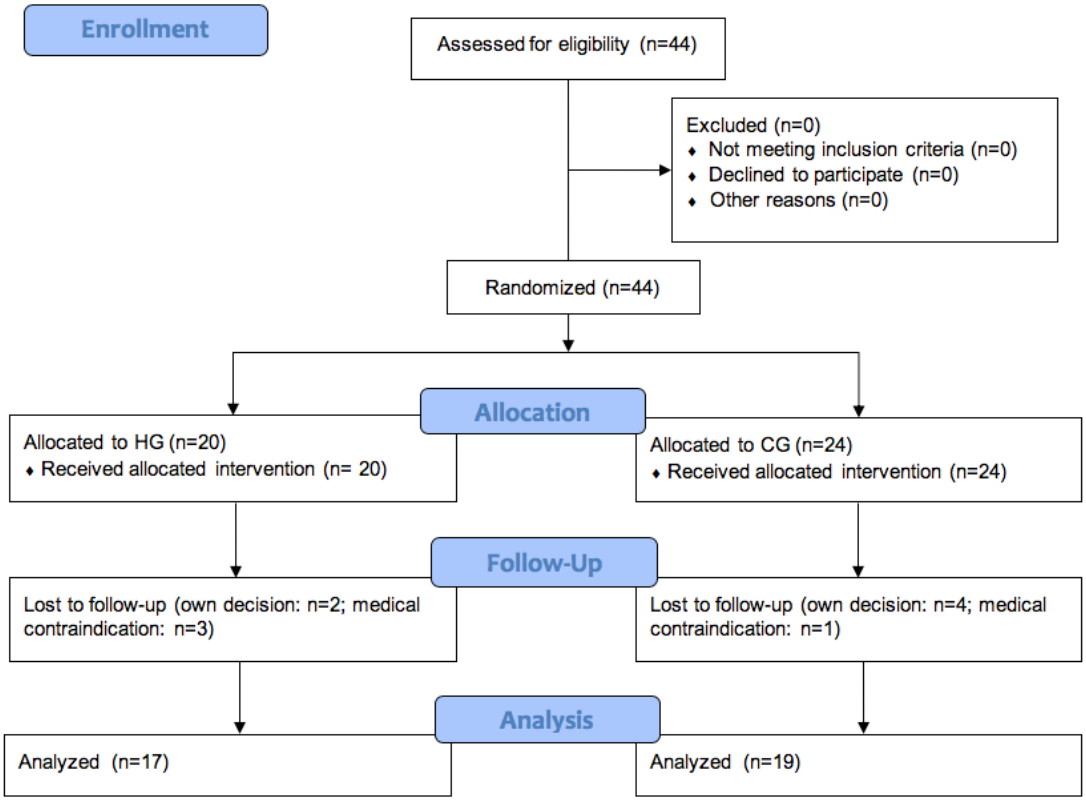


Table 1. Baseline and demographic characteristics of the study population.

Characteristic	Control group	Intervention group	*P* value
Age in years, mean (SD)	68.1 (7.0)	68.1 (6.6)	.97
Gender			.34
Male, n (%)	13 (59.1)	9 (40.9)	—[a]
Female, n (%)	6 (42.9)	8 (57.1)	—
Weight (kg), mean (SD)	68.4 (15.3)	70.1 (10.1)	.68
Height (cm), mean (SD)	162.9 (9.8)	161.7 (6.7)	.91
BMI (kg/m^2), mean (SD)	26.07 (4.2)	26.50 (4.1)	.77
GOLD[b] classification			.28
Level II	2 (1M/1F)	3 (1M/2F)	—
Level III	14 (10M/4F)	9 (4M/5F)	—
Level IV	3 (2M/1F)	5 (4M/1F)	—
FEV_1[c] (%), mean (SD)	43.1 (13.6)	45.0 (15.3)	.95
FVC[d] (%), mean (SD)	72.6 (24.4)	78.6 (22.9)	.44
FEV_1/FVC (%), mean (SD)	44.5 (12.2)	49.06 (12.0)	.81
Oxygen users, n (%)	10 (52.6)	9 (52.9)	.52
Oxygen hours per day, mean (SD)	10.8 (10.8)	10.0 (11.2)	.98
Smokers, n (%)	5 (26.3)	3 (17.6)	.43
Exsmokers, n (%)	11 (57.9)	11 (64.7)	.77

[a]: not applicable.

[b]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

[c]FEV_1: forced expiratory volume in the first second of expiration.

[d]FVC: forced vital capacity.

Adherence

The results of adherence to the maintenance program measured with the CAP questionnaire and adherence to physical activity using the Morisky-Green questionnaire are detailed in Table 2.

Cuestionario Adherencia Percepción Questionnaire

Total CAP dimension results showed an improvement in adherence during the follow-up period, which was statistically significant at 12 months after pulmonary rehabilitation. Differences observed at 6 months were 45.3 (SD 15.0) and 53.6 (SD 5.4) in the control group and HappyAir group, respectively. Differences at 12 months were 44 (SD 13.6) and 56.1 (SD 4.0) in the control group and HappyAir group, respectively. The results showed significant differences between the groups at 12 months (*P*=.004; Figure 3).

Morisky-Green Questionnaire

In the HappyAir group, 25% of patients were adherent at 12 months of follow-up versus 11% of patients in the control group. The intergroup factor analysis showed statistically significant differences in adherence to physical activity at 12-month follow-up (*P*=.049). The intragroup factor analysis of exercise adherence showed no statistically significant differences over time (Figure 4).

Table 2. Adherence and perception effects of the home rehabilitation program.

Variable	Control group (n=19), mean (SD)		HappyAir group (n=17), mean (SD)		Intragroup *P* value (95% CI), 12 m[a] vs 6 m		Intergroup *P* value	
	6 m	12 m	6 m	12 m	Control	HappyAir	6 m	12 m
CAP[b]								
Total	45.3 (15)	44 (13.6)	53.6 (5.4)	56.1 (4)	.69 (–5.3 to 7.9)	.05 (–5.01 to 0.075)	.16	.004
Adherence	21.7 (6.2)	21.1 (5.5)	24.7 (2.9)	25.7 (1.8)	.64 (–2.2 to 3.4)	.19 (–2.5 to 0.57)	—[c]	—
Perception	23.5 (9)	22.9 (8.1)	28.9 (4.2)	30.4 (2.5)	.74 (–3.3 to 4.5)	.14 (–3.4 to 0.54)	—	—
Morisky Green								
Adherence PA[c] (%)	35.70	30.80	64.30	69.20	—	—	.11	.049

[a]m: month.

[b]CAP: Cuestionario Adherencia Percepción

[c]Not applicable.

[d]PA: physical activity.

Figure 3. Adherence to HappyAir program: total dimension of CAP questionnaire.

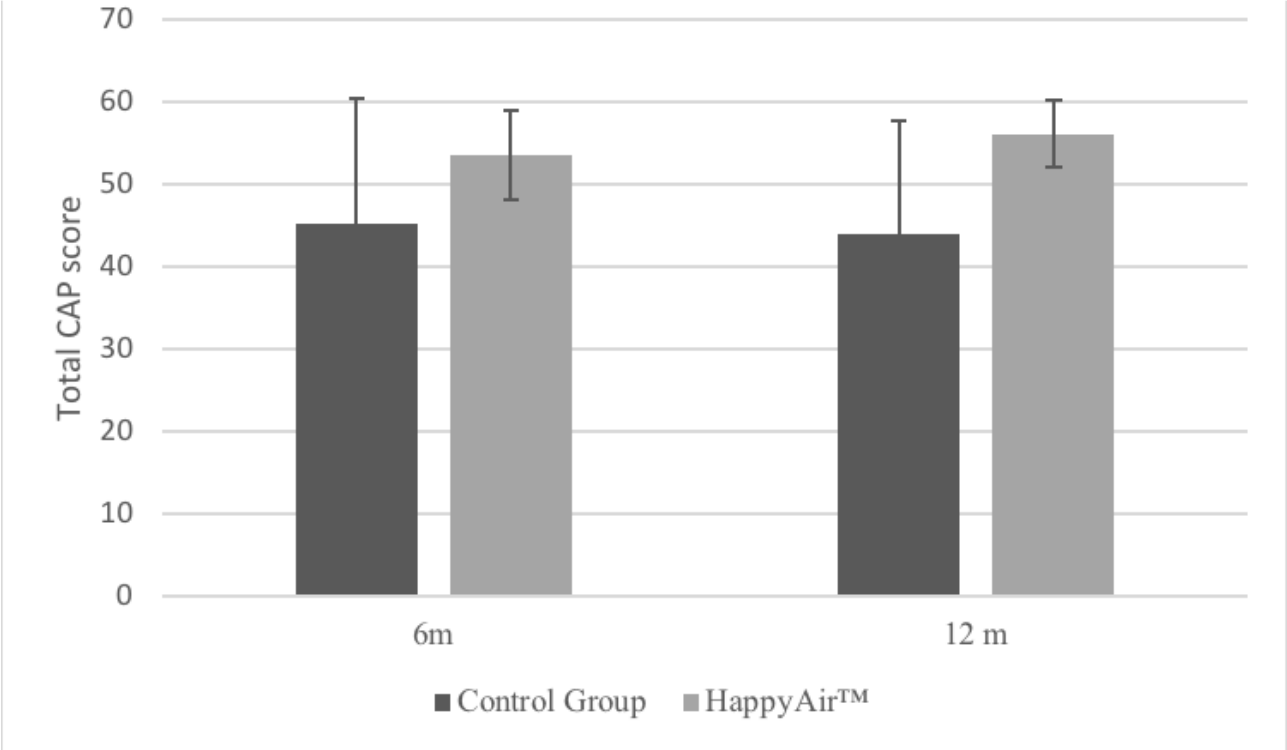
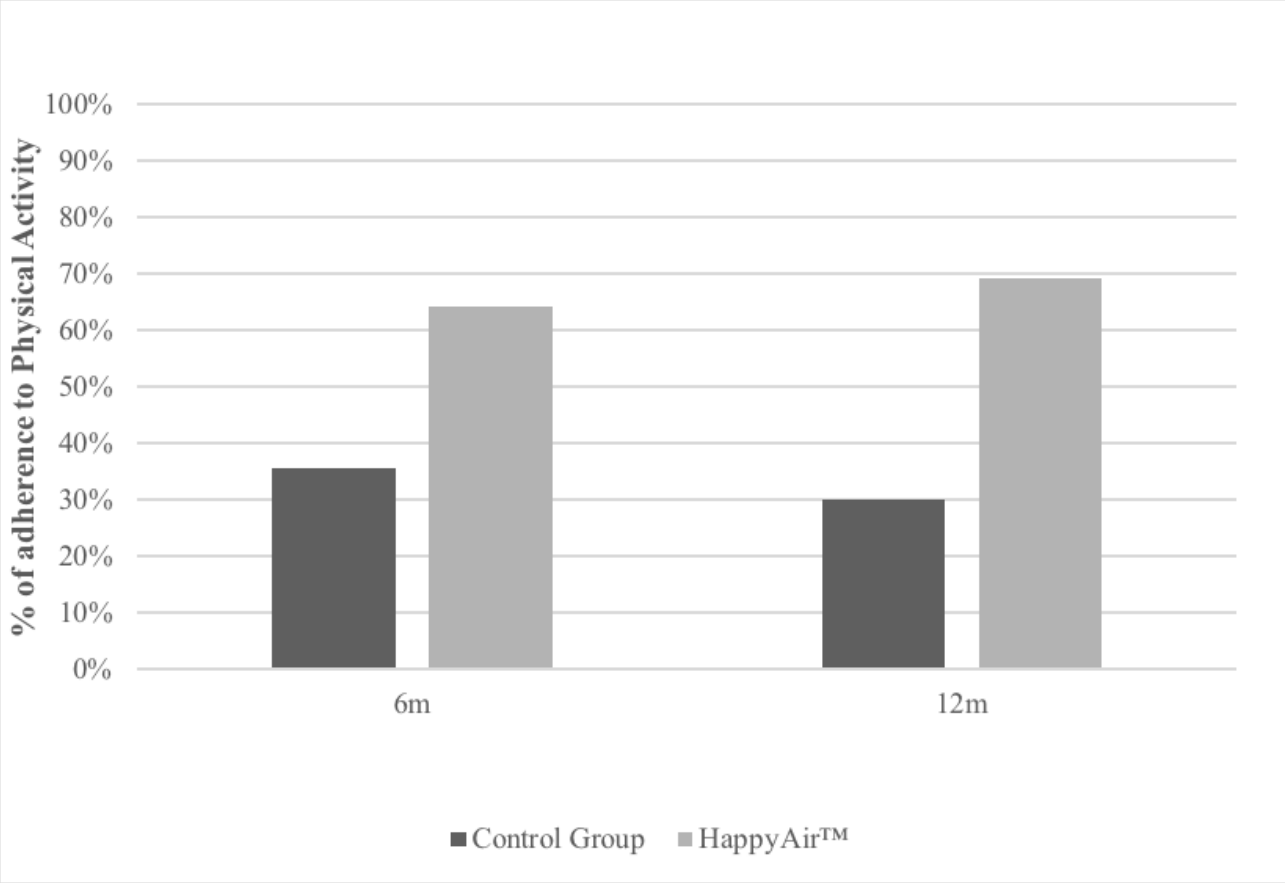


Figure 4. Percentage of patients adherent to physical activity: Morisky Green questionnaire.



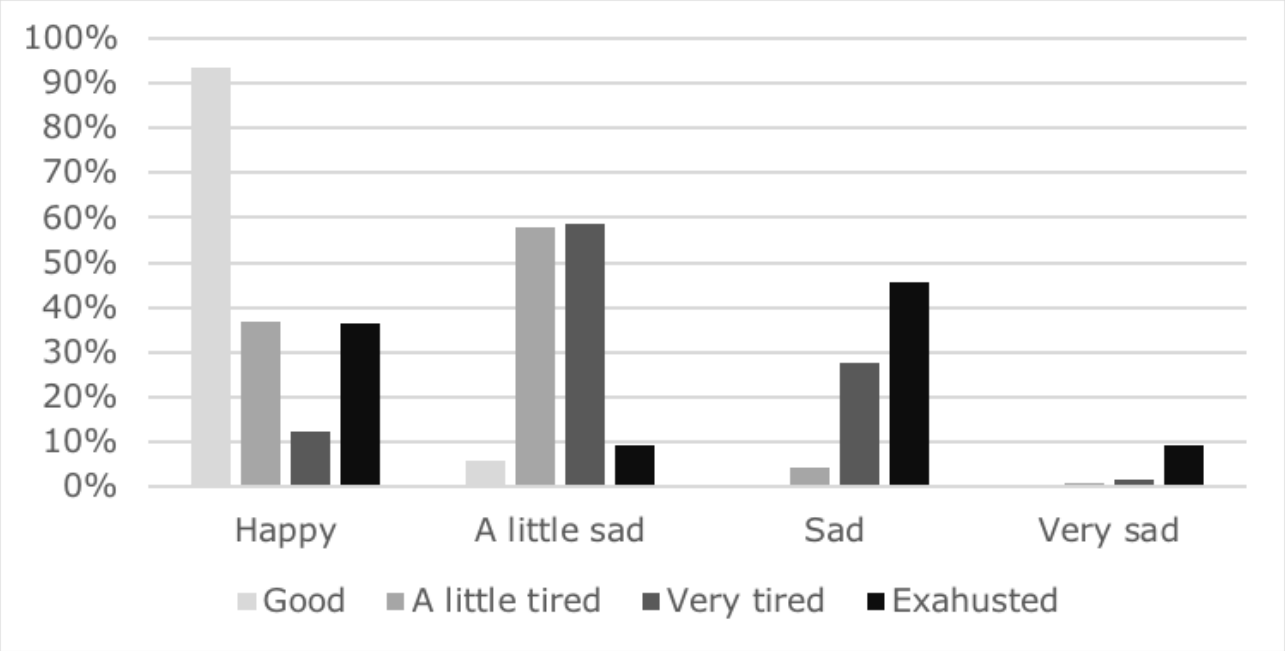
Usability of the HappyAir App

Most of the patients were able to start using the app and almost all in the HappyAir group managed to be skilled and 100% autonomous 15 days after beginning to use the platform. We observed that the average number of physical exercise records was 242 records per patient during the 10 months of follow-up, which is almost a daily record. Records showed that approximately 92% of patients from the HappyAir group exercised daily, which reflects the high rate of use of the app.

Physical Activity, Mood, and Fatigue

Patients in the HappyAir group performed a mean of 66 (SD 37.43) minutes of daily physical activity (95% CI 65.03 to 67.21). Most of the patients analyzed showed a relationship between the feeling of tiredness experienced at the end of the exercises and mood, being less tired, in general, in those patients who were happy. Otherwise, when they finished tired, they felt sad or very sad. The Kruskal-Wallis test showed a significant effect (P=.001) between both variables (Figure 5).

Figure 5. Relationship between mood state and physical activity.



Quality of Life

Patients showed an improvement in quality of life in the CAT questionnaire, with a difference at 6 months during the follow-up period compared with a baseline of 2.2 (SD 0.3) and 4.2 (SD 0.4) in the control group and HappyAir group, respectively. This improvement was only significant in the HappyAir group (*P*=.001; 95% CI –6.6 to –1.6).

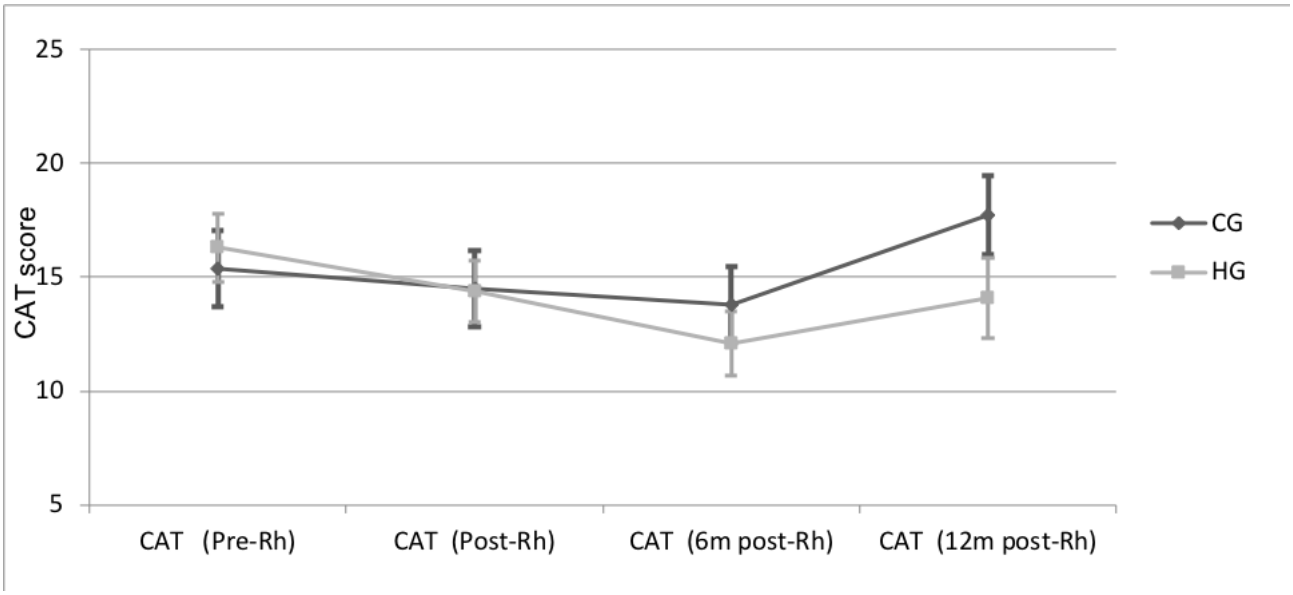
No statistically significant differences between the two groups were observed in the CAT at 6 months (*P*=.53; 95% CI –5.8 to 3.1) or at 12 months (*P*=.21; 95% CI –8.1 to 1.9). However, the mean evolution graph (Figure 6) shows a difference of 3.6 points between the HappyAir group and control group after 12 months of follow-up, denoting a better quality of life in the HappyAir group with respect to the baseline evaluation, since a 2-point difference in the CAT questionnaire is established as the clinically relevant difference in quality of life outcomes.

Regarding the SGRQ, patients in the HappyAir group showed an improvement in quality of life after pulmonary rehabilitation and over time in symptoms, impact, and total dimensions but did not show significant differences in the activities dimension. No statistically significant differences between the two groups were observed in the SGRQ at 6 months (*P*=.77; 95% CI –10.2 to 13.9) or at 12 months (*P*=.79; 95% CI –13.2 to 17.0).

The differences in symptoms at 6 months follow-up compared to baseline were 15.8 (SD 3.3) and 12.1 (SD 2.2) in the control group and HappyAir group, respectively. These differences were significant both for the control group (*P*=.002; 95% CI –27.7 to 4.8) and HappyAir group (*P*=.04; 95% CI –23.8 to –0.2). The control group also showed differences in symptoms at 6-month follow-up after pulmonary rehabilitation compared with post–pulmonary rehabilitation results (14.9 [SD 2.1]; *P*=.01; 95% CI –26.9 to –1.6). The differences in impact at 12-month follow-up after pulmonary rehabilitation were 9.6 (SD 0.8) and 12 (SD 6.9) in the control group and HappyAir group, respectively, compared with baseline. These differences were only significant for the HappyAir group (*P*=.001; 95% CI –22.1 to –1.1). The control group also showed differences in impact at 12-month follow-up after pulmonary rehabilitation compared to results at 6-month follow-up (10 [SD 0.7]; *P*=.04; 95% CI –19.8 to –0.09).

Total dimension results showed an improvement in quality of life observed at both 6- and 12-month follow-up compared with baseline. The difference at 6-month follow-up was 3.2 (SD 1) and 6.4 (SD 1.2) in the control group and HappyAir group, respectively. Only differences in the HappyAir group were significant (*P*=.05; 95% CI –13.5 to 0.6). The difference at 12-month follow-up was 6.3 (SD 0.8) and 7.5 (SD 7.7) in the control group and HappyAir group, respectively. Only differences in the HappyAir group were significant (*P*=.04; 95% CI –16.1 to 1.2). The EuroQOL-5D questionnaire showed no statistically significant intragroup improvements and no statistically significant differences between groups (Multimedia Appendix 2).

Figure 6. Quality of life: Chronic Obstructive Pulmonary Disease Assessment Test questionnaire.



Exercise Capacity, Perceived Dyspnea, and Fatigue

Patients showed an improvement in walking distance after pulmonary rehabilitation, with differences in the 6MWT of 29 (SD 3) meters versus 42 (SD 14) meters in the control group and HappyAir group, respectively. Only differences in the HappyAir group were significant after pulmonary rehabilitation (*P*=.001; 95% CI 6.8 to 79.0). At 12-month follow-up, a decrease in walking distance was observed in both groups. However, the HappyAir group, although not statistically significant, was able to maintain a distance walked above the baseline values (Multimedia Appendix 2). No statistically significant differences between the two groups were observed at 6 months (*P*=.38; 95% CI –34.8 to 88.3) or 12 months (*P*=.58; 95% CI –47.6 to 82.6).

Regarding dyspnea after pulmonary rehabilitation, patients in both groups showed less dyspnea, with improvements of 0.4 (SD 0.5) and 1.1 (SD 0) in the control group and HappyAir group, respectively, which were not statistically significant (*P*=.05). Patients in both groups also showed less fatigue, with improvements at 6-month follow-up compared with a post–pulmonary rehabilitation of 0.2 (SD 0.7) versus 0.8 (SD 0.5) in the control group and HappyAir group, respectively, which were not statistically significant (*P*=.05). Lung function values showed no statistically significant differences between the two groups (Multimedia Appendix 2).

Discussion

Principal Findings

This study examined the effectiveness of a pulmonary care app designed to improve adherence to an integral 1-year maintenance program applied after pulmonary rehabilitation. The challenge was to design an effective intervention to maintain the effects obtained in the initial treatment. Our main hypothesis that the use of this web-based app would improve adherence to the maintenance program was confirmed, in addition to improvement in other variables such as quality of life, behavioral change, and adherence to physical activity.

Adherence, Perception, and Mood

The key to making integral care plans in COPD work, either in a traditional way or through new technologies, is to consider in its design, in addition to the functional dimension, an emotional dimension [28]. This is because adherence is linked to a large number of affective factors such as personalized follow-up programs, social support, the patient's state of mind, and even marital status [21,28-30], so it would be very interesting to take them into account to facilitate integration by patients.

Like other studies [31-33], our study showed a very positive perception of the integrated care plan in the HappyAir group compared with the control group, which is fully consistent with the adherence results in the same group. Those patients who believed that engaging in their treatment would help improve their health were active and responsible in managing their care plan. Their perception about the plan and its value in being continued over time was very positive, and they showed good physical ability, followed their daily care routines, had a positive attitude, and became confident in self-management, which was favorable to long-term adherence to their plan.

Thus, after analyzing the results of their adherence to the care plan using a questionnaire and the online web-based app that connected with their coach for additional support, we discovered both study groups, the control group and HappyAir group, progressed significantly at the beginning of the evaluation period with a positive trend in the engagement in their integrated care plan. What is significant in this study, is that, during the 10 months of monitoring, the control group no longer maintained that difference; however, in contrast, a clear improvement in adherence was observed in the HappyAir group.

These observations lead us to conclude that the integrated care plan, designed to improve adherence, managed to establish a more proactive and responsible attitude toward self-care and improved adherence. We believe that this holistic approach in integrated care with coach support is what has favored this important change in patients from a passive attitude to taking active responsibility in managing their care plan and following

their treatment, a factor that is probably linked to behavior change and adherence. Therefore, remote care can be a very important and effective part of maintenance programs but only if it is accompanied by the behavioral change necessary to promote patient adherence [34].

Apart from that, it is important to consider COPD patients' perceptions of the use of telehealthcare systems. Many patients encounter barriers, such as limited health literacy, difficulties in using technologies, and insufficient clinical support; they feel that mHealth systems could never completely replace face-to-face visits from a health care professional [35]. On the other hand, there are patients who consider technology a facilitator in the management of their disease, due to having a greater responsibility for their own health and for integrating personalized education [36,37].

In any case, when developing mHealth tools, it is important to consider factors such as age, information technology experience, education level, and possible comorbidities. Engagement with patients in the design and testing is essential to make sure the intervention is easy for older people [38-40].

Relation Between Mood and Physical Activity Level

Mood has a strong influence on adherence, so enjoying therapy and overcoming depressive or anxiety states also facilitates success and patient involvement. The feeling of achievement improves self-efficacy [41,42]. Thus, in our study, the HappyAir group was monitored for 1-year using the web-based app. Daily exercise was recorded. This showed that the patients completed an average of 66.12 minutes of physical activity daily, even exceeding the recommendations of the World Health Organization and American College of Sports Medicine [43,44]: a minimum of 150 minutes of moderate intensity exercise per week. In general, our patients exceeded the recommendation, which could be interpreted as a positive motivational effect of the HappyAir program. In addition, we observed an important relationship between mood and tiredness expressed by the patients at the end of the physical activity. This observation showed that the better the mood was, the longer the exercise period lasted; additionally, patients felt less or not tired at all when their mood was positive. This shows that the emotional and psychological states of the patients may be more important factors in achieving adherence to physical activity than the increase in physical capacity itself, corroborating the theory of the study of Mantoani [5].

Social Support and the Role of Therapeutic Educator

Regarding the role of therapeutic educator in maintenance programs, several studies have concluded that this is a very important predictor of adherence [45,46], but not all maintenance programs include it. An important and differentiating aspect of our follow-up program was the figure of the respiratory educator/physiotherapist whom patients could contact at any time. All consultations made were classified into two categories: a need for social support or technical issues, such as with the computer web-based app, connections, or their mobile device. Motivation and support provided by the therapeutic educator were key to successful therapy and patient adherence [31].

Behavioral Change and Its Influence on Adherence

The key that determines the difference in adopting and maintaining any habit lies in the way the initial decision making is carried out, which must be a process shared with the patient: a reasoned discussion about a process that allows that person to be aware of their current health status and ascertains their desire to maintain their health positively in the future. This happens in a dialogue and analysis about the current situation and the undesired alternative. To maintain an adequate level of physical activity in unsupervised periods, when it is very easy to become sedentary, patients must adopt behavioral changes associated with physical abilities. In that sense, the therapeutic educator has an important role to favor the commitment and maintenance of the level of physical activity adopted during the rehabilitation in the long term [34].

Both health professionals and patients consider face-to-face appointments necessary and irreplaceable by technology. It should be noted that the relationship established between the two, also known as therapeutic alliance, largely determines the success of the treatment. Scientific evidence supports the idea that the quality of the health professional–patient relationship is strongly related to patient satisfaction [12]. Thus, the HappyAir web-based app used in this study included the option of contacting the therapeutic educator, as needed, to achieve the necessary personalized support. An analysis of the records showed that approximately 65% of direct contacts requested were for assistance in resolving technical issues and were not related to clinical or social support on disease management issues.

The results were favorable in terms of adopting positive habits in self-care and from the patients' point of view, about the support they received in clinical, social, and therapeutic guidance. Some recent studies, such as by Boer and colleagues [47], highlighted patients' perception of the usability and support offered by mHealth systems. However, as in some of the studies reviewed [48], the therapeutic educator reported an increased workload compared with studies of periodic calls or more time-spaced follow-up.

Quality of Life

According to published literature, during the development of maintenance programs, clinical improvements in quality of life last about 6 to 9 months [9,49]. Thus, our study conducted a follow-up of greater duration than the usual ones in the literature, where the usual follow-up period ranges from 3 to 6 months [15,50-53], and was able to verify that at 1-year follow-up the HappyAir group studied showed significant differences in health-related quality of life over time from the values presented prior to pulmonary rehabilitation until the evaluations carried out at 6 and 12 months, with a gradual improvement in their values, as shown objectively with the SGRQ and CAT questionnaires.

Exercise Capacity, Perceived Dyspnea, and Fatigue

Patients in the HappyAir group showed a clinical improvement in exercise capacity using the 6MWT at 6-month follow-up, and they were the only group to maintain differences in this outcome from baseline to 12 months. The effect of improvement

at 6 months could be attributed to our integrated care plan, since the control group did not present this outcome. But we cannot know exactly because rehabilitation itself can sometimes be responsible for this long-term effect, as shown in several studies [9,10,49]. However, the differences maintained at 12-month follow-up could be directly related to the comprehensive maintenance program, since pulmonary rehabilitation programs consisting of exercise and education have not improved the quality of life and physical capacity after 9 months of follow-up [30,54].

Future Research

After analyzing the results of this study and comparing them with similar studies, it can be concluded that the direction to follow for the design of programs that improve adherence in the management of COPD should be focused on generating behavioral changes and better perceptions in patients. So, the idea is to act on the most essential aspect of the patients, their person and their disease, to achieve awareness of it, generate a proactive attitude, and empower the patient to make them responsible for their care.

The development and use of mHealth systems and innovative technology must be implemented and advanced, as they indicate a promising future. They should be considered as the resources needed to improve patient support and monitoring to learn more about the individual or apply a personalized approach. Moreover, they can generate better outcomes for each person in health and social support, as well as provide further information to apply innovative care and therapeutic techniques in patient management by health care teams, offering resources to those people who live in remote locations with low access to ongoing care management.

Limitations

One of the limitations of this study is the small number of patients included. Although we calculated the sample size, the number of patients included in the control group was too small to extrapolate the results to a more heterogeneous population. However, the sample was in concordance with the majority of similar studies. Another limitation is not having assessed exercise capacity with a gold standard test, such as an incremental exercise test. We chose instead to use simple field tests commonly used in the clinical setting in order to facilitate the implementation of the program in the future. In relation to this, the long-term results achieved by patients does not reflect the increase in daily physical activity observed by the HappyAir integrated care plan. Finally, this study was designed for those patients who would adapt to using a mobile phone. If they had low-level technical skills, this would be a limitation that would need addressing with other technology or support (eg, voice technology).

Conclusions

This study showed the development of the HappyAir integrated care plan after pulmonary rehabilitation, which uses a web-based app accessible by a mobile device and involves periodic therapeutic educator (coach) support. This has been shown to be effective in improving patient adherence to their self-care plan and treatment and consequently their state of health and attitude, with a resulting change in their perception of the disease and their engagement on their care, key factors to achieve positive health outcomes. Internet-enabled and telehealth web-based apps can serve as a means to transform and reinvent the way patients and health care professionals interact. However, this study shows the development and preliminary evaluation of a novel mHealth web-based platform in a reduced sample, which limits the generalization of our results. Further research is needed to integrate HappyAir into larger study populations with COPD.

Authors' Contributions

BJR designed and developed the project and wrote and reviewed the manuscript. EM supervised the collection of data. LJM, MSC, JLRH, MCR, MTHC, and MLM contributed to the study protocol, recruitment of participants, development of the intervention, and follow-up. SA performed the statistical analysis and interpretation of the data. AMP contributed to the data analyses and the writing of the paper and critically reviewed the paper. SF contributed to the development of web-based platform and app. JV contributed to the design of the project, the data analyses, and the writing of the paper and critically reviewed and approved the final version of the paper.

Conflicts of Interest

This study was financed by the Board of Lovexair Foundation (HappyAir in intellectual property of the Lovexair Foundation); SF is its president. EM is an employee of the Lovexair Foundation.

This randomized study was only retrospectively registered, explained by authors with the originally conceived project being the creation of a clinical tool rather than a trial; the completion of the protocol at different stages of the pilot study, some not initially planned, meant that the clinical trial was not registered in due course. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
HappyAir app supplementary information.
[PDF File (Adobe PDF File), 516 KB - mhealth_v8i7e18465_app1.pdf]

Multimedia Appendix 2
Pulmonary function, exercise capacity, and quality of life effects along the study.
[DOCX File , 16 KB - mhealth_v8i7e18465_app2.docx]

Multimedia Appendix 3
Consort checklist.
[PDF File (Adobe PDF File), 1533 KB - mhealth_v8i7e18465_app3.pdf]

References

1. Celli BR, MacNee W, ATS/ERS Task Force. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J 2004 Jun;23(6):932-946 [FREE Full text] [doi: 10.1183/09031936.04.00014304] [Medline: 15219010]
2. Fau LF. [The IBERPOC project: an epidemiological study of COPD in Spain. Scientific Committee of the IBERPOC study]. Arch Bronconeumol 1997 Jun;33(6):293-299. [Medline: 9289325]
3. Alcázar-Navarrete B, Trigueros JA, Riesco JA, Campuzano A, Pérez J. Geographic variations of the prevalence and distribution of COPD phenotypes in Spain: the ESPIRAL-ES study. Int J Chron Obstruct Pulmon Dis 2018;13:1115-1124 [FREE Full text] [doi: 10.2147/COPD.S158031] [Medline: 29692606]
4. Miravitlles M. Guía de Práctica Clínica para el Diagnóstico y Tratamiento de Pacientes con Enfermedad Pulmonar Obstructiva Crónica (EPOC)—Guía Española de la EPOC (GesEPOC). Arch Bronconeumol 2012 Jan;48:2-58. [doi: 10.1016/S0300-2896(12)70035-2]
5. Mantoani LC, Rubio N, McKinstry B, MacNee W, Rabinovich RA. Interventions to modify physical activity in patients with COPD: a systematic review. Eur Respir J 2016 Jul;48(1):69-81 [FREE Full text] [doi: 10.1183/13993003.01744-2015] [Medline: 27103381]
6. Miravitlles M, Soler-Cataluña JJ, Calle M, Molina J, Almagro P, Antonio Quintano J, et al. Spanish Guideline for COPD (GesEPOC). Update 2014. Arch Bronconeumol (English Edition) 2014;50:1-16. [doi: 10.1016/s1579-2129(14)70070-9]
7. Pleguezuelos E, Miravitlles M. Pulmonary rehabilitation… yes but no. Arch Bronconeumol 2019 Dec;55(12):617-618. [doi: 10.1016/j.arbres.2019.07.018] [Medline: 31676192]
8. Global Strategy for Prevention, Diagnosis and Management of COPD. URL: https://goldcopd.org/wp-content/uploads/2018/11/GOLD-2019-v1.7-FINAL-14Nov2018-WMS.pdf [accessed 2020-07-02]
9. Nishiyama O, Taniguchi H, Kondoh Y, Kimura T, Ogawa T, Watanabe F, et al. Factors in maintaining long-term improvements in health-related quality of life after pulmonary rehabilitation for COPD. Qual Life Res 2005 Dec;14(10):2315-2321. [doi: 10.1007/s11136-005-7710-y] [Medline: 16328910]
10. Heerema-Poelman A, Stuive I, Wempe JB. Adherence to a maintenance exercise program 1 year after pulmonary rehabilitation: what are the predictors of dropout? J Cardiopulm Rehabil Prev 2013;33(6):419-426. [doi: 10.1097/HCR.0b013e3182a5274a] [Medline: 24029813]
11. Beauchamp MK, Francella S, Romano JM, Goldstein RS, Brooks D. A novel approach to long-term respiratory care: results of a community-based post-rehabilitation maintenance program in COPD. Respir Med 2013 Aug;107(8):1210-1216 [FREE Full text] [doi: 10.1016/j.rmed.2013.04.014] [Medline: 23706779]
12. Rubeis G, Schochow M, Steger F. Patient autonomy and quality of care in telehealthcare. Sci Eng Ethics 2018 Feb;24(1):93-107. [doi: 10.1007/s11948-017-9885-3] [Medline: 28281148]
13. Criner RN, Han MK. COPD care in the 21st century: a public health priority. Respir Care 2018 Apr 24;63(5):591-600. [doi: 10.4187/respcare.06276]
14. Pedone C, Lelli D. Systematic review of telemonitoring in COPD: an update. Pneumonol Alergol Pol 2015;83(6):476-484 [FREE Full text] [doi: 10.5603/PiAP.2015.0077] [Medline: 26559801]
15. Bolton CE, Waters CS, Peirce S, Elwyn G. Insufficient evidence of benefit: a systematic review of home telemonitoring for COPD. J Eval Clin Pract 2011 Dec;17(6):1216-1222. [doi: 10.1111/j.1365-2753.2010.01536.x] [Medline: 20846317]
16. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Dec 23;5:CD011425. [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
17. Korpershoek YJG, Vervoort SCJM, Trappenburg JCA, Schuurmans MJ. Perceptions of patients with chronic obstructive pulmonary disease and their health care providers towards using mHealth for self-management of exacerbations: a qualitative study. BMC Health Serv Res 2018 Oct 04;18(1):757 [FREE Full text] [doi: 10.1186/s12913-018-3545-4] [Medline: 30286761]

18. Güell Rous MR, Díaz Lobato S, Rodríguez Trigo G, Morante Vélez F, San Miguel M, Cejudo P, Sociedad Española de Neumología y Cirugía Torácica (SEPAR). Pulmonary rehabilitation. Sociedad Española de Neumología y Cirugía Torácica (SEPAR). Arch Bronconeumol 2014 Aug;50(8):332-344. [doi: 10.1016/j.arbres.2014.02.014] [Medline: 24845559]
19. Cebrià IIM, Del CNT, Cañas N, Martínez M, Froilán C, Vilaró-Casamitjana J. Physiotherapy adherence and perception in patients with cystic fibrosis: a descriptive study. Eur Respir J (suppl 59) 2015;46. [doi: 10.1183/13993003.congress-2015.pa2793]
20. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [Medline: 3945130]
21. Rodríguez Chamorro MA, García-Jiménez E, Amariles P, Rodríguez Chamorro A, Faus MJ. [Review of the test used for measuring therapeutic compliance in clinical practice]. Aten Primaria 2008 Aug;40(8):413-418 [FREE Full text] [doi: 10.1157/13125407] [Medline: 18755102]
22. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline LN. Development and first validation of the COPD Assessment Test. Eur Respir J 2009 Sep;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
23. Ferrer M, Alonso J, Prieto L, Plaza V, Monsó E, Marrades R, et al. Validity and reliability of the St George's Respiratory Questionnaire after adaptation to a different language and culture: the Spanish example. Eur Respir J 1996 Jun;9(6):1160-1166 [FREE Full text] [doi: 10.1183/09031936.96.09061160] [Medline: 8804932]
24. Güell R, Casan P, Sangenís M, Morante F, Belda J, Guyatt GH. Quality of life in patients with chronic respiratory disease: the Spanish version of the Chronic Respiratory Questionnaire (CRQ). Eur Respir J 1998 Jan;11(1):55-60 [FREE Full text] [doi: 10.1183/09031936.98.11010055] [Medline: 9543270]
25. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [Medline: 10109801]
26. Güell R, Perpiñá M. Calidad de vida relacionada con la salud en neumología. Arch Bronconeumol 1997 May;33(5):257-258. [doi: 10.1016/s0300-2896(15)30619-0]
27. Singh SJ, Puhan MA, Andrianopoulos V, Hernandes NA, Mitchell KE, Hill CJ, et al. An official systematic review of the European Respiratory Society/American Thoracic Society: measurement properties of field walking tests in chronic respiratory disease. Eur Respir J 2014 Dec;44(6):1447-1478 [FREE Full text] [doi: 10.1183/09031936.00150414] [Medline: 25359356]
28. Moullec G, Ninot G. An integrated programme after pulmonary rehabilitation in patients with chronic obstructive pulmonary disease: effect on emotional and functional dimensions of quality of life. Clin Rehabil 2010 Feb;24(2):122-136. [doi: 10.1177/0269215509346088] [Medline: 20026578]
29. Jolly K, Sidhu MS, Hewitt CA, Coventry PA, Daley A, Jordan R, et al. Self management of patients with mild COPD in primary care: randomised controlled trial. BMJ 2018 Jun 13;361:k2241 [FREE Full text] [doi: 10.1136/bmj.k2241] [Medline: 29899047]
30. Spencer LM, Alison JA, McKeough ZJ. Do supervised weekly exercise programs maintain functional exercise capacity and quality of life, twelve months after pulmonary rehabilitation in COPD? BMC Pulm Med 2007 May 16;7:7 [FREE Full text] [doi: 10.1186/1471-2466-7-7] [Medline: 17506903]
31. Arnold E, Bruton A, Ellis-Hill C. Adherence to pulmonary rehabilitation: a qualitative study. Respir Med 2006 Oct;100(10):1716-1723 [FREE Full text] [doi: 10.1016/j.rmed.2006.02.007] [Medline: 16554147]
32. Lilholt PH, Witt UF, Ehlers L, Hejlesen OK. Telehealthcare for patients suffering from chronic obstructive pulmonary disease: effects on health-related quality of life: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open 2017 May 09;7(5):e014587 [FREE Full text] [doi: 10.1136/bmjopen-2016-014587] [Medline: 28490555]
33. Trappenburg JCA, Niesink A, de Weert-van Oene GH, van der Zeijden H, van Snippenburg R, Peters A, et al. Effects of telemonitoring in patients with chronic obstructive pulmonary disease. Telemed J E Health 2008 Mar;14(2):138-146. [doi: 10.1089/tmj.2007.0037] [Medline: 18361703]
34. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key elements of mhealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
35. Ding H, Fatehi F, Maiorana A, Bashi N, Hu W, Edwards I. Digital health for COPD care: the current state of play. J Thorac Dis 2019 Oct;11(Suppl 17):S2210-S2220 [FREE Full text] [doi: 10.21037/jtd.2019.10.17] [Medline: 31737348]
36. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. A qualitative study of chronic obstructive pulmonary disease patient perceptions of the barriers and facilitators to adopting digital health technology. Digit Health 2019;5:2055207619871729 [FREE Full text] [doi: 10.1177/2055207619871729] [Medline: 31489206]
37. Gorst SL, Armitage C, Hawley M, Coates E. Exploring patient beliefs and perceptions about sustained use of telehealth. Int J Integr Care 2013 Nov 20;13(7):1-2. [doi: 10.5334/ijic.1393]
38. Dobson R, Herbst P, Candy S, Brott T, Garrett J, Humphrey G, et al. Understanding end-user perspectives of mobile pulmonary rehabilitation (mPR): cross-sectional survey and interviews. JMIR Form Res 2019 Dec 20;3(4):e15466 [FREE Full text] [doi: 10.2196/15466] [Medline: 31859681]

39. Gordon N, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 04;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
40. Sobnath DD, Philip N, Kayyali R, Nabhani-Gebara S, Pierscionek B, Vaes AW, et al. Features of a mobile support app for patients with chronic obstructive pulmonary disease: literature review and current applications. JMIR Mhealth Uhealth 2017 Feb 20;5(2):e17 [FREE Full text] [doi: 10.2196/mhealth.4951] [Medline: 28219878]
41. Almagro P, Castro A. Helping COPD patients change health behavior in order to improve their quality of life. Int J Chron Obstruct Pulmon Dis 2013;8:335-345 [FREE Full text] [doi: 10.2147/COPD.S34211] [Medline: 23901267]
42. 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]
43. Global recommendations on physical activity for health. World Health Organization. URL: https://www.who.int/dietphysicalactivity/publications/9789241599979/en/ [accessed 2020-07-06]
44. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
45. Fahy BF. Pulmonary rehabilitation for chronic obstructive pulmonary disease: a scientific and political agenda. Respir Care 2004 Jan;49(1):28-36 [FREE Full text] [Medline: 14733620]
46. Liacos A, Burge AT, Cox NS, Holland AE. Promoting physical activity using the internet: is it feasible and acceptable for patients with chronic obstructive pulmonary disease and bronchiectasis? J Aging Phys Act 2018 Jul 01;26(3):372-381. [doi: 10.1123/japa.2017-0123] [Medline: 28952865]
47. Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, et al. A smart mobile health tool versus a paper action plan to support self-management of chronic obstructive pulmonary disease exacerbations: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 09;7(10):e14408 [FREE Full text] [doi: 10.2196/14408] [Medline: 31599729]
48. Brunton L, Bower P, Sanders C. The contradictions of telehealth user experience in chronic obstructive pulmonary disease (COPD): a qualitative meta-synthesis. PLoS One 2015 Oct;10(10):e0139561 [FREE Full text] [doi: 10.1371/journal.pone.0139561] [Medline: 26465333]
49. Troosters T, Gosselink R, Decramer M. Short- and long-term effects of outpatient rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. Am J Med 2000 Aug 15;109(3):207-212. [doi: 10.1016/s0002-9343(00)00472-1] [Medline: 10974183]
50. Lewis KE, Annandale JA, Warm DL, Rees SE, Hurlin C, Blyth H, et al. Does home telemonitoring after pulmonary rehabilitation reduce healthcare use in optimized COPD? A pilot randomized trial. J COPD 2010 Feb;7(1):44-50. [doi: 10.3109/15412550903499555] [Medline: 20214462]
51. Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. BMC Med Inform Decis Mak 2015 Jun 18;15:46 [FREE Full text] [doi: 10.1186/s12911-015-0171-5] [Medline: 26084626]
52. Román M, Larraz C, Gómez A, Ripoll J, Mir I, Miranda EZ, et al. Efficacy of pulmonary rehabilitation in patients with moderate chronic obstructive pulmonary disease: a randomized controlled trial. BMC Fam Pract 2013 Feb 11;14:21 [FREE Full text] [doi: 10.1186/1471-2296-14-21] [Medline: 23399113]
53. Kwon H, Lee S, Jung EJ, Kim S, Lee J, Kim DK, et al. An mHealth management platform for patients with chronic obstructive pulmonary disease (efil breath): randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10502 [FREE Full text] [doi: 10.2196/10502] [Medline: 30143475]
54. Busby AK, Reese RL, Simon SR. Pulmonary rehabilitation maintenance interventions: a systematic review. Am J Health Behav 2014 May;38(3):321-330. [doi: 10.5993/AJHB.38.3.1] [Medline: 24636028]

Abbreviations

6MWT: 6-Minute Walk Test
CAP FISIO: respiratory physiotherapy adherence self-report questionnaire (Spanish, Cuestionario Adherencia Percepción Physio)
CAT: Chronic Obstructive Pulmonary Disease Assessment Test
COPD: chronic obstructive pulmonary disease
GOLD: Global Initiative for Chronic Obstructive Lung Disease
mHealth: mobile health
SGRQ: St. George's Respiratory Questionnaire

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by R Torres, C Bentley; comments to author 20.04.20; revised version received 12.05.20; accepted 03.06.20; published 31.07.20.

Please cite as:
Jiménez-Reguera B, Maroto López E, Fitch S, Juarros L, Sánchez Cortés M, Rodríguez Hermosa JL, Calle Rubio M, Hernández Criado MT, López M, Angulo-Díaz-Parreño S, Martín-Pintado-Zugasti A, Vilaró J
Development and Preliminary Evaluation of the Effects of an mHealth Web-Based Platform (HappyAir) on Adherence to a Maintenance Program After Pulmonary Rehabilitation in Patients With Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e18465
URL: http://mhealth.jmir.org/2020/7/e18465/
doi:10.2196/18465
PMID:32513646

Original Paper

Barriers and Facilitators to the Adoption of Mobile Health Among Health Care Professionals From the United Kingdom: Discrete Choice Experiment

Simon Leigh[1], BSc, MSc; Liz Ashall-Payne[1], BSc, MA; Tim Andrews[1], BA

The Organisation for the Review of Care and Health Applications, Daresbury, United Kingdom

Corresponding Author:
Simon Leigh, BSc, MSc
The Organisation for the Review of Care and Health Applications
Sci-Tech Daresbury, Vanguard House, Keckwick Lane
Daresbury, WA4 4AB
United Kingdom
Phone: 44 07503955592
Email: simon.leigh@orcha.co.uk

Abstract

Background: Despite the increasing availability of mobile health services, clinical engagement remains minimal.

Objective: This study aims to identify and weight barriers to and drivers of health app use among health care professionals (HCPs) from the United Kingdom.

Methods: A discrete choice experiment was conducted with 222 HCPs using a web-based survey between March 2019 and June 2019. Participants were recruited to take part via social media and asked to choose their preferred option of 2 hypothetical health apps to prescribe to a hypothetical patient or to prescribe neither. Choices were characterized by differing levels of patient age, cost, published evidence bases, whether they had a National Health Service (NHS) stamp of approval, personal familiarity with the technology, and whether they were recommended by a fellow HCP. The results were analyzed using a mixed logit model, with subgroup analyses to account for heterogeneity.

Results: We received 230 responses, a total of 96.5% (n=222/230) of respondents understood the survey task and passed the test of rationality. The median age was between 36 and 45 years, and 62.6% (n=139/222) of the health care providers responding to the survey had previously recommended the use of health apps to patients. Health apps were most likely to be prescribed to patients if they had an NHS stamp of approval or if they were recommended by another HCP (both $P<.001$). Published studies detailing clinical effectiveness were important ($P<.001$), but it would take five published studies to have the same impact on prescribing behavior as an NHS stamp of approval and two studies to be as convincing as having used the technology personally. Increasing patient age and costs resulted in significant reductions in digital health prescribing ($P<.001$), none more so than among allied health professionals. Willingness-to-pay for health apps increased by £124.61 (US $151.14) if an NHS stamp of approval was present and by £29.20 (US $35.42) for each published study. Overall, 8.1% (n=18/222) of respondents were reluctant to use health apps, always choosing the I would prescribe neither option, particularly among older HCPs, nurses, and those who do not use health apps personally. Subgroup analyses revealed significant differences in preferences among HCPs of differing ages and clinical backgrounds.

Conclusions: An NHS stamp of approval, published studies, and recommendations from fellow HCPs are significant facilitators of digital prescribing, whereas increasing costs and patient age are significant barriers to engagement. These findings suggest that demonstrating assurances of health apps and supporting both the dissemination and peer-to-peer recommendation of evidence-based technologies are critical if the NHS is to achieve its long-term digital transformation ambitions.

(JMIR Mhealth Uhealth 2020;8(7):e17704) doi:10.2196/17704

KEYWORDS

digital health; mHealth; discrete-choice; preferences; mobile phone

Introduction

Increasing public expectations of National Health Service (NHS) care, a rapidly aging population, and an ever-increasing prevalence of long-term and comorbid conditions mean that health care systems are working harder than ever. Adapting the way patients and health care professionals (HCPs) communicate and collaborate in promoting health and well-being is, therefore, essential if the future expectations of high-quality patient-centric care are to be realized.

New digital technologies, which are generally low cost and widely accessible to 95% of adults in the United Kingdom who own a smartphone [1], have the potential to be a valuable strategic tool in this transition. These technologies have the potential to deliver significant improvements in disease management by empowering patients to monitor and manage their long-term conditions [2], encourage the promotion of health and well-being, and provide care outside of traditional health care settings. Furthermore, the highly personalized real-world data captured by digital health technologies have the potential not only to improve service planning by aligning capacity more closely with demand [3] but also provide a first-of-a-kind opportunity to understand real-world condition management, enabling both informed decision making and the delivery of personalized care [4-6].

From 2024, patients in England will have the *right* to access digital primary care services under the NHS long-term plan [7]. By the end of the 10-year period covered by the plan, it is envisaged that digital care will become standard care, with people increasingly being cared for and supported at home using remote monitoring (via wearable devices), electronic services, and digital tools. Although the rollout of e-services including the electronic prescription service (EPS) has widely been a success, with 93% of England's 7300 general practitioner (GP) practices enrolled and more than 67% of prescriptions delivered via EPS [7], clinical engagement with digital health technologies, despite a suggested appetite for digital health [8-10], is to date minimal [11,12]. A survey of nurses in Catalonia demonstrated that only 6.5% of those included frequently recommended digital technologies as part of routine patient care [11], with similar studies in France [13], Ireland [14], Norway [15], Germany [16], Belgium [17], and Australia [18] also highlighting a hesitance among health professionals to engage with and recommend digital health technologies in routine practice.

Being the gatekeepers of health services, understanding the perspectives of HCPs is crucial to achieving the long-term plans of the NHS for digital transformation, and the effective delivery and uptake of safe and high-quality digital solutions, which have proven to be clinically effective. However, to date, a paucity of evidence regarding HCP preferences for digital health solutions limits our ability to realize any potential such technologies may deliver.

On the basis of the combined results of a qualitative pilot study [19] and a targeted literature review [8-18,20,21], we conducted a discrete choice experiment (DCE) among 222 HCPs in the United Kingdom to determine priorities, trade-offs, and willingness-to-pay (WTP) for several common characteristics of digital health technologies.

Methods

Ethics Approval and Consent to Participate

The study did not require ethical approval as the form of the study was opinion seeking for the purpose of market research, with the subject matter limited to topics that are strictly within the professional competence of the participants. In addition, no vulnerable groups were included, the data were completely anonymous, there was no risk of disclosure, and the data collected were not sensitive in nature. We also submitted an inquiry to the NHS Health Research Authority, which confirmed that ethical approval was not required.

Availability of Data and Materials

The data that support the findings of this study are available from the authors upon reasonable request.

Face-to-Face Discussions and Ranking Exercise

We followed methodological guidelines from the International Society for Pharmacoeconomics and Outcomes Research [22]. First, a pragmatic literature review was conducted to identify stated preferences regarding the use of digital health. There were no time or geographical restrictions applied. The search terms used are provided in Multimedia Appendix 1. Subsequently, we invited 10 HCPs from the Northwest of England to participate in face-to-face discussions, of which 8 agreed. We asked respondents a single question, "What matters to you when considering the use of health-apps with your patients?" Combining the findings with the results of our targeted literature review, we created a ranking exercise provided in Multimedia Appendix 2. In total, 30 HCPs were invited via email and social media to take part in the ranking exercise, of which 28 accepted the invitation (93% participation rate). During the ranking exercise, 9 attributes, each identified through the initial face-to-face discussions and literature review, were ranked on a scale from 1 (most important) to 9 (least important) by a cohort of 85 HCPs from the United Kingdom. In addition to the face-to-face discussions, we also invited HCPs to provide responses on the web, of which 57 HCPs from the United Kingdom contributed. Further details of this exercise are provided elsewhere [19]. Following the conclusion of the ranking exercise, the 6 highest ranked attributes were selected for inclusion in the discrete choice experiment (DCE).

Discrete Choice Experiment

A DCE, designed to elicit the preferences of HCPs for digital health prescribing, and including the attributes identified during the literature review, face-to-face discussions, and ranking exercise [19], was provided to HCPs between March and June 2019 (sample survey instrument is provided in Multimedia Appendix 3). The DCE survey was provided digitally, using Google Forms (Google, Alphabet Inc) and disseminated using social media (LinkedIn, Twitter, and Facebook). All participants consented to participation by submitting the survey on the web after being provided with a study information sheet. Demographic information was collected for all respondents.

DCE methodology is well described [23-25], and given the increasing emphasis on shared decision making and value-based care, it is used extensively to measure the preferences of both patients and HCPs for the delivery of health care. In DCEs, respondents are given a hypothetical scenario, typically comparing one option to another, and asked to choose which they prefer. This process is repeated with the values (levels) of the characteristics (attributes) randomly changing each time. The choices respondents make can be used to infer preferences for each level of each of the attributes included. In our DCE, all attributes and levels were identified during the face-to-face discussions and ranking exercises and are provided in Table 1.

All respondents were HCPs working in the United Kingdom. We included qualified nursing and medical staff of all grades (nurse, GP, secondary care physician, and allied health professional), from primary, secondary, and tertiary care settings. We stated clearly in the description of the survey and advertisements that the survey was only to be answered by HCPs based in the United Kingdom; however, we also included a screening question as the first question of the survey, where respondents were asked to provide their job title. Responses from those who were not HCPs were removed at this point. Each respondent received 16 discrete choice tasks plus a test of rationality to gauge their understanding of the survey. In the test of rationality, one option was clearly superior to the alternative in every domain, including a lower price, a greater number of studies, and an NHS stamp of approval. If respondents failed the test of rationality, by selecting the inferior option, their responses were excluded from the formal analysis. Each question asked respondents to choose between 2 digital health technologies, each characterized by different levels of the attributes included (Multimedia Appendix 3), in addition to an opt-out *I would prescribe neither* option. This option was important to highlight those who were consistently reluctant to engage with digital health technologies, regardless of the characteristics of the technologies under consideration. As the full factorial design, whereby each respondent is provided with each possible combination of levels for each attribute, would have necessitated ($3^3 \times 4^3 = 1728$) choices per respondent, a D-optimal design was used with 2 blocks of 16 possible questions. This approach was used to maximize statistical power while minimizing the cognitive burden to respondents. The surveys were written in language that could be understood by a 10-year old. The DCE was pilot tested in a sample of 10 HCPs not involved in the main study to gauge interpretation and response times, during which period a researcher was available to answer any questions. Minor changes were subsequently made to clarify wording and graphics.

Table 1. Attributes and levels of the discrete choice experiment.

Attribute	Levels
Number of studies concerning safety and effectiveness	0, 1, 2, or 3
Does the app have an NHS[a] stamp of approval?	No or yes
Cost of the technology to the NHS, £	0, 5, 25, or 75
You have personally used the app yourself?	No or yes
Age of the patient (years)	18, 35, 55, or 75
The app has been recommended by other HCPs[b]	No or yes

[a]NHS: National Health Service.

[b]HCPs: health care professionals.

Data Analysis

We used a mixed logit model to estimate preferences for the different levels of attributes associated with digital health technologies, thereby determining which increased or decreased utility and subsequently increased/decreased the likelihood of recommending these technologies to patients. Dummy coding was used for all categorical variables, with the number of published studies, patient age, and the cost of the app coded as linear continuous variables. We first estimated the main effects model and then estimated the effects for subgroups based on factors such as years of experience as an HCP, current digital engagement, and clinical role (eg, GP, secondary care physician, and allied health professional). WTP analyses were performed to determine how HCPs were willing to trade-off one attribute for another. CIs for WTP estimates were estimated via joint distributed bootstrapping. All analyses were performed using Stata 14 (StataCorp LP) and deemed statistically significant at the 5% level (ie, $P<.05$). Although sample size calculations represent a technical challenge in DCEs, we used Johnson and Orme's approach [26] to estimate our minimum sample size, which was equal to 42 respondents per block.

Results

Characteristics of Participants

Between March 2019 and June 2019, 250 responses were received, of which 20 were excluded because of not being from HCPs. A further 8 respondents were excluded because of failing the test of rationality, suggesting an understanding rate of 96.5% (n=222/230), and a complete dataset of 222 respondents. Table 2 demonstrates the demographics of those completing the DCE.

Table 2. Characteristics of health care professionals completing the discrete choice experiment (N=222).

Characteristic	Value, n (%)
Age (years)	
<26	5 (2.3)
26-35	34 (15.3)
36-45	74 (33.3)
46-55	76 (34.2)
56-65	27 (12.2)
66-75	5 (2.3)
>75	1 (0.5)
Have you used health apps personally?	
Yes	169 (76.1)
No	53 (23.9)
Have you used health apps with patients?	
Yes	139 (62.6)
No	83 (37.4)
Role	
Allied health professional	86 (38.7)
Community caregiver	5 (2.3)
Dentist	3 (1.4)
General practitioner	32 (14.4)
Nurse	27 (12.2)
Pharmacist	1 (0.5)
Secondary care physician	40 (18)
Other	28 (12.6)

Barriers and Facilitators to Health Care Professionals Prescribing Digital Health Technologies

In the discrete choice analysis, all attributes were statistically different from 0, suggesting importance with respect to digital health prescribing and the decision to prescribe to patients. Table 3 illustrates the net effect on the preferences of HCPs for each characteristic. A positive coefficient represents a facilitator, and a negative coefficient represents a barrier to digital health prescribing.

Having a stamp of approval from the NHS was the most important factor in encouraging mobile health (mHealth) prescribing (β=2.36, 95% CI 2.08-2.64), followed by a recommendation from a fellow HCP (β=1.28, 95% CI 1.07-1.49) and having used health apps personally (β=1.04, 95% CI 0.83 to 1.26). Although having published studies to demonstrate safety and effectiveness was important (β=.55, 95% CI 0.44-0.67), it would take 5 published studies to be as convincing as an NHS stamp of approval and 3 to be as convincing as a recommendation from a fellow HCP. Patient age (β=−0.02, 95% CI −0.01 to −0.02) and the cost of the app (β=−0.02, 95% CI −0.02 to −0.02) were both statistically significantly associated with a reduced likelihood of prescribing digital health technologies, suggesting that as patient age (per year) and cost (per £1) increase, prescribing of digital health technologies can be expected to fall. Finally, the opt-out option was also statistically significant; regardless of a high number of clinical studies, recommendations by HCPs, or having a stamp of approval, 8.1% (n=18/222) of respondents chose to prescribe neither app, suggesting a reluctance to utilize health apps among a considerable number of respondents.

Table 3. Preferences for digital health technologies as reported by health care professionals.

Attribute	Coefficient (β), mean (SD)	95% CI
NHS[a] stamp of approval (yes)	2.36[b] (1.51)	2.09-2.64
Health app recommended by HCP[c] (yes)	1.28[b] (0.76)	1.07-1.49
Used health apps personally (yes)	1.04[b] (0.82)	0.83-1.26
Published study (per additional study)	.555[b] (0.29)	0.44-0.67
Patient age (per additional year)	−0.018[b] (−0.01)	−0.02 to −0.01
Cost (per additional £1)	−0.019[b] (−0.02)	−0.02 to −0.02
Alternative specific constant	1.25[b] (1.58)	0.84-1.65
Observations	10,656 (N/A[d])	N/A
Log likelihood	−1226.3 (N/A)	N/A

[a]NHS: National Health Service.

[b]Significant at 5% level. The table represents beta coefficients and CIs from mixed logit regression. The regression coefficients for each attribute level represent the mean part-worth utility of that attribute level in the respondent sample. A positive value denotes utility/satisfaction, and a negative value denotes disutility/dissatisfaction.

[c]HCP: health care professional.

[d]Not applicable.

Differences in the Preferences of Health Care Professionals for Digital Health Prescribing: Subgrouping by Clinical Role and Digital Familiarity

Multimedia Appendix 4 demonstrates the varying perceptions toward digital health prescribing among HCPs with varying clinical roles, whereas Multimedia Appendix 5 breaks down perceptions by digital familiarity and current use of digital health prescribing.

Increasing patient age had a stronger negative impact on digital prescribing among allied health professionals when compared with secondary care physicians and GPs, whereas among nurses, patient age did not impact digital health prescribing behavior at all. Similarly, a recommendation to use digital health prescribing from another HCP was shown to be highly influential in promoting digital prescribing among nurses and secondary care physicians, but significantly less so among GPs and allied health professionals (Table 4), which provides the relative attribute importance for each attribute and for each subgroup of HCPs under consideration and shows that an NHS stamp of approval, a single published study, and a recommendation from other HCPs were 25, 21, and 17 times more likely to impact digital prescribing than patient age among nurses.

Having an NHS stamp of approval was the most influential factor in promoting digital prescribing across all subgroups. Having published studies to demonstrate safety and clinical effectiveness was also important to respondents, and exceptionally so among younger professionals and those unfamiliar with prescribing digital health technologies to patients. However, such studies were less important to allied health professionals who believe that cost is the most important factor in the decision to prescribe or not to prescribe health apps to patients.

Finally, reluctance to use health apps varied considerably among the HCP groups under analysis. Older clinicians (aged >46 years), nurses, allied health professionals, and those who do not currently use health apps personally, all things being equal, are far more likely to opt out of prescribing digital health technologies to patients. Conversely, the most digitally enabled groups were secondary care physicians, GPs, and those who use digital health technology to manage their own health and well-being (Multimedia Appendix 5).

Table 4. Relative attribute importance[a].

Group	Value, n	Characteristics of health apps					
		Published study	NHS[b] stamp of approval	Cost	Used health apps personally	Patient age	Recommended by another HCP[c]
All	222	7	10	6	4.4	4.3	5.4
Use health apps personally							
Yes	169	7.2	10	5.8	4.8	4.6	5.5
No	53	7.5	10	3.9	1.6	3.1	3.9
Previously prescribed health apps to patient							
Yes	139	6.2	10	6.3	4.5	4.3	4.8
No	83	8.8	10	6	3.5	3.3	5.1
Age (years)							
>46	113	5.8	10	7.1	4	4.1	5.2
<46	109	9.1	10	5.4	4.6	5.3	5.4
Role							
General practitioner	32	6.8	10	5.6	2.2	3.6	3
Allied health professional	86	5.1	10	6.2	4.3	6.3	4.7
Secondary care physician	40	6.7	10	2.6	6.3	5.8	7.4
Nurse	27	8.3	10	0.5	3.3	0.4	6.6

[a]Standardized relative attribute importance (RAI) for each attribute was calculated across the subgroups to allow for across subgroups comparisons. First, an RAI was calculated for each attribute by taking the difference between the most and least preferred levels. The RAI was then standardized across subgroups by dividing it by the RAI of the most important attribute across the subgroups (NHS stamp of approval) and multiplying it by 10. The resulting number indicates the relative importance of each attribute across the subgroups (where a higher number indicates a relatively more important attribute).

[b]NHS: National Health Service.

[c]HCP: health care professional.

Trade-Offs: Willingness-To-Pay for the Attributes of Digital Health Technologies

HCPs expressed a WTP of £124.61 (US $152.02) for a digital health technology with an NHS stamp of approval; however, this varied from £119.14 (US $145.35) among those already familiar with prescribing digital health technology to patients, up to a maximum of £1616 (US $1971.45) among nurses, as shown in Table 5. The fact that nurses are willing to trade-off up to £1616 (US $1971.45) of NHS funds for a digital health technology with an NHS stamp of approval signifies the relatively low importance placed on cost and the high ranking of a stamp of approval by nurses. Respondents were willing to pay £29.20 (US $35.62) for each published study of effectiveness or safety, which again varied from £20.79 (US $25.36; allied health professionals) to £449.50 (US $548.37; nurses). Finally, HCPs had a WTP of £67.30 US $82.10) for digital health technologies that came recommended by other HCPs, varying from £39.84 (US $ 48.60) among GPs, to £1061.50 (US $1294.99) among nurses. Using the WTP estimates provided, it is possible to estimate WTP for various types of digital health technologies with varying features. For example, a technology with no stamp of approval that is recommended by a fellow HCP and has a single study behind it would have a WTP of (£124.61 [US $152.02]×0)+(£67.30 [US $82.10])+(£29.20 [US $35.62]×1)=£96.50 (US $117.73). Similarly, a health app with an NHS stamp of approval, with an evidence base consisting of 3 studies, but which is not recommended by a HCP would have a WTP of (£124.61 [US $152.02]×1)+(£67.30 [US $82.10]×0)+(£29.20 [US $35.62]×3)=£212.21 (US $258.89).

Table 5. Willingness-to-pay by health care professional subgroup.

Group	Value, n	Willingness-to-pay according to characteristics of health apps, £ (US $)			
		Published study	NHS[a] stamp of approval	Used health apps personally	Recommended by another HCP[b]
All	222	29.20 (35.62)	124.61 (152.02)	54.81 (66.87)	67.30 (82.10)
Use health apps personally					
Yes	169	31.11 (37.95)	129.44 (157.91)	62.11 (75.77)	71.28 (86.96)
No	53	48.56 (59.24)	194.38 (237.14)	31.19 (38.05)	76.44 (93.25)
Previously prescribed health apps to patients					
Yes	139	24.48 (29.86)	119.14 (145.35)	53.43 (65.18)	57.57 (70.23)
No	83	36.89 (45.00)	125.42 (153.01)	43.37 (52.91)	64.11 (78.21)
Age (years)					
>46	113	20.25 (24.70)	105.00 (128.10)	41.75 (50.93)	54.21 (66.13)
<46	109	42.35 (51.67)	140.00 (170.79)	64.59 (78.80)	76.29 (93.07)
Role					
General practitioner	32	30.28 (36.94)	134.44 (164.01)	29.56 (36.06)	39.84 (48.60)
Allied health professional	86	20.79 (25.36)	121.38 (148.08)	52.79 (64.40)	57.38 (70.00)
Secondary care physician	40	63.50 (77.47)	284.75 (347.38)	178.13 (217.31)	209.75 (255.89)
Nurse	27	449.50 (548.37)	1616.00 (1971.45)	526.00 (641.70)	1061.50 (1294.99)

[a]NHS: National Health Service.

[b]HCP: health care professional.

Discussion

Principal Findings

In this first-of-its-kind study, examining the barriers and drivers of digital health prescribing among a broad sample of HCPs from the United Kingdom, we found that the factors most influential in digital prescribing behaviors are an NHS stamp of approval, a published evidence base, cost of the technology, and recommendation by other HCPs. Respondents expressed a WTP of more than £100 for technologies with a stamp of approval from the NHS and were willing to pay approximately an extra £30 for every additional published study. The strength of this preference varied significantly among our heterogenous cohort, influenced by clinical role, age, and current level of digital literacy. This suggests that a one-size-fits-all approach to increasing digital health adaptation is unlikely to be successful in increasing the uptake of digital health in routine practice.

A recommendation from a HCP can go a long way in encouraging patients to use digital health technology. Although previous research suggests there is an appetite for digital health technologies among HCPs [8-10], engagement is minimal [11,12], with 2 studies in Catalonia and Belgium, respectively, giving utilization rates of 0% [11] and 33% [17], and others from across the globe suggesting a similar picture [13-16]. Our finding that 62.6% of respondents have to date recommended digital health technology to patients in routine practice and that 92% would rather prescribe a digital health technology than not is an interesting and atypical finding in the context of existing literature. One explanation may be the recent advocation for digital medicine by the English pharmaceutical regulator, the National Institute for Health and Care Excellence (NICE), such as recommending digital technology as a first-line treatment for children with mild depression [27] and increasing advocacy for digital prescribing from patient groups and charities, such as Age UK's recommendation for digital services for adults experiencing from loneliness [28].

The latter is of interest, given the strong tendency of our sample to reduce digital prescribing as patient age increased, in line with a recent study conducted among chronic obstructive pulmonary disorder specialists in Canada [21]. This behavior prevails despite evidence suggesting that older individuals are increasingly accessing digitally enabled services [29-32]. If these findings are indicative of the underlying preferences of HCPs, additional effort may be required to promote digital engagement in elderly patients, particularly among allied health professionals, secondary care physicians, and younger HCPs (<46 years), where a negative impact of increasing patient age was most significant.

We also found that the age of a HCP may also predict digital engagement, with those over the age of 46 years considerably more likely to opt out of providing digital health technology, all things being equal, than any other group. A mixed methods cross-sectional study conducted in the Czech Republic observed a similar pattern [20]. The findings of these studies combined suggest that older professionals, much like older members of the general population [33], may be more skeptical of mHealth and may, therefore, require additional support to ensure the NHS long-term digital vision becomes a reality. Combined with nurses and those who do not currently use health apps personally, if the NHS digital strategy is to become a reality,

these groups could represent an ideal target for workforce development. Examples include continuous professional development, the development of specialist digital skills, such as those provided by NHS Digital's digital academy, and the creation of digital champions, which have previously been shown to be highly effective in expanding and supporting the use of digital health technologies in routine practice [34,35].

A 2019 study of Irish chronic obstructive pulmonary disorder specialists found a need for a strong evidence before considering the adaptation digital health technology in clinical practice, with published studies seen as a surrogate for safety [14], whereas findings similar to that from this study were observed among Dutch dermatologists [10], Norwegian GPs [15], and Belgian family physicians [17]. The same was observed in this study. However, despite the assertion from regulators that mHealth randomized controlled trials (RCTs) are not the only method available for evaluating digital health technologies [36,37], part of the preference for published studies that we observed may have been rooted in a desire for *gold-standard* evidence in the form of an RCT. A limitation of our study is that we did not specify which *type* of evidence was provided with health apps, which some respondents may have assumed to be RCTs, as opposed to naturalistic studies, case-control, or qualitative studies. Each of these study types is now considered acceptable forms of evidence for the *right type* of health app under NICE's evidence standards framework. However, it is unclear whether this notion of *not all apps being equal* [19] is accepted by HCPs who have the ultimate responsibility for patient well-being. This is a question that future research should aim to address before developers of health apps to invest too heavily in evidence generation, particularly considering the WTP estimates suggested by this analysis (£29.20 extra for each additional study and £124.61 for an NHS stamp of approval). Advocates of digital health such as the Organisation for the Review of Care and Health Applications (ORCHA), NHS Digital, NICE, and the Medicines and Healthcare Products Regulatory Agency are likely to play a key role in this process, ensuring that HCPs are well informed of the benefits that health apps and digital health more broadly can deliver, know-how and where to access technologies that have been proven to be effective, and are aware of what level of evidence is appropriate for each kind of technology available.

Our pilot study suggested that HCPs value an NHS stamp of approval above all else [19], a finding echoed in a study among Catalonian nurses [11]. This was also observed in this analysis, with every group of HCPs, regardless of age, professional role, or personal digital literacy, prioritizing this characteristic above all else. This finding, although both consistent and significant, should be interpreted with caution, as it is not abundantly clear what respondents of the survey consider as an NHS stamp of approval, which could vary from inclusion on a clinical commissioning group website, NHS choices or the NHS apps library, or inclusion within Egton medical information systems or SystmOne. Historically, it has been a difficult task for developers to achieve inclusion in the NHS apps library, and to date, only 76 digital health technologies have been listed. This may be explained by a misalignment in the evidentiary requirements for these technologies, which until recently have applied a one-size-fits-all approach to evidence, making approval difficult. As existing assessment processes such as the NHS Digital Assessment Portal begin to incorporate proportionate evidentiary requirements [36-38], these *stamps of approval* may be expected to increase. However, it is uncertain whether this is what HCPs value and, ultimately, whether this will have any impact on the digital prescribing behaviors of HCPs. Further research should aim to identify precisely what an NHS stamp of approval should look like. It is also important that we understand how this can be conveyed to HCPs to provide reassurance that technologies with an NHS stamp of approval are both safe and effective.

Our study has several strengths. First, to the best of our collective knowledge, the combination of qualitative pilot testing followed up by the quantitative assessment of relative preferences using the DCE methodology make this research the first-of-its-kind when considering attitudes toward digital health. Second, the low rate of failures during the rationality test suggests that participant understanding of the survey was very high, with just 3.5% failing to progress. This lends support for the responses received being the true beliefs of the HCPs involved, rather than stochastic variation, which can be expected in the event of misunderstanding the survey. Finally, although our participants could be considered an anomalous group of highly digitally engaged HCPs, our subgroup analyses, which adjusted for digital familiarity and personal perceptions around the use of digital technologies, found no significant differences between those who personally use health apps and those who do not, and more importantly, those who currently recommend health apps to patients and those who do not. Therefore, the findings of this study can be considered representative of underlying preferences as vignettes from all levels of digital engagement were included.

The findings of our study should also be viewed in the context of several limitations. First, although we rigorously followed methodological guidelines for eliciting preferences, our study was not large enough to capture all relevant attributes that factor into the decision to provide or not to provide health apps to patients. For example, a systematic review conducted in 2018 highlighted that excessive data creation and burden in the analysis would be a significant deterrent to GPs recommending health apps to patients [12]. Similarly, a survey of Belgian family physicians suggested that additional time taken to download and explain the technologies to patients may limit the ability of HCPs to recommend health apps [17]. As neither of these attributes were included in this study, our results cannot be considered definitive, and further research is required to understand the role of these attributes relative to those examined in this study. Second, the sample sizes for our subgroup comparisons were limited, which makes robust, precise conclusions surrounding the respondents' WTP or comparisons between subgroups difficult. Third, a stamp of approval may, in practical terms, be many things, from a listing on NHS choices to recommendations via local trust websites or an actual NHS badge provided on app stores such as Google Play, App Store, and ORCHA. There is currently no standard for an NHS stamp of approval, and further research is required on this subject to determine how this may look in practice. Similarly,

as mentioned previously, our lack of clarity regarding the details of what we meant by a published study, and the fact that not all studies in mHealth are RCTs, may have led people to value this more highly than if explained in greater detail. Finally, although stated preference techniques are a useful tool in understanding the perspectives of HCPs, nothing provides greater insights than real-world utilization. This study should, therefore, be seen as indicative until a sufficient body of real-world evidence has been generated that can adequately demonstrate facets of health apps used most commonly in routine practice.

Conclusions

This is the first study to quantitatively determine factors associated with health app prescribing among HCPs in the United Kingdom. The findings suggest that having an NHS stamp of approval, published studies, and recommendations to use digital health technology from fellow HCPs are the greatest facilitators of digital prescribing among HCPs, whereas increasing costs and patient age are significant barriers to engagement. These findings suggest that demonstrating assurances and supporting both the dissemination and peer-to-peer recommendation of evidence-based technologies that meet health challenges are critical if the NHS is to achieve its digital transformation ambitions.

Acknowledgments

The authors would like to thank the many HCPs who gave their time to answer the survey. This research was funded by ORCHA Health Limited.

Authors' Contributions

SL devised the study and acted as a guarantor for the paper, and SL and LA collected the data. SL and TA planned statistical analyses, and SL performed all statistical analyses. SL, LA, and TA wrote the first draft of the manuscript and revised and approved the final manuscript as submitted. All authors helped draft the manuscript and approved the final submitted version.

Conflicts of Interest

SL is a paid contractor, and both TA and LA are paid directors of ORCHA Health Limited, a company specializing in the review of digital health technologies. There are no other financial or other interests to declare.

Multimedia Appendix 1
Search terms for literature review.
[PDF File (Adobe PDF File), 54 KB - mhealth_v8i7e17704_app1.pdf]

Multimedia Appendix 2
Ranking exercise to determine most important characteristics of health-apps to health care professionals.
[PDF File (Adobe PDF File), 75 KB - mhealth_v8i7e17704_app2.pdf]

Multimedia Appendix 3
Sample survey instrument.
[PDF File (Adobe PDF File), 100 KB - mhealth_v8i7e17704_app3.pdf]

Multimedia Appendix 4
Variation in health care professional preferences for digital health prescribing, subgrouped by age and clinical role.
[PNG File , 20 KB - mhealth_v8i7e17704_app4.png]

Multimedia Appendix 5
Variation in health care professional preferences for digital health prescribing, subgrouped by digital familiarity and literacy.
[PNG File , 86 KB - mhealth_v8i7e17704_app5.png]

References

1. Deloitte US. 2017. UK Public Are 'Glued to Smartphones' as Device Adoption Reaches New Heights URL: https://www2.deloitte.com/uk/en/pages/press-releases/articles/uk-public-glued-to-smartphones.html [accessed 2019-10-17]
2. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]

3. Imison C, Castle-Clarke S, Watson R, Edwards N. The Nuffield Trust. 2016. Delivering the Benefits of Digital Health Care URL: https://www.nuffieldtrust.org.uk/files/2017-01/delivering-the-benefits-of-digital-technology-web-final.pdf [accessed 2019-07-10]
4. Chung AE, Sandler RS, Long MD, Ahrens S, Burris JL, Martin CF, et al. Harnessing person-generated health data to accelerate patient-centered outcomes research: the crohn's and colitis foundation of America PCORnet patient powered research network (CCFA partners). J Am Med Inform Assoc 2016 May;23(3):485-490 [FREE Full text] [doi: 10.1093/jamia/ocv191] [Medline: 26911821]
5. McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 May 23;5:CD011425 [FREE Full text] [doi: 10.1002/14651858.CD011425.pub2] [Medline: 28535331]
6. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The asthma mobile health study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
7. The NHS Long Term Plan. 2019. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2019-08-19]
8. 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):62-9.e1. [doi: 10.1016/j.jneb.2017.06.004] [Medline: 29325664]
9. Berkowitz CM, Zullig LL, Koontz BF, Smith SK. Prescribing an app? Oncology providers' views on mobile health apps for cancer care. JCO Clin Cancer Inform 2017 Nov;1:1-7 [FREE Full text] [doi: 10.1200/CCI.17.00107] [Medline: 30657404]
10. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and facilitators to ehealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 5;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
11. Mayer MA, Blanco OR, Torrejon A. Use of health apps by nurses for professional purposes: web-based survey study. JMIR Mhealth Uhealth 2019 Nov 1;7(11):e15195 [FREE Full text] [doi: 10.2196/15195] [Medline: 31682587]
12. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
13. El Amrani L, Engberink AO, Ninot G, Hayot M, Carbonnel F. Connected health devices for health care in French general medicine practice: cross-sectional study. JMIR Mhealth Uhealth 2017 Dec 21;5(12):e193 [FREE Full text] [doi: 10.2196/mhealth.7427] [Medline: 29269336]
14. Slevin P, Kessie T, Cullen J, Butler MW, Donnelly SC, Caulfield B. Exploring the barriers and facilitators for the use of digital health technologies for the management of COPD: a qualitative study of clinician perceptions. QJM 2020 Mar 1;113(3):163-172. [doi: 10.1093/qjmed/hcz241] [Medline: 31545374]
15. Fagerlund AJ, Holm IM, Zanaboni P. General practitioners' perceptions towards the use of digital health services for citizens in primary care: a qualitative interview study. BMJ Open 2019 May 5;9(5):e028251 [FREE Full text] [doi: 10.1136/bmjopen-2018-028251] [Medline: 31061056]
16. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018 Dec 4;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
17. Mutebi I, Devroey D. Perceptions on mobile health in the primary healthcare setting in Belgium. Mhealth 2018;4:44 [FREE Full text] [doi: 10.21037/mhealth.2018.09.01] [Medline: 30363764]
18. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 3;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
19. Leigh S, Ashall-Payne L. The role of health-care providers in mhealth adoption. Lancet Digit Health 2019 Jun;1(2):e58-e59. [doi: 10.1016/s2589-7500(19)30025-1]
20. Klocek A, Šmahelová M, Knapová L, Elavsky S. GPs' perspectives on ehealth use in the Czech Republic: a cross-sectional mixed-design survey study. BJGP Open 2019 Oct;3(3):- [FREE Full text] [doi: 10.3399/bjgpopen19X101655] [Medline: 31344683]
21. Alwashmi MF, Fitzpatrick B, Davis E, Gamble J, Farrell J, Hawboldt J. Perceptions of health care providers regarding a mobile health intervention to manage chronic obstructive pulmonary disease: qualitative study. JMIR Mhealth Uhealth 2019 Jun 10;7(6):e13950 [FREE Full text] [doi: 10.2196/13950] [Medline: 31199330]
22. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value Health 2016 Jun;19(4):300-315 [FREE Full text] [doi: 10.1016/j.jval.2016.04.004] [Medline: 27325321]
23. Tinelli M, Ryan M, Bond C. What, who and when? Incorporating a discrete choice experiment into an economic evaluation. Health Econ Rev 2016 Dec;6(1):31 [FREE Full text] [doi: 10.1186/s13561-016-0108-4] [Medline: 27472943]
24. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 2019 Feb;37(2):201-226 [FREE Full text] [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]

25. Louviere J, Hensher D, Swait J. Stated Choice Methods: Analysis and Applications. Cambridge, UK: Cambridge University Press; 2000.
26. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. Patient 2015 Oct;8(5):373-384 [FREE Full text] [doi: 10.1007/s40271-015-0118-z] [Medline: 25726010]
27. The National Institute for Health and Care Excellence. 2019. Offer Digital CBT to Young People With Mild Depression, NICE Says URL: https://www.nice.org.uk/news/article/offer-digital-cbt-to-young-people-with-mild-depression-nice-says [accessed 2019-01-02]
28. Age UK Mobility. 2017. Useful Apps and Websites for Older People URL: https://www.ageukmobility.co.uk/mobility-news/article/useful-apps-and-websites-for-older-people [accessed 2019-01-01]
29. Burch S, Preston C, Bateup S, Hina F. The use of internet-enabled cognitive behavioral therapy in the treatment of depression and anxiety amongst older people. J Aging Soc Change 2017;8(1):1-11. [doi: 10.18848/2576-5310/cgp/v08i01/1-11]
30. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
31. Espie CA, Kyle SD, Williams C, Ong JC, Douglas NJ, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. Sleep 2012 Jun 1;35(6):769-781 [FREE Full text] [doi: 10.5665/sleep.1872] [Medline: 22654196]
32. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 1;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
33. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 4;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
34. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
35. Terry J, Davies A, Williams C, Tait S, Condon L. Improving the digital literacy competence of nursing and midwifery students: a qualitative study of the experiences of NICE student champions. Nurse Educ Pract 2019 Jan;34:192-198. [doi: 10.1016/j.nepr.2018.11.016] [Medline: 30579041]
36. National Institute for Health And Care Excellence. 2019. Evidence Standards Framework for Digital Health Technologies URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2020-05-07]
37. Nicholas J, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. Evid Based Ment Health 2016 May;19(2):33-34. [doi: 10.1136/eb-2015-102278] [Medline: 27044849]
38. Greaves F, Joshi I, Campbell M, Roberts S, Patel N, Powell J. What is an appropriate level of evidence for a digital health intervention? Lancet 2019 Dec 22;392(10165):2665-2667. [doi: 10.1016/S0140-6736(18)33129-5] [Medline: 30545779]

Abbreviations

DCE: discrete choice experiment
EPS: electronic prescription service
HCP: health care professional
mHealth: mobile health
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
ORCHA: Organisation for the Review of Care and Health Applications
RCT: randomized controlled trial
WTP: willingness-to-pay

Edited by G Eysenbach; submitted 06.01.20; peer-reviewed by O Byambasuren, S Wilhelm, O Harrison; comments to author 18.02.20; revised version received 10.04.20; accepted 26.04.20; published 06.07.20.

Please cite as:
Leigh S, Ashall-Payne L, Andrews T
Barriers and Facilitators to the Adoption of Mobile Health Among Health Care Professionals From the United Kingdom: Discrete Choice Experiment
JMIR Mhealth Uhealth 2020;8(7):e17704
URL: https://mhealth.jmir.org/2020/7/e17704
doi:10.2196/17704
PMID:32628118

Original Paper

Potential of Mobile Technology to Relieve the Urgent Mental Health Needs in China: Web-Based Survey

Yuxi Tan[1,2,3], MD; Ziwei Teng[1,2,3], MD; Yan Qiu[1,2,3], MD; Hui Tang[1,2,3], MD, PhD; Hui Xiang[1,2,3], MD, PhD; Jindong Chen[1,2,3], MD, PhD

[1]Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, China

[2]National Clinical Research Center for Mental Disorders, Changsha, China

[3]Institute of Mental Health, Central South University, Changsha, China

Corresponding Author:
Jindong Chen, MD, PhD
Department of Psychiatry
The Second Xiangya Hospital
Central South University
139 Renmin Middle Road
Changsha, 410011
China
Phone: 86 0731 85292158
Email: chenjindong@csu.edu.cn

Abstract

Background: With the rapid development of information technology and mobile devices, an increasing number of mobile medical services and platforms have emerged. However, China's current mental health situation necessitates further discussion and research on how to provide more patient-centered services in the face of many challenges and opportunities.

Objective: This study aims to explore the attitudes and preferences of mental health service stakeholders regarding mobile mental health services and discuss the challenges and opportunities faced by mobile technology developers in China.

Methods: A web-based survey was conducted by following the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist. A total of 586 valid questionnaires were collected. Respondents included 184 patients or their family members, 225 mental health professionals, and 177 people from the general population. Data analysis was completed using SPSS 24.0.

Results: Among the various problems perceived regarding the current mental health medical environment, difficulty in finding appropriate psychologists and limited visit times ranked highest. Social media (n=380/586, 64.9%) was the most preferred platform among all participants, whereas professionals showed a higher preference for smartphone apps (n=169/225, 75.1%). Professional instruction, psychological consultation, and mental health education (ranked top 3) were the most commonly identified needs. Mental health professionals generally emphasized more on treatment-related mobile mental health service needs, especially medication reminders (χ^2_2=70.7; P<.001), symptom monitoring (χ^2_2=24.0; P<.001), and access to mental health resources (χ^2_2=38.6; P<.001). However, patients and their family members focused more on convenient web-based prescriptions (χ^2_2=7.7; P=.02), with the general population interested in web-based psychological consultation (χ^2_2=23.1; P<.001) and mental health knowledge (χ^2_2=9.1; P=.01). Almost half of the participants regarded mobile mental health services as highly acceptable or supported their use, but less than 30% of participants thought mobile mental health services might be very helpful. Concerns about mobile mental health mainly focused on information security. Service receivers also suspected the quality and professionalism of content, and mental health professionals were worried about time and energy consumption as well as medical safety.

Conclusions: In terms of service flow, mobile services could be used to expand service time and improve efficiency before and after diagnosis. More individualized mobile mental health service content in more acceptable forms should be developed to meet the various needs of different mental health stakeholders. Multidisciplinary training and communication could be incorporated to facilitate the integration and cooperation of more well-rounded service teams. A standard medical record system and data format would better promote the development of future intelligent medical care. Issues such as ensuring service quality, solving safety risks, and better integrating mobile services with regular medical workflows also need to be addressed.

(JMIR Mhealth Uhealth 2020;8(7):e16215) doi:10.2196/16215

KEYWORDS

mental health; mobile health; mobile phone; mobile app; needs

Introduction

Background

Of the 1.39 billion people in China, more than 16 million are affected by severe mental disorders [1]. The number of people who require professional intervention for any mental disorder is even higher, estimated at 230 million [1], that is, approximately 1 in 8 people in China have psychological problems and would likely benefit from psychological counseling. The mental health system in China is also facing many challenges, such as a low service utilization rate, unevaluated treatment effectiveness, limited access, and unbalanced allocation of resources [2].

There has been an explosive growth of app markets worldwide since 2015, with Android and iOS platforms releasing more than 165,000 health-related apps, approximately 7% of which target mental health issues [3]. The worldwide emergence of mobile apps in the field of mental health can be divided into 4 categories: evaluation, tracking or monitoring, treatment, and multitargets [4]. Many studies have reported the benefits of mobile mental health apps, such as improving symptoms and reducing recurrence [5]. However, there are still a number of problems related to service quality, data privacy, interoperability, etc. Stronger cooperation with medical institutions and more rigorously validated evidence regarding its effectiveness in addressing significant clinical symptoms are also necessary [6]. More research is needed to promote the formulation of standards and suggestions for better mobile mental health services, especially in the large untapped market of China.

In April 2018, the general office of the State Council of China promulgated the *Opinions on Promoting The Development of Internet + Medical Health*, and encouraged medical institutions to expand the space and content of medical services using the internet and other advanced information technologies. To improve the efficiency of medical services, this policy aims to build a smoother service process that covers prevention, diagnosis, and recovery while integrating web-based and offline medical services and to explore typical mobile health demonstrations based on artificial intelligence (AI) and intelligent health equipment that could realize real-time monitoring and evaluation of personal health conditions, such as chronic disease screening, early warning, and proactive intervention [7]. In December 2018, 10 ministries and committees of China jointly issued the *National Pilot Program for the Construction of a Psychosocial Service System*, which explicitly calls for psychological service networks and assistant service platforms for the whole society and encourages the use of information technologies, such as hotlines, networks, apps, and public accounts on social media platforms, to carry out science publicity and mental health assessment of common mental disorders, including depression and anxiety [8].

Objectives

As mentioned, the use of advanced devices and technologies to provide more convenient mental and psychological services are not only urgent needs of the government and the public, but also inevitable trends of social development. However, whether the mobile services provided by enterprises, technology companies, or medical institutions can truly fulfill the needs of patients, family members, physicians, and the general population remains to be determined. Current studies on attitude or acceptance are mainly from countries with advanced electronic health (eHealth) experience, such as the United States, Australia, or Canada, which cannot represent the situation in China [9]. This study aimed to explore the needs and preferences for mobile mental health services of different stakeholders, including patients and family members, mental health professionals, and the general population, and to discuss challenges and opportunities for better mental health services based on China's mental health situation.

Methods

Questionnaire Survey

An open web-based survey was designed following the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist [10], with an expert panel consisting of 4 clinical psychiatry professionals (HT, HX, YQ, and JC), and 2 academic mental health researchers (ZT and YT). The study was approved by the Institutional Review Board of the Second Xiangya Hospital, Central South University. The web-based survey was developed and tested by a well-known web-based survey tool called *Survey Stars*, which generated Quick Response (QR) codes and links to the questionnaire for distribution. Two versions of the questionnaire could be generated by the first appearing question asking what the participant's role is: 1 questionnaire comprising 14 items targeted patients, family members, and the general population and 1 questionnaire comprising 12 items targeted mental health professionals. Each questionnaire started with a request for informed consent, a brief introduction of the researchers and aims, a notice of less than 5-min time consumption for participation, and information regarding privacy, which stated that no identifiable personal data would be collected and that all data would be under strict protection and accessible only to the researchers. The content of the questionnaire was designed based on existing mobile mental health services in China and the designers' clinical experience, which consisted of 3 parts: (1) demographic characteristics, including gender, age, marital status, education, and economic status of the patients, family members, and the general population; mental illnesses that the patients, family members, and the general population focus on or are interested in; sex, age, and job title of mental health professionals; mobile device usage frequency and attitude toward current mental health services for all participants; (2) preference for web-based service platforms and service categories; and (3) perceived acceptance

and helpfulness of, willingness to pay for or provide, and concerns about mobile mental health care (Multimedia Appendix 1). Single-choice, multiple-choice, and ranking questions were designed to fit the content. Ranking questions were automatically presented in a random order in the questionnaire to prevent bias.

Participants and Procedure

With the aim of understanding the differences in needs and preferences between different stakeholders to help bridge discrepancies and increase acceptance for future mobile services, the sample was divided into 2 major groups: service providers and service demanders. Service providers consisted of mental health professionals, who are the most direct providers and are considered to have some authority and knowledge concerning what kinds of services are necessary and beneficial. Service demanders were divided into 2 types: (1) patients with mental illness from mental health institutes and their family members; and (2) individuals in the general population. Recent epidemiological data suggest that most mental disorders have become increasingly common across China in the past 30 years [1]. However, less than 7% of patients with mental disorders have pursued mental health care, which indicates that there are a multitude of unmet mental health needs among the general population and suggests the necessity to research their potential needs and preferences [11]. On the other hand, patients in contact with mental health care generally have severe symptoms and represent the largest proportion of mental health service demanders [11]. For the above reason, the attitudes of different service demanders were collected separately, with the assumption that there may be possible differences in needs and preferences between patients who have identified the need to see psychiatric doctors and the general population that might also have unidentified mental health needs.

Participants were recruited voluntarily, and the questionnaire was accessible to all visitors who opened the link or scanned the QR code. Incomplete responses and multiple entries (via Internet Protocol Address) were automatically checked by the *Survey Star* website to guarantee valid responses. A total of 586 completed questionnaires with the correct participation time were collected in July 2018 from the most popular social media platform in China, WeChat. The respondents included 184 patients and family members, 225 mental health professionals, and 177 people from the general population. Data from mental health professionals and the general population were mainly collected by sharing internet links on the authors' WeChat. Data from patients and family members were mainly collected at a psychiatric outpatient department of a large general hospital through researchers' face-to-face invitation to scan the QR code and fill out the questionnaire after participants demonstrated a willingness to participate and gave consent. A primary analysis demonstrated very similar results between patients and family members; thus, further analysis was conducted with these 2 groups of data merged as one.

Statistical Analysis

Data analyses were performed using IBM SPSS version 24.0 (IBM Corp). Descriptive statistics were used to describe the study sample. Chi-square tests for single- and multiple-choice questions were applied to analyze whether the reported differences between the 3 groups were significant. Independent sample median tests were used for the preference ranking questions.

Results

Demographic Characteristics and Mobile Product Use

The basic information of all participants is presented in Tables 1 and 2. The general population tended to have a high proportion of women, young adults, higher education, good financial conditions, and expressed interest in common mental health issues, such as depression, anxiety, obsessive-compulsive disorder, and sleep disorders, whereas patients and family members were more concerned about severe illnesses, such as schizophrenia and mood disorders.

With regard to mobile device use, 83.2% (n=133/184) of patients and family members, 92.4% (n=208/225) of mental health professionals, and 91.5% (n=162/177) of respondents from the general population reported high or relatively frequent use. A significant difference of $P<.001$ between patients and family members and the 2 other groups was detected. The results are shown in Table 3.

Table 1. Demographic characteristics of the patients, family members, and the general population (N=361).

Characteristic	Patients and family members (n=184), n (%)	General population (n=177), n (%)	Chi-square (*df*)	*P* value
Gender			**7.5 (1)**	**.006**
Male	83 (45.1)	55 (31.1)		
Female	101 (54.9)	122 (68.9)		
Age (years)			**21.1 (6)**	**.001**
<18	9 (4.9)	1 (0.6)		
18-25	33 (17.9)	51 (28.8)		
26-30	48 (26.1)	62 (35.0)		
31-40	40 (21.7)	33 (18.6)		
41-50	39 (21.2)	18 (10.2)		
51-60	14 (7.6)	11 (6.2)		
>60	1 (0.5)	1 (0.6)		
Marital status			**7.0 (3)**	**.06**
Single	76 (41.3)	94 (53.1)		
Married or living with partner	102 (55.4)	74 (41.8)		
Divorced	5 (2.7)	7 (4.0)		
Widowed	1 (0.5)	2 (1.1)		
Education level			**69.5 (5)**	**<.001**
Elementary school and below	3 (1.6)	0 (0.0)		
Junior high school	29 (15.8)	7 (4.0)		
High school	43 (23.4)	9 (5.1)		
Vocational school	40 (21.7)	26 (14.7)		
Bachelor degree	54 (29.3)	85 (48.0)		
Postgraduate and above	15 (8.2)	50 (28.2)		
Economic status			**29.4 (4)**	**<.001**
Very poor	22 (12.0)	6 (3.4)		
Fairly poor	30 (16.3)	7 (4.0)		
Moderate	106 (57.6)	138 (78.0)		
Fairly good	23 (12.5)	25 (14.1)		
Very good	3 (1.6)	1 (0.6)		
Disease focused/interested			**N/A[a]**	**N/A**
Schizophrenia	62 (33.7)	60 (33.9)		
Bipolar disorder	32 (17.4)	37 (20.9)		
Depressive disorder	29 (15.8)	128 (72.3)		
Anxiety disorder	19 (10.3)	113 (63.8)		
Obsessive-compulsive disorder	8 (4.3)	72 (40.7)		
Phobia	1 (0.5)	33 (18.6)		
Panic disorder	3 (1.6)	18 (10.2)		
Eating disorder	0 (0.0)	10 (5.6)		
Sleep disorder	13 (7.1)	86 (48.6)		
Paranoid mental disorder	3 (1.6)	38 (21.5)		
Schizoid affective disorder	3 (1.6)	16 (9.0)		

Characteristic	Patients and family members (n=184), n (%)	General population (n=177), n (%)	Chi-square (*df*)	*P* value
Mental disorder caused by epilepsy	1 (0.5)	13 (7.3)		
Mental retardation with mental disorders	1 (0.5)	15 (8.5)		
Other	9 (4.9)	8 (4.5)		

[a]Not applicable.

Table 2. Demographic characteristics of mental health professionals (n=225).

Characteristic	Mental health professionals, n (%); (n=225)
Gender	
Male	85 (37.8)
Female	140 (62.2)
Age (years)	
<18	0 (0.0)
18-25	22 (9.8)
26-30	53 (23.6)
31-40	82 (36.4)
41-50	44 (19.6)
51-60	23 (10.2)
>60	1 (0.4)
Job title	
Chief physician/chief nurse	17 (7.6)
Deputy chief physician/deputy chief nurse	25 (11.1)
Attending physician/supervisor nurse	64 (28.4)
Resident physician/nurse	57 (25.3)
Assistant physician/nurse	22 (9.8)
Medical related major student	28 (12.4)
Other	12 (5.3)

Table 3. Frequency of mobile device use and perceived problems of mental health services in China (N=586).

Characteristic	Patients and family members (n=184), n (%)	Mental health professionals (n=225), n (%)	General population (n=177), n (%)	Chi-square (*df*)	*P* value
Frequency of mobile device use				**33.8 (8)**	**<.001**
Very often	75 (40.8)	131 (58.2)	106 (59.9)		
Fairly often	78 (42.4)	77 (34.2)	56 (31.6)		
Moderate	24 (13.0)	13 (5.8)	15 (8.5)		
Not often	7 (3.8)	1 (0.4)	0 (0.0)		
Almost not use	0 (0.0)	3 (1.3)	0 (0.0)		
Perceived difficulties of mental health services in China				**142.2 (18)**	**<.001**
Difficult to find a suitable psychological counselor or institution	49 (26.6)	105 (46.7)	115 (65.0)		
Limited visit time with the doctor	80 (43.5)	99 (44.0)	84 (47.5)		
Heavy economic burden	67 (36.4)	84 (37.3)	64 (36.2)		
Heavy transportation burden for nonlocal patients	50 (27.2)	100 (44.4)	34 (19.2)		
Difficult to make an appointment to a fixed doctor	29 (15.8)	71 (31.6)	42 (23.7)		
Lack of simpler procedures for regular medicine purchase	40 (21.7)	77 (34.2)	23 (13.0)		
Extremely long waiting time	47 (25.5)	39 (17.3)	32 (18.1)		
Difficult to make appointments	46 (25.0)	35 (15.6)	28 (15.8)		
Other	8 (4.4)	16 (7.1)	4 (2.3)		
Preference for platforms				**83.0 (12)**	**<.001**
Social media (WeChat or QQ)	134 (72.8)	123 (54.7)	123 (69.5)		
Smartphone apps	78 (42.4)	169 (75.1)	104 (58.8)		
Text message	25 (13.6)	19 (8.4)	9 (5.1)		
Phone call	35 (19.0)	45 (20.0)	16 (9.0)		
Websites	21 (11.4)	36 (16.0)	30 (16.9)		
Other	5 (2.7)	6 (2.7)	5 (2.8)		

Perceived Problems With Mental Health Services

Regarding the current inconvenience of or the unmet needs for mental health services, the difficulty of finding a suitable psychologist (n=269/586, 45.9%) and the limited visit time with the doctor (n=263/586, 44.9%) were most frequently mentioned (χ^2_{18}=142.2; *P*<.001). Mental health professionals were more concerned about finding proper psychologists (n=105/225, 46.7%), the difficulty for patients from rural areas to visit the doctor (n=100/225, 44.4%), the limited visit time (n=99/225, 44.0%), the economic burden (n=84/225, 37.3%), the inconvenience of regular medicine purchase (n=77/225, 34.2%), and the difficulty of regularly seeing the same doctor (n=71/225, 31.6%). Patients and family members complained more about the limited visit time (n=80/184, 43.5%) and economic burden (n=67/184, 36.4%). Compared with the 2 other groups, they also expressed more dissatisfaction with the length of the waiting time (n=47/184, 25.5%) and the difficulty of making appointments (n=46/184, 25.0%). The general population emphasized psychological consultation (n=115/177, 65.0%), limited visit time in the clinic (n=84/177, 47.5%), and economic burden (n=64/177, 36.2%). The results are shown in Table 3.

Preference for Web-Based Service Platforms

Among the frequently used platforms for web-based health services in China, there are social media such as WeChat and QQ, smartphone apps, text messages, phone calls, and websites. Social media (n=380/586, 64.8%) and apps (n=351/586, 59.9%) were most commonly chosen by the respondents. A total of 72.8% (n=134/184) of patients and family members reported a willingness to use social media and 42.4% (n=78/184) for apps, whereas 54.7% (n=123/225) of mental health professionals were more interested in the former and 75.1% (n=169/225) in the latter; the figures were 69.5% (n=123/177) and 58.8% (n=104/177), respectively, for the general population. The results are shown in Table 3.

There were significant differences in the preference toward mainstream social media among the 3 groups of participants (χ^2_2=17.0; *P*<.001). The higher interest in specific apps among professionals was statistically significant compared with other

participants (χ^2_2=45.3; *P*<.001). With regard to more traditional means of communication, a higher proportion of patients and family members chose text messages compared with other groups (χ^2_2=8.1; *P*=.02), and the general population was less interested in using phone calls for mental health services (χ^2_2=10.5; *P*=.01). No significant difference in attitude toward websites was detected (χ^2_2=2.6; *P*=.28).

Preference for Categories of Mobile Mental Health Services

A multichoice ranking question was provided to determine whether there are differences among respondent groups regarding various categories of mobile mental health services. A total of 11 options were provided for ranking in the questionnaire including 9 categories of services, *others*, and not selected as the eleventh option. Table 4 presents the statistical differences of the priority given by the 3 groups for each category of mobile mental health service. Figure 1 provides a more visualized comparison of the median of each category's ranking among the 3 groups. Numbers on the X axis correspond to the categories listed in Table 4. A smaller ranking number in the table or figure indicates a higher priority. Among all categories, the need for web-based professional instruction, web-based psychological consultation, mental health knowledge, and guidance for a healthy lifestyle ranked high, on average.

Except for guidance for a healthy lifestyle, all other items were significantly different among the 3 groups. Mental health professionals generally emphasized many items, especially medication reminders (χ^2_2=70.7; *P*<.001), symptom monitoring (χ^2_2=24.0; *P*<.001), mental health resources (χ^2_2=38.6; *P*<.001), and peer support (χ^2_2=19.0; *P*<.001). The general population attached much more importance to web-based psychological consultation (χ^2_2=23.1; *P*<.001) and mental health knowledge (χ^2_2=9.1; *P*=.01) but less importance to other demands, whereas patients and family members tended to focus more on convenient web-based prescription (χ^2_2=7.7; *P*=.02).

Table 4. Preferences for categories of mobile mental health services among the 3 groups.

Categories	Ranking[a]			Chi-square (*df*)	*P* value
	25th percentile	Median	75th percentile		
Web-based professional instruction				**6.6 (2)**	**.04**
Patients and family members	2	5[b,c]	11		
Mental health professionals	2	4[b]	9		
General population	3	6[c]	11		
Web-based psychological consultation				**23.1 (2)**	**<.001**
Patients and family members	2	6[c]	11		
Mental health professionals	3	7[c]	11		
General population	2	3[b]	11		
Mental health knowledge				**9.1 (2)**	**.01**
Patients and family members	3	7[c]	11		
Mental health professionals	2	5[b]	8		
General population	2	4[b]	11		
Guidance for healthy lifestyle				**1.9 (2)**	**.38**
Patients and family members	3	6	11		
Mental health professionals	3	6	11		
General population	2	5	11		
Medication reminder and side effects monitoring				**70.7 (2)**	**<.001**
Patients and family members	3	7[c]	11		
Mental health professionals	2	4[b]	7		
General population	6	11[d]	11		
Symptom monitoring				**24.0 (2)**	**<.001**
Patients and family members	3	9[c]	11		
Mental health professionals	3	5[b]	9		
General population	4	9[c]	11		
Collection of mental health resources				**38.6 (2)**	**<.001**
Patients and family members	5	11[c]	11		
Mental health professionals	3	6[b]	11		
General population	5	11[c]	11		
Web-based prescription				**7.7 (2)**	**.02**
Patients and family members	3	7[b]	11		
Mental health professionals	4	9[b]	11		
General population	5	11[c]	11		
Web-based peer support				**19.0 (2)**	**<.001**
Patients and family members	5	11[c]	11		
Mental health professionals	5	8[b]	11		
General population	5	11[c]	11		

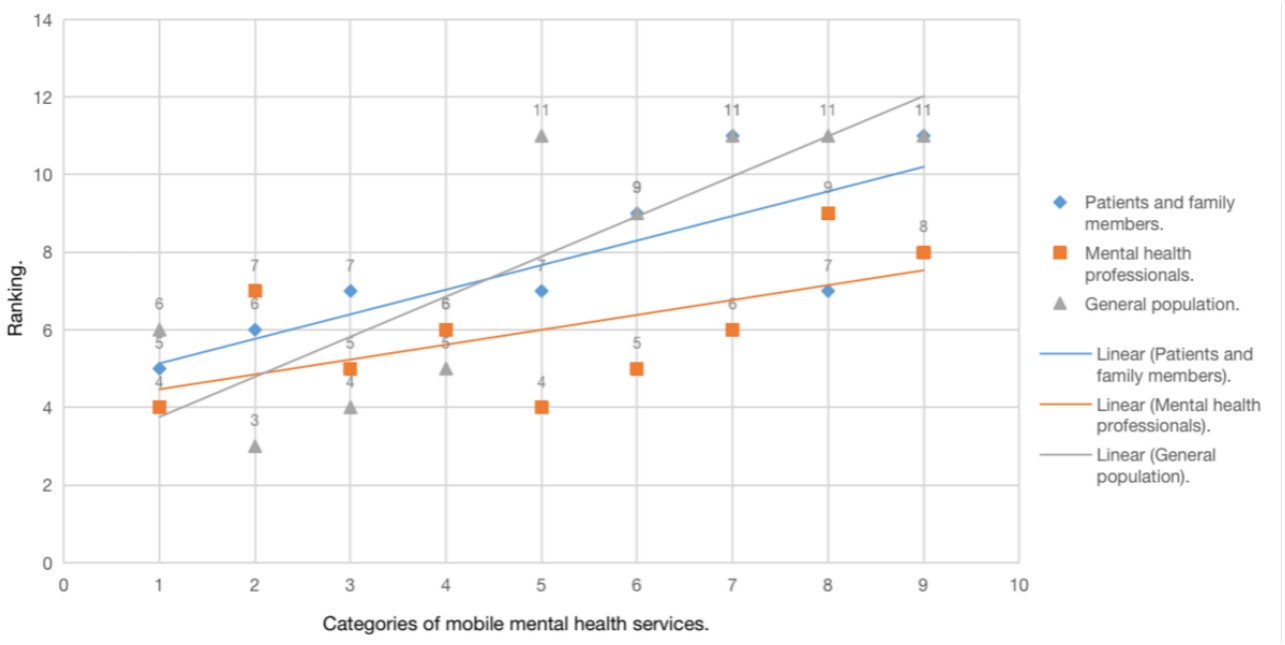
[a]The presence of b, c, and d of the median column indicates whether there are statistically significant differences in the median ranking among the 3 groups in each category. The same letter indicates no significant difference between groups; otherwise, there are significant differences. No b, c, or d means no significant difference was detected among all 3 groups.

[b]No significant differences between groups for these numbers.

[c]No significant differences between groups for these numbers.

[d]No significant differences between groups for these numbers.

Figure 1. Median comparison of the ranking of different categories of mobile mental health services among the 3 groups. The X-axis: 1. Web-based professional instruction; 2. Web-based psychological consultation; 3. Mental health knowledge; 4. Guidance for healthy lifestyle; 5. Medication reminder and side effects monitoring; 6. Symptom monitoring; 7. Collection of mental health resources; 8. Web-based prescription; 9.Web-based peer support. The Y-axis represents the median ranking of each category of the 3 groups.



Acceptability and Concerns

Tables 5 and 6 show the acceptance and concerns regarding mobile mental health services. Nearly half of the participants regarded mobile mental health services as highly acceptable or recommended their use. However, in regard to perceived helpfulness, <30% of participants thought mobile mental health services might be very helpful, and almost half of them considered such services fairly helpful.

Most service receivers had an above moderate level of willingness to pay for mobile mental health services. The risk of leaked personal information and the professionalism of web-based services were frequently chosen as concerns with regard to these web-based services.

A total of 76.0% (n=171/225) of mental health professionals believed that mobile mental health services can be fairly or very beneficial to their clinical work. Given proper payment, most mental health professionals showed willingness to provide these services. However, almost half of the mental health professionals were concerned about the potential disadvantages of web-based services that were listed on the questionnaire, especially the danger of leaked personal information, insufficient time and energy investment, and medical safety.

Table 5. Acceptance of and concerns regarding web-based services of patients, family members, and the general population (n=361).

Characteristic	Patients and family members (n=184), n (%)	General population (n=177), n (%)
Acceptance		
Very willing to accept	89 (48.4)	74 (41.8)
Fairly willing to accept	50 (27.2)	53 (29.9)
Indifferent	38 (20.7)	42 (23.7)
Tend to not accept	7 (3.8)	7 (4.0)
Not accept	0 (0.0)	1 (0.6)
Perceived helpfulness		
Very helpful	50 (27.2)	46 (26.0)
Fairly helpful	88 (47.8)	84 (47.5)
Indifferent	39 (21.2)	41 (23.2)
Tend to not be helpful	7 (3.8)	5 (2.8)
Cannot be helpful	0 (0.0)	1 (0.6)
Willingness to pay		
Very willing	33 (17.9)	29 (16.4)
Fairly willing	64 (34.8)	60 (33.9)
Indifferent	68 (37.0)	64 (36.2)
Tend to not be willing	18 (9.8)	19 (10.7)
Not willing	1 (0.5)	5 (2.8)
Concerns		
Leakage of personal information	105 (57.1)	127 (71.8)
Difficult to learn on mobile devices	60 (32.6)	77 (43.5)
Difficult to keep recording and learning	66 (35.9)	84 (47.5)
Not able to truly solve my problems	86 (46.7)	82 (46.3)
Not professional or authoritative enough	71 (38.6)	98 (55.4)
Other	6 (3.3)	7 (4.0)

Table 6. Acceptance of and concerns with web-based services among mental health professionals (n=225).

Characteristic	Mental health professionals (n=225), n (%)
Support to use	
Very support	103 (45.8)
Fairly support	69 (30.7)
Indifferent	32 (14.2)
Tend to not support	15 (6.7)
Not support	6 (2.7)
Helpfulness to patients and family members	
Very helpful	65 (28.9)
Fairly helpful	104 (46.2)
Neither helpful nor unhelpful	51 (22.7)
Tend to not be helpful	3 (1.3)
Cannot be helpful	2 (0.9)
Helpfulness to work	
Very helpful	67 (29.8)
Fairly helpful	104 (46.2)
Neither helpful nor unhelpful	49 (21.8)
Tend to not be helpful	3 (1.3)
Cannot be helpful	2 (0.9)
Willingness to provide with payment	
Very willing	80 (35.6)
Fairly willing	62 (27.6)
Indifferent	72 (32.0)
Tend to not be willing	8 (3.6)
Not willing	3 (1.3)
Concerns	
Leakage of personal information	125 (55.6)
Increased workload	105 (46.7)
Not enough time and energy	145 (64.4)
Medical safety is not guaranteed	132 (58.7)
Unnecessary disturbance	107 (47.6)
Other	4 (1.8)

Discussion

What service providers prioritize may not be urgent from the perspective of service receivers. However, few studies have shed light on the needs concerning electronic-mental health (e-mental health) services from different stakeholders' perspectives. Considering the current status of mental health in China and the fast-emerging eHealth trend worldwide, in-depth discussion is needed regarding the underlying contradiction between supply and demand in the current Chinese mental health services as well as the opportunities to solve these challenges and provide better patient-centered services through eHealth technologies.

Acceptance and Preferences

In comparison with previous studies on acceptance and preferred mobile mental health services, the findings of this study shared some similarities, as expected. First, e-mental health services have been proven to be generally acceptable with a high willingness to use, even though they are not perceived to be as helpful as face-to-face services [12]. The results of this survey also indicate a high willingness to try such services alongside a relatively prudent expectation of e-mental health service helpfulness. Second, professional support has proved to be the most critical facilitator of user acceptability because both face-to-face and professionally guided web-based services are preferred over unguided self-help services [12]. The results of

this study also show the high demand for services involving professional participation, such as professional medical instructions and psychological counseling.

This study adds to the body of literature in a few substantial ways. First, some studies have shown that low e-mental health awareness (e-awareness) and the digital divide of patients with severe mental illness might hinder the future use and perceived helpfulness of mobile services [13,14]. However, other studies have demonstrated that the internet is a significant resource for seekers of mental health knowledge and that the internet behavior of patients with psychosis is not very different from that of the general population [15]. The results of this survey supported the latter finding. The frequency of mobile device usage by patients and family members in this study was lower than that in the other 2 groups, but the respondents indicating above-average usage still accounted for 83.2%, and no significant difference in acceptance or willingness to use was found. There is still much space for developers to explore how to take full advantage of the cyber world and create more acceptable services for mental health seekers, regardless of the severity of their illness.

Second, previous studies have found that the vast majority of respondents preferred traditional face-to-face services to mobile mental health services, which put traditional medical services and mobile medical services on opposing sides [16]. However, the advantage of mobile medical services lies in their ability to assist and optimize traditional services rather than replace them. According to the reflection on current medical problems in China from this study, there are still many unmet needs, especially in terms of accessibility, service time, economic burden, and distance to institutions. This indicates that mobile mental health services have great potential to solve these medical needs that cannot be met at present, such as by offering more professional medical instruction and health education. Therefore, mobile health services need not be considered in opposition to traditional medical treatment but as advantageous to complement existing mental health services.

Third, to our knowledge, this is the first study on the preference for different types of mobile medical services and the first to compare among different stakeholders. Therefore, it is difficult to compare our findings with those arising from other countries' situations, but the results of this study still provide valuable indications for the future development of mobile mental health services. This survey showed that mental health professionals generally place more emphasis on symptoms and treatment, such as monitoring and access provision, whereas service demanders are more focused on psychological counseling and convenient prescription. The emphasis of professionals on medical issues might be due to their tendency to diagnose and provide treatment plans based on symptoms, whereas psychological counseling services are not commonly integrated with current mental health institutions in China [2]. The higher requirements of patients and their family members for convenient prescription might be the result of the imperfect hierarchical diagnosis and treatment system in China [2], which leads to large provincial medical institutions often being crowded and inconvenient for rural patients to access. The obvious lower requirements of service demanders for monitoring and access provision could be due to the low adherence to medication, as shown in other research studies [17], but might also be due to the fact that service demanders believe they can monitor their symptoms and medications without additional mobile tools and instead give higher priority to items that they cannot accomplish alone, such as professional instruction and counseling. In short, this result reflects that the functions professionals think are important for mental health might not be considered of the same importance by patients and family members. This indicates that service developers should pay more attention to the different needs of different stakeholders and try to meet these differentiated needs in a more acceptable manner.

Challenges

The urgent needs for professional instruction and psychological therapy seems to be intrinsic for patients and family members with mental illnesses, but many underlying problems in the current Chinese mental health environment provide no simple solution to these seemingly simple needs. eHealth services seem to be a promising solution, but there are some systemic challenges to be solved that are critical for their application and further development.

First, mental health human resources in China are limited. Due to the lack of social workers and family physicians and the imperfect hierarchical diagnosis and treatment system, patients are used to crowding into first-class hospitals for treatment. However, limited human resources and heavy work burdens make it difficult for physicians and nurses in hospitals to provide personalized and continuous disease management, which leads to the urgent need for more professional instruction, as shown in this study [2].

Telemedicine services seem to be a promising solution to help patients who lack convenient access to local services to obtain high-quality mental health services. The authors conducted a rough search on 2 App Store optimization websites, *Kuchuan* [18], and *Chandashi* [19] to obtain the number of recent mobile mental health apps available at the time of writing this paper. The results shown there are 222 of these apps in the App Store and 621 in the Android markets on *psychology (心理 in Chinese)*, and 212 and 121,560, respectively, on *psychological counseling (心理咨询 in Chinese)*. These results are in accordance with an industry report, where most of the high-ranked mental health service platforms provide web-based counseling [20]. These platforms continue to gather well-known psychiatrists and psychologists as much as possible to improve their popularity and professionalism [20]. There are also clinical trials demonstrating the effectiveness of mental health services provided on social media, such as WeChat [21].

However, there are several challenges that require further attention. On the one hand, the time consumption and workload disruption might hinder the sustainable willingness of more professionals to provide such services, as Granja and Janssen noted [22], which was also shown in this study as a major concern of professionals. On the other hand, the current lack of systematic training and standardized entry threshold for psychological counselors in China also leads to a lack of professional psychological counselors, which consequently

makes psychological therapy an urgent need for the general population [2]. Additionally, there is no explicit regulation on the pricing of the psychological counseling industry in China. Current web-based psychological consultation fees are mostly hundreds of yuan, some even more than 2000 yuan (or approximately US $285) per hour, and none of these services are currently covered by medical insurance [23]. As a newly emerging industry, the medical and ethical safety of the web-based services provided by commercial companies is also awaiting further perfection of relevant laws and regulations. As a result, it is still doubtful whether a high-threshold unguaranteed psychological counseling can truly alleviate so many patients' urgent needs.

Second, professional mental health publicity and education resources for the general population are insufficient. The various kinds of mental health information that flood the internet are not necessarily effective or safe. First, nonindividualized health education information is not as helpful as what personal physicians or counselors can provide for patients to solve their mental health problems [24]. Second, some negative unregulated psychology websites may completely misguide patients and even lead to death [25]. Service providers need to come up with more individualized and authorized health education, which can not only guarantee the effectiveness of information but also help information seekers avoid being inundated and misled by the massive amount of information.

Third, the inconformity of medical information impedes the development of medical big data. For a long time, the inconsistency of data standards and the irregularity of data formats have been important factors hindering the development of big data in China [26]. Medical systems in different regions and hospitals have different standards and formats in China, resulting in many data deficiencies and great difficulties in data cleaning. These information islands also hinder physicians from gathering information about the medical history of patients, which consequently affects treatment effectiveness. In addition, the lack of records of symptoms and medication use from daily life, such as mood fluctuations of patients with bipolar disorder and metabolic indicators of patients with schizophrenia, make it difficult to understand the course of patients' disease in the short visit time with inconsistent information [27]. These are also urgent needs for professionals, as shown in this study. Although there have been many mobile apps aiming at offering ecological momentary assessment (EMA), which monitors adverse events and might help physicians understand the course of the disease, whether patients will accept such apps and how to guarantee adherence are problems that need to be resolved [27]. Nonetheless, if information from this monitoring system is summarized and reported to the health information systems in medical institutes in a unified and continuous manner, it might greatly facilitate diagnosis, treatment adherence, and rehabilitation.

Opportunities

The survey showed that all groups generally have high acceptance for mobile mental health services, with different preferences. However, the application of mobile mental health in China still faces many challenges along with great opportunities to improve the current service flow, service content, and service quality. The following are some suggestions for the application of mobile mental health in China based on the worldwide trends of mobile health and China's national conditions.

Optimize Service Flow for Different Stakeholders and Different Needs

To reduce the threshold and obstacles for the public to positively seek and accept mental health services, professional mobile platforms could be equipped to establish more coherent service flows. For the general population, there could be platforms providing mental health education, early symptom screening, self-help psychological intervention, and smooth referral services to professional psychological counseling or psychiatric institutions. For patients with severe disorders who need more systematic diagnosis and treatment, mobile devices can be used to provide pretreatment guidance before a formal visit by asking patients to fill in disease-related information in advance, which could not only provide more referential information for physicians in the very limited outpatient visit time but also expand the content of service and enhance the sense of being cared for. After the first visit with a clear diagnosis, web-based prescription and drug delivery services after video conferences could be used to change repeated brief follow-ups to high-quality web-based instruction, with all the information gathered from previous visits and daily EMA data on mobile platforms. Transforming the repetitive and tedious routine through web-based services might help reduce the travel time and economic burden of patients and relieve the crowded environment in large medical institutions.

It should be emphasized that to ensure the acceptance and perceived usability among professionals, all of these mobile services should be optimized in the current mental health service mode without disturbing the regular workflow or increasing the burden of staff [20]. In addition, mobile mental health services should be integrated into people's daily life habits as much as possible through common social media such as WeChat or QQ, which are preferred platforms for service demanders. The preference of professionals for using specialized smartphone apps rather than social media might be due to privacy, energy, and time concerns, which calls for careful design of service platforms. For example, a special processing mechanism that allows service requests to be transferred to professionals only when the background robot or assistants are not capable of dealing with them can be established. With specialized methods for information receiving channels and processes, structured information gathering and convenient disease management could be ideally realized to meet both scientific and clinical research needs.

Provide Individualized Service Content in Acceptable Forms

Psychological Consultation

Psychological intervention is one of the services of most concern for the public, patients, and family members. Web-based psychological intervention services may be self-helped or professionally guided, synchronous, or asynchronous [28].

Recently, another promising method that combines psychological consultation with AI-supported chatbots has become popular, providing companionship, listening, interpersonal skills, cognitive correction, and behavioral activation, even though there is a long way before such chatbots become natural and humanized enough [29,30]. With limited accessible services in China, self-help services might be cost-efficient, but human-supported services may have higher acceptance and adherence [24]. One possible solution worth exploring might be to integrate the convenience of a chatbot with auxiliary human support from trained service assistants without creating more workload to the currently limited professional mental health human resources. In addition, there should be different forms of services for illnesses of different severities. For common mild psychological problems in the general population, priority could be given to easier access to self-help services, such as mindfulness meditation exercises, self-help cognitive behavioral therapy, and telepsychological counseling. What is important is that these platforms are regulated properly, and referral links to reliable mental health institutions should be provided. For patients with more severe problems, there should be more disease-specific or even individualized psychological therapy services that are provided by authorized organizations and closely linked with stepped mental health systems from the community to the hospital level. Technologies integrated with virtual reality (VR), augmented reality (AR), or serious games can also be used to provide more vivid therapy scenes, such as for patients with cognitive impairment or specific phobia.

Psychological Education

At present, there are many search engines on the internet that provide a variety of mental health knowledge, but most of them are delivered either as a professional reference or with unguaranteed quality, which might lead to inaccurate self-diagnosis [28]. Additionally, website information is usually provided passively by users searching for possible relevant content and does not target individuals, thus having limited potential to solve problems properly. Smartphone apps, emails, or text messages can be easily ignored, and may not be interesting enough to ensure adherence without further human interaction. More individualized and professional education methods should be emphasized. For the general population, educational knowledge related to a healthy lifestyle and mental health could be actively provided using public web-based spaces such as social media and official group chats or chatrooms that belong to schools, organizations, or institutions, to increase the dissemination of mental health knowledge and decrease the sensitivity to discussing such issues. Content combined with current news or interesting practical skills or provided in the form of short videos or games may be more acceptable. For patients with more health care requirements, mobile platforms provided by strictly selected mental health institutions can be used to provide systematic, individualized content for patients in different stages of disease.

Professional Instruction and Peer Support

The lack of mental health human resources and limited visit time at clinics in China leads to individualized professional instruction for disease-related questions being the most important requirement of patients and family members. A stepped health care system could be introduced to mental health institutions by integrating AI technology and human professionals. After learning and training, the AI robot could be used to complete the screening and identification work and automatically answer some simple professional questions, but as a guarantee, medical assistants could be equipped to solve more complex problems that the robot cannot address. Important decision-making problems, such as prescription or emergencies, including suicide or self-injury, can be referred to a physician. With the help of mobile technology and AI, the burden of physicians could be reduced, and more personalized services could be provided.

Peer support has been considered an important way of improving mental health that provides more perspectives and support for patients and family members. Mobile devices are widely used social tools and educational platforms for peer support in synchronous or asynchronous forms, but the activation of, commitment to, access to, and effectiveness of mobile peer support apps await further research [31,32]. The unfamiliarity of peer support and its benefit, deeply rooted stigma, and relatively limited authority might have caused the lower preference in this study, and more peer-based apps and risks need to be discussed.

Symptom Management and Service Information

Symptom monitoring and medication reminders provided by mobile tools were not as prioritized by patients and family members as they were by mental health professionals. Professionals might habitually pay more attention to symptoms and side effects and hope to have a more comprehensive understanding of the variation of medical indicators. However, patients might think these are redundant and even reminders of pain. A mutually beneficial approach is necessary. Motivational research has proven that external and intrinsic factors, such as the sense of accomplishment and control, increase of personal benefit, and formation of habits, could facilitate adherence to mobile services [33]. This reminds us that these monitoring functions could be delivered in more interesting, interactive, and rewarding forms, such as games, or integrated into the conversational context on a regular basis, rather than simply requiring patients to provide information and complete complex missions. A passive data collection function embedded in mobile phones or wearable devices could also be utilized to ensure service adherence and acceptance.

In the past 15 years, China's mental health services have greatly improved with the conduction of a national program named *Central Government Support for the Local Management and Treatment of Serious Mental Illness Project* or the *686 Program* that was initiated in 2004 [2]. Many financial subsidy policies have also been put forward for patients with severe mental diseases, especially for impoverished patients, to obtain affordable, even free treatment and rehabilitation services [2]. However, the accessibility of this service information is the key factor that determines whether patients can truly obtain help. In this study, service providers prioritized this, but service demanders did not. This might be because service demanders believe that they have the ability to freely and conveniently

access the internet and obtain all needed information. The internet helps patients gain relevant information quickly, but providing this information in more authoritative and integrated platforms might improve the efficiency of accessing high-quality medical services without being misled by advertisements or unrelated information in the broad cyber world.

Increase the Quantity of Mental Health Services

Considering the current divided training system between psychiatry and psychology in China [34], web-based continuing education could help provide more convenient multidisciplinary training in fields such as psychiatry, psychological therapy, social work, and vocational rehabilitation. Technological innovations, such as AR and VR, can also contribute to more vivid and realistic education scenarios and help increase learning efficiency. Multidisciplinary communication facilitated by mobile technologies could also be incorporated to facilitate the integration and cooperation of more well-rounded service teams. Another important aspect of quality improvement depends on the interconnection of various health data, especially medical records from different medical information systems. It is necessary to establish unified clinical and scientific data standards that are consistent with electronic medical records and other major health information systems to increase the efficiency of big data scientific research and promote the continuous development of intelligent medical services. In addition, mobile mental health services should be combined with medical records as much as possible to avoid information islands, help promote more coherent service processes, and increase adherence. With regard to the suspicions of service quality as reflected in this survey, it is vital to continuously research the acceptability, usability, and effectiveness of all fast-emerging types of exploration in this constantly evolving world of technology.

Limitations

There are some limitations to this study. First, respondents representing mental health professionals and the general population were collected from the social media accounts of the researchers, and the respondents representing patients and family members are mainly collected from one mental health department of a provincial hospital. The similar background of the participants limits this survey in gathering perspectives from community-level medical institutes or patients with limited resources or literacy. Multicentered research is needed for a more well-rounded understanding. Second, this study is conducted with the prerequisite that all participants have smartphones and use WeChat and thus cannot address the needs of people who do not frequently use smartphones or WeChat. However, with the rapidly increasing number of cyber citizens and the large number of WeChat users in China, understanding the needs of this part of the public first is more feasible and promising. Third, it was unpractical to conduct an assessment on cognitive ability before consent was obtained for a questionnaire, and there is a possibility that some patients did not fully understand the questionnaire items. However, as the results of patients and family members are very similar, we tend to believe that there is only a minor possible influence.

Conclusions

This study investigated the preferences, acceptance, and concerns regarding mobile mental health services of different mental health stakeholders and discussed the potential of mobile health services to relieve the urgent mental health needs of current China. Psychological therapy, professional instruction, and mental health education are most needed, which reflects the current problems of mental health services in China, such as the severely imbalanced supply and demand, doubtful and insufficient public education materials, incoherent mental health services flow, and the lack of standardization of medical information. However, challenges always accompany opportunities. With the high acceptability of mobile health services, mobile technologies have the potential to build a smoother mental health workflow, enrich urgently needed service categories, and improve the overall service quality. More practice and research are needed in the future to continuously maximize the advantages of technology while avoiding privacy risks and medical security issues.

Acknowledgments

The authors thank all participants for providing their perspectives in this study. They are also thankful for the assistance in language editing provided by American Journal experts, and further language editing by Katherine Cohen from Northwestern University.

This study was funded by the National Key R&D Program of China (2017YFC1311102) and Fundamental Research Funds for the Central Universities of Central South University (2019zzts361).

Authors' Contributions

YT drafted the questionnaire and manuscript and analyzed the data. ZT and YQ helped with the questionnaire and analysis of the data. HT and HX helped with the questionnaire and revised the manuscript. JC supervised and assisted with the entire process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey of preferences for mobile mental health services.

[DOCX File , 20 KB - mhealth_v8i7e16215_app1.docx]

References

1. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry 2019 Mar;6(3):211-224. [doi: 10.1016/S2215-0366(18)30511-X] [Medline: 30792114]
2. Liang D, Mays VM, Hwang W. Integrated mental health services in China: challenges and planning for the future. Health Policy Plan 2018 Jan 1;33(1):107-122 [FREE Full text] [doi: 10.1093/heapol/czx137] [Medline: 29040516]
3. Aitken M. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance. IQVIA: The Human Data Science Company . 2015. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf?la=en&hash=527005489AB6870674491DFEB6A22B03DB111D99 [accessed 2019-04-20]
4. van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
5. Ebert DD, van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. Eur Psychol 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
6. Frank E, Pong J, Asher Y, Soares CN. Smart phone technologies and ecological momentary data: is this the way forward on depression management and research? Curr Opin Psychiatry 2018 Jan;31(1):3-6. [doi: 10.1097/YCO.0000000000000382] [Medline: 29084010]
7. General Office of the State Council: Opinions on Promoting the Development of 'Internet + Medical Health'. Government of China. 2018. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2019-04-19]
8. Notice on Printing and Distributing the National Social Psychological Service System Pilot Work Plan. National Health Commission. 2018. URL: http://www.nhc.gov.cn/jkj/s5888/201812/f305fa5ec9794621882b8bebf1090ad9.shtml [accessed 2019-04-18]
9. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
10. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
11. Patel V, Xiao S, Chen H, Hanna F, Jotheeswaran AT, Luo D, et al. The magnitude of and health system responses to the mental health treatment gap in adults in India and China. Lancet 2016 Dec 17;388(10063):3074-3084. [doi: 10.1016/S0140-6736(16)00160-4] [Medline: 27209149]
12. Apolinário-Hagen J, Kemper J, Stürmer C. Public acceptability of e-mental health treatment services for psychological problems: a scoping review. JMIR Ment Health 2017 Apr 3;4(2):e10 [FREE Full text] [doi: 10.2196/mental.6186] [Medline: 28373153]
13. Apolinário-Hagen J, Vehreschild V, Alkoudmani RM. Current views and perspectives on e-mental health: an exploratory survey study for understanding public attitudes toward internet-based psychotherapy in Germany. JMIR Ment Health 2017 Feb 23;4(1):e8 [FREE Full text] [doi: 10.2196/mental.6375] [Medline: 28232298]
14. Berry N, Bucci S, Lobban F. Use of the internet and mobile phones for self-management of severe mental health problems: qualitative study of staff views. JMIR Ment Health 2017 Nov 1;4(4):e52 [FREE Full text] [doi: 10.2196/mental.8311] [Medline: 29092809]
15. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec 19;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
16. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
17. Geretsegger C, Pichler E, Gimpl K, Aichhorn W, Stelzig R, Grabher-Stoeffler G, et al. Non-adherence to psychotropic medication assessed by plasma level in newly admitted psychiatric patients: prevalence before acute admission. Psychiatry Clin Neurosci 2019 Apr;73(4):175-178 [FREE Full text] [doi: 10.1111/pcn.12809] [Medline: 30552718]
18. Kuchuan.: Beijing Kuchuan Technology Co., Ltd; 2020. URL: https://www.kuchuan.com/ [accessed 2020-05-22]
19. Chandashi.: Xiamen Limayao Network Technology Co, Ltd; 2020. URL: https://www.chandashi.com/ [accessed 2020-05-22]
20. Li Y. The Medical Consultation Rate in This 100 Billion Market is Less Than 10%. Can Internet + Psychological Services Allow Users to Pay? [Mental Health + Case Series Topics]. VC Beat. 2018. URL: https://vcbeat.top/ZWU1ZDc4MTEwYTkzMDY0NjFhNzYyMmZhNjNiM2Y3M2Q= [accessed 2019-04-18]
21. Wang XW, Nina. Analysis of application effect and compliance of self-management in depression patients based on the WeChat platform. China Health Industry 2018(2):177-178.
22. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]

23. Zhang A, Wang Z, Huang X. Current situation of fee of counseling and psychotherapy in China. Chin Ment Health J 2017;31(1):40-45 [FREE Full text] [doi: 10.3969/j.issn.1000?6729.2017.01.008]
24. Scholten MR, Kelders SM, van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Nov 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]
25. Sousa DF, Filho JD, Cavalcanti RD, Santos AB, Neto ML. The impact of the 'Blue Whale' game in the rates of suicide: short psychological analysis of the phenomenon. Int J Soc Psychiatry 2017 Dec;63(8):796-797. [doi: 10.1177/0020764017732595] [Medline: 28936914]
26. Xu P, Huang K. The status, problems and countermeasures of the big data of health care in China. Chin Digit Med 2017;12(5):24-26. [doi: 10.3969/j.issn.1673-7571.2017.05.008]
27. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From ehealth to ihealth: transition to participatory and personalized medicine in mental health. J Med Internet Res 2018 Jan 3;20(1):e2 [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]
28. Hilty DM, Chan S, Hwang T, Wong A, Bauer AM. Advances in mobile mental health: opportunities and implications for the spectrum of e-mental health services. Mhealth 2017;3:34 [FREE Full text] [doi: 10.21037/mhealth.2017.06.02] [Medline: 28894744]
29. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
30. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory Group. Can your phone be your therapist' young people's ethical perspectives on the use of fully automated conversational agents (ChatBots) in mental health support. Biomed Inform Insights 2019;11:1178222619829083 [FREE Full text] [doi: 10.1177/1178222619829083] [Medline: 30858710]
31. Thomas N, Bless JJ, Alderson-Day B, Bell IH, Cella M, Craig T, et al. Potential applications of digital technology in assessment, treatment, and self-help for hallucinations. Schizophr Bull 2019 Feb 1;45(45 Suppl 1):S32-S42 [FREE Full text] [doi: 10.1093/schbul/sby103] [Medline: 30715539]
32. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
33. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
34. Su X. The Incidence of Mental Illness in China is 17.5%. Only 30,000 Professional Doctors. China Mobile. 2017. URL: http://www.china.com.cn/guoqing/2017-05/04/content_40744090.htm [accessed 2019-04-18]

Abbreviations

AI: artificial intelligence
AR: augmented reality
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
eHealth: electronic health
e-mental health: electronic-mental health
e-awareness: e-mental health awareness
EMA: ecological momentary assessment
QR code: Quick Response code
VR: virtual reality

Edited by G Eysenbach; submitted 11.09.19; peer-reviewed by E Lattie, A Errazuriz; comments to author 03.10.19; revised version received 22.11.19; accepted 28.03.20; published 07.07.20.

Please cite as:
Tan Y, Teng Z, Qiu Y, Tang H, Xiang H, Chen J
Potential of Mobile Technology to Relieve the Urgent Mental Health Needs in China: Web-Based Survey
JMIR Mhealth Uhealth 2020;8(7):e16215
URL: https://mhealth.jmir.org/2020/7/e16215
doi:10.2196/16215
PMID:32673239

Original Paper

Application and Preliminary Outcomes of Remote Diagnosis and Treatment During the COVID-19 Outbreak: Retrospective Cohort Study

Luwen Liu[1], MD, PhD; Jianqin Gu[1], MD; Fengmin Shao[1], MD; Xinliang Liang[1], MD; Lixia Yue[1], MD; Qiaomei Cheng[1], MD; Lianzhong Zhang[1], MD, PhD

Zhengzhou University People's Hospital, Henan Provincial People's Hospital, Zhengzhou, China

Corresponding Author:
Lianzhong Zhang, MD, PhD
Zhengzhou University People's Hospital
Henan Provincial People's Hospital
No. 7 Weiwu Road
Zhengzhou
China
Phone: 86 13598896699
Email: llw0624@gs.zzu.edu.cn

Abstract

Background: The coronavirus disease (COVID-19) pandemic, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has resulted in the self-quarantine of countless people due to possible infection. This situation makes telemedicine necessary as it can overcome geographical barriers, increase the number of people served, and provide online clinical support for patients. However, the outcomes of telemedicine have not yet been evaluated.

Objective: The aim of our study is to describe the epidemiological features and clinical symptoms of patients receiving remote diagnosis and treatment at the online outpatient clinic of our hospital, as well as to analyze the outcomes and advantages of telemedicine, during the COVID-19 pandemic.

Methods: Data from patients receiving remote diagnosis and treatment via consultation services for COVID-19 concerns at the online outpatient clinic of Henan Provincial People's Hospital from January 24 to February 17, 2020, were collected. A retrospective analysis was performed on epidemiological features, clinical symptoms, and preliminary outcomes.

Results: Online inquiry, consultation, and suggestions were provided for patient concerns related to COVID-19. Our hospital also offered offline noncontact drug delivery services following online ordering and payment. A total of 4589 patients receiving remote diagnosis and treatment were recruited. The daily number of online outpatient visits initially increased and then decreased, reaching its peak on January 28 when the daily number of online outpatient visits totaled 612. Of 4589 patients, 1940 (42.3%) were males and 2649 (57.7%) were females (age range: 78 days to 85 years). Most patients were aged 20-39 years (n=3714, 80.9%) and came from Henan Province (n=3898, 84.9%). The number of patients from other provinces was 691 (15.1%). During the online consultations, patients discussed the following symptoms: fever (n=2383), cough (n=1740), nasal obstruction (n=794), fatigue (n=503), and diarrhea (n=276). A total of 873 orders of noncontact drug delivery following online payment was completed. The daily number of such orders gradually stabilized after the initial, steady increase. For offline drug delivery orders, the median (IQR) was 36 (58). An online satisfaction survey was filled out postconsultation by patients; of the 985 responses received, 98.1% (n=966) of respondents were satisfied with the service they received.

Conclusions: Remote diagnosis and treatment offered via online outpatient consultations effectively reduced the burden on hospitals, prevented overcrowding, reduced the risk of cross-infection, and relieved patients' anxiety during the COVID-19 outbreak. This plays an essential role in pandemic management.

(JMIR Mhealth Uhealth 2020;8(7):e19417) doi:10.2196/19417

KEYWORDS

coronavirus disease; COVID-19; remote diagnosis and treatment; telemedicine; online outpatient visit; offline drug delivery; pandemic management; China; Henan Province

Introduction

Novel coronavirus (2019-nCoV)–infected pneumonia is caused by the 2019 novel coronavirus [1]. On February 11, 2020, the World Health Organization (WHO) announced the names “coronavirus disease 2019” (COVID-19) for the disease and “severe acute respiratory syndrome coronavirus 2” (SARS-CoV-2) for the virus. According to existing studies, SARS-CoV-2 is closely related to severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS CoV). All three viruses can cause severe symptoms of pneumonia. SARS-CoV-2 is usually found in respiratory secretions and transmitted by droplets (eg, by sneezing and saliva) but can also spread by contact. Its virulence is weaker than SARS CoV, but its transmissibility is higher than SARS CoV. SARS-CoV-2 can spread from person to person, and the general population is susceptible to infection. It is also contagious during the latency period [2-5]. According to the National Health Commission of the People’s Republic of China, it had received reports of 72,528 cumulative confirmed cases of COVID-19 from 31 provinces (autonomous regions and municipalities directly under the central government) and Xinjiang Production and Construction Corps, by 24:00 on February 17, 2020. Other countries, including Thailand, Japan, Korea, and the United States had also reported COVID-19 cases [6-9]. It is difficult to monitor COVID-19 due to its high transmissibility, unclear route of transmission, and atypical symptoms. COVID-19 has posed enormous challenges to China's public health [2].

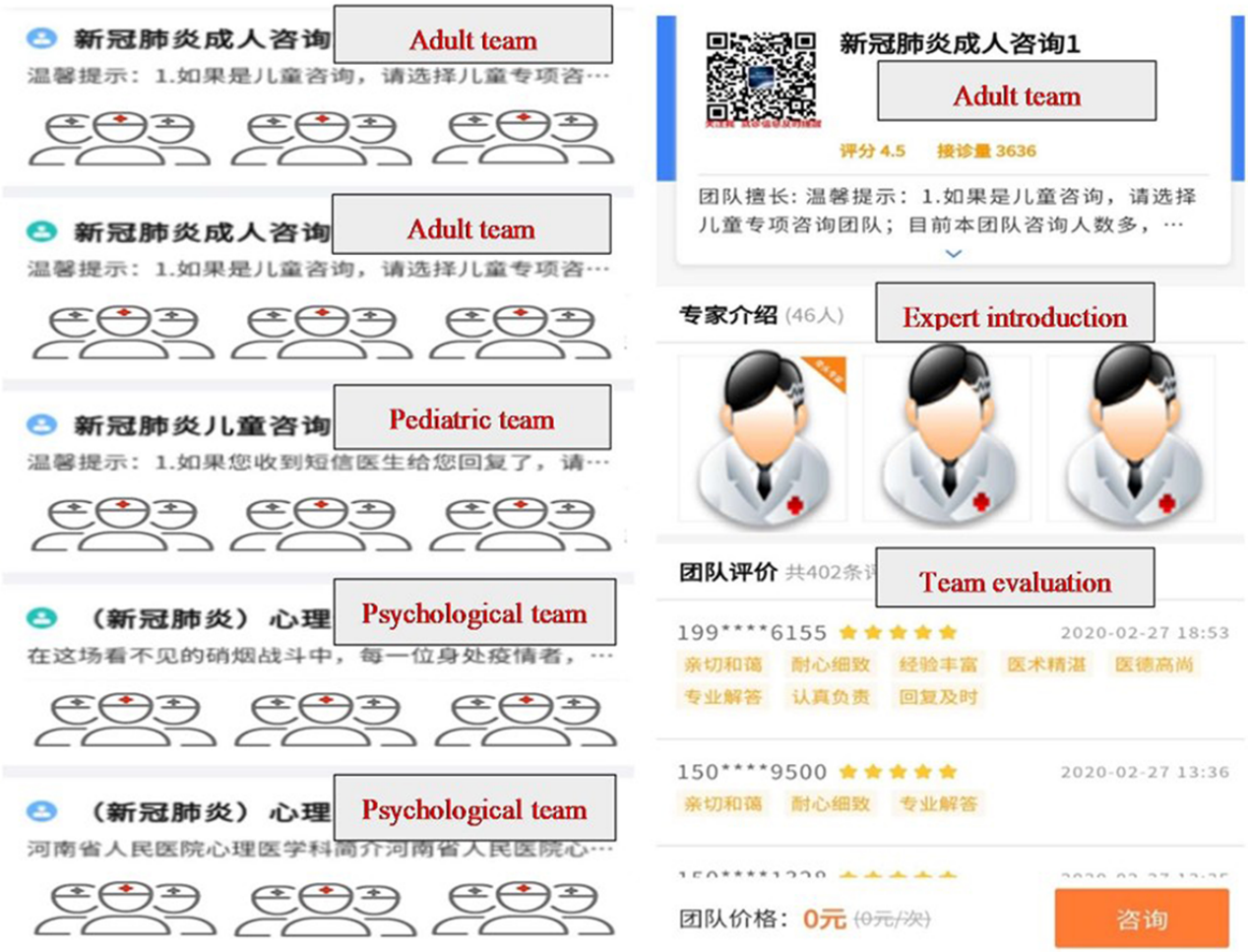
During the COVID-19 outbreak, nonurgent visits may cause overcrowding in hospitals [10], which not only adds to the burden of hospital staff but also dramatically increases the risk of infection spread. There may be severely adverse consequences for both patients and hospitals, making pandemic control even more difficult. On the other hand, as a major public health event, the COVID-19 pandemic not only endangers people's life and health but also causes psychological stress and anxiety. Active prevention and monitoring of the public’s psychological behavior is an integral part of a reasonable response to public health emergencies. Telemedicine is a health service discipline that combines modern communication, electronic technology, computer networks, and medicine [11]. Telemedicine can overcome geographical barriers, increase the number of people served, and provide online clinical support for patients [12,13]. This novel health service model has great potential for disease prevention and treatment as well as patient nursing during epidemiologic outbreaks [14].

Henan Provincial People's Hospital has opened the COVID-19 online outpatient clinic via the connected smart health service center platform, which provides smart health solutions for patients. The platform has already achieved 5G coverage and combines the internet with medical technology to successfully connect Henan Provincial People's Hospital to online hospitals. Online health services provided to patients include appointment, diagnosis and treatment, payment, quality control, follow-up, hierarchical diagnosis and treatment, emergency rescue, healthy management, public services, logistics, health education, and remote vital signs monitoring. Here, a retrospective analysis was performed to analyze the outcomes and advantages of telemedicine in the context of the COVID-19 outbreak in China.

Methods

A retrospective analysis was carried out using data from all patients who received remote diagnosis and treatment via online consultation services provided by our hospital from January 24 to February 17, 2020. During this period, 132 clinicians from Henan Provincial People's Hospital provided online outpatient consultation services to address COVID-19 concerns. These clinicians were from the Department of Infectious Diseases, the Respiratory Department, the Department of Critical Care Medicine, and the Department of Psychology and Psychiatry. They were divided into 5 teams: 2 teams provided consultation services for adult patients (adult team), 1 team for pediatric patients (pediatric team), and 2 teams for patients with psychological problems (psychological team; Figure 1). Online inquiry, consultation, and suggestions were provided to patients for concerns related to COVID-19 and the services were free of charge. Meanwhile, the hospital offered offline noncontact drug delivery services following online ordering and payment.

Figure 1. Expert clinician teams.



In order to consult with a clinician, patients can scan the official QR (Quick Response) code or follow the WeChat public account HNFYSY1904, which directs them to the expert consultation interface and allows access to the online outpatient consultation. Alternatively, patients can directly search our health service website to access the consultation page. Following the prompts, the patients click on the “Consult” button in the lower right corner, then input personal information and a description of symptoms, and start the consultation (Figure 2). The patients can choose real-time communication or online messaging for inquiry and consultation. The patients can interact with the clinicians online via voice, text, photo, and video (Figure 3).

During telemedicine consultations, we defined mild illness as follows: fever below 38˚C and no history of epidemiological exposure. Patients with a fever greater than 38˚C or with a history of epidemiological exposure were advised go to the hospital immediately. For mildly symptomatic patients and those with symptoms like cough, nasal congestion, fatigue, diarrhea, etc, our procedure was to provide medication guidance, recommend home isolation, and regular temperature monitoring. In cases where the temperature exceeds 38˚C or symptoms worsen, we recommended patients to visit a hospital for treatment.

At the end of the remote consultation, the platform will automatically open an electronic questionnaire, which the patients have the option to fill. The questionnaire includes star ratings and open answers. A minimum of 1 star represents high dissatisfaction, and a maximum of 5 stars represents high satisfaction. Four stars and above represent satisfaction. Following this, patients can input their own opinions and suggestions.

Since the online outpatient consultation service first began on January 24, 2020, online remote diagnosis and treatment and offline drug delivery services have been provided nonstop every day from 8:00 to 22:00. Patients’ personal information is stored on the connected smart health service center’s cloud-based platform, which ensures data safety. All statistical analyses were performed using SPSS 18.0 software (IMB Corporation). Categorical variables were expressed as absolute and relative frequencies and percentages.

The study protocol was approved by the ethics committee of Henan Provincial People's Hospital (Zhengzhou, China).

Figure 2. Flow chart of remote diagnosis and treatment via the online outpatient clinic. The platform supports consultations to be conducted under any network conditions. COVID-19: coronavirus disease; QR: Quick Response.

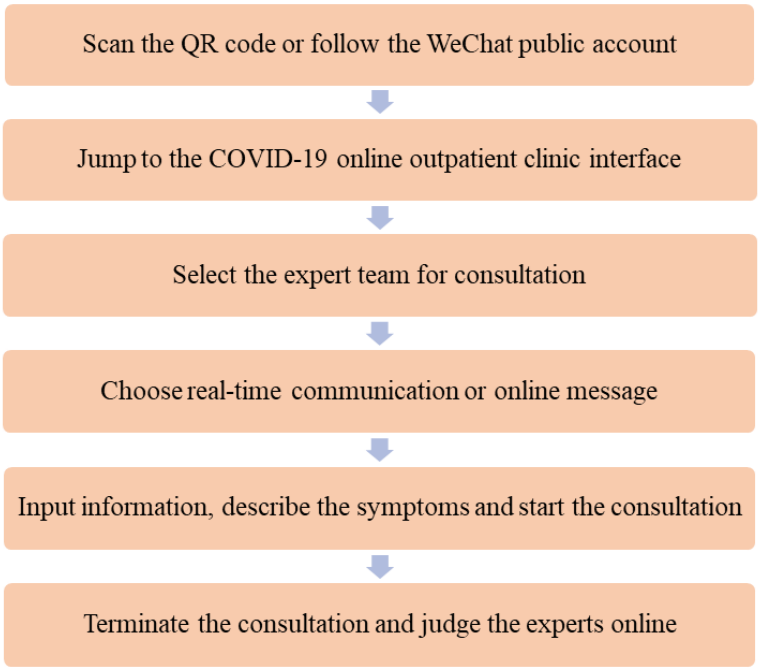


Figure 3. The patient-doctor dialogue interface for online outpatient consultation.

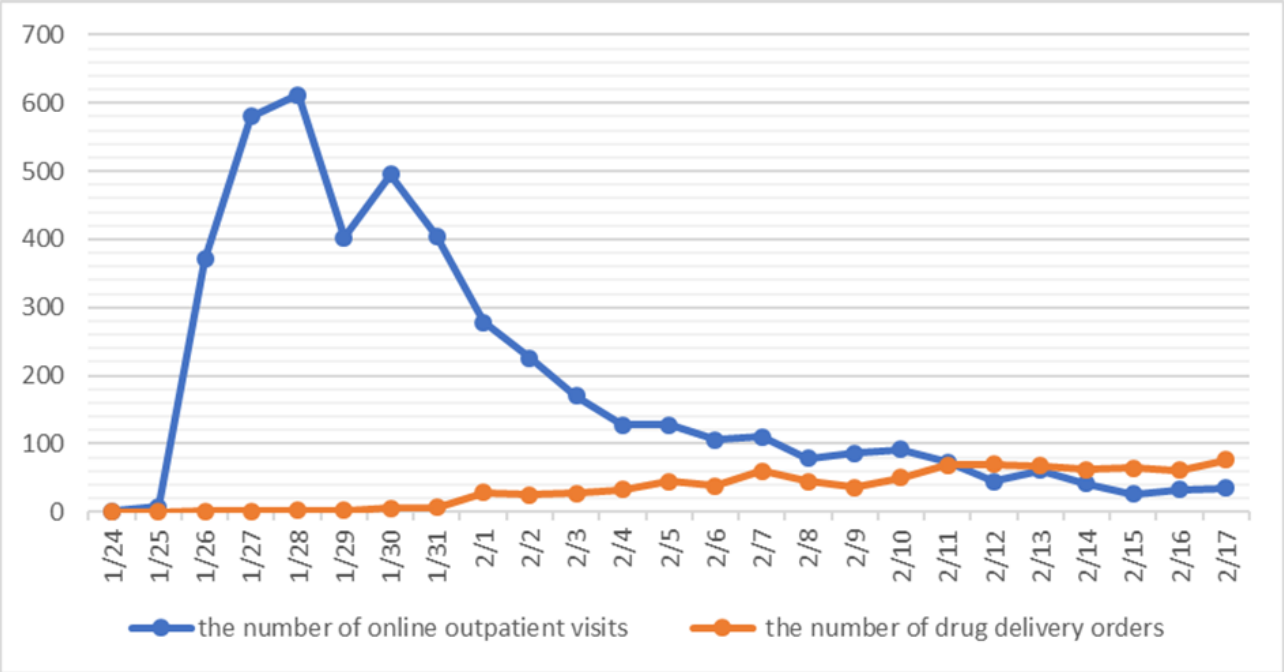
Results

In total, 4589 patients received remote diagnosis and treatment via the COVID-19 online outpatient clinic. Among them, there were 1940 (42.3%) males and 2649 (57.7%) females. The youngest patient was aged 78 days, and the eldest was 85 years old. In terms of age, 320 patients were aged <20 years, 2007 patients were aged 20-29 years, and 1707 were aged 30-39 years; the latter 2 groups accounted for 80.9% (n=3714) of the sample. Additionally, 315 patients were aged 50-59 years, and 50 patients were ≥60 years. Most patients (n=3898, 84.9%) came from Henan Province; 691 (15.1%) patients came from other provinces. During the epidemic, all patients with suspected COVID-19 received by a hospital were referred to a fever clinic. A total of 4561 patients were treated in the fever clinic of our

hospital. Among them, there were 2375 (52.1%) male patients and 2186 (47.9%) female patients.

From January 24 to February 17, 2020, the daily number of online outpatient visits first increased and then decreased, reaching its peak on January 28 when the daily number of online outpatient visits reached 612. Following this, the second peak occurred on January 30 when visits totaled 495. This number gradually decreased afterward. Between February 15-17, the daily number of online outpatient visits remained at 25-35, with minimal change. The number of offline noncontact drug delivery orders following online payment first increased and then stabilized. The total number of offline noncontact drug delivery orders was 873. As shown in Figure 4, after February 12, the daily number of noncontact drug delivery orders exceeded that of online outpatient visits before gradually stabilizing.

Figure 4. The daily number of online outpatient visits and offline noncontact drug delivery orders.



By February 17, the 5 expert teams had provided remote consultation services for 4589 patients (Table 1). In total, the 2 adult teams provided consultations to 4399 adult patients, accounting for 95.86% of the sample; the pediatric team provided consultations to 80 (1.74%) pediatric patients; and the 2 psychological teams provided psychological counseling for 110 (2.40%) patients (Table 1). The content of the consultations was classified according to the symptoms concerned (one patient might have one or more symptoms), including fever, headache, cough, fatigue, nasal obstruction, sore throat, nasal obstruction, and diarrhea. The most common symptom that patients initiated a consultation for was fever (n=2383 patients), followed by cough (n=1740), nasal obstruction (n=794 patients), fatigue (n=503), and diarrhea (n=276). Among the 110 patients who received psychological counseling, 7 patients reported psychological stress and anxiety due to their history of epidemiological exposure. The remaining 103 patients had no history of epidemiological exposure.

Table 1. Total number of online outpatient visits handled by clinicians for concerns related to coronavirus disease (COVID-19) (N=4589).

Consultation type	Clinicians, n	Online outpatient visits, n (%)
Consultation for adult patients	77	4399 (95.86)
Consultation for pediatric patients	28	80 (1.74)
Psychological counseling	27	110 (2.40)

Among the 4589 patients who received remote consultation services, 310 patients were advised to go to the hospital for further examination due to fever greater than 38℃, and 301 patients were advised to visit a hospital for further examination because of a history of epidemiological exposure and other symptoms. In total, 985 patients responded to the satisfaction questionnaire, of whom 98.1% (n=966) were satisfied with the service (rated as 4 stars or above). They praised the doctors for their rich experience, professional answers, dedication to their work, timely response, and patience. In the 110 cases of psychological counseling, 47 cases provided feedback, and the proportion of satisfaction ratings greater than 4 stars was 100%.

Discussion

Principal Findings

The COVID-19 outbreak occurred suddenly and spread rapidly and extensively. The disease is characterized by a long latency period and is even contagious during latency. The WHO attributed a “very high” risk level to COVID-19 on January 28,

2020. Cutting off sources of infection and reducing cross-infection through quarantine, disinfection, and personal protection are the top priorities of countries around the world [15]. Markwell et al [16] showed that the risk of cross-infection in hospitals could be decreased by reducing the visitor flow rate and chances of virus exposure in hospitals. Such countermeasures have already been widely adopted during the outbreak of many other contagious diseases [16-18]. Our hospital launched the online outpatient clinic services to address COVID-19 concerns via telemedicine. This online outpatient service platform can provide patients prehospital guidance, support online payment and offline drug distribution, reduce the possibility of cross-infection caused by nonurgent visits to offline clinics, and relieve the patient's mental health problems. It is of great significance for the prevention and treatment of pandemics.

In the present study, 80.9% of the online outpatient visits were from patients in the 20-39–year age group (n=3714), whereas patients aged ≥60 years made up only 1.1% (n=50) of the sample. The following reason is proposed to explain this phenomenon: young people are better adapted to use the internet. In the information era, young people prefer to use the internet to solve less complex health problems [14]. According to the latest Chinese statistical report on internet development released by the China Internet Network Information Center in 2019 [19], combined with the age distribution of the Chinese population, internet users aged 10-39 years account for 65.1% of the total number of internet users. Within this group, internet users aged 20-29 years account for the highest proportion (24.6%), and about 91.3% of the total population of the same age group; internet users aged 30-39 years account for 23.7%, and about 94.1% of the total population of the same age group; internet users aged ≥60 years account for 6.9%, and about 23.2% of the total population of the same age group. These data suggest that the internet is indeed more widely used by young people, which is an important reason for the increased use of telemedicine in this cohort.

During the pandemic, the number of online outpatient remote consultations and offline fever clinic visits was 4589 and 4561, respectively. The number of remote consultations via the online clinic was comparable to the total number of cases that the fever clinic handled in the same period. Although these data cannot directly show that telemedicine has reduced the total number of cases in hospitals, it is certain that telemedicine has reduced the burden on hospitals to some extent. Fewer outpatient visits allow hospitals to avoid overcrowding while reducing the risk of hospital cross-infection.

As shown in Figure 4, the number of patients who received remote consultation regarding COVID-19 concerns first increased and then decreased. After the first peak on January 28, a minor peak occurred on January 30. Later, the daily number of online outpatient visits gradually decreased. In the meantime, the daily number of offline noncontact drug delivery orders following online payment steadily increased. After February 12, the daily number of noncontact drug delivery orders exceeded that of online outpatient visits and gradually stabilized. According to National Health Commission statistics [20], Henan Province had a cumulation of 32 confirmed cases by January 24. At this time, the online outpatient consultation service had just been launched, and few people knew about the service. The daily number of online outpatient visits was only 1, and offline drug delivery was not yet available. By January 28, Henan Province had 206 confirmed cases [20]. This was also the day when the daily number of online outpatient visits reached a peak (n=612). On the contrary, the daily number of offline drug delivery remained low, since many people were afraid of personal contact. The daily number of newly confirmed cases increased considerably in Henan Province in the following days [20], and the general public began to gain more knowledge about the COVID-19 outbreak. Those with more severe symptoms would go to a hospital for further examination, resulting in a mild increase in the daily number of offline drug delivery orders. From February 13-17, the daily number of newly confirmed cases was not above 20 in Henan Province [20]. Most people had already been practicing self-quarantine for over 14 days. The daily number of online outpatient visits fluctuated at a low level, and the daily number of offline noncontact drug delivery orders stabilized.

According to the *Diagnosis and Treatment Protocol for Novel Coronavirus-Infected Pneumonia (Version 6)* released by the National Health Commission on February 19, 2020, and the latest research findings, fever, fatigue, and dry cough are the primary symptoms of COVID-19 [21]. Nasal obstruction, runny nose, and other upper respiratory symptoms are rare [22]. Gastrointestinal symptoms are also uncommon [23]. This study demonstrates that, among all remote consultations, 2383 cases exhibited fever symptoms and 1740 cases had cough symptoms, which is in agreement with the fact that fever and cough are the primary clinical symptoms of COVID-19. Nasal obstruction, a rare symptom of COVID-19, was the third frequent symptom that patients consulted for, and fatigue ranked fourth in the number of consultations; this is inconsistent with reported primary symptoms of COVID-19. The fifth most common symptom was diarrhea, which is also uncommon among COVID-19 patients. It is assumed that the symptoms consulted for by patients are more common symptoms of infection, and there is no precise connection with the main symptoms of COVID-19.

Previous research has demonstrated that over 58% of the respondents have psychological health problems amidst public health emergencies and have a strong need for psychological intervention [24-26]. As a result, some people may suffer from acute stress disorder, posttraumatic stress disorder, depression, other psychological disorders, or even commit suicide [27]. Most patients with psychological disorders may not be willing to go to the hospital for consultation due to the risks of cross-infection [28]. Telemedicine provides a way for patients to alleviate their psychological problems. In total, 110 patients received professional psychological counseling through our hospital's online outpatient consultation service. The clinicians taught them simple methods to relieve anxiety, including relaxation training through online tutorials, deep breathing exercises, soothing music, etc. In total, 47 patients provided feedback on the psychological services they received. All 47 patients provided ratings greater than 4 stars and reported that after the remote consultation, they had eliminated their doubts,

reduced their anxiety, and were satisfied with the consulting doctor. The data suggest that remote counseling can be helpful in relieving anxiety.

Limitations

The biggest limitation of the present study is the absence of follow-up due to the already heavy burden of medical staff. Therefore, there is no information about follow-up, which is not conducive for further studies. Secondly, because the satisfaction questionnaire was filled out voluntarily by patients rather than obtained by random sampling, the results may be biased.

Conclusion

During the COVID-19 outbreak, telemedicine can quickly establish online services within a short timeframe. This can reduce the burden of hospital personnel and the risk of cross-infection caused by offline treatment visits, as well as relieve the psychological burden and anxiety of patients to a certain extent. Telemedicine, therefore, plays a crucial role in the prevention and management of the pandemic.

Acknowledgments

We express our most sincere gratitude for members of the expert panel providing online outpatient consultation services and all frontline medical workers fighting the COVID-19 outbreak. We also thank the participants of the present study.

Conflicts of Interest

None declared.

References

1. Novel Coronavirus – China. World Health Organization. 2020 Jan 12. URL: https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ [accessed 2020-01-30]
2. Tian H. [2019-nCoV: new challenges from coronavirus]. Chinese Journal of Preventive Medicine 2020 Mar 06;54(3):235-238. [doi: 10.3760/cma.j.issn.0253-9624.2020.03.002] [Medline: 32187929]
3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb 15;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
5. Diagnosis and treatment plan for pneumonitis caused by new coronavirus (trial version 5). National Health Committee of the People's Republic of China. 2020 Jan 05. URL: http://www.gov.cn/zhengce/2020-02/05/content_5474852.htm [accessed 2020-03-30]
6. Novel Coronavirus – Thailand (ex-China). World Health Organization. 2020 Jan 14. URL: http://www.who.int/csr/don/14-january-2020-novel-coronavirus-thailand/en [accessed 2020-01-19]
7. Novel Coronavirus – Japan (ex-China). World Health Organization. 2020 Jan 17. URL: http://www.who.int/csr/don/17-january-2020-novel-coronavirus-japan-ex-china/en [accessed 2020-01-19]
8. Novel Coronavirus – Republic of Korea (ex-China). World Health Organization. 2020 Jan 21. URL: http://www.who.int/csr/don/21-january-2020-novel-coronavirus-republic-of-korea-ex-china/en [accessed 2020-01-19]
9. First Travel-related Case of 2019 Novel Coronavirus Detected in United States. Centers of Disease Control and Prevention. 2020 Jan 21. URL: https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html [accessed 2020-01-23]
10. Giunta DH, Pedretti AS, Elizondo CM, Grande Ratti MF, González Bernaldo de Quiros F, Waisman GD, et al. Descripción de las características del fenómeno Crowding en la Central de Emergencia de Adultos, en un hospital universitario de alta complejidad: estudio de cohorte retrospectiva. Rev. méd. Chile 2017 May;145(5):557-563. [doi: 10.4067/s0034-98872017000500001]
11. Liu L, Duan S, Zhang Y, Wu Y, Zhang L. Initial Experience of the Synchronized, Real-Time, Interactive, Remote Transthoracic Echocardiogram Consultation System in Rural China: Longitudinal Observational Study. JMIR Med Inform 2019 Jul 08;7(3):e14248 [FREE Full text] [doi: 10.2196/14248] [Medline: 31287062]
12. Lugn NE. Global health care-bridging the gap. J Telemed Telecare 2006 Jun 22;12(3):109-110. [doi: 10.1258/135763306776738576] [Medline: 16638230]
13. Britton N, Miller MA, Safadi S, Siegel A, Levine AR, McCurdy MT. Tele-Ultrasound in Resource-Limited Settings: A Systematic Review. Front Public Health 2019 Sep 4;7:244 [FREE Full text] [doi: 10.3389/fpubh.2019.00244] [Medline: 31552212]
14. Frid S, Ratti M, Pedretti A, Pollan J, Martínez B, Abreu AL, et al. Telemedicine for Upper Respiratory Tract Infections During 2018 Epidemiological Outbreak in South America. Stud Health Technol Inform 2019 Aug 21;264:586-590. [doi: 10.3233/SHTI190290] [Medline: 31437991]

15. Wei Q, Ren Z. Disinfection measures for the epidemic focus of COVID-19. Chinese Journal of Disinfection 2020:1-4. [doi: 10.11726/j.issn.1001-7658.2020.01.019]
16. Markwell H, Godkin D. Visitor Restrictions during a Public Health Emergency: Ethical Issues and Guidelines for Policy Development. In: Ontario Ministry of Health and Long-Term Care 2004. URL: https://www.deslibris.ca/ID/200785
17. Currie K, Curran E, Strachan E, Bunyan D, Price L. Temporary suspension of visiting during norovirus outbreaks in NHS Boards and the independent care home sector in Scotland: a cross-sectional survey of practice. J Hosp Infect 2016 Mar;92(3):253-258. [doi: 10.1016/j.jhin.2015.10.018] [Medline: 26654470]
18. Yin H, Wang A. Discussion on the visiting system for preventing the input and transmission of severe acute respiratory syndrome. Shanghai Nursing 2004;4(2):49-50. [doi: 10.1142/5469]
19. The 44th China statistical report on Internet development. Cyberspace Adminstration of China. 2019 Jul 30. URL: http://www.cac.gov.cn/2019-08/30/c_1124938750.htm [accessed 2020-05-26]
20. Epidemic prevention and control trends. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/wjw/index_gzbd.shtml [accessed 2020-02-19]
21. National Health Commission of the People's Republic of China, National Administration of Traditional Chinese Medicine. Diagnosis & Treatment Scheme for Novel Coronavirus Pneumonia (Trial) 6th Edition. NHC. 2020 Jan 19. URL: http://www.nhc.gov.cn/xcs/zhengcwj/202002/8334a8326dd94d329df351d7da8aefc2.shtml [accessed 2020-02-19]
22. Deng S, Peng H. Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China. J Clin Med 2020 Feb 20;9(2) [FREE Full text] [doi: 10.3390/jcm9020575] [Medline: 32093211]
23. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
24. Xiao JQ, Wu QH, Hao YH, Gao LJ, Sun H, Ning N. A study on the mental health status, influence factors and intervention among residents in Harbin during during SARS epidemic. Chinese Health Economics 2007;26(3):20-23. [doi: 10.3969/j.issn.1003-0743.2007.03.006]
25. Lv J, Tang D. Knowledge, psychology, and behaviors about avian influenza among part of the population in Daxing District, Beijing. Chinese Journal of Health Education 2004;20(10):924-925. [doi: 10.3969/j.issn.1002-9982.2004.10.021]
26. Xiang H, Huang XY, Wang RK, Pu SW, He YZ. Crisis intervention in Mianyang City after the Wenchuan Earthquake. Journal of Evidence-Based Medicine 2008;8(11):918-921. [doi: 10.1136/ebmh.11.2.38]
27. Wang C, Wang X. COVID-19 epidemic, nosocomial infection, and psychological intervention. Chinese General Nursing 2020;18(03):309-310. [doi: 10.12104/j.issn.1674-4748.2020.03.016]
28. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]

Abbreviations

2019-nCoV: novel coronavirus
COVID-19: coronavirus disease
MERS CoV: Middle East respiratory syndrome coronavirus
QR: Quick Response
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.04.20; peer-reviewed by R Ohannessian, L Seuren; comments to author 24.05.20; revised version received 02.06.20; accepted 21.06.20; published 03.07.20.

Please cite as:
Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, Zhang L
Application and Preliminary Outcomes of Remote Diagnosis and Treatment During the COVID-19 Outbreak: Retrospective Cohort Study
JMIR Mhealth Uhealth 2020;8(7):e19417
URL: https://mhealth.jmir.org/2020/7/e19417
doi:10.2196/19417
PMID:32568722

Original Paper

Effects of a WhatsApp-Delivered Education Intervention to Enhance Breast Cancer Knowledge in Women: Mixed-Methods Study

Antonio Augusto Claudio Pereira[1*], BA; Juliana Regina Destro[1*], BA; Marcelo Picinin Bernuci[1,2*], PhD; Lucas França Garcia[1,2*], PhD; Tiago Franklin Rodrigues Lucena[3*], PhD

[1]Universidade Cesumar, Maringá, Brazil

[2]ICETI-Cesumar Institute of Science, Technology, and Innovation, Maringá, Brazil

[3]Departamento de Fundamentos da Educação, Universidade Estadual de Maringá, Maringá, Brazil

[*]all authors contributed equally

Corresponding Author:
Tiago Franklin Rodrigues Lucena, PhD
Departamento de Fundamentos da Educação
Universidade Estadual de Maringá
Av Colombo, 5790 - Bloco I-12
Maringá, 87020-900
Brazil
Phone: 55 44991034777
Email: tfrlucena2@uem.br

Abstract

Background: Breast cancer is the leading cause of cancer-related death in the female population. Health education interventions based on the use of mobile technologies enable the development of health self-care skills and have emerged as alternative strategies for the control of breast cancer. In previous studies, WhatsApp has stood out as a useful tool in health education strategies; however, it has not yet been applied for breast cancer education.

Objective: This study aimed to analyze the potential of WhatsApp as a health education tool used to improve women's knowledge on the risk reduction of breast cancer. It also aimed to understand how women feel sensitized within the WhatsApp group throughout the intervention and how they incorporate information posted to improve knowledge about early detection and risk reduction methods.

Methods: The study involved a pre-post health educational intervention with 35 women (aged 45-69 years) included in a WhatsApp group to share information (audio, video, text, and images) over 3 weeks on the early detection and risk reduction of breast cancer. Data were collected through questionnaires on topics related to risk reduction, as well as qualitative content analysis of group interactions. Effectiveness and feasibility were analyzed through conversations and the comparison of the scores obtained in the questionnaires before and after the intervention.

Results: A total of 293 messages were exchanged (moderator 120 and users 173). The average scores of the participants were 11.21 and 13.68 points before and after the educational intervention, respectively, with sufficient sample evidence that the difference was significant ($P<.001$). The intervention enabled women to improve their knowledge on all topics addressed, especially “myths and truths,” “incidence,” “clinical manifestations,” and “protective factors.” Some themes emerged from the interactions in the group, including group dynamics, general doubts, personal narratives, religious messages, daily news, and events.

Conclusions: The use of groups for women in WhatsApp for health education purposes seems to be a viable alternative in strategies on breast cancer control, especially as it provides a space for the exchange of experiences and disinhibition. However, the need for a moderator to answer the questions and the constant distractions by members of the group represent important limitations that should be considered when improving this strategy.

(JMIR Mhealth Uhealth 2020;8(7):e17430) doi:10.2196/17430

KEYWORDS

mHealth; WhatsApp; cancer education; breast cancer

Introduction

Breast cancer represents the main cause of death by cancer among women around the world [1]. The mortality rates associated with this malignancy are higher in low- and middle-income countries owing to early diagnosis deficiencies and low access to treatment [2,3]. Mortality reduction in many developed countries has been occurring after the implementation of organized screening of precursor lesions associated with more effective therapies [4-6]. However, the adoption of such strategies requires high infrastructure and human resource investments, which are below the capacity of low- and middle-income countries [7,8].

Considering financial constraints, it is suggested that the most prudent action to be applied in cancer control is the guarantee of early diagnosis and access to treatment without delay [9,10]. Nevertheless, the limitations of population awareness regarding the relevance of early diagnosis and preventive behavior constitute major barriers to the effectiveness of these actions [11-13]. Educational interventions mediated by mobile technologies represent alternatives to be explored in the practice of health promotion for their convenience and ubiquity, and for minimizing the barriers of distance, cost, and time [14-16].

The wide availability and functionality of mobile devices have made them great tools for planning, executing, and evaluating health interventions [17,18]. Indeed, smartphones and their apps have been used to spread health information [19-22], including information for breast cancer care [23]. Evidence for individuals who use health self-management apps points to positive results [24-26] and reinforces the viability of using this technology in health promotion strategies.

Among the apps with the potential for use in health education interventions, WhatsApp (WhatsApp Inc) stands out [27,28]. It is an instant messaging app for smartphones that enables users to send voice, text, video, or location using the group communication feature [29]. The app offers additional social information to its users when compared with SMS text messaging. Contacts are able to see when other users are online and typing, as well as their last access. Moreover, it is the most downloaded app in the world [30], is free, and has a user-friendly interface, and it accounts for about 20% of total smartphone use [31].

Despite the popularity of WhatsApp and its promising potential in health education and health care strategies [27,32-34], only few studies have evaluated the effectiveness of using this app to convey health information for user empowerment purposes [33,35-38]. Therefore, this study aimed to provide evidence for the utility and effectiveness of this mobile health technology to improve women's knowledge regarding the risk reduction and early detection of breast cancer. The primary aim was to evaluate women's engagement and the viability of content delivery on breast cancer education through a WhatsApp group. An additional goal was to evaluate the effects of a social interaction channel on key knowledge about breast cancer through WhatsApp. We hypothesized that after 3 weeks of women engagement on the WhatsApp group, knowledge about the risk reduction and early detection of breast cancer would improve.

Methods

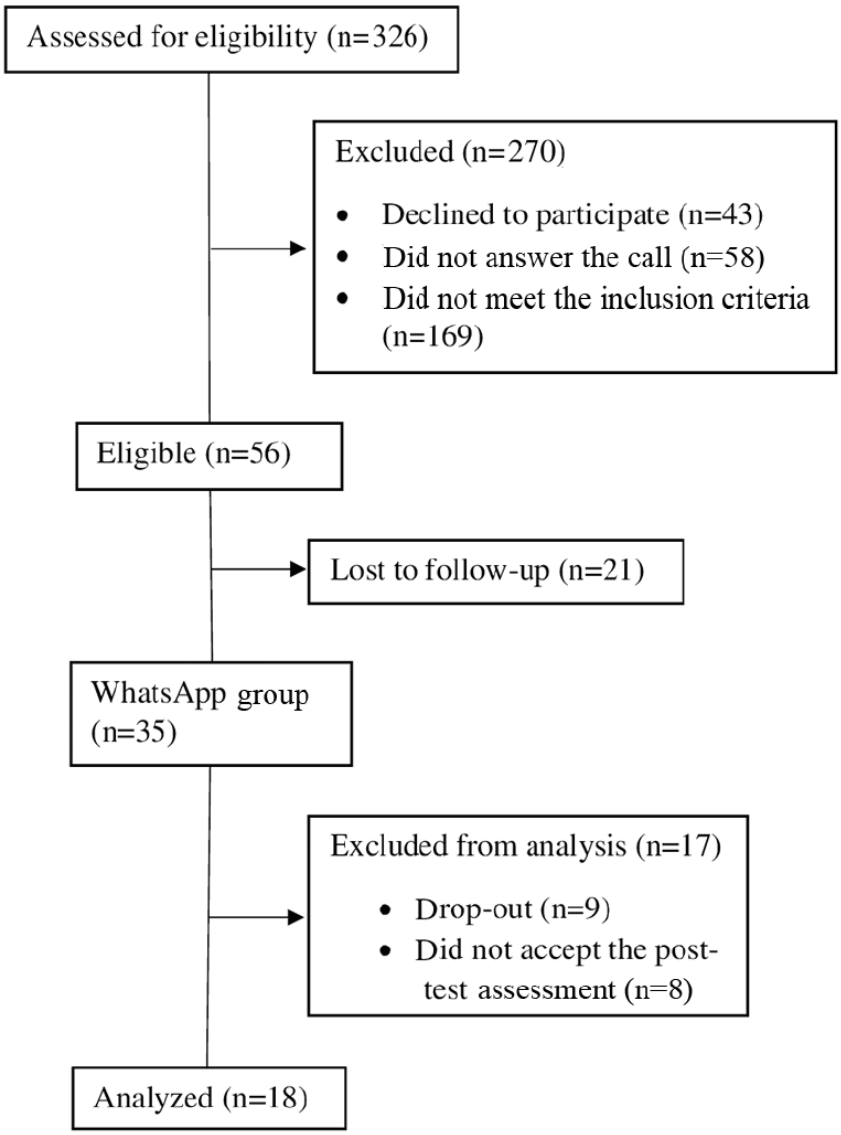
Study Design

A 3-week pre-post intervention study was conducted among women enrolled in the Brazilian Unified Health System, and a breast cancer screening control study design was adopted to evaluate the feasibility of WhatsApp as an educational tool to increase women's knowledge about the risk reduction and early detection of breast cancer. The study was performed at a basic health unit in the city of Maringá-PR, south Brazil, from July to September 2017. The basic health unit offers free primary care services, including clinical breast examination and mammography, and the annual census reports care for approximately 8000 patients per month. The patient data can be accessed on an electronic information system. The study protocol was approved by the local research ethics committee (protocol number 2.015.313). Written informed consent was obtained from each participant.

Participants and Recruitment

Participants were adult women aged 45 to 69 years, who were chosen because this is the age group in which the breast cancer incidence is the highest [39]. A sample of 326 women was calculated from a population of 2116 women registered in the information system at the basic health unit. The sample selection process was performed in four steps (Figure 1).

Figure 1. Flow of participants throughout the study.



In the first step, stratified randomization was performed by age group. The number of women necessary to compose the sample (n_g) in each regional strata, corresponding to areas of coverage, was calculated according to the following equation:

The level of significance considered was α=5%, and the maximum error allowed between the estimate and the real parameter value was e=0.05 (ie, five percentage points). The number of women registered in the *gth* area of coverage is represented by N_g. Additionally, N is the total number of women (2116), and p_g is the prevalence of the characteristics to be researched, which was established as 0.5 for all groups, since there is no previous information.

After calculation, the finite population correction factor was used as follows:

Thus, the sample size obtained with the help of the R Statistical Environment (version 3.2.2, R Core Team) was 326 women.

In the second step, a list was generated from the electronic records of the 326 women in order to identify those who had a cell phone number. After applying this criterion, 110 women were excluded. In the third step, a telephone call was made to the remaining 216 women with the aim to invite them to a face-to-face meeting to explain the study objectives. Among these women, 58 did not answer the calls after three attempts on different days, 45 had incorrect telephone numbers registered, 43 did not show interest in participating in the study, and 14 did not have or use WhatsApp (totaling 160 exclusions). In the fourth step, only 35 of the remaining 56 women attended the meeting, and sociodemographic data were collected and a questionnaire was applied to identify the knowledge level on breast cancer–related topics.

A WhatsApp group, named “AllAboutBreastCancer” (in Portuguese), was created to allocate the 35 women participants, two observers, and one content mediator, who was responsible for sending messages and clarifications. Throughout the

intervention, nine women left the WhatsApp group (dropped out of the study) and eight did not answer the questionnaire to identify the knowledge level on breast cancer after 4 weeks of the intervention. Thus, the data of a total of 18 women were analyzed.

Health Education Content

Topics related to breast cancer were identified on the Brazilian National Cancer Institute (NCI) website, and they corresponded to risk factors, protective factors, definitions, incidence, clinical signs and symptoms, diagnostic examinations, mammography, myths and truths, and places to seek support. It is a publicly accessible website. However, both the direct sources of content and the information on the existence of the questionnaire to be used were not disclosed at any time during the intervention.

For each identified topic, we selected specific media (video, audio, or image) and freely available media databases on NCI and Brazilian Ministry of Health websites. In total, two videos, 23 images, and four audios were utilized in the media database.

Life Habits, Sociodemographic Factors, and Breast Cancer Knowledge

The data were collected at two time points (pre and post intervention) through the application of a two-block structured questionnaire based on information available in the Breast Cancer Early Detection Guidelines, Ministry of Health, and materials from the NCI [40]. In the first block, sociodemographic and life habit characteristics (age, marital status, educational level, religion, and occupation) were collected. In the second block, the knowledge level on breast cancer was assessed, utilizing 18 assertive questions distributed in nine topics as follows: (1) risk factors (questions 1 and 2); (2) protective factors (question 3); (3) definitions (questions 4 and 18); (4) incidence (questions 1 and 4); (5) clinical signs and symptoms (question 5); (6) diagnostic examinations (questions 6-13); (7) mammography (questions 13-15); (8) myths and truths (question 16); and (9) places to seek support (question 17).

The score attributed to each question ranged from 0 to 1, with a maximum score of 18 points. The scores of questions 1, 2, 3, 5, 7, 14, 15, and 16, for which all listed alternatives were correct, were calculated as the number of marked alternatives divided by the total number of alternatives. The score of question 4, for which one alternative was incorrect and the other three alternatives were correct, was calculated as the number of correctly marked alternatives divided by the total number of correct alternatives minus one-third if the incorrect alternative was marked, or was considered to equal 0 if the incorrect alternative was the only one marked. The scores of questions 6, 8, 9, 10, 11, 12, 13, 17, and 18, for which only one listed alternative was correct, was calculated as 1 if the correct alternative was marked or 0 if the incorrect alternative was marked.

Educational Intervention

The intervention consisted of daily sessions of text, figure, video, and voice messages during 3 weeks. The mediator followed a content submission protocol to send messages between 9 AM and 9 PM. The content was distributed over the period according to the topics. Additional media were created or adapted by the team, according to the demands and dynamics of the group.

Statistical Analysis

Data were analyzed using the R statistical software (version 3.3.1, R Development Core Team). For the classification of the level of education, we considered the years of education in Brazil, with middle school (minimum 9 years of education), high school (minimum 12 years of education), and higher education (minimum 14 years of education) levels.

To assess the intervention effectiveness, the nonparametric paired Wilcoxon test was applied, with the significance level set at 5%, and the scores obtained before and after the intervention were compared for each individual. The test used is a nonparametric alternative to the paired *t* test when its assumptions are not accepted. In this case, the score distribution was not normal (one of the assumptions), so the Wilcoxon test was chosen.

Content Analysis

Inductive content analysis [41] was applied for the data extracted from the WhatsApp conversations (exported as a .doc file). The content was converted to an excel file with the following three columns: date/time, user, and message (each message, for analysis purposes, is equivalent to one entry in WhatsApp). The file was imported to be classified and analyzed using NVIVO 11 Pro version (QSR International) for Windows. The messages were initially separated into two categories according to the sender as follows: (1) messages sent by the moderator and (2) messages sent by the users. Emojis were not considered as content to be analyzed.

The preanalysis initially focused on assessment of the 30 most frequent words present only in the user messages, with the goal of identifying the themes that emerged from the group conversations in the original language (Brazilian Portuguese). The parameters used for the preanalysis through the most frequent words in the WhatsApp conversations were as follows: (1) words with four or more characters, (2) words grouped by synonyms, according to the NVIVO 11 Portuguese language dictionary, and (3) exclusion of words that did not have any direct relation to the project content (that, people, during, any, etc). Two of the authors (TFRL and LFG) applied the analysis, and this step focused on building thematic categories that emerged during the process. The following five categories were created: (1) group dynamics; (2) general questions; (3) personal narratives; (4) religious messages; and (5) daily news and events. The data inference, treatment, and interpretation steps focused on the analysis of the text excerpts of each participant within each category, so that they were characterized in the theme. After the categorization process, a translator assisted with the translation of all sentences in each category from Brazilian Portuguese to English.

Results

Sociodemographic Characterization and Life Habits

Regarding the sociodemographic characterization of the women who participated in the study, all were over 30 years old, and 9 out of 18 (50%) participants were aged 50 years. For the ethnic-racial profile, 16 out of 18 (89%) participants declared themselves to be white people. Additionally, 10 out of 18 (56%) participants reported a lower degree of education than a completed higher education, 11 out of 18 (61%) declared being married and having children, and 12 out of 18 (67%) declared living with more than two adults. Regarding family income, 6 out of 18 (33%) participants declared monthly income below Brazilian Real 3,520 (approximately US $850). Regarding religious belief, 17 women declared themselves Christian and 1 Buddhist (data not shown).

Concerning life habits, 11 out of 18 (61%) participants reported consuming high-fat foods at least three times a week and consuming fruits and vegetables daily, 15 out of 18 (83%) reported not consuming alcohol or smoking, and 12 out of 18 (67%) reported not practicing physical activity. In addition, 9 out of 18 (50%) participants reported being postmenopausal, 12 out of 18 (67%) reported not using oral contraceptives (however, 11 out of 18 [61%] used them in the past), and 16 out of 18 (89%) reported not taking hormone replacement therapy (however, 12 out of 18 [67%] did take it in the past). Finally, 12 out of 18 (67%) participants did not report any cases of breast cancer in the family and 13 out of 18 (72%) underwent mammography (data not shown).

Content Analysis

A total of 293 messages were exchanged (moderator 120 and users 173). From the preanalysis step, it was possible to identify five main thematic categories and three subcategories (Table 1).

The category “group dynamics” addressed group participants’ interventions regarding group functioning. In this sense, the following three subcategories stand out in relation to group dynamics: (1) feedback, in which participants express their feelings about the shared content; (2) self-regulation, in which participants express their discomfort regarding content not related to the theme of the group, thus agreeing on some parameters about the group discussions; and (3) coexistence (mostly short messages), in which participants introduce themselves to the group, wish “good morning,” “good afternoon,” or “good night,” and express some expectations about what the group can offer.

The category “general doubts” deals with questions and doubts from the participants mostly about the topic “myths and truths” related to breast cancer. Among the main doubts, there were doubts regarding breastfeeding, contraceptive use, deodorant use, and signs and symptoms that may or may not be related to cancer, as well as lifestyle issues, such as the practice of physical exercise.

The category “personal narratives” deals with participants’ personal reports on their knowledge of cancer cases in their social circle, especially in the family.

The category “religious messages” is about biblical excerpts shared by users without apparent relation to the group work content.

The category “daily news and events” is about the sharing of a public safety event in the city.

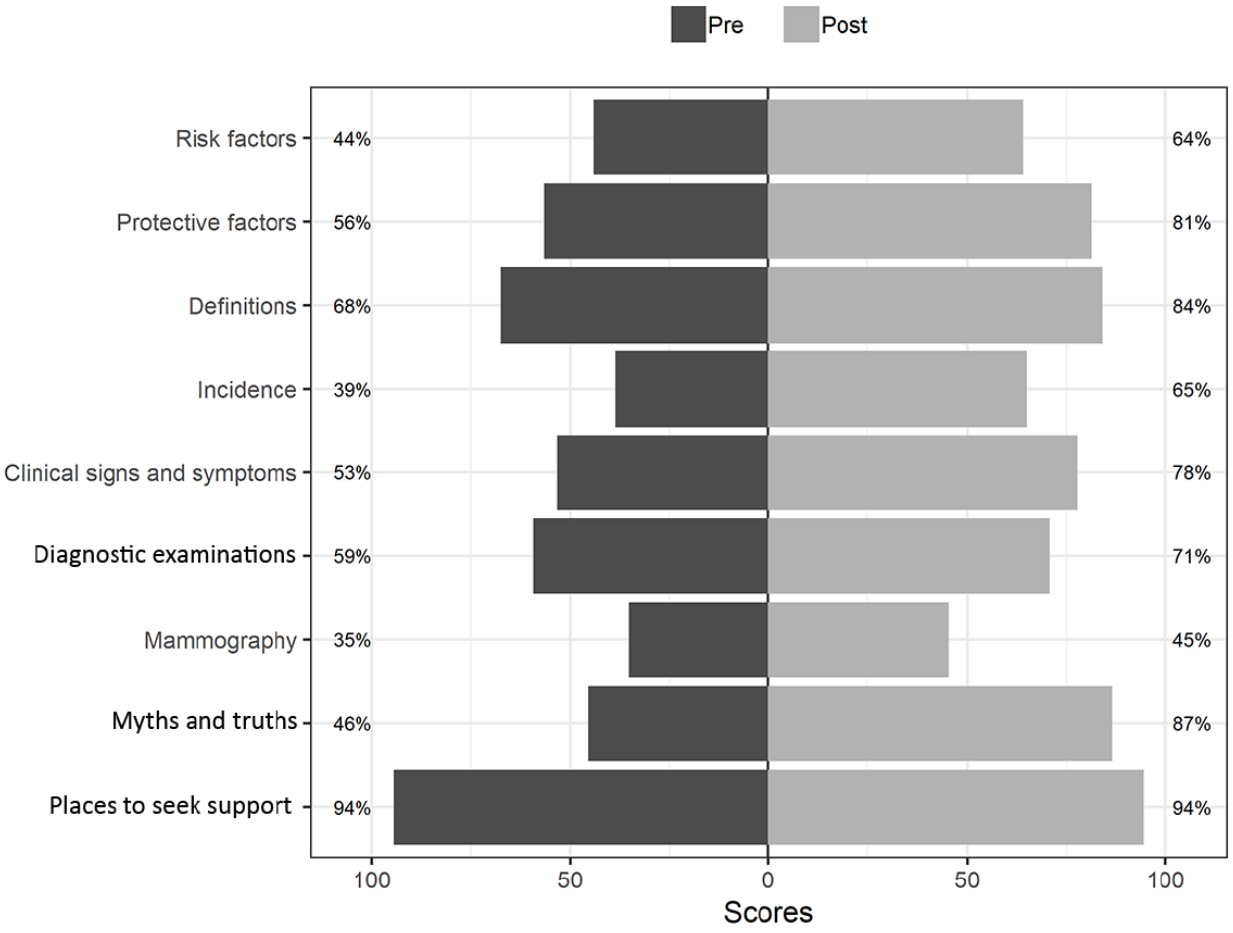
Table 1. Thematic categories (translated from Portuguese to English).

Topic/description	Illustrative quote
1. Group dynamics (90 messages)	
1.1. Feedback (27 messages)	"I should start doing a sport, lol. Thanks!"
	"Wow, I'm loving the questions, is clearing up all my doubts"
	"Hi Antonio, I'm also enjoying it, it's been very helpful to me"
	"I'm reading it, liking it and learning, and I'm already talking to my friends to share what I've learned"
	"Very good! Useful information and very important for prevention! I'm passing it along"
1.2. Self-regulation (35 messages)	"Hi, what's up? I don't think it's cool to start forwarding chain messages in here. The group has a purpose, let's just talk about it. It's better this way"
	"The group's purpose is to talk about health, I think that is very clear, at least for me"
	"Girls, I think everything is relevant, but if this group doesn't stay limited to scientific information about the subject matter, I will leave the group, sorry."
	"We have to stay, let's not make this one slide put all the work in jeopardy. Focus on the discussions"
1.3. Coexistence (28 messages)	"Good evening, I'm Bethe, we will learn a lot with this project"
	"Good evening, healthy family! I'm Iracema, it's an honor to be part of this group, I thank the teachers and academics in acquiring knowledge for a lifetime, taking care of our health. We are here to help each other"
2. General doubts (18 messages)	"I didn't know that the lack of sport could be one of the causes of breast cancer. Can we say that the intake of pesticides/industrialized products can influence?"
	"What about the deodorant? I've heard rumors that can cause it, is it true?"
	"Is it true or myth about the breastfeeding?"
	"True or myth about the use of birth control?"
	"My daughter with 10 years old had her first period, that concerns me, I heard that when the first period comes the girl stops to grow up, is it myth or true?"
	"Antonio [moderator], I already had breast cancer and lately I'm overweight, is it true that this is one of the factors that can cause cancer recurrence?"
	"About the breastfeeding… is there a time for this protection? I breastfed my first child for 5 months and now I'm breastfeeding my second child, but I think I won't be able to do it for much more than 6 months…"
	"My sister heard that a spot on the toenail can be a sign of breast cancer. It this correct?"
	"And the women with less than 50 years old? Can or should do the mammogram?"
	"Good evening, Antonio. My daughter took the first dose of HPV vaccine with 10 years old, the second dose would be after 6 months and she hasn't taken for 10 months now, can she still take the second dose?"
3. Personal narratives (7 messages)	"I also have cases in my family, my sister decided not to have children and had breast cancer, she found out in the beginning and treated it… but she travels to Sao Paulo every 6 months to have a follow-up care!"
	"My mother-in-law had breast cancer 3 years ago, and now has been diagnosed with liver and lung metastasis."
	"My maternal grandmother had it. Today she is with 94 years old, this was 20 years ago."
	"My mother had it with 45 years old. It's been 17 years."
4. Religious messages (3 messages)	"What are the commandments? Asked the young man. Jesus said to him, "Do not kill, do not commit adultery, do not steal." "Do not bear false witness. Honor Father and Mother. Love your neighbor as yourself. Matthew 19, 16-22"
5. Daily news and events (1 message)	"Everyone, I'm sorry to send this message right now, but almost 90 prisoners escaped the Maringá prison. It broadcasting in media news alerting residents and people in towns nearby. Be very careful. #scary"

Breast Cancer Knowledge

Figure 2 presents data regarding the scores obtained by participants in the nine domains of the questionnaire that was designed to analyze the level of knowledge on breast cancer and was applied before and after the educational intervention. The topics presenting a lower score gain in the posttest analysis were "definitions" (16 points gained), "examinations" (12 points gained), and "mammography" (10 points gained), whereas the topics with the highest gains were "myths and truths" (41 points gained), "incidence" (26 points gained), "clinical manifestations" (25 points gained), and "protective factors" (25 points gained). It was observed that beside "places to seek support," which remained stable, there were score gains in all 80 domains or topics of the questionnaire in the pre-post intervention.

Figure 2. Profile of scores obtained in each domain of the questionnaire designed to analyze the level of knowledge about breast cancer.



Regarding factors that increase the chances of developing breast cancer, before the intervention, 9 out of 18 participants mentioned only family history of breast cancer, smoking, and alcohol consumption. After the intervention, all listed factors had a higher frequency of response, and factors, such as consumption of high-fat foods, first pregnancy after 30 years of age, first period before 12 years of age, and end of menstruation after 55 years of age, were mentioned by less than half of the women.

Regarding family history, before the intervention, 15 out of 18 participants indicated that having at least one first-degree relative diagnosed with cancer in one breast increases the risk of developing this type of cancer in women under the age of 50 years, which did not change after the intervention. Similarly, most of the participants responded that the risk also includes a family history of at least one first-degree relative diagnosed with cancer in both breasts or ovarian cancer in any age group, which also did not change after the intervention. The increased risk for family history of male breast cancer was only mentioned after the educational intervention by one-third of the participants.

Concerning protective factors, before the intervention, 15 out of 18 participants mentioned only diet with many fruits and vegetables, which did not change after the intervention. However, all the other listed protective factors had a higher mentioning frequency after the educational intervention, with breastfeeding as the most frequently mentioned protective factor (16 out of 18 participants).

Regarding knowledge about the concept of the disease, before the intervention, 8 out of 18 participants incorrectly pointed out that breast cancer can be a benign tumor. After the intervention, this answer was chosen by only three participants. The most mentioned signs and symptoms before the intervention were small nodules in the underarm (armpit) or neck region and the appearance of a fixed usually painless lump, which did not change after the educational intervention.

Regarding the early detection of breast cancer, before the intervention, the three recommendations listed as ways of contributing to early detection were mentioned by 9 out of 18 (50%) participants, with mammography as the main examination for lesion tracking. After the intervention, all of these items were mentioned more frequently. In addition, after the intervention, all of the participants stated that doing the breast self-examination or being asymptomatic does not exclude the necessity for a clinical breast examination or mammography.

Nevertheless, after the intervention, most women indicated the correct concept of clinical breast examination and pointed out the recommendation of age range and periodicity for the examination. Lastly, it was observed that all the participants continued to point out that the basic health unit was an ideal place to look for assistance and information regarding doubts or any initial manifestations. After the intervention, all women mentioned “breast cancer is curable.”

Regarding mammography, most participants had already scored correctly in the examination concept, maintaining the score after the intervention. However, the same was not observed with the recommendations regarding periodicity, age group, and details about the examination, as most women did not score correctly.

After the intervention, all the listed statements were considered as myths with more frequency, and the most mentioned was "large breasts represent a higher risk of developing the disease."

Individual data regarding the effect of the educational intervention on improving knowledge about breast cancer are shown in Figure 3. The score regarding the knowledge of breast cancer tended to increase from before to after the educational intervention for each study participant. The average scores of the participants were 11.21 and 13.68 points before and after the intervention, respectively, and there was sufficient sample evidence that the difference was significant, according to the paired Wilcoxon test results, with a significance level of 5% (P<.001) (Table 2).

Figure 3. Profile of individual scores obtained in the nine domains of the questionnaire designed to analyze the level of knowledge about breast cancer.

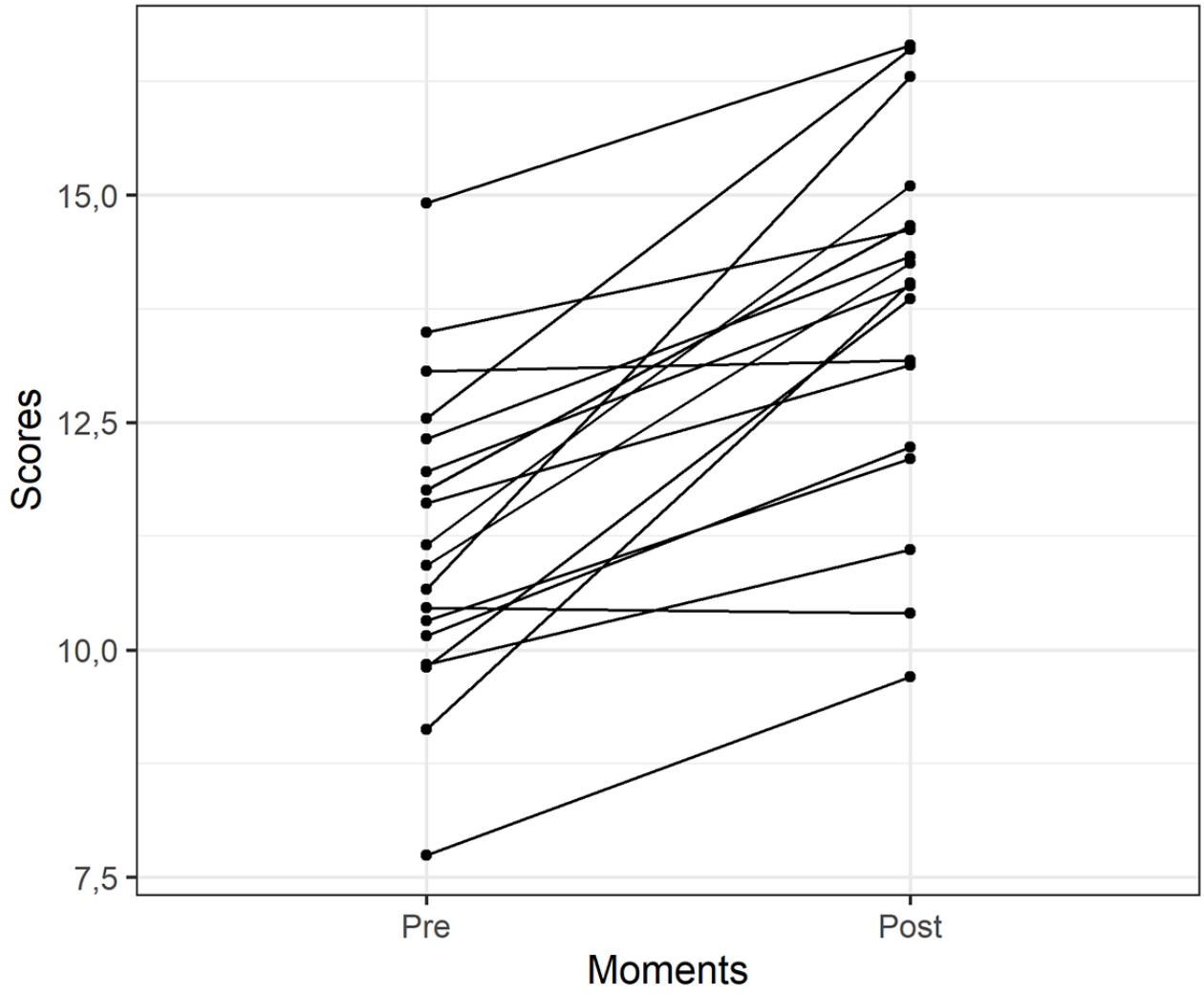


Table 2. Descriptive measures of the scores obtained in the nine domains of the questionnaire designed to analyze the level of knowledge on breast cancer applied before and after the educational intervention, according to the paired Wilcoxon test.

Moment	Mean	SD	Median	Q_1-Q_3	*P* value
Preintervention	11.21	1.70	11.04	10.20-12.23	<.001
Postintervention	13.68	1.99	14.02	12.46-14.65	

Discussion

Principal Findings

To our knowledge, this is the first study to use WhatsApp as a health education tool for Brazilian women from the public health system, with a focus on the risk reduction and early detection of breast cancer. The 3-week intervention, which used the app as an online space to mediate content selected from the NCI and the Brazilian Ministry of Health websites, enabled women to improve their knowledge in all topics addressed, especially in "myths and truths," "incidence," "clinical manifestations,"

and "protective factors." The app, which is known by women as a daily tool to interact with friends and relatives, allowed a receptive space to solve doubts not commonly discussed during presential care. In general, the results obtained herein suggest that WhatsApp could be incorporated as an additional tool in traditional health communication actions and campaigns.

Use of WhatsApp

No woman reported having difficulties when interacting with the device or app, or asked for training on how to use it. Many aspects may have contributed to the gains in pre- and postintervention scores. The combination of text, image, video, and audio may partly explain the gain in knowledge. This multimedia aspect differentiates WhatsApp interventions from other mobile interventions that exclusively use SMS text messaging. Mayer [42] has talked about "multimedia learning" and has argued that well-designed multimedia content (the combination of video, audio, and text) can enhance the learning process when compared with an educational process that uses only one modality (eg, only text). In addition, the possibility to send messages directly to each woman has been highlighted as a positive aspect, since some messages were designed for each doubt. Both health communication [43] and neuroscience researchers [44] have demonstrated that tailoring health behavior intervention materials increases the effectiveness of health messages.

The data confirm the viability of the app as a health learning tool. However, the majority of articles that describe this type of gain involve health students or health professionals [28]. For example, Khanna et al [45] used the app to optimize communication within a group of residents in the orthopedic sector. Graziano et al [46] used it for neurosurgeons in Italy, and Gulacti et al [47] used it in the emergency sector. Johnston et al [48] applied the app effectively among health professionals. Regarding patients and professionals, the app is described as a communication channel, but is not linked to an educational process, as observed in the report by Petruzzi and De Benedittis [49] (use of WhatsApp for oral health care) and by Krynski et al [50] (pediatricians and patients in Argentina). However, data exchange or tele-consultation is still a matter of debate, since messages exchanged in commercial apps represent a risk for data that should be more confidential [51]. According to Watson et al [52], in response to Johnston et al [48], even after the adoption of encryption technology, the app does not meet the standards required for clinical data transfer, and consequently, it is not recommended for use in a clinical environment. Even as an education and intervention tool, which was the case in this study, we understand that the app is a commercial platform with financial interests and that all interactions are converted into data to analyze user behavior patterns, which can be used in digital marketing strategies. Facebook Inc, the owner of the app, leads this commercial usage.

Regarding the advantages and disadvantages of the use of WhatsApp in health care settings, two positions seem to coexist in the scientific debate. First, one that exposes and underlines all of the positive aspects of the phenomenon as follows: improvement of communication, no requirement for a computer, time saving, possibility of an immediate response, improvement of health worker performance, reduction of consultation time, and smoothening of hierarchy between health groups. Second, one that highlights the negative aspects linked in particular to the clinical risks for patients, data security, and privacy protection [53].

In another promising direction, the app was used in social marketing campaigns to increase the uptake of free mammography for underprivileged women in Malaysia [54].

When compared with other types of media, the advantage of using this timeless tool is always mentioned. Web-based forums on websites are alternatives to debate about breast cancer [55,56], and even online social networks, such as Facebook, have been used before as an educational channel between women and health workers [57,58]. However, WhatsApp is a more pervasive, ubiquitous, and ready-to-hand tool than forums or websites. Before the popularity of this app, previous research in Brazil pointed out that television was the most used source by interviewees to acquire knowledge about breast cancer [59]. In 2018, WhatsApp was mentioned as a communication channel that 66.5% of women with cancer in Ecuador would like to use to receive information about the pathology [17].

In our case, the group dynamics and feedback shared by the users confirmed that the app is an informal environment of learning. During the intervention process, three women mentioned that they added value to the content and shared the knowledge and media to friends and relatives. This easy possibility to share content with the tool can also extend the impact of our intervention beyond the intended primary target [60]. The participants made the following comments:

> *I'm reading it, liking it and learning, and I'm already talking to my friends to share what I've learned.*
>
> *…I'm passing it on*
>
> *…Participating in this project was really good, besides from adding knowledge I also pass it on a lot to my relatives, Antonio.*

In addition to the informal environment, the diversity of women and the quality of grouping them (since they were not a group before the intervention) may have contributed to some conflicts during the intervention (see coexistence and self-regulation messages). There were 35 messages dedicated to self-regulation, where the participants negotiated what kind of content can be shared. Some of the messages were as follows:

> *Hi, what's up? I don't think it's cool to start forwarding chain messages in here. The group has a purpose, let's just talk about it. It's better this way*
>
> *Girls, I think everything is relevant, but if this group doesn't stay limited to scientific information about the subject matter, I will leave the group, sorry.*
>
> *The chain messages are good, but let's focus on our health.*

Deviation of the main theme is a natural phenomenon from a diverse group, and the dos and don'ts inside the community were expected to avoid information overload. However, WhatsApp allows copying and forwarding messages of other

groups, which has been described by Church and De Oliveira [61] and was perceived as a problem between some users in our educational intervention using the app.

During the intervention process, the creation of a friendly and harmonious interaction environment (a persistence online space) enabled the users to continually handle situations and slowly establish social norms and community parameters. They decided what could and could not be shared between them. In this case, the qualitative data showed two topics that emerged during the intervention, which were not directly related to the initial planned themes, as follows: (1) religious messages and (2) daily news. These deviation topics triggered a heated regulation discussion, causing the drop-out of three women from the group (nine during all the intervention processes). Although they belonged to the same neighborhood, the fact that the intervention was preventive was not enough to create a sense of identity as in studies involving patients already with the disease. It is relevant to remember that researchers who deal with online social networking emphasize that groups are assembled by common interests and not by geographic location [62,63]. Likewise, in the first face-to-face meeting, we observed that the initial moment in the intervention (when participants introduce themselves) was not able to connect them, and just like in-person groups, the social ties in an online environment also need interaction time to be effective [64].

In the first case of one of the themes shared among the participants (religious messages), the literature shows that cancer is usually and symbolically associated with death and religious topics [65]. Indeed, religious faith is praised as a provider of hope, optimism, and empowerment for diagnosed patients [66]. However, in our intervention, this topic was not related to breast cancer. An entire Bible quotation was sent in the group (cut-and-paste content of over 3000 characters), and some of the users did not enjoy the content or even found it boring. Another religious video was identified as fake (voice over the original content) by one of the women.

The other topic "daily news" occurred after the news "prisoners that escaped the Maringá prison." The concern about security was shared in the group, advising others to stay safe at home. For an alert (in the sense of "breaking news"), WhatsApp was perceived as a better tool to share information, overpassing the primary goal of the group.

Mammography

We believe that the topics "definitions," "examinations," and "mammography" corresponded to the content with the lowest score gain because they are of technical or instrumental nature and require more cognitive resources to memorize concepts, such as age range, conceptual differences, and high-risk criteria. A study performed with 914 women also identified limited knowledge about mammography [67]. The diversity of access to information about examinations disseminated by media could pass on information that may cause confusion among recipients, clarifying the central theme without making it easier to understand the related issues [67].

In our intervention, information was delivered with emphasis on the recommended age range for the examination and its technical characteristics. Exposure to breast cancer content traditionally puts women in a passive position [68]. Thus, the low gain in knowledge on this topic can be explained by the way we approached it, without rescuing the mammography experience of the participants [69].

In addition, a study analyzing the online search behavior of women on mammography identified that there is a greater interest during the month in which the national campaign called "Pink October" occurs every year [70]. In our case, WhatsApp enabled timeless contact about the theme, thus extending the talk about the disease beyond "Pink October."

Myths

Myths and truths about cancer received more interaction and more questions, and similarly, it was the topic that showed gain in the posttest period. We understand that health professionals preferably provide medical information to avoid misinformation. In our case, the tutor was responsible for managing the content by provoking discussions about it.

We did not find any studies that discussed women's knowledge gain related to the deconstruction of popular untruths common in daily life. Given the frequency at which these issues are debated among laypeople and the impact on the behaviors of individuals, such as wearing a bra, using deodorant, and using breast prosthesis, it is assumed that the level of interest and attention to this topic was a variable that contributed to the gain.

The doubts formulated by the participants can be compared as "fake news" contents, a key topic of discussion nowadays, which is alarming public health authorities because of its potential to spread quickly [71] and the requirement of an approach to counteract such misinformation [72]. WhatsApp has been accused of being a protagonist channel of misinformation, including during the presidential elections in Brazil [73]. Doubts revolved around topics seeking to understand breastfeeding, deodorant use, human papillomavirus vaccine, and obesity. However, some questions were straightforward as follows:

> *Can you say that the intake of pesticides/industrialized products can also influence?*

Additionally, there were concerns about protective factors as follows:

> *About the breastfeeding… is there a time for this protection? I breastfed my first child for 5 months and now I'm breastfeeding my second child, but I think I won't be able to do it for much more than 6 months...*

Sharing Personal Narratives

The cancer theme is indeed very complex. Doubts indicate the need to increase the health literacy of women, and even with the investments in health communication strategies during the month of the campaign "Pink October," frequent doubts remain.

"Pink October" is a set of communication strategies integrated by the NCI since 2010 as part of the National Breast Cancer Control Program, with adherence to several television programs. However, despite the national effort, it was noted in a study from 2013 that the campaign was able to legitimate an audience

(people were able to interpret and judge the message conveyed as positive and important), but was not effective for behavior change [74]. It is worth remembering that the increase in the number of cancer cases is due to a lack of regularity in prevention campaigns. Additionally, the fact that campaigns do not reach the entire country and the difficulties of representing and demystifying cancer as a fatal disease have been mentioned [68].

A positive moment of interaction was when participants shared family history of cancer. The narratives and descriptions were an opportunity for the mediator to later make some comments about risk factors. The stories were about close relatives as follows:

> *I also have cases in my family, my sister decided not to have children and had breast cancer, she found out in the beginning and treated it…*  *but she travels to Sao Paulo every 6 months to have a follow-up care!*

Even stories concerning people with no direct genetically bonds were shared as follows:

> *My mother-in-law had breast cancer 3 years ago, and has now been diagnosed with liver and lung metastasis*  .

Another woman recognized the risk factor by considering the genetical factor as follows:

> *I don't have any genetic factors or cases in the family, but I am already 55 years old and very sedentary. Take some action, walk!*  .

Additionally, previous stories of survival were shared as follows:

> *My maternal grandmother had it. Today she is 94 years old, this was 20 years ago.*
>
> *My mother had it with 45 years old. It's been 17 years.*

No death stories were shared. This aspect may have been influenced by the usage context of WhatsApp, which is seen by many users as more informal when compared with SMS text messaging. In a study by Church and De Oliveira [61], users reported that the tool allows a more fluid and open conversation.

Regarding the protective factors, before the intervention, 15 out 18 (83%) women only mentioned eating many fruits and vegetables, which did not change after the intervention. However, all other listed protective factors had a higher frequency of occurrence after the educational intervention, with breastfeeding being the most frequently mentioned protective factor (16/18, 89%). Not coincidentally, the participants formulated many interactions and questions about breastfeeding as a protective factor during the intervention.

Even with formulated doubts, the participants waited for expert clarification. None of the women in the group tried to solve the doubts or answer them. A more dialogical approach, in which women participate more and do not wait for an "expert voice," can add another perspective. Our intervention may have reinforced a receptive profile to messages rather than the creation of an active approach, recognizing the role in the search of information. These challenges had already been identified in previous group research, which pointed out that women preferred to expose themselves to content already created by health authorities (particularly in the figure of a doctor). This can be the result of the following two aspects: (1) women did not feel as part of a group or had the willingness to collaborate and try to find answers for themselves and (2) there is a passive posture for receiving information about cancer.

Despite the "Pink October" campaign in the country, breast cancer still represents a challenge for the public health sector, which needs more efficient screening techniques for early diagnosis. Thus, communication challenges and health education actions encourage alternative communication tools that add up to traditional campaigns. The use of WhatsApp is promising because it is a low investment tool and can act on a massive scale, especially in developing countries, such as Brazil, where there are higher rates, mainly due to the lack of early detection and adequate treatment [75].

Limitations

The authors acknowledge that the mediation that directed the content may have contributed to the maintenance of a passive posture by the participants on receiving messages. However, during the intervention, owing to the high necessity of the mediator to suggest and deliver content, we perceived a challenge to provide a top-down model of health information release and to place individuals as participants in the teaching process and health learning approach.

In addition, we recognize that a more expressive sample is important to perform a more meaningful analysis. Our initial analysis was based on a representative sample size of the target population of the study. However, participant loss was greater than expected. The reasons were diverse but consistent with the peculiarities of the investigated population. Nevertheless, we believe it is relevant to present the results of the statistical analysis, as they enable a better understanding of the phenomena that occurred throughout the intervention when integrated with the qualitative analysis.

Lastly, we highlight that even though knowledge gain can create a change in behavior, this is a multifactorial aspect of complex measurement. In our initial proposal, there was a desire to follow the participants in order to observe if the group would continue with new behaviors to reduce the risk of breast cancer after the end of the study. However, owing to the high dropout rate at the beginning of the study and recruitment difficulties, we chose to implement the analysis of knowledge gain.

Conclusions

This study explored the use of WhatsApp as a tool to facilitate knowledge exchange for the risk reduction and early detection of breast cancer between women and a content moderator. The results confirm our hypothesis that the app is a useful tool that can be incorporated into health education strategies that focus on breast cancer. Although the intervention improved the knowledge of the participants with regard to breast cancer, they expected that the moderator would send content and solve their doubts, suggesting a need for new strategies directed at the encouragement of dialogic communication. Nevertheless, the

results obtained demonstrate that WhatsApp is a feasible online space for women to seek answers for general doubts that are not covered in communication campaigns on breast cancer.

Acknowledgments

This work was supported by a research grant from Cesumar Institute of Science, Technology, and Innovation (ICETI) to MPB, LFG, and TFRL.

Conflicts of Interest

None declared.

References

1. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, CONCORD Working Group. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018 Mar 17;391(10125):1023-1075 [FREE Full text] [doi: 10.1016/S0140-6736(17)33326-3] [Medline: 29395269]
2. Knaul FM, Arreola-Ornelas H, Rodriguez NM, Méndez-Carniado O, Kwete XJ, Puentes-Rosas E, et al. Avoidable Mortality: The Core of the Global Cancer Divide. J Glob Oncol 2018 Jul;4:1-12 [FREE Full text] [doi: 10.1200/JGO.17.00190] [Medline: 30096010]
3. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global Cancer in Women: Burden and Trends. Cancer Epidemiol Biomarkers Prev 2017 Apr;26(4):444-457 [FREE Full text] [doi: 10.1158/1055-9965.EPI-16-0858] [Medline: 28223433]
4. Helvie MA, Bevers TB. Screening Mammography for Average-Risk Women: The Controversy and NCCN's Position. J Natl Compr Canc Netw 2018 Nov;16(11):1398-1404. [doi: 10.6004/jnccn.2018.7081] [Medline: 30442738]
5. Jensen M, Ejlertsen B, Mouridsen H, Christiansen P, Danish Breast Cancer Cooperative Group. Improvements in breast cancer survival between 1995 and 2012 in Denmark: The importance of earlier diagnosis and adjuvant treatment. Acta Oncol 2016 Jun;55 Suppl 2:24-35. [doi: 10.3109/0284186X.2015.1128119] [Medline: 26797010]
6. Morrell S, Taylor R, Roder D, Robson B, Gregory M, Craig K. Mammography service screening and breast cancer mortality in New Zealand: a National Cohort Study 1999-2011. Br J Cancer 2017 Mar 14;116(6):828-839 [FREE Full text] [doi: 10.1038/bjc.2017.6] [Medline: 28183141]
7. Gyawali B, Shimokata T, Honda K, Tsukuura H, Ando Y. Should low-income countries invest in breast cancer screening? Cancer Causes Control 2016 Nov;27(11):1341-1345. [doi: 10.1007/s10552-016-0812-8] [Medline: 27680017]
8. Schiller-Fruehwirth I, Jahn B, Einzinger P, Zauner G, Urach C, Siebert U. The Long-Term Effectiveness and Cost Effectiveness of Organized versus Opportunistic Screening for Breast Cancer in Austria. Value Health 2017 Sep;20(8):1048-1057 [FREE Full text] [doi: 10.1016/j.jval.2017.04.009] [Medline: 28964436]
9. Basu P, Ponti A, Anttila A, Ronco G, Senore C, Vale DB, et al. Status of implementation and organization of cancer screening in The European Union Member States-Summary results from the second European screening report. Int J Cancer 2018 Jan 01;142(1):44-56 [FREE Full text] [doi: 10.1002/ijc.31043] [Medline: 28940326]
10. Teh Y, Tan G, Taib NA, Rahmat K, Westerhout CJ, Fadzli F, et al. Opportunistic mammography screening provides effective detection rates in a limited resource healthcare system. BMC Cancer 2015 May 15;15:405 [FREE Full text] [doi: 10.1186/s12885-015-1419-2] [Medline: 25972043]
11. Altobelli E, Rapacchietta L, Angeletti P, Barbante L, Profeta F, Fagnano R. Breast Cancer Screening Programmes across the WHO European Region: Differences among Countries Based on National Income Level. Int J Environ Res Public Health 2017 Apr 23;14(4) [FREE Full text] [doi: 10.3390/ijerph14040452] [Medline: 28441745]
12. Islam R, Billah B, Hossain M, Oldroyd J. Barriers to Cervical Cancer and Breast Cancer Screening Uptake in Low-Income and Middle-Income Countries: A Systematic Review. Asian Pac J Cancer Prev 2017 Jul 27;18(7):1751-1763 [FREE Full text] [doi: 10.22034/APJCP.2017.18.7.1751] [Medline: 28749101]
13. Migowski A, Dias MB, Nadanovsky P, Silva GA, Sant'Ana DR, Stein AT. Diretrizes para detecção precoce do câncer de mama no Brasil. III - Desafios à implementação. Cad. Saúde Pública 2018 Jun 25;34(6). [doi: 10.1590/0102-311x00046317]
14. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth Assessment: Conceptualization of a Global Framework. JMIR Mhealth Uhealth 2017 May 02;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
15. Faiola A, Papautsky EL, Isola M. Empowering the Aging with Mobile Health: A mHealth Framework for Supporting Sustainable Healthy Lifestyle Behavior. Curr Probl Cardiol 2019 Aug;44(8):232-266. [doi: 10.1016/j.cpcardiol.2018.06.003] [Medline: 30185374]
16. Oreskovic NM, Huang TT, Moon J. Integrating mHealth and Systems Science: A Combination Approach to Prevent and Treat Chronic Health Conditions. JMIR Mhealth Uhealth 2015 Jun 02;3(2):e62 [FREE Full text] [doi: 10.2196/mhealth.4150] [Medline: 26036753]

17. Cherrez Ojeda I, Vanegas E, Torres M, Calderón JC, Calero E, Cherrez A, et al. Ecuadorian Cancer Patients' Preference for Information and Communication Technologies: Cross-Sectional Study. J Med Internet Res 2018 Feb 20;20(2):e50 [FREE Full text] [doi: 10.2196/jmir.8485] [Medline: 29463492]
18. Davis SW, Oakley-Girvan I. mHealth Education Applications Along the Cancer Continuum. J Cancer Educ 2015 Jun;30(2):388-394. [doi: 10.1007/s13187-014-0761-4] [Medline: 25482319]
19. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-Management Education Through mHealth: Review of Strategies and Structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771 [FREE Full text] [doi: 10.2196/10771] [Medline: 30341042]
20. 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]
21. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
22. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
23. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-Tested Mobile Apps for Breast Cancer Care: Systematic Review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
24. 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]
25. Holmen H, Wahl A, Torbjørnsen A, Jenum AK, Småstuen MC, Ribu L. Stages of change for physical activity and dietary habits in persons with type 2 diabetes included in a mobile health intervention: the Norwegian study in RENEWING HEALTH. BMJ Open Diabetes Res Care 2016;4(1):e000193 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000193] [Medline: 27239317]
26. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth App Intervention for Persons With Type 2 Diabetes and its Associations With Initial Self-Management: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
27. 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]
28. Ganasegeran K, Renganathan P, Rashid A, Al-Dubai SA. The m-Health revolution: Exploring perceived benefits of WhatsApp use in clinical practice. Int J Med Inform 2017 Jan;97:145-151. [doi: 10.1016/j.ijmedinf.2016.10.013] [Medline: 27919374]
29. WhatsApp - Features. WhatsApp. 2019. URL: https://www.whatsapp.com/features/ [accessed 2019-07-04]
30. Nelson R. The Top Mobile Apps, Games, and Publishers of Q1 2018. Sensor Towers Data Digest. 2018. URL: https://sensortower.com/blog/top-apps-games-publishers-q1-2018 [accessed 2018-10-16]
31. Montag C, Błaszkiewicz K, Sariyska R, Lachmann B, Andone I, Trendafilov B, et al. Smartphone usage in the 21st century: who is active on WhatsApp? BMC Res Notes 2015 Aug 04;8:331 [FREE Full text] [doi: 10.1186/s13104-015-1280-z] [Medline: 26238512]
32. Barhoumi C. The Effectiveness of WhatsApp Mobile Learning Activities Guided by Activty Theory on Students' Knowldege Management. CONTEMP EDUC TECHNOL 2015;6(3). [doi: 10.30935/cedtech/6151]
33. Nayak PP, Nayak SS, Sathiyabalan D, Aditya N, Das P. Assessing the Feasibility and Effectiveness of an App in Improving Knowledge on Oral Cancer-an Interventional Study. J Cancer Educ 2018 Dec;33(6):1250-1254. [doi: 10.1007/s13187-017-1239-y] [Medline: 28612324]
34. Cetinkaya L. The Impact of Whatsapp Use on Success in Education Process. IRRODL 2017 Nov 29;18(7). [doi: 10.19173/irrodl.v18i7.3279]
35. Alanzi T, Bah S, Alzahrani S, Alshammari S, Almunsef F. Evaluation of a mobile social networking application for improving diabetes Type 2 knowledge: an intervention study using WhatsApp. J Comp Eff Res 2018 Sep;7(9):891-899. [doi: 10.2217/cer-2018-0028] [Medline: 29943647]
36. Tang YH, Chong MC, Chua YP, Chui PL, Tang LY, Rahmat N. The effect of mobile messaging apps on cardiac patient knowledge of coronary artery disease risk factors and adherence to a healthy lifestyle. J Clin Nurs 2018 Dec;27(23-24):4311-4320. [doi: 10.1111/jocn.14538] [Medline: 29777560]
37. Sap S, Kondo E, Sobngwi E, Mbono R, Tatah S, Dehayem M, et al. Effect of patient education through a social network in young patients with type 1 diabetes in a Sub-Saharan context. Pediatr Diabetes 2019 May;20(3):361-365. [doi: 10.1111/pedi.12835] [Medline: 30779272]
38. Li WH, Ho KY, Lam KK, Wang MP, Cheung DY, Ho LL, et al. A study protocol for a randomised controlled trial evaluating the use of information communication technology (WhatsApp/WeChat) to deliver brief motivational interviewing (i-BMI)

in promoting smoking cessation among smokers with chronic diseases. BMC Public Health 2019 Aug 09;19(1):1083 [FREE Full text] [doi: 10.1186/s12889-019-7417-6] [Medline: 31399047]
39. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, et al. Cancer Incidence in Five Continents Vol XI. International Agency for Research on Cancer. 2017. URL: http://ci5.iarc.fr [accessed 2019-07-08]
40. Migowski A, Silva G, Dias M, Diz M, Sant'Ana D, Nadanovsky P. Diretrizes para detecção precoce do câncer de mama no Brasil. II - Novas recomendações nacionais, principais evidências e controvérsias. Cad. Saúde Pública 2018 Jun 21;34(6):1-15 [FREE Full text] [doi: 10.1590/0102-311x00074817]
41. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
42. Mayer RE. The promise of multimedia learning: using the same instructional design methods across different media. Learning and Instruction 2003 Apr;13(2):125-139 [FREE Full text] [doi: 10.1016/S0959-4752(02)00016-6]
43. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
44. Chua HF, Ho SS, Jasinska AJ, Polk TA, Welsh RC, Liberzon I, et al. Self-related neural response to tailored smoking-cessation messages predicts quitting. Nat Neurosci 2011 Apr;14(4):426-427 [FREE Full text] [doi: 10.1038/nn.2761] [Medline: 21358641]
45. Khanna V, Sambandam SN, Gul A, Mounasamy V. "WhatsApp"ening in orthopedic care: a concise report from a 300-bedded tertiary care teaching center. Eur J Orthop Surg Traumatol 2015 Jul;25(5):821-826. [doi: 10.1007/s00590-015-1600-y] [Medline: 25633127]
46. Graziano F, Maugeri R, Basile L, Guli C, Giugno A, Iacopino DG. WhatsAPP in neurosurgery: the best practice is in our hands. Acta Neurochir (Wien) 2017 Apr;159(4):601. [doi: 10.1007/s00701-017-3088-1] [Medline: 28144777]
47. Gulacti U, Lok U, Hatipoglu S, Polat H. An Analysis of WhatsApp Usage for Communication Between Consulting and Emergency Physicians. J Med Syst 2016 Jun;40(6):130. [doi: 10.1007/s10916-016-0483-8] [Medline: 27083574]
48. Johnston MJ, King D, Arora S, Behar N, Athanasiou T, Sevdalis N, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. Am J Surg 2015 Jan;209(1):45-51. [doi: 10.1016/j.amjsurg.2014.08.030] [Medline: 25454952]
49. Petruzzi M, De Benedittis M. WhatsApp: a telemedicine platform for facilitating remote oral medicine consultation and improving clinical examinations. Oral Surg Oral Med Oral Pathol Oral Radiol 2016 Mar;121(3):248-254. [doi: 10.1016/j.oooo.2015.11.005] [Medline: 26868466]
50. Krynski L, Goldfarb G, Maglio I. Technology-mediated communication with patients: WhatsApp Messenger, e-mail, patient portals. A challenge for pediatricians in the digital era. Arch Argent Pediatr 2018 Aug 01;116(4):e554-e559 [FREE Full text] [doi: 10.5546/aap.2018.eng.e554] [Medline: 30016032]
51. O'Sullivan DM, O'Sullivan E, O'Connor M, Lyons D, McManus J. WhatsApp Doc? BMJ Innov 2017 Oct 24;3(4):238-239. [doi: 10.1136/bmjinnov-2017-000239]
52. Watson L, Pathiraja F, Depala A, O'Brien B, Beyzade S. Ensuring safe communication in health care: a response to Johnston et al on their paper "Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams". Am J Surg 2016 Jan;211(1):302-303. [doi: 10.1016/j.amjsurg.2015.04.017] [Medline: 26184352]
53. De Benedictis A, Lettieri E, Masella C, Gastaldi L, Macchini G, Santu C, et al. WhatsApp in hospital? An empirical investigation of individual and organizational determinants to use. PLoS One 2019;14(1):e0209873 [FREE Full text] [doi: 10.1371/journal.pone.0209873] [Medline: 30633754]
54. Muniasamy M, Hashim HM, Lim C, Junazli NB, Choo S, Low K. Would You Like A Free Mammogram? Effectiveness of Utilizing WhatsApp As a Social Marketing Tool in Improving Uptake for a Free Mammography Service in Kuala Lumpur Malaysia. JGO 2018 Oct 01;4(Supplement 2):54s-54s. [doi: 10.1200/jgo.18.29100]
55. Beisecker AE, Hayes J, Ashworth JK, Reese PL. Providing information about breast cancer via public forums. Cancer Detect Prev 1997;21(4):370-379. [Medline: 9232329]
56. McRoy S, Rastegar-Mojarad M, Wang Y, Ruddy KJ, Haddad TC, Liu H. Assessing Unmet Information Needs of Breast Cancer Survivors: Exploratory Study of Online Health Forums Using Text Classification and Retrieval. JMIR Cancer 2018 May 15;4(1):e10 [FREE Full text] [doi: 10.2196/cancer.9050] [Medline: 29764801]
57. Abramson K, Keefe B, Chou WS. Communicating about cancer through Facebook: a qualitative analysis of a breast cancer awareness page. J Health Commun 2015;20(2):237-243. [doi: 10.1080/10810730.2014.927034] [Medline: 25495200]
58. Bender JL, Jimenez-Marroquin M, Jadad AR. Seeking support on facebook: a content analysis of breast cancer groups. J Med Internet Res 2011 Feb 04;13(1):e16 [FREE Full text] [doi: 10.2196/jmir.1560] [Medline: 21371990]
59. Godinho ER, Koch HA. Fontes utilizadas pelas mulheres para aquisição de conhecimentos sobre câncer de mama. Radiol Bras 2005 Jun;38(3):169-173. [doi: 10.1590/s0100-39842005000300004]
60. Jagosh J, Bush PL, Salsberg J, Macaulay AC, Greenhalgh T, Wong G, et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. BMC Public Health 2015 Jul 30;15:725 [FREE Full text] [doi: 10.1186/s12889-015-1949-1] [Medline: 26223523]
61. Church K, de Oliveira R. What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. In: Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services

- MobileHCI. New York, USA: Association for Computing Machinery; 2013 Presented at: 15th International Conference on Human-Computer Interaction with Mobile Devices and Services; August 27-30, 2013; Munich, Germany p. 352-361 URL: https://dl.acm.org/doi/10.1145/2493190.2493225 [doi: 10.1145/2493190.2493225]
62. Castells M. The Rise of the Network Society: Volume I: The Information Age: Economy, Society, and Culture. 2nd ed. West Sussex, UK: Blackwell Publishing; 2010.
63. Rheingold H. The Virtual Community: Homesteading on the Electronic Frontier. London, UK: MIT Press; 2000.
64. Turkle S. Life on the Screen: Identity in the Age of the Internet. New York, USA: Simon & Schuster; 1997.
65. Costa DT, Silva DM, Cavalcanti ID, Gomes ET, Vasconcelos JL, Carvalho MV. Religious/spiritual coping and level of hope in patients with cancer in chemotherapy. Rev Bras Enferm 2019 Jun 27;72(3):640-645 [FREE Full text] [doi: 10.1590/0034-7167-2018-0358] [Medline: 31269127]
66. Teixeira JJ, Lefèvre F. [The meaning of medical intervention and religious faith for the elderly cancer patient]. Cien Saude Colet 2008;13(4):1247-1256 [FREE Full text] [doi: 10.1590/s1413-81232008000400021] [Medline: 18813624]
67. Schneider IJ, Corseuil MW, Boing AF, d'Orsi E. Knowledge about mammography and associated factors: population surveys with female adults and elderly. Rev Bras Epidemiol 2013 Dec;16(4):930-942 [FREE Full text] [doi: 10.1590/s1415-790x2013000400013] [Medline: 24896598]
68. Ramos C, Coin de Carvalho JE, Corrêa Mangiacavalli MA. [Impact and (im)mobilization: a study of cancer prevention campaigns]. Cien Saude Colet 2007;12(5):1387-1396 [FREE Full text] [doi: 10.1590/s1413-81232007000500036] [Medline: 18813473]
69. Galpin A, Meredith J, Ure C, Robinson L. "Thanks for Letting Us All Share Your Mammogram Experience Virtually": Developing a Web-Based Hub for Breast Cancer Screening. JMIR Cancer 2017 Oct 27;3(2):e17 [FREE Full text] [doi: 10.2196/cancer.8150] [Medline: 29079555]
70. Vasconcellos-Silva PR, Carvalho DB, Trajano V, de La Rocque LR, Sawada AC, Juvanhol LL. Using Google Trends Data to Study Public Interest in Breast Cancer Screening in Brazil: Why Not a Pink February? JMIR Public Health Surveill 2017 Apr 06;3(2):e17 [FREE Full text] [doi: 10.2196/publichealth.7015] [Medline: 28385679]
71. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
72. Merchant RM, Asch DA. Protecting the Value of Medical Science in the Age of Social Media and "Fake News". JAMA 2018 Dec 18;320(23):2415-2416. [doi: 10.1001/jama.2018.18416] [Medline: 30453315]
73. Ferreira R. Rede de mentiras: a propagação de fake news na pré-campanha presidencial brasileira. OBS* 2018(Special Issue):139-162 [FREE Full text] [doi: 10.15847/obsobs12520181272]
74. Martins AF, Barbosa TR, Cezar LC. Análise da campanha Outubro Rosa de prevenção do câncer de mama em Viçosa, MG/Does the “Pink Ribbon” breast cancer prevention campaign in Viçosa, MG. Revista de Ciências Humanas 2014;14(2):539-556 [FREE Full text]
75. Anderson BO, Cazap E, El Saghir NS, Yip C, Khaled HM, Otero IV, et al. Optimisation of breast cancer management in low-resource and middle-resource countries: executive summary of the Breast Health Global Initiative consensus, 2010. Lancet Oncol 2011 Apr;12(4):387-398. [doi: 10.1016/S1470-2045(11)70031-6] [Medline: 21463833]

Abbreviations

NCI: (Brazilian) National Cancer Institute

Edited by G Eysenbach; submitted 13.12.19; peer-reviewed by L O'Riordan, K Ganasegeran; comments to author 25.01.20; revised version received 28.03.20; accepted 03.04.20; published 21.07.20.

Please cite as:
Pereira AAC, Destro JR, Picinin Bernuci M, Garcia LF, Rodrigues Lucena TF
Effects of a WhatsApp-Delivered Education Intervention to Enhance Breast Cancer Knowledge in Women: Mixed-Methods Study
JMIR Mhealth Uhealth 2020;8(7):e17430
URL: http://mhealth.jmir.org/2020/7/e17430/
doi:10.2196/17430
PMID:32706726

Original Paper

Improving a Web-Based Tool to Support Older Adults to Stay Independent at Home: Qualitative Study

Mirjam Marjolein Garvelink[1,2], PhD; Titilayo Tatiana Agbadjé[1,2,3], MSc; Adriana Freitas[1,2,3], PhD; Lysa Bergeron[1,2,4], BSc; Thomas Petitjean[5], BSc, MDM; Michèle Dugas[2,3,6], MSc; Louisa Blair[1,2,3], DPTh; Patrick Archambault[2,7,8,9,10], MSc, MD; Noémie Roy[1,2,11], BSc, MA; Allyson Jones[12], PhD; France Légaré[1,2,3,10], BSc, MSc, PhD, MD

[1]Canada Research Chair in Shared Decision Making and Knowledge Translation, Laval University, Québec, QC, Canada

[2]Centre de recherche en santé durable (VITAM), Québec, QC, Canada

[3]Centre intégré universitaire de santé et services sociaux de la Capitale-Nationale (CIUSSS-CN), Québec, QC, Canada

[4]Department of Social and Preventive Medicine, Faculty of Medicine, Laval University, Québec, QC, Canada

[5]Centre for Digital Media, Simon Fraser University, Vancouver, BC, Canada

[6]Health and Social Services Systems, Knowledge Translation and Implementation Component of the Quebec SPOR SUPPORT Unit, Laval University, Québec, QC, Canada

[7]Centre intégré de santé et de services sociaux de Chaudière-Appalaches, Ste-Marie, QC, Canada

[8]Centre de recherche intégré pour un système apprenant en santé et services sociaux, Lévis, QC, Canada

[9]Division of Critical Care Medicine, Department of Anesthesiology and Critical Care Medicine, Université Laval, Québec, QC, Canada

[10]Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Québec, QC, Canada

[11]School of Architecture, Faculty of Planning, Architecture, Arts and Design, Université Laval, Québec, QC, Canada

[12]Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
France Légaré, BSc, MSc, PhD, MD
Canada Research Chair in Shared Decision Making and Knowledge Translation
Laval University
2525, chemin de la Canardière
Québec, QC, G1J 0A4
Canada
Phone: 1 (418) 663 5919
Email: france.legare@fmed.ulaval.ca

Abstract

Background: Older adults desire to stay independent at home for as long as possible. We developed an interactive website to inform older adults and caregivers about ways to achieve this.

Objective: This study aimed to perform an in-depth exploration among potential end users about how to improve the interactive website to better inform older adults and caregivers about ways to stay independent at home.

Methods: To complement the results of a quantitative survey on the usability and acceptability of the website before implementation, we conducted a qualitative descriptive study. Using multiple recruitment strategies, we recruited a purposeful sample of older adults (aged ≥65 years) and caregivers of older adults struggling to stay independent at home. We conducted face-to-face or telephonic interviews in either English or French. In addition, we collected sociodemographic characteristics, other characteristics of participants (eg, health, digital profile, and perception of retirement homes), and experiences with using the website (factors facilitating the use of the website, barriers to its use, and suggestions for improvement). Interviews were audio recorded, transcribed verbatim, and thematically analyzed by two researchers.

Results: We recruited 15 participants, including 5 older adults (mean age 75 years, SD 6) and 10 caregivers (mean age 57 years, SD 14). The mean interview time was 32 min (SD 14). Most older adults had either mobility or health problems or both, and many of them were receiving home care services (eg, blood pressure measurement and body care). Overall, participants found the website easy to navigate using a computer, reassuring, and useful for obtaining information. Barriers were related to navigation (eg, difficult to navigate with a cellphone), relevance (eg, no specific section for caregivers), realism (eg, some resources presented are not state funded), understandability (eg, the actors' accents were difficult to understand), and accessibility (eg, not adapted for low digital literacy). Suggestions for improvement included a needs assessment section to direct users to the support appropriate

to their needs, addition of information about moving into residential care, a section for caregivers, distinction between state-provided and private support services, simpler language, expansion of content to be relevant to all of Canada, and video subtitles for the hearing impaired.

Conclusions: Users provided a wealth of information about the needs of older adults who were facing a loss of autonomy and about what such a website could usefully provide. The request for less generic and more personalized information reflects the wide range of needs that electronic health innovations, such as our interactive website, need to address. After integrating the changes suggested, the new website—Support for Older Adults to Stay Independent at Home (SUSTAIN)—will be implemented and made available to better assist older adults and caregivers in staying independent at home.

(JMIR Mhealth Uhealth 2020;8(7):e16979) doi:10.2196/16979

KEYWORDS

internet-based intervention; frail elderly; caregivers; decision making; personal autonomy; housing for the elderly

Introduction

Background

Worldwide, the proportion of older adults is increasing dramatically. In Canada, the number of adults aged 65 years and older is 17.5% of the population (2019), and by 2031, it will represent 22.7% of the population [1]. Currently, 92% of all Canadian adults aged 65 years and older live in private households [2], and most of them desire to stay independent at home for as long as possible. Studies have shown, however, that as age increases, functional and cognitive impairment, the presence of chronic diseases, a diminishing social network, and a low level of physical activity make staying at home very difficult [3-6]. As a consequence, at some point, older adults and their caregivers may face the decision about whether to remain at home (with or without assistance) or move to another location that better meets either their physical or social needs or both.

Multiple options for supporting older adults to remain independent at home are currently available [7-9]. For instance, regular home care or home visiting promotes health and delivers preventive care to older adults [8]. However, older adults and their caregivers are not always aware of these options [10,11]. This lack of knowledge can result in an hasty decision to relocate older adults, for example, to a nursing home. It is important that older adults be aware of all the options available to them—the advantages and disadvantages associated with each option—and that they can weigh up all this information and make informed decisions according to what matters most to them [12]. Research has shown that good levels of knowledge about services and support, as well as convenient housing, are associated with the likelihood of continuing to live in the community [13].

Technology can provide an easy and fast way to gain access to this information. In a previous study [14], we used a user-centered design with older adults, caregivers, and health professionals to develop an interactive decision support website called *Supporting Seniors and Caregivers to Stay Mobile at Home* (SPINACH) for older adults, caregivers, and health professionals in two Canadian provinces (Quebec and Alberta). We define *interactive* as providing a 2-way information flow between the user and the site [15]. The SPINACH website consists of 3 web pages: a home page, a video page, and a resource page [16]. On the home page, visitors have the option of selecting a language (English or French) and then the option of choosing whether to consult the video page or the resource page. The video page provides information on how different providers (1 video per provider type) can help older adults to stay independent at home. The resource page provides additional information (in text form) for staying independent at home [14]. Users can either watch only the videos relevant to their current needs or watch others if their needs change over time. They can also submit comments and information. According to an earlier quantitative usability survey, the website was deemed acceptable and potentially helpful for all kinds of end users [14]. However, it also required modifications, and we sought further insights from users as to what these might entail. In this study, we aimed to complement our quantitative survey results with a qualitative usability study to explore in-depth views of potential end users on how to improve the website, renamed Support for Older Adults to Stay Independent at Home (SUSTAIN), before its implementation.

Methods

Research Team

Our team is a multidisciplinary group of experts in shared decision making, primary care, rehabilitation, architecture, intensive care, and caregiving. We have been working together for 6 years on research studies aiming to develop tools and strategies to facilitate the engagement of older adults and caregivers in shared decision-making processes related to housing decisions [13,14,17-21].

Study Design and Context

We conducted a qualitative descriptive study to improve the content of an interactive website [14,22]. Initially, our study focused on 2 English-speaking provinces in western Canada (British Columbia and Alberta) and 2 provinces in eastern Canada (Ontario [English speaking] and Quebec [French speaking]). Due to recruiting difficulties within the allotted time, we extended our recruitment to another country (France) where older adults face similar issues regarding remaining independent at home [23,24]. Although the resource information referred to resources in Canada, we believed that the French participants could provide useful perspectives on the general usability of the website in a broader range of contexts. This study was approved by the Centre Hospitalier Universitaire de Quebec—Université Laval Ethics Committee (no 2018-3751)

and the University of Alberta Health Research Ethics Board (Pro00055678). We used the Consolidated Criteria for Reporting Qualitative Research checklist [25] to report the findings.

Participants

Eligibility Criteria

Older adults were eligible to participate if they were aged 65 years or older and cognitively capable of indicating their informed consent to participate in an individual interview (face-to-face or over the phone). We defined *cognitively capable* as not having been diagnosed with any disorder affecting reasoning. Formal or informal caregivers were eligible to participate in individual interviews (face-to-face or telephone) if they cared for an older adult struggling to remain independent at home.

Both older adults and caregivers had to be available to consult the SPINACH website before the interview. Participants were asked to navigate through the website and to explore each section at their own pace. No minimum consultation duration was defined.

Recruitment and Procedures

We used multiple recruitment strategies. We contacted 4 associations of caregivers and older adults (in Quebec and Alberta) to seek their support for disseminating recruitment information. They were asked to put an advertisement on their website or in their newsletter. One association never responded, 2 promised their support but did not follow through, and the fourth association posted the information on their website newsletter but without success. Finally, because of time constraints, we opted for recruiting through the social and professional networks of our research team and the snowballing method. Persons interested in participating in the study gave their first and last name, email, and telephone number to the person who recruited them, who, in turn, forwarded it to the project coordinator. One of the trained research assistants was then assigned to follow up with the participants by evaluating their eligibility and availability for the interview. The research assistant then emailed participants the link to the SPINACH website and instructed them to explore the website and watch the videos at their own pace and convenience at least one week before the interview date. We stressed the need to consult the website before the interview, whose purpose was to capture their experience of the SPINACH website and ask for suggestions for improving it. All participants were informed that they would receive financial compensation of Can $20 (US $14.89).

Sample Size

We recruited a purposeful sample of caregivers and older adults [26]. Guided by the model of information power, suggesting factors to be taken into account for sample size determination, we assumed that a small number of participants were needed to reach saturation for this study because (1) the aim of this study concerned a specific experience (of consultation and navigation on the SPINACH website) and participants with specific characteristics; (2) the interviewer was experienced and knew the website well, predicting a high quality of dialog; and (3) we planned to perform an in-depth exploration of narratives [27]. Other qualitative studies conducted in the older adult population relating to the decision to relocate [28], the opportunity to make independent decisions [29], and the use of a networking website [30] conducted their studies with relatively small samples (11, 12, and 6 participants, respectively) of the population of interest. Furthermore, using data from a study involving 60 in-depth interviews, Guest et al [31] found that data saturation occurred within the first 12 interviews, but the basic elements for meta-themes were present in as early as 6 interviews. Thus, we planned to recruit at least 12 participants, while making sure to reach data saturation.

Data Collection

We collected data through individual interviews (face-to-face or over the phone) from October 2017 to January 2018. Exceptionally, we conducted 2 dyadic interviews: (1) an older adult and his caregiver, in this case, his son, and (2) a couple, both of whom were caregivers who wished to be interviewed together. Face-to-face interviews were conducted in a room provided for this purpose in our research center or in a nearby affiliated center. Participants signed an informed consent form before the interview. Interviews were conducted by a trained female research assistant (MD, MSc in Public Health) and a trained male research assistant (TP, Master in Digital Media). Interviewers had no personal attachment to the SPINACH website and were open to all comments. We conducted the interviews in English or French, according to the preference of the participant. After each interview, the interviewers reported back to the coordinator with supporting field notes. On the basis of this, they decided whether data saturation had been reached [31].

At the beginning of the interview, the research assistant greeted the participant, introduced himself/herself, and reiterated the objectives of the interview. Each participant completed an individual sociodemographic questionnaire (Table 1). The interview grid was based on the results of a previous survey of end users on the usability and acceptability of the website [14], that is, the survey results provided the hypotheses regarding what elements of usability and acceptability should be further explored. The interview began by asking participants about their health profile, digital profile, sources of health information, and perceptions of long-term care facilities. After this, the topics discussed were related to their experience with the website and suggestions for improving it. Interviews were audio recorded with the consent of participants. The interviewer took notes on any additional relevant remarks made during each encounter.

Table 1. Characteristics of end users (N=15).

Characteristics	Values
Caregivers, n (%)	10 (67)
Age (years)	
Mean (SD)	56.9 (14)
Range	37-70
Sex, n (%)	
Female	6 (60)
Male	4 (40)
Education, n (%)	
University degree	10 (100)
Relationship with the older adult, n (%)	
Son/daughter	8 (80)
Spouse	1 (10)
Other	1 (10)
Older adults, n (%)	5 (34)
Age (years)	
Mean (SD)	74.6 (6)
Range	66-83
Sex, n (%)	
Female	4 (80)
Male	1 (20)
Education, n (%)	
High school diploma	1 (20)
College diploma	1 (20)
University degree	3 (60)
Older adults including those cared for by participating caregivers, n (%)	
Region/country	
Western Canada	4 (27)
Eastern Canada	9 (60)
Europe (France)	2 (13)
Setting	
Urban	9 (60)
Semiurban	2 (13)
Rural	4 (27)
Housing situation	
Home without home care	8 (53)
Home with medical home care	3 (20)
Private residence with services	4 (27)
Are facing a housing decision?	
Yes	5 (33)
No (the decision is already made)	5 (33)
No	4 (27)
Maybe	1 (7)

Data Analysis

Interviews were audio recorded and transcribed verbatim. The interview grid, based on the earlier survey results, provided the initial nodes. Overall, 2 authors independently performed deductive thematic analyses of verbatim transcripts [26,32] using qualitative data analysis software (NVivo version 12). First, the authors independently read the transcripts to familiarize themselves with the data. Each analyst proceeded with individual coding by refining and developing the pre-established nodes (including subnodes and node formulations). Afterward, coders met for a consensus meeting for 3 hours to cross-check their coding, analyze the nodes and the links between them, and categorize them. Discrepancies were discussed and resolved. The authors produced a report of relevant themes from the analysis and related quotations (Table 2 and Multimedia Appendix 1). Data saturation was reached for the presented themes.

Table 2. Factors facilitating the use of the website and illustrative quotations.

Theme and facilitators	Quotations
Navigation	
Navigation easy	"Oh. Piece of cake, really nice, it's really clear, big obvious menus. Cause in some you really have to hunt for the link you want." (Caregiver 8)
Information easy to find	"I found it useful, it was easy to manage, to find information, I would say." (Caregiver 9)
Relevance	
Helpful for decision making about housing	"It offers you a lot of links and people that you can talk to, to make the decision, because it's a difficult one. You know you'd be happy in your own home, but you're not safe there." (Caregiver 8)
Increases knowledge and potential for more	"The more relevant information you add that responds to people's immediate needs, the more useful it will be. I learned new things and I consider myself relatively educated." (Caregiver 5)[a]
Reassurance (about doing the right thing)	"To have the support of somebody else saying, yes you're doing the right thing and giving you places to look and people to talk to and that support you. Well it would have been really helpful for me." (Caregiver 8)
Reassurance (about others experiencing the same thing)	"It was nice, the comments that made you feel less guilty, like you can get tired of doing the cooking or that there are incontinence problems, it's normal. Everyone has those problems. It's a good way to reassure people." (Caregiver 5)[a]
Understandability	
Simple language	"I didn't find that very complicated. No." (Caregiver 2)[a]
Transcripts useful for the hearing impaired	"What I also found useful was that there were transcripts, you could see the video, so you can also read... because some people have hearing problems or they're not able to understand, they can read it as well." (Caregiver 9)
Interactive (videos)	"The interactive part like that, with the little videos—that draws people in...I'd never seen that before." (Caregiver 1)[a]
Realism	
True to life even in another country	"I'm in a French context but things are quite similar to what you have in Canada. The type of resources, the needs, it's the same. The organizations aren't exactly the same. But what we're looking for is the same, i.e. the help in the medical sector and all the more social things, like meals, home help, presence at home etc.; it's all there." (Caregiver 1)[a]

[a]Original in French.

Results

Quantifiers

We reported data using the graded quantifiers *few*, *some*, *many*, and *most* [33]. On the basis of a study by Chang et al [34], we used *few* when 1 or 2 participants commented on a theme, *some* when 3 to 5 commented, *many* when 6 to 9 commented, and *most* when 9 to 15 participants commented.

Sociodemographic Characteristics of the Participants

Between October 2017 and January 2018, we interviewed 15 end users: 60% (n=9/15) from eastern Canada, 27% (n=4/15) from western Canada, and 13% (n=2/15) from France. Two-thirds (n=10/15, 67%) of the participants were caregivers and 34% (n=5/15) were older adults. The mean interview time was 32 minutes and 35 seconds (SD 14).

Of the 5 older adults participating in the study, 60% (n=3/5) had a university degree with an average age of 75 years (SD 6). Most were living in urban areas (n=3/5, 60%,) and at home (n=4/5, 80%,). Of the 10 caregivers participating in the study, 60% (n=6/10) were women, with an average age of 57 years (SD 14); all were highly educated (n=10); and approximately 90% (n=9/10) were natural caregivers taking care of their parent (n=8) or their spouse (n=1).

Furthermore, 4 participants (27%) stated that a decision about whether to stay living at home or move to another place had already been made, whereas 5 participants (33%) were expecting to make this decision in the near future and one participant (7%) was possibly facing a housing decision (Table 1).

Other Characteristics of the Participants

Health Profile of Older Adults

Most older adults involved in this study, including those cared for by a participant caregiver, were currently living at home. Many older adults were receiving general home care services (eg, body care, walking aid, and grocery shopping aid). Some of them were receiving medical home care services (eg, blood samples, medication aid, and blood pressure measurement). Most older adults had either mobility or health problems or both. Some of them had started thinking about moving to another place because of autonomy loss. Some of them were experiencing difficulties related to this decision, for example:

> *I looked into it...I have a friend whose mother has been placed in a residence, and she told me what she had to go through and it sounds like a nightmare.* [Caregiver 6]

Digital Profiles of Participants

The most commonly used digital devices among participants were, in order of importance, computers, cellphones, and iPads. The search engine that they used most was Google. Some participants (whether older adults or caregivers) reported using these technologies daily. At the same time, a few participants reported not being comfortable with new technologies, for example:

> *A lot of seniors don't own a computer here, because it's like... why should we?...we've done our thing...we're old, we want to talk to people, you know.* [Older adult 5]

Sources of Information and Perceptions of Retirement Homes

Participants said that their main sources of information about options for staying independent at home or moving to a nursing home were, in order of importance, (1) local resources, such as community centers and health and social services; (2) their personal social network; (3) the internet; and (4) their health providers. Some participants had a negative impression of nursing homes. Negative impressions were linked to high costs, isolation, a restricted social environment, accounts of abuse in nursing homes, and loss of one's health care team, for example:

> *Some family physicians will no longer see a patient after they transfer to long term care...you know you're going to a new environment, but also your traditional healthcare team goes away too.* [Caregiver 10]

Experiences With the Website

Factors Facilitating the Use of the Website

Overall, many participants found the website helpful (eg, diversity of resources available with their contact information) for obtaining information about how to stay independent at home. Many participants found the website acceptable in terms of the content, especially the videos. They clarified that the presentation of the various scenarios in the videos was creative, reassuring (allowed them to recognize their own situation through the scenarios and the experience sharing), and helpful for understanding the roles of the various people who can help them. Many participants liked the length of the videos and found them a good way to present information. The participants mentioned specific aspects of the website that they liked, such as its clarity, its interactivity, the diversity of the resource people, and the ease of understanding the information. Some of them found the website easy to navigate when using a computer. Participants appreciated having the transcripts of the videos on the website, which they considered especially useful for people with hearing impairment (Table 2).

Factors Hindering the Use of the Website

Participants also discussed factors that could limit or hinder the use of the website (Multimedia Appendix 1). These factors and their solutions generally fell into the following categories: (1) navigation (eg, difficulties using a cellphone); (2) relevance (eg, insufficient information for caregivers or about cognitive impairment); (3) interactivity (eg, out-of-date information and dead links); (4) realism (eg, lack of ethnic diversity among actors); (5) understandability (eg, print too small and language too complex); (6) accessibility (eg, unwillingness to use computers); and (7) esthetics (eg, unattractive website design).

Proposed Modifications to the SPINACH Website

Participants proposed several improvements that could be made to the website for each of the categories. They suggested simplifying instructions on how to use the website. They suggested adding information for those deciding about a move to a nursing home, a specific section relevant to caregivers, information relevant to people with increasing cognitive impairment, and information about safety (eg, resources for people experiencing elder abuse). They also suggested better differentiation between public and private resources, more ethnic diversity among actors, shortening the videos, and adding subtitles for the hearing impaired. Full details of the barriers and proposed modifications with illustrative quotations are presented in Multimedia Appendix 1.

Other themes that emerged from the interviews were related to the types of care received at home (general and medical), the home care equipment used, and community resources available and used.

Discussion

Principal Findings

Aiming to improve an interactive website for older adults and caregivers developed in a previous study [14], we asked potential end users for in-depth feedback on how to improve the website to better address their needs related to staying independent at home. Overall, participants rated the SPINACH interactive website as a useful tool for helping them obtain information about options for staying independent at home. They also listed barriers to using the website (eg, information too generic and lack of a specific section for caregivers) and made several suggestions for improving its content. These results lead us to make the following observations.

Tailored Information

First, participants showed great interest in having tailored information. For instance, they wanted the option of specifying their city and province (eg, on a map) so that resources could be suggested based on their place of residence. This is consistent with other findings that older adults and caregivers want a personalized and flexible approach to their care (or the care of their loved ones) and their decision-making process, one that respects them as individuals [35] and provides support to help them prioritize the needs associated with their multiple conditions [36]. Older adults' autonomy changes over time [17,19], and thus, their need for information changes too. Indeed, in line with previous research, although the SPINACH website focuses on the many options for staying independent at home, some of our participants requested more information about moving to a nursing home [37]. A decision support tool has been developed specifically to support older adults in the decision-making process about housing options [18], which could be integrated into the website to meet the needs of older adults making other choices than to stay at home.

Supporting Caregivers

Second, the website does not yet have a specific section for caregivers. Although caregiver participants found the website helpful, they also wanted to know how to find the help they might need for themselves when caring for an older adult losing autonomy. They were also interested in knowing how to manage difficult conversations with their loved ones. This confirms the results of a recent study showing that it is difficult for caregivers to find a balance between the needs of their loved ones and their own needs [38]. In fact, this factor is associated with the burden of care felt by family caregivers caring for an older adult facing housing decisions [39]. According to a literature review about caregiver involvement in any decision with their loved one, caregivers often feel uninformed and unsupported in making informed decisions congruent with their personal values, which leads to negative feelings after decision making [40]. Moreover, a recent study showed that caregivers experience more decisional conflict than their loved ones and the same level of decisional regret [21]. Thus, creating a section for caregivers on the website with information to help them face the challenges of caregiving (eg, better address the decision-making needs of their loved ones), including navigating the health care system to get the help they need [41], will reduce the risk of decisional conflict and regret. This will also indirectly help older adults get the care they need, especially if they are cognitively impaired.

The Challenge of Balancing Tailoring With Scaling Up

Third, participants' comments raised interesting issues relevant to the challenges of scaling up. Although wanting the information to be more tailored, they also suggested expanding the website to be relevant to a wider population, for example, different user groups in different geographical areas, and to different decision-making needs. The unexpected French participants in this study helped us understand that French older adults and caregivers face challenges similar to those faced by Canadian older adults and caregivers in terms of housing decision making and homecare. This finding as well as the growing potential of health care technologies to deal with complex choices suggest that our platform could be scaled up to the rest of Canada and other developed countries with aging populations [42]. In the future, we plan to improve the SPINACH website in line with participants' comments. We also plan to integrate a GPS tracker system that will provide real-time information on older adults' outdoor mobility to promote self-management and to support them in staying at home for as long as possible. In addition, the website could further respond to users' concerns by integrating collaborative writing apps for adding regional sources [43]. Other methods, such as algorithms for self-assessment that direct users to appropriate resources, and assigning a webmaster to update the website weekly, could also be explored.

Limitations

Our study has some limitations. We used various strategies to recruit participants; however, our sample size was small, and some populations were not well represented, notably those with lower education levels and older adults with cognitive impairment, who could have provided a different perspective for the future format of the website. Some interviews were interrupted, and we had to contact the participants again. In addition, the sound quality was sometimes poor when interviews had to be conducted over the telephone, which had an impact on the quality of the interview. However, field notes were taken by the interviewers, which filled the gaps. Finally, we did not specify a minimum consultation time for the website or track the time participants spent on the website, so it is possible that some participants did not consult every section of the site, which may have affected their evaluation.

Conclusions

We consulted end users to improve SPINACH, an interactive website for older adults and caregivers about options for staying independent at home [14]. Users provided a wealth of information on what such a website could usefully provide. The request for less generic and more personalized information reflects the wide range of needs such a website needs to address and raises the technological issues about how to achieve this in all electronic health innovations. Requests for more information for caregivers reflect the key role of caregivers in sustaining older people independently at home. Suggestions for more information on residential care suggest that needs can change quickly and that people prefer to be informed about the full range of options. Once the suggested changes have been made, the new SUSTAIN website will be implemented in Canada. Subsequently, a thorough analysis of the scalability of this innovation is required before it can be adapted to other contexts and cultures.

Acknowledgments

This work was funded by the Improving Cognitive and Joint Health Network of the Canadian Institute of Health Research. FL holds a Tier 1 Canada Research Chair in shared decision making and knowledge translation. MMG has received a postdoctoral fellowship from the Canadian Institutes of Health Research, funding reference number MFE-140842.

Authors' Contributions

The study was conceived and designed by PA, L Blair, MMG, AJ, FL, and NR. The project was administered and data collection was coordinated by AF and MMG. Data were collected and analyzed by TTA, L Bergeron, MD, and TP. Formal analysis and supervision were performed by AF, MMG, and FL. The paper was drafted by AF and TTA and was reviewed, edited, and approved for submission by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Barriers to use of the website, suggestions for improvement, and illustrative quotations.
[DOCX File , 35 KB - mhealth_v8i7e16979_app1.docx]

References

1. Canada's Population Estimates: Age and Sex, July 1, 2019. Data - Statistics Canada. 2012. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/190930/dq190930a-eng.htm [accessed 2020-02-28]
2. Study: Emerging Trends in Living Arrangements and Conjugal Unions for Current and Future Seniors, 1981 to 2011. Data - Statistics Canada. 2014. URL: http://www.statcan.gc.ca/daily-quotidien/140224/dq140224a-eng.htm [accessed 2020-02-28]
3. Perissinotto CM, Stijacic Cenzer I, Covinsky KE. Loneliness in older persons: a predictor of functional decline and death. Arch Intern Med 2012 Jul 23;172(14):1078-1083 [FREE Full text] [doi: 10.1001/archinternmed.2012.1993] [Medline: 22710744]
4. Luppa M, Luck T, Weyerer S, König HH, Brähler E, Riedel-Heller SG. Prediction of institutionalization in the elderly. A systematic review. Age Ageing 2010 Jan;39(1):31-38. [doi: 10.1093/ageing/afp202] [Medline: 19934075]
5. Gaugler J, Duval S, Anderson K, Kane R. Predicting nursing home admission in the US: a meta-analysis. BMC Geriatr 2007 Jun 19;7:13 [FREE Full text] [doi: 10.1186/1471-2318-7-13] [Medline: 17578574]
6. Stuck AE, Walthert JM, Nikolaus T, Büla CJ, Hohmann C, Beck JC. Risk factors for functional status decline in community-living elderly people: a systematic literature review. Soc Sci Med 1999 Feb;48(4):445-469. [doi: 10.1016/s0277-9536(98)00370-0] [Medline: 10075171]
7. Kim H, Ahn YH, Steinhoff A, Lee KH. Home modification by older adults and their informal caregivers. Arch Gerontol Geriatr 2014;59(3):648-656. [doi: 10.1016/j.archger.2014.07.012] [Medline: 25109810]
8. Elkan R, Kendrick D, Dewey M, Hewitt M, Robinson J, Blair M, et al. Effectiveness of home based support for older people: systematic review and meta-analysis. Br Med J 2001 Sep 29;323(7315):719-725 [FREE Full text] [doi: 10.1136/bmj.323.7315.719] [Medline: 11576978]
9. Beswick AD, Rees K, Dieppe P, Ayis S, Gooberman-Hill R, Horwood J, et al. Complex interventions to improve physical function and maintain independent living in elderly people: a systematic review and meta-analysis. Lancet 2008 Mar 1;371(9614):725-735 [FREE Full text] [doi: 10.1016/S0140-6736(08)60342-6] [Medline: 18313501]
10. Publications du Ministère de la Santé et des Services Sociaux. 2008. URL: https://publications.msss.gouv.qc.ca/msss/fichiers/ainee/F-5149-MSSS.pdf [accessed 2020-02-28]
11. Lord K, Livingston G, Robertson S, Cooper C. How people with dementia and their families decide about moving to a care home and support their needs: development of a decision aid, a qualitative study. BMC Geriatr 2016 Mar 21;16:68 [FREE Full text] [doi: 10.1186/s12877-016-0242-1] [Medline: 27001704]
12. Légaré F, O'Connor AC, Graham I, Saucier D, Côté L, Cauchon M, et al. Supporting patients facing difficult health care decisions: use of the Ottawa decision support framework. Can Fam Physician 2006 Apr;52:476-477 [FREE Full text] [Medline: 17327891]
13. Roy N, Dubé R, Després C, Freitas A, Légaré F. Choosing between staying at home or moving: a systematic review of factors influencing housing decisions among frail older adults. PLoS One 2018;13(1):e0189266 [FREE Full text] [doi: 10.1371/journal.pone.0189266] [Medline: 29293511]
14. Garvelink MM, Jones CA, Archambault PM, Roy N, Blair L, Légaré F. Deciding how to stay independent at home in later years: development and acceptability testing of an informative web-based module. JMIR Hum Factors 2017 Dec 14;4(4):e32 [FREE Full text] [doi: 10.2196/humanfactors.8387] [Medline: 29242178]

15. Aoki K. Taxonomy of Interactivity on the Web. Semantic Scholar. 2000. URL: https://pdfs.semanticscholar.org/ed39/6d88541685ec738d1b9f5ac859e377036aa7.pdf?_ga=2.131429962.435462905.1591382291-732732824.1591382291 [accessed 2020-02-28]
16. What Can I Do to Stay Independent at Home as Long as Possible? SPINACH Project. 2016. URL: https://spinach-project.blogspot.com/ [accessed 2020-02-28]
17. Légaré F, Brière N, Stacey D, Bourassa H, Desroches S, Dumont S, et al. Improving decision making on location of care with the frail elderly and their caregivers (the DOLCE study): study protocol for a cluster randomized controlled trial. Trials 2015 Feb 12;16:50 [FREE Full text] [doi: 10.1186/s13063-015-0567-7] [Medline: 25881122]
18. Garvelink MM, Emond J, Menear M, Brière N, Freitas A, Boland L, et al. Development of a decision guide to support the elderly in decision making about location of care: an iterative, user-centered design. Res Involv Engagem 2016;2:26 [FREE Full text] [doi: 10.1186/s40900-016-0040-0] [Medline: 29062524]
19. Légaré F, Brière N, Stacey D, Lacroix G, Desroches S, Dumont S, et al. Implementing shared decision-making in interprofessional home care teams (the IPSDM-SW study): protocol for a stepped wedge cluster randomised trial. BMJ Open 2016 Nov 24;6(11):e014023 [FREE Full text] [doi: 10.1136/bmjopen-2016-014023] [Medline: 27884857]
20. Boland L, Légaré F, Perez MM, Menear M, Garvelink MM, McIsaac DI, et al. Impact of home care versus alternative locations of care on elder health outcomes: an overview of systematic reviews. BMC Geriatr 2017 Jan 14;17(1):20 [FREE Full text] [doi: 10.1186/s12877-016-0395-y] [Medline: 28088166]
21. Adekpedjou R, Stacey D, Brière N, Freitas A, Garvelink MM, Turcotte S, et al. 'Please listen to me': a cross-sectional study of experiences of seniors and their caregivers making housing decisions. PLoS One 2018;13(8):e0202975 [FREE Full text] [doi: 10.1371/journal.pone.0202975] [Medline: 30161238]
22. Witteman HO, Dansokho SC, Colquhoun H, Coulter A, Dugas M, Fagerlin A, et al. User-centered design and the development of patient decision aids: protocol for a systematic review. Syst Rev 2015 Jan 26;4:11 [FREE Full text] [doi: 10.1186/2046-4053-4-11] [Medline: 25623074]
23. Lord S, Després C, Ramadier T. When mobility makes sense: a qualitative and longitudinal study of the daily mobility of the elderly. J Environ Psychol 2011 Mar;31(1):52-61 [FREE Full text] [doi: 10.1016/j.jenvp.2010.02.007]
24. Lord S, Luxembourg N. The mobility of elderly residents living in suburban territories. J Hous Elderly 2007 Mar 12;20(4):103-121 [FREE Full text] [doi: 10.1300/j081v20n04_07]
25. 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. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
26. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2015.
27. Malterud K, Siersma V, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
28. Peyrovi H, Mortazavi H, Joolaee S. Challenges of Iranian family members in decision making to put elderly people with chronic diseases in nursing homes: a qualitative study. Arch Sci 2012;65(10):178-188 [FREE Full text]
29. Breitholtz A, Snellman I, Fagerberg I. Older people's dependence on caregivers' help in their own homes and their lived experiences of their opportunity to make independent decisions. Int J Older People Nurs 2013 May;8(2):139-148 [FREE Full text] [doi: 10.1111/j.1748-3743.2012.00338.x]
30. Ballantyne A, Trenwith L, Zubrinich S, Corlis M. 'I feel less lonely': what older people say about participating in a social networking website. Qual Age Older Adult 2010 Sep 13;11(3):25-35 [FREE Full text] [doi: 10.5042/qiaoa.2010.0526]
31. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
32. Ritchie J, Lewis J, Nicholls CM, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Thousand Oaks, CA: Sage Publications; 2019.
33. Sandelowski M. Real qualitative researchers do not count: the use of numbers in qualitative research. Res Nurs Health 2001 Jun;24(3):230-240. [doi: 10.1002/nur.1025]
34. Chang Y, Voils CI, Sandelowski M, Hasselblad V, Crandell JL. Transforming verbal counts in reports of qualitative descriptive studies into numbers. West J Nurs Res 2009 Nov;31(7):837-852 [FREE Full text] [doi: 10.1177/0193945909334434] [Medline: 19448052]
35. Ottmann G, Allen J, Feldman P. A systematic narrative review of consumer-directed care for older people: implications for model development. Health Soc Care Community 2013 Nov;21(6):563-581 [FREE Full text] [doi: 10.1111/hsc.12025]
36. Boeckxstaens P, de Graaf P. Primary care and care for older persons: position paper of the European forum for primary care. Qual Prim Care 2011;19(6):369-389 [FREE Full text] [Medline: 22340900]
37. Légaré F, Stacey D, Brière N, Robitaille H, Lord M, Desroches S, et al. An interprofessional approach to shared decision making: an exploratory case study with family caregivers of one IP home care team. BMC Geriatr 2014 Jul 2;14:83 [FREE Full text] [doi: 10.1186/1471-2318-14-83] [Medline: 24985335]

38. Liu H, Yang C, Wang Y, Hsu W, Huang T, Lin Y, et al. Balancing competing needs mediates the association of caregiving demand with caregiver role strain and depressive symptoms of dementia caregivers: a cross-sectional study. J Adv Nurs 2017 Dec;73(12):2962-2972 [FREE Full text] [doi: 10.1111/jan.13379] [Medline: 28816364]
39. Boucher A, Haesebaert J, Freitas A, Adekpedjou R, Landry M, Bourassa H, et al. Time to move? Factors associated with burden of care among informal caregivers of cognitively impaired older people facing housing decisions: secondary analysis of a cluster randomized trial. BMC Geriatr 2019 Sep 9;19(1):249 [FREE Full text] [doi: 10.1186/s12877-019-1249-1] [Medline: 31500590]
40. Garvelink MM, Ngangue PA, Adekpedjou R, Diouf NT, Goh L, Blair L, et al. A synthesis of knowledge about caregiver decision making finds gaps in support for those who care for aging loved ones. Health Aff (Millwood) 2016 Apr;35(4):619-626 [FREE Full text] [doi: 10.1377/hlthaff.2015.1375] [Medline: 27044961]
41. Eloranta S, Arve S, Isoaho H, Welch A, Viitanen M, Routasalo P. Perceptions of the psychological well-being and care of older home care clients: clients and their carers. J Clin Nurs 2010 Mar;19(5-6):847-855 [FREE Full text] [doi: 10.1111/j.1365-2702.2009.02994.x] [Medline: 20500328]
42. Charif AB, Hassani K, Wong ST, Zomahoun HT, Fortin M, Freitas A, et al. Assessment of scalability of evidence-based innovations in community-based primary health care: a cross-sectional study. CMAJ Open 2018;6(4):E520-E527 [FREE Full text] [doi: 10.9778/cmajo.20180143] [Medline: 30389751]
43. Archambault P, van de Belt TH, Kuziemsky C, Plaisance A, Dupuis A, McGinn C, et al. Collaborative writing applications in healthcare: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2017 May 10;5:CD011388 [FREE Full text] [doi: 10.1002/14651858.CD011388.pub2] [Medline: 28489282]

Abbreviations

SPINACH: Supporting Seniors and Caregivers to Stay Mobile at Home
SUSTAIN: Support for Older Adults to Stay Independent at Home

Edited by G Eysenbach; submitted 08.11.19; peer-reviewed by M Khetani, Y Zheng, M Kolotylo-Kulkarni; comments to author 06.01.20; revised version received 28.02.20; accepted 14.05.20; published 22.07.20.

Please cite as:
Garvelink MM, Agbadjé TT, Freitas A, Bergeron L, Petitjean T, Dugas M, Blair L, Archambault P, Roy N, Jones A, Légaré F
Improving a Web-Based Tool to Support Older Adults to Stay Independent at Home: Qualitative Study
JMIR Mhealth Uhealth 2020;8(7):e16979
URL: https://mhealth.jmir.org/2020/7/e16979
doi:10.2196/16979
PMID:32412908

Original Paper

Evaluating the Effects of a Mobile Health App on Reducing Patient Care Needs and Improving Quality of Life After Oral Cancer Surgery: Quasiexperimental Study

Tze-Fang Wang[1], RN, MSN; Rou-Chen Huang[1], RN, MSN; Su-Chen Yang[2]; Chyuan Chou[3], DDS, MD, PhD; Lee-Chen Chen[2], RN, MSN

[1]School of Nursing, National Yang-Ming University, Taipei City, Taiwan

[2]Department of Nursing, Far Eastern Memorial Hospital, New Taipei City, Taiwan

[3]Excellent Dental Center, Taipei City, Taiwan

Corresponding Author:
Lee-Chen Chen, RN, MSN
Department of Nursing
Far Eastern Memorial Hospital
No. 21, Section 2, Nanya South Road
Banqiao District
New Taipei City, 220
Taiwan
Phone: 886 2 77281610
Email: leegan@mail.femh.org.tw

Abstract

Background: Intervention with a mobile Health (mHealth) app can improve the efficacy of early detection of oral cancer and the outcomes for patients taking oral anticancer medications. The quality of life of oral cancer patients is significantly reduced within three months after surgery; also, their needs for nursing care and health information increase, mainly due to side effects and associated psychological problems.

Objective: This study aimed to evaluate changes in the care needs and quality of life of patients with oral cancer after receiving the intervention of a newly developed mHealth app.

Methods: After surgery, oral cancer patients were divided into an experimental group (n=50) who received the mHealth app intervention and a control group (n=50) who received routine health care and instruction. After 3 months of intervention, survey questionnaires were used to assess the patients' quality of life, nursing care needs, and acceptance of the mHealth app.

Results: The physiological care needs were significantly decreased in the experimental group compared with the control group (P<.05). Although the differences were not statistically significant, the psychological needs, communication needs, and care support needs all improved after the mHealth app intervention. The overall improvement in quality of life was higher in the experimental group than in the control group (–7.24 vs –4.36). In terms of intention to use, perceived usefulness, and perceived ease of use, the acceptability scores of the mHealth app were significantly increased after 3 months of intervention (P<.05).

Conclusions: Compared with routine health care and instruction, for patients after surgery, the education/information intervention using the mHealth app significantly reduced their nursing care needs, improved their quality of life, and increased their acceptance of using an mHealth app on a mobile device. These findings can provide a theoretical basis for future health care app design and improvement. This study suggests that an mHealth app should be incorporated into the routine care of oral cancer patients to provide medical information quickly and improve their self-management abilities, thereby reducing the patients' need for physiological care and improving their quality of life.

Trial Registration: ClinicalTrials.gov NCT04049968; https://www.clinicaltrials.gov/ct2/show/NCT04049968

(JMIR Mhealth Uhealth 2020;8(7):e18132) doi:10.2196/18132

KEYWORDS

care needs; health information; mobile health app; oral cancer; technology acceptance; quality of life

Introduction

Oral cancer ranks sixth in incidence among cancers worldwide; among men, it ranks eighth in incidence rate (3.8/100,000) and mortality [1]. The World Health Organization (WHO) estimated that 657,000 new cases of oral cancer are diagnosed each year and that more than 330,000 people die annually from this disease [2]. Oral cancer is a serious issue in Taiwan; it is ranked fifth in number of cancer deaths and fourth among men in cancer morbidity and mortality. Indeed, the morbidity and mortality rates of oral cancer among men are almost 10 times those among women [3]. Although the five-year survival rate for oral cancer has increased to 70% to 80% [4,5], oral cancer patients still experience comorbidities, psychological distress, and reduced quality of life [6-8].

A prospective survey analyzed the quality of life of 83 oral cancer patients and found that physiological status (such as fatigue, pain, nausea, and vomiting symptoms) and physical activities (such as resuming sustainable work and leisure activities) significantly deteriorated within three months after surgery [9]. The patients' social functioning, body image, and financial status also declined significantly after surgery. The incidence of anxiety/depression in patients with oral cancer after surgery reached 25% in patients surveyed within six months to six years [10]. Furthermore, pain, facial changes, and social activity were negatively correlated with psychological outcomes and quality of life. Even with the use of neoadjuvant treatments and observed improvement in the first year after surgery, the quality of life of patients within three months of surgery remained significantly reduced [11]. Therefore, there is still an urgent need for new approaches to improve the quality of life of patients with oral cancer after surgery.

The decline in the quality of life of patients after oral cancer treatment increases the demand for nursing care and health information, mainly due to the side effects and resulting psychological problems after treatment [12]. For example, compared with cancer patients who did not receive radiotherapy, cancer patients who received radiotherapy may report that their care needs and communication needs are not being met [13]. Similarly, cancer patients receiving chemotherapy may experience nausea, fatigue, decreased physical function, and emotional problems, all of which require care and attention [14]. Postoperative oral cancer patients have a series of nursing needs, including psychological needs, professional medical care needs, and particularly support needs in terms of interpersonal communication, including disease information, treatment options, and medication options, such as pain management [15-17]. However, most health-related information in manuals or books fails to meet these needs because the desired information may take too long to find and the level of writing may be too specialized for patients to understand. Instead, patients may want to quickly obtain specific information at any time and place [18]. When patients receive sufficient support to meet their needs, they can more effectively cope with their negative emotions and disease symptoms.

The popularity in many countries of smart devices such as mobile phones and tablets has led to increasing use of mobile health (mHealth) apps to quickly and efficiently transmit medical information and provide health care services to patients [19]. The WHO Global Observatory for eHealth defines mHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices [20]." An mHealth app is a software program that runs on mobile phones, tablets, or other mobile devices (eg, smartwatches, wristbands) for health care and disease prevention [21]. In addition to monitoring health, an mHealth app can encourage healthy behaviors and provide patients with effective ways to manage disease [22,23]. Several mHealth apps have been developed for disease management approaches such as blood glucose control (diabetes) [24], hypertension control [25], depression treatment [26], remote cancer surveillance [27], and medication monitoring [28,29].

In this study, we developed an mHealth app for patients after oral cancer surgery. To the best of our knowledge, no study has investigated whether an intervention based on an mHealth app can affect the daily needs and quality of life of patients after oral cancer surgery. This study aimed to investigate whether the medical information and education provided through the developed mHealth app can reduce the care needs of patients and improve their quality of life after the intervention. We believe that patients and their families may benefit from this convenient physical and social support system, which enables patients to quickly obtain information to help them cope with their disease, reduce their anxiety, and improve their quality of life.

Methods

Study Design and Sample

This study used a quasiexperimental research design and convenience sampling. The study protocol was reviewed and approved by the internal review board of the Research Ethics Review Committee of Far Eastern Memorial Hospital (No. 105110-E). Study participants were recruited from the departments of Oral Surgery, Otolaryngology, Hematology Oncology, and Radiation Oncology at the Far Eastern Memorial Hospital in New Taipei City, Taiwan. The inclusion criteria were patients who were diagnosed with oral cancer by a physician and who underwent oral cancer surgery within one week; patients who were conscious and able to communicate in Mandarin or Taiwanese; and patients who agreed to participate in the study and possessed a smartphone. The exclusion criteria were patients who did not have oral cancer or patients who had oral cancer but underwent oral cancer surgery one week previously; patients who could not communicate in Mandarin or Taiwanese; patients who were unconscious or unable to answer the questions in the questionnaire; and patients who had cognitive impairment, dementia, or intellectual disability. After the well-trained researchers explained the study to the eligible participants who met the inclusion criteria, the participants who agreed and who provided signed informed consent were included in the study. Each patient was individually instructed by the investigator prior to discharge, and the content of the instruction focused on

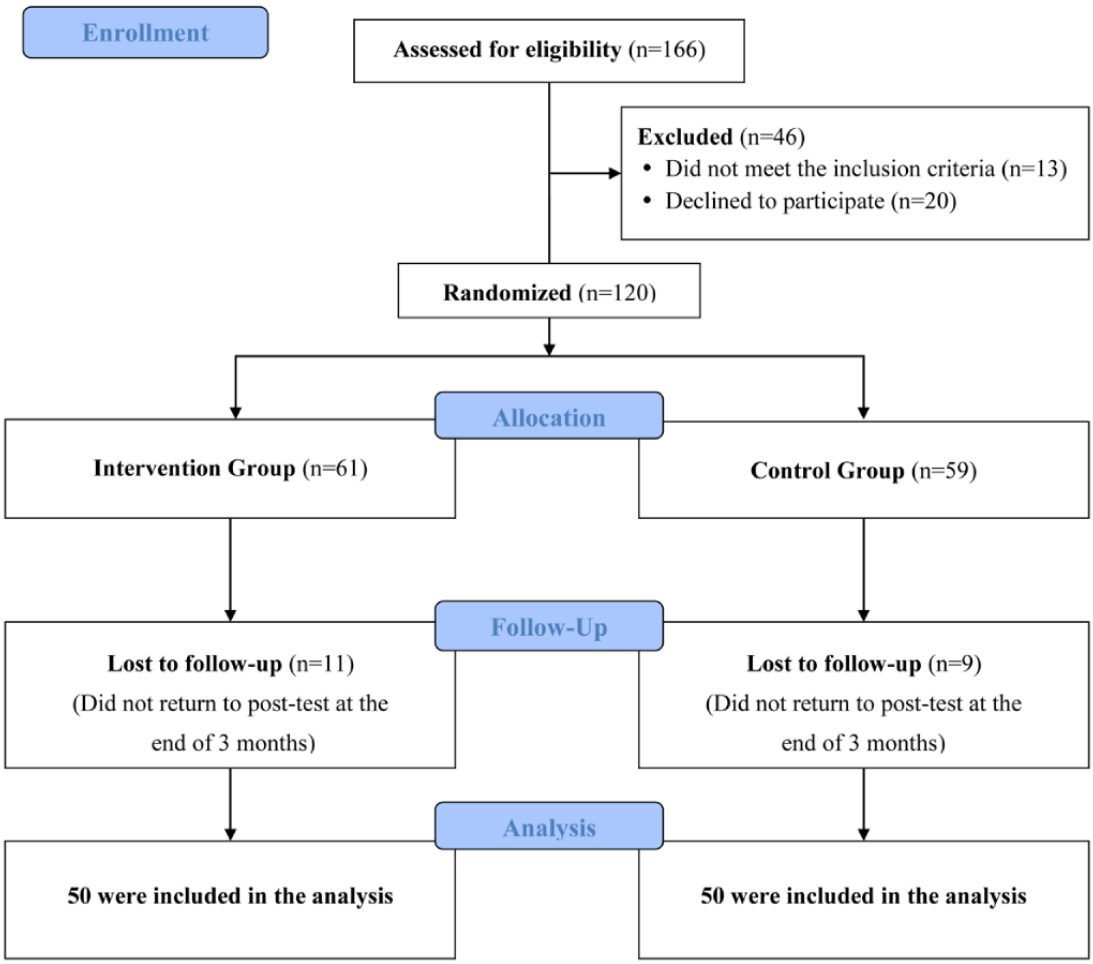
treatment messages, oral care, and social resource delivery. The intervention outcomes of the patients in both groups were surveyed after 3 months using questionnaires. The study was registered prior to launch (ClinicalTrials.gov NCT04049968).

Data Collection and Features of the mHealth App

All included patients were randomly divided into an experimental group and a control group (Figure 1). The patients in the control group received routine care and education, while the patients in the experimental group received 20 minutes of mHealth app education and guidance before being discharged from the hospital. The educational content included helping the participants to download, install, and use the mHealth app. It also included instructions to teach patients how to use the mHealth app at home after discharge to provide education about oral cancer treatment (surgical treatment, chemotherapy, radiotherapy, rehabilitation exercise) and included links to videos about self-recording of symptoms, available support groups, and other applicable information to help meet the needs of the patients. The two groups of patients received the same routine care before discharge. The patients in the experimental group and the control group returned to the hospital 3 months later to complete the questionnaires.

The mHealth app has four main interfaces: Latest News, Medical Information, Self-Recording, and Revisit Reminders. The Latest News interface provides the latest exhortations for patients after surgery and provides YouTube links to oral health education and head and neck rehabilitation videos. In addition, the interface contains a link to join to other patient groups through the LINE app. Therefore, patients with oral cancer after surgery can share information about their lives and treatment experiences after treatment. The Medical Information interface provides information about oral cancer, oral cancer treatment, pain information, hospice care, and any other supporting personnel or cancer treatment institutions. The Self-Recording interface enables patients to record their own postoperative information and symptoms, including date, body temperature, pain levels, oral ulcer, vomiting, skin reactions, and diarrhea. On their next return visit, patients can provide this information to their physician for reference and discuss the response of their disease to treatment. Finally, the Revisit Reminder interface provides a reminder function for patients to remember to return to the hospital. Screenshots of the interfaces are shown in Multimedia Appendix 1.

Figure 1. Flowchart of study sample selection.



Questionnaires

The questionnaires used in this study included a self-administered questionnaire containing questions about the demographic characteristics of the patients (age, gender, marital status, education level, financial status, and religion) and their clinical characteristics (cancer stage, surgical procedure, follow-up treatment, age at onset, and duration of illness); and

a care needs scale, namely the short-form Cancer Needs Questionnaire (CNQ-SF), a 32-item self-administered questionnaire that evaluates 5 domains of patient needs, including psychological, health information, physical and daily living, patient care and support, and interpersonal communication needs. The CNQ-SF score ranges from 0 to 100, where 0 means no need and 100 means the highest need [15]. The questionnaires also included the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) and Head and Neck Module (EORTC QLQ-H&N35) [30] and the Science and Technology Acceptance Model (TAM) scale, which was designed based on the information systems theory developed by Davis [31] in 1989 to evaluate patients' responses to health information technology. The Cronbach α values of the care needs scale and quality of life scale were .94 to .77 and .92, respectively. The content validity index of the TAM scale was between 0.92 and 1.00, with an average of 0.96 (Cronbach α=.97). The questionnaires used in this study are shown in Multimedia Appendix 2.

Statistical Analysis

Continuous variables are presented as mean (SD); categorical variables are presented as n (%). Differences in categorical variables were examined using the chi-square test or Fisher exact test, and the differences in the continuous variables were examined using the independent *t* test. The paired *t* test was used to examine the differences between measurements before and after the intervention. After adjusting for age and sex, multivariate linear regression was used to assess the associations between the patients' care needs and quality of life before and after using the app. All statistical assessments were two-sided, and $P<.05$ was considered statistically significant. Statistical analyses were performed using SPSS version 22 for Windows (IBM Corp).

Results

Analytical Sample

A total of 120 participants were enrolled in this study. The experimental group included 61 postoperative patients who used the mHealth app within three months of discharge, while the control group included 59 postoperative patients who received routine care and instruction but did not use the app. After 3 months of follow-up, 11/61 patients in the experimental group (18%) and 9/59 patients in the control group (15%) did not return to the hospital to fill out the questionnaires. The final sample therefore included 50 patients in the experimental group and 50 patients in the control group (Figure 1).

Background Characteristics of the Study Population

The study enrolled 100 postoperative oral cancer patients, including 92 men (92.0%) and 8 women (8.0%), with a mean age of 57.01 years (SD 8.87). Most patients were married (68/100, 68.0%), had a middle school education (37, 37.0%), were unemployed (61, 61.0%), earned less than 20,000 NT$ (US $677.13) per month (59, 59.0%), had a religious affiliation (87, 87.0%), and were self-caregivers (60, 60.0%). Most of the 100 patients were diagnosed with stage I (32, 32.0%) or stage II (32, 32.0%) cancer, without cancer metastasis (65, 65.0%), and the primary cancer was located at a buccal site in most cases (53, 53.0%). Of the 100 patients, 92 (92.0%) received tumor resection, 66 (66.0%) received radiation therapy, and 38 (38.0%) received chemotherapy. In follow-up treatment, 35/100 patients (35.0%) received radiation therapy and 3 patients (3.0%) received chemotherapy (Table 1).

Table 1. Baseline demographics and clinical characteristics of the patients in the study (N=100).

Variable	Total (N=100)	Experimental group (n=50)	Control group (n=50)	*P* value
Demographics				
Sex , n (%)				.72
Male	92 (92.0)	47 (94.0)	45 (90.0)	
Female	8 (8.0)	3 (6.0)	5 (10.0)	
Age (years), mean (SD)	57.01 (8.87)	58.7 (7.56)	55.32 (9.79)	.06
Marital status, n (%)				.39
Married	68 (68.0)	32 (64.0)	36 (72.0)	
Other (unmarried/widowed/divorced)	32 (32.0)	18 (36.0)	14 (28.0)	
Education, n (%)				.10
Below primary school	28 (28.0)	18 (36.0)	10 (20.0)	
Middle school	37 (37.0)	19 (38.0)	18 (36.0)	
Above high school	35 (35.0)	13 (26.0)	22 (44.0)	
Employment status, n (%)				.54
No	61 (61.0)	32 (64.0)	29 (58.0)	
Yes	39 (39.0)	18 (36.0)	21 (42.0)	
Income (NT$)[a], n (%)				.45
Less than 20,000	59 (59.0)	32 (64.0)	27 (54.0)	
20,000-39,999	20 (20.0)	10 (20.0)	10 (20.0)	
More than 40,000	21 (21.0)	8 (16.0)	13 (26.0)	
Religion, n (%)				.77
No	13 (13.0)	6 (12.0)	7 (14.0)	
Yes	87 (87.0)	44 (88.0)	43 (86.0)	
Primary caregiver, n (%)				.41
Self	60 (60.0)	32 (64.0)	28 (56.0)	
Other (spouse/child/caregiver)	40 (40.0)	18 (36.0)	22 (44.0)	
Clinical characteristics, n (%)				
Cancer stage				.34
I	32 (32.0)	19 (38.0)	13 (26.0)	
II	32 (32.0)	17 (34.0)	15 (30.0)	
III	15 (15.0)	5 (10.0)	10 (20.0)	
IV	21 (21.0)	9 (18.0)	12 (24.0)	
Tumor metastasis				.06
No	65 (65.0)	28 (56.0)	37 (74.0)	
Yes	35 (35.0)	22 (44.0)	13 (26.0)	
Primary site				
Lip	14 (14.0)	9 (18.0)	5 (10.0)	.25
Buccal side	53 (53.0)	28 (56.0)	25 (50.0)	.55
Hard palate	34 (34.0)	14 (28.0)	20 (40.0)	.21
Posterior molar region	6 (6.0)	2 (4.0)	4 (8.0)	.68
Tongue	44 (44.0)	23 (46.0)	21 (42.0)	.69
Previous treatment				
Tumor resection	92 (92.0)	48 (96.0)	44 (88.0)	.27

Variable	Total (N=100)	Experimental group (n=50)	Control group (n=50)	*P* value
Radiation therapy	66 (66.0)	33 (50.0)	33 (50.0)	>.99
Chemotherapy	38 (38.0)	23 (46.0)	21 (42.0)	.68
Follow-up treatment				.87
Radiation therapy	35 (35.0)	17 (34.0)	18 (36.0)	
Chemotherapy	3 (3.0)	1 (2.0)	2 (4.0)	

[a]NT $1=US $0.034.

The mHealth App Intervention Improved Quality of Life and Reduced Care Needs

Table 2 shows the statistical results of the patients' scores on the global quality of life scale (EORTC QLQ-H&N35). The lower the score on the scale, the higher the patient's satisfaction with their quality of life. At baseline, the total quality of life scores in the experimental group and control group were 32.15 and 28.99, respectively. After 3 months of intervention, the total quality of life scores in the experimental group and the control group were reduced to 24.91 and 24.63, respectively. Although the changes in the total scores between the two groups were statistically insignificant, the overall improvement in the intervention group was greater than that in the control group (–7.24 vs –4.36).

Table 2. Quality of life scores of patients in the study (N=100) before and after the mHealth app intervention measured with the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

Variable	Before			After		
	Control group, mean (SD)	Experimental group, mean (SD)	*P* value[a]	Control group, mean (SD)	Experimental group, mean (SD)	*P* value
Overall quality of life score	28.99 (16.40)	32.15 (18.65)	.37	24.63 (16.97)	24.91 (17.13)	.94
Change of overall quality of life score	N/A[b]	N/A	N/A	–4.36 (10.26)	–7.24 (12.77)	.22
Quality of life scores						
Pain	19.50 (20.10)	22.50 (22.40)	.48	19.33 (18.24)	18.99 (20.89)	.93
Swallowing	25.50 (26.69)	31.50 (29.03)	.29	26.16 (24.80)	30.83 (26.09)	.36
Teeth	40.00 (34.99)	42.00 (37.38)	.78	38.66 (33.23)	40.00 (34.99)	.85
Opening mouth	41.33 (35.99)	48.00 (35.74)	.36	39.99 (31.58)	43.33 (33.16)	.61
Dry mouth	46.67 (33.67)	40.00 (36.89)	.35	45.33 (29.16)	34.66 (26.90)	.06
Sticky saliva	43.33 (35.16)	37.33 (35.41)	.40	39.99 (29.35)	23.33 (25.42)	.003[c]
Senses problems	18.33 (21.09)	18.33 (27.20)	>.99	15.66 (17.94)	14.66 (18.63)	.79
Coughing	17.33 (24.50)	24.00 (27.80)	.21	27.33 (94.44)	17.33 (24.50)	.47
Felt ill	16.67 (27.15)	21.33 (27.57)	.40	13.99 (27.01)	11.99 (21.03)	.68
Trouble with social eating	29.50 (28.78)	30.83 (27.63)	.81	25.66 (27.39)	29.66 (28.23)	.47
Speech problems	22.00 (21.59)	23.55 (28.36)	.76	18.22 (21.38)	16.22 (22.47)	.65
Trouble with social contact	16.27 (20.16)	15.33 (20.03)	.82	14.26 (20.47)	13.60 (18.75)	.87
Less sexuality	11.33 (18.27)	17.00 (24.86)	.20	7.66 (15.50)	12.66 (22.22)	.20
Pain killers	54.00 (50.35)	56.00 (50.14)	.84	40.00 (40.00)	44.00 (50.14)	.69
Nasogastric tube feeding	42.00 (49.86)	62.00 (49.03)	.046[c]	32.00 (47.12)	56.00 (50.14)	.015[c]
Nutritional supplements	28.00 (33.75)	45.00 (35.36)	.016[c]	23.00 (32.27)	35.00 (35.35)	.08
Weight loss	22.00 (41.85)	30.00 (46.29)	.37	6.00 (23.98)	6.00 (23.98)	>.99
Weight gain	28.00 (45.36)	14.00 (35.05)	.09	10.00 (30.30)	0	.024[c]

[a]*P* value was used to identify statistical significance between the experimental group and control group.

[b]N/A: not applicable.

[c]*P*<.05.

The CNQ-SF measures baseline postoperative care needs (Table 3). The higher the scores on the scale, the higher the patient's need for care. Before the intervention, the mean scores of the 5 care domains in the experimental group and control group were 26.33 vs 21.33 for physiological needs, 24.55 vs 26.27 for psychological needs, 13.50 vs 16.75 for communication needs, 26.92 vs 19.58 for care support needs, and 64.0 vs. 60.29 for health information needs, respectively. After 3 months of intervention, the mean scores for the experimental group and control group were 20.67 vs 20.25 for physiological needs, 18.18 vs 23.14 for psychological needs, 8.25 vs 12.75 for communication needs, 23.75 vs 17.67 for care support needs, and 63.86 vs 57.0 for health information needs, respectively. These results show that the experimental group (mHealth app intervention) had significantly reduced physiological needs compared to the control group (*P*=.015, Table 3). Although the results were not statistically significant, the experimental group had more obvious reductions in psychological needs, communication needs, and care support needs than the control group. Multivariate linear regression analysis also confirmed that after adjusting for age and sex variables, the experimental group had significantly greater improvement in physiological needs compared to the control group (*P*=.022, Table 4).

Table 3. Care needs of patients in the study (N=100) before and after the mHealth app intervention measured with the short-form Cancer Needs Questionnaire.

Variable	Before			After			Change		
	Control group, mean (SD)	Experimental group, mean (SD)	*P* value[a]	Control group, mean (SD)	Experimental group, mean (SD)	*P* value	Control group, mean (SD)	Experimental group, mean (SD)	*P* value
Physiological needs	21.33 (18.36)	26.33 (20.03)	.20	20.25 (15.95)	20.67 (15.45)	.90	–1.08 (7.80)	–5.67 (10.47)	.015[b]
Psychological needs	26.27 (23.07)	24.55 (23.98)	.71	23.14 (20.40)	18.18 (17.29)	.19	–3.14 (8.04)	–6.36 (13.88)	.16
Communication needs	16.75 (22.53)	13.50 (21.70)	.46	12.75 (21.94)	8.25 (17.97)	.27	–4.0 (12.74)	–5.25 (9.81)	.58
Care support needs	19.58 (21.33)	26.92 (18.52)	.07	17.67 (15.92)	23.75 (15.48)	.05	–2.0 (12.88)	–3.17 (7.79)	.59
Health information needs	60.29 (25.72)	64.00 (28.78)	.50	57.00 (25.95)	63.86 (23.61)	.17	–3.29 (19.06)	–0.14 (13.15)	.34

[a]*P* value was used to identify statistical significance between the experimental group and control group.

[b]*P*<.05.

Table 4. Multivariate linear regression analysis of patients' care needs after the mHealth app intervention.

Model[a]	App use	
	β (95% CI)	*P* value
EORTC QLQ-H&N35[b]	–3.34 (–7.83 to 1.15)	.15
CNQ-SF[c]		
Physiological needs	–4.24 (–7.88 to –0.60)	.022[d]
Psychological needs	–2.75 (–7.22 to 1.73)	.23
Communication needs	–0.99 (–5.49 to 3.51)	.67
Care support needs	–0.80 (–4.99 to 3.39)	.71
Health information needs	3.21 (–3.27 to 9.70)	.33

[a]Model adjusted for sex and age.

[b]EORTC QLQ-H&N35: Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 and Head and Neck Module.

[c]CNQ-SF: short-form Cancer Needs Questionnaire.

[d]*P*<.05.

Patient Acceptance of the mHealth App

Patient acceptance of the mHealth app was measured by intention to use, perceived usefulness, and perceived ease of use (TAM scale). At baseline, the mean scores in the experimental group for intention to use, perceived usefulness, and perceived ease of use were 2.54, 2.52, and 2.32, respectively. In the control group, the mean scores for the 3 aspects were 2.68, 2.49, and 2.49, respectively. After 3 months of the mHealth app intervention, the mean scores for the three aspects increased in the intervention group to 3.02, 2.95, and

3.01, respectively. All three aspects of app acceptability significantly increased after the intervention (P<.01, Table 5).

Table 5. Acceptability of the mHealth app by patients based on the Science and Technology Acceptance Model scale.

Variable	Before			After		
	Control group, mean (SD)	Experimental group, mean (SD)	P value[a]	Control group, mean (SD)	Experimental group, mean (SD)	P value[b]
Intention to use	2.68 (1.12)	2.54 (1.05)	.52	N/A[c]	3.02 (0.87)	.002
Perceived usefulness	2.49 (1.14)	2.52 (1.09)	.90	N/A	2.95 (0.99)	.004
Perceived ease of use	2.49 (1.04)	2.32 (0.77)	.37	N/A	3.01 (0.90)	<.001

[a]Independent t test with P value was used to identify statistical differences between the experimental group and the control group.

[b]Paired t test with P value was used to identify statistical differences before and after the mHealth app intervention.

[c]N/A: not applicable.

Discussion

Principal Findings

Previous studies showed that the physical function and oral function of oral cancer patients deteriorates significantly 3 months after surgery, accompanied by poorer body image and less social contact [9]. Birur et al [27] used a remote mHealth-based approach to establish an effective oral cancer screening program and found that the intervention could improve the efficiency of early detection of oral cancer. In addition, an intervention with a personalized mobile phone–based self-management app could improve outcomes for patients taking oral anticancer medications [28]. In particular, education and information received through mHealth apps has been shown to improve prevention and posttreatment outcomes in various clinical situations [19]. In addition to supporting these studies, this study further showed that compared with routine health care and instruction, the education/information intervention provided by the mHealth APP indeed reduced care needs; the study also showed that patients had a higher degree of acceptance of using mobile devices to learn about and manage their disease.

Quality of life scores have been linked to predicted survival of head and neck cancer patients [32]. Initiating supportive care as early as possible, with measures such as encouraging optimal nutritional intake and improving oral function by reducing symptoms, can help improve the quality of life of head and neck cancer patients after surgery [2]. In this study, although the results were not statistically significant, patients reported improved quality of life after the mHealth app intervention. We note that quality of life is a long-term state, and more than three months may be needed to observe statistically significant differences between control and intervention groups, as in other studies [33]. In this study, the intervention lasted only 12 weeks; thus, further investigation is needed to determine the ideal intervention time to measure the differences in outcomes.

Interestingly, most patients in the control group claimed that they had received medical care and sufficient information from the medical staff; however, more than half of these patients stated during revisits that they had forgotten or only remembered part of this information. This result may be partly due to the fact that the health education leaflets provided by the nursing staff are not easy to carry around or that the contents are relatively boring (patients' statement), which may have reduced patients' willingness to read the leaflets. In contrast, the mHealth app can enable patients to immediately access information on their medical condition, medications and dosages, or changes in symptoms whenever they want. It can also provide a useful reminder system to help patients manage their treatment schedules [33].

In general, patients with oral cancer require chemotherapy or radiotherapy in addition to surgery. However, most patients are very unfamiliar with and fearful of subsequent chemotherapy and radiotherapy; also, they may be concerned about the side effects of these treatments. Therefore, patients will want to understand their disease during treatment, strive for self-care, adapt to life changes as early as possible, and understand the disease response strategies appropriate to their situation [3]. Patients undergoing cancer treatment report high levels of physical care needs [34]. In the present study, the mHealth app intervention did reduce the physiological needs, psychological needs, communication needs, and care support needs of patients who received it more than the routine health care and instruction provided by nurse caregivers. However, compared to the control group, the health information needs in the intervention group did not improve (–3.29 vs –0.14). This result may have occurred because when patients found they could obtain more health information from the app, their demand for health information also grew. When refining the mHealth app, more questionnaires should be used to determine the needs of patients for other health information content, particularly at different points in disease progression.

Before using the mHealth app, the acceptance of such apps in the 2 groups of patients was low; however, the scores for the three acceptance variables (intention to use, perceived usefulness, and perceived ease of use) all increased significantly after 3 months of the mHealth app intervention. This result indicates that familiarity with the mHealth app reduced the uncertainty and increased the acceptance of using it. Similarly, other studies have found that well-designed smartphone apps help to enhance compliance with oral anticancer medications, even for patients who were not previously compliant [19,27,33]. A nurse-led prechemotherapy education intervention (ChemoEd) using DVDs for pretreatment consultation also demonstrated that a DVD-based intervention can significantly reduce the

incidence and severity of sensory, psychological, and procedural concerns as well as of vomiting [14]. A review study also supported our findings that technology-based interventions can have positive effects on pain, depression, and quality of life in cancer patients [19]. The results of this and the above studies indicate that mHealth app interventions may also provide health education benefits for postoperative oral cancer patients.

Limitations and Recommendations

The present study has several limitations. Due to time constraints, the number of trained researchers, budget constraints, and other factors, patients could only be followed for 3 months. Despite the positive results observed during this period, we recommend conducting longer intervention studies in the future, such as those including 3-month, 6-month, 9-month, and 12-month intervals. We believe that some statistically insignificant results will improve if the study is expanded to a longer time frame (eg, psychological needs and communication needs). In addition, all the oral cancer patients in this study were recruited from one medical center. Therefore, the results may not be applicable to oral cancer patients from other medical centers or patients treated at nonmedical centers. In the future, patients from different medical centers or nonmedical centers should be included. In addition, treatment plans for each patient can be included in the analysis to provide researchers with an understanding of the association of treatment plans with the care needs of postoperative patients. Based on participants' feedback, rehabilitation videos and oral cancer support groups provided the most useful information, and the participants suggested that doctors and nursing staff should be invited to join the mHealth app to provide immediate consultation.

Conclusions

An mHealth app intervention can significantly reduce physiological needs in postoperative oral cancer patients, and use of the mHealth app was highly accepted by patients. These data may also provide health care professionals with a better understanding of the optimal course of patient care after surgery. The main results of this study indicated that the mHealth app can be easily incorporated into routine care of postoperative oral cancer patients to conveniently provide medical information and improve patients' self-management abilities, thereby reducing their physiological care needs and promoting better health.

Acknowledgments

We would like to thank Convergence CT for language editing assistance. This study was supported by the National Yang-Ming University-Far Eastern Memorial Hospital Joint Research Program (106DN01).

Authors' Contributions

LCC and TFW conceptualized the study and conducted the formal analysis and investigation. LCC and RHC wrote the first draft of the paper. LCC, RHC, SCY, and CC conducted the investigation. TFW provided resources and helped write, review, and edit the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the interfaces of the app.
[PDF File (Adobe PDF File), 5614 KB - mhealth_v8i7e18132_app1.pdf]

Multimedia Appendix 2
Questionnaires used in the study.
[DOCX File , 52 KB - mhealth_v8i7e18132_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V. 1. 6. 1).
[PDF File (Adobe PDF File), 355 KB - mhealth_v8i7e18132_app3.pdf]

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011 Mar;61(2):69-90 [FREE Full text] [doi: 10.3322/caac.20107] [Medline: 21296855]
2. de Leeuw J, van den Berg MGA, van Achterberg T, Merkx MAW. Supportive care in early rehabilitation for advanced-stage radiated head and neck cancer patients. Otolaryngol Head Neck Surg 2013 Apr;148(4):625-632. [doi: 10.1177/0194599812474797] [Medline: 23348870]

3. Whelan TJ, Mohide EA, Willan AR, Arnold A, Tew M, Sellick S, et al. The supportive care needs of newly diagnosed cancer patients attending a regional cancer center. Cancer 1997 Oct 15;80(8):1518-1524. [doi: 10.1002/(sici)1097-0142(19971015)80:8<1518::aid-cncr21>3.0.co;2-7] [Medline: 9338478]
4. Shah JP, Gil Z. Current concepts in management of oral cancer--surgery. Oral Oncol 2009;45(4-5):394-401 [FREE Full text] [doi: 10.1016/j.oraloncology.2008.05.017] [Medline: 18674952]
5. Coleman JP. Cancer in the oral area. Tex Dent J 2012 May;129(5):442. [Medline: 22779197]
6. Stark DP, House A. Anxiety in cancer patients. Br J Cancer 2000 Nov;83(10):1261-1267 [FREE Full text] [doi: 10.1054/bjoc.2000.1405] [Medline: 11044347]
7. Zainal N, Hui K, Hang T, Bustam A. Prevalence of distress in cancer patients undergoing chemotherapy. Asia-Pacific J Clinical Oncology 2007 Oct 11;3(4):219-223. [doi: 10.1111/j.1743-7563.2007.00114.x]
8. Saevarsdottir T, Fridriksdottir N, Gunnarsdottir S. Quality of life and symptoms of anxiety and depression of patients receiving cancer chemotherapy: longitudinal study. Cancer Nurs 2010;33(1):E1-E10. [doi: 10.1097/NCC.0b013e3181b4adb5] [Medline: 20010331]
9. Schliephake H, Jamil MU. Prospective Evaluation of Quality of Life After Oncologic Surgery for Oral Cancer. Int J Oral Maxillofac Surg 2002 Aug;31(4):427-433 [FREE Full text] [doi: 10.1054/ijom.2001.0194] [Medline: 12361079]
10. Hassanein KAM, Musgrove BT, Bradbury E. Psychological outcome of patients following treatment of oral cancer and its relation with functional status and coping mechanisms. J Craniomaxillofac Surg 2005 Dec;33(6):404-409 [FREE Full text] [doi: 10.1016/j.jcms.2005.05.005] [Medline: 16253509]
11. Kessler PA, Bloch-Birkholz A, Leher A, Neukam FW, Wiltfang J. Evaluation of quality of life of patients with oral squamous cell carcinoma. Comparison of two treatment protocols in a prospective study. Radiother Oncol 2004 Mar;70(3):275-282 [FREE Full text] [doi: 10.1016/j.radonc.2003.11.017] [Medline: 15064013]
12. Waller A, Forshaw K, Bryant J, Mair S. Interventions for preparing patients for chemotherapy and radiotherapy: a systematic review. Support Care Cancer 2014 Aug;22(8):2297-2308 [FREE Full text] [doi: 10.1007/s00520-014-2303-3] [Medline: 24906837]
13. Butow PN, Phillips F, Schweder J, White K, Underhill C, Goldstein D, Clinical Oncological Society of Australia. Psychosocial well-being and supportive care needs of cancer patients living in urban and rural/regional areas: a systematic review. Support Care Cancer 2012 Jan;20(1):1-22. [doi: 10.1007/s00520-011-1270-1] [Medline: 21956760]
14. Aranda S, Jefford M, Yates P, Gough K, Seymour J, Francis P, et al. Impact of a novel nurse-led prechemotherapy education intervention (ChemoEd) on patient distress, symptom burden, and treatment-related information and support needs: results from a randomised, controlled trial. Ann Oncol 2012 Jan;23(1):222-231. [doi: 10.1093/annonc/mdr042] [Medline: 21471562]
15. Cossich T, Schofield P, McLachlan SA. Validation of the cancer needs questionnaire (CNQ) short-form version in an ambulatory cancer setting. Qual Life Res 2004 Sep;13(7):1225-1233. [doi: 10.1023/B:QURE.0000037496.94640.d9] [Medline: 15473501]
16. Chen SC, Yu WP, Chu TL, Hung HC, Tsai MC, Liao CT. Prevalence and correlates of supportive care needs in oral cancer patients with and without anxiety during the diagnostic period. Cancer Nurs 2010;33(4):280-289. [doi: 10.1097/NCC.0b013e3181d0b5ef] [Medline: 20467308]
17. Yun YH, Shon EJ, Yang AJ, Kim SH, Kim YA, Chang YJ, et al. Needs regarding care and factors associated with unmet needs in disease-free survivors of surgically treated lung cancer. Ann Oncol 2013 Jun;24(6):1552-1559 [FREE Full text] [doi: 10.1093/annonc/mdt032] [Medline: 23471105]
18. Ziegler L, Newell R, Stafford N, Lewin R. A literature review of head and neck cancer patients information needs, experiences and views regarding decision-making. Eur J Cancer Care (Engl) 2004 May;13(2):119-126. [doi: 10.1111/j.1365-2354.2004.00449.x] [Medline: 15115467]
19. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015;3(2):e33 [FREE Full text] [doi: 10.2196/mhealth.3789] [Medline: 25842282]
20. Ryu S. Book Review: mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
21. Dounavi K, Tsoumani O. Mobile Health Applications in Weight Management: A Systematic Literature Review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
22. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
23. Helbostad JL, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile Health Applications to Promote Active and Healthy Ageing. Sensors (Basel) 2017 Mar 18;17(3) [FREE Full text] [doi: 10.3390/s17030622] [Medline: 28335475]
24. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]

25. Carrasco MP, Salvador CH, Sagredo PG, Márquez-Montes J, González DMMA, Fragua JA, et al. Impact of patient-general practitioner short-messages-based interaction on the control of hypertension in a follow-up service for low-to-medium risk hypertensive patients: a randomized controlled trial. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):780-791. [doi: 10.1109/TITB.2008.926429] [Medline: 19000959]
26. Bexelius C, Löf M, Sandin S, Lagerros YT, Forsum E, Litton JE. Measures of physical activity using cell phones: validation using criterion methods. J Med Internet Res 2010;12(1):e2 [FREE Full text] [doi: 10.2196/jmir.1298] [Medline: 20118036]
27. Birur PN, Sunny SP, Jena S, Kandasarma U, Raghavan S, Ramaswamy B, et al. Mobile health application for remote oral cancer surveillance. J Am Dent Assoc 2015 Dec;146(12):886-894. [doi: 10.1016/j.adaj.2015.05.020] [Medline: 26610833]
28. 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]
29. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, 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]
30. Singer S, Wollbrück D, Wulke C, Dietz A, Klemm E, Oeken J, et al. Validation of the EORTC QLQ-C30 and EORTC QLQ-H&N35 in patients with laryngeal cancer after surgery. Head Neck 2009 Jan;31(1):64-76. [doi: 10.1002/hed.20938] [Medline: 18972411]
31. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
32. Karvonen-Gutierrez CA, Ronis DL, Fowler KE, Terrell JE, Gruber SB, Duffy SA. Quality of life scores predict survival among patients with head and neck cancer. J Clin Oncol 2008 Jun 01;26(16):2754-2760. [doi: 10.1200/JCO.2007.12.9510] [Medline: 18509185]
33. Ali EE, Leow JL, Chew L, Yap KYL. Patients' Perception of App-based Educational and Behavioural Interventions for Enhancing Oral Anticancer Medication Adherence. J Cancer Educ 2018 Dec;33(6):1306-1313. [doi: 10.1007/s13187-017-1248-x] [Medline: 28707206]
34. Newell S, Sanson-Fisher RW, Girgis A, Ackland S. The physical and psycho-social experiences of patients attending an outpatient medical oncology department: a cross-sectional study. Eur J Cancer Care (Engl) 1999 Jun;8(2):73-82. [doi: 10.1046/j.1365-2354.1999.00125.x] [Medline: 10476109]

Abbreviations

CNQ-SF: short-form Cancer Needs Questionnaire
EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
EORTC QLQ-H&N35: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 and Head and Neck Module
mHealth: mobile health
PDA: personal digital assistant
TAM: Technology Acceptance Module
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.02.20; peer-reviewed by B Chaudhry, C Jacob; comments to author 06.04.20; revised version received 26.05.20; accepted 03.06.20; published 27.07.20.

Please cite as:
Wang TF, Huang RC, Yang SC, Chou C, Chen LC
Evaluating the Effects of a Mobile Health App on Reducing Patient Care Needs and Improving Quality of Life After Oral Cancer Surgery: Quasiexperimental Study
JMIR Mhealth Uhealth 2020;8(7):e18132
URL: http://mhealth.jmir.org/2020/7/e18132/
doi:10.2196/18132
PMID:32716303

Original Paper

A Tool for Rating the Value of Health Education Mobile Apps to Enhance Student Learning (MARuL): Development and Usability Study

Tehmina Gladman[1], PhD; Grace Tylee[2]; Steve Gallagher[3], PhD; Jonathan Mair[4], MBChB; Sarah C Rennie[1], PhD; Rebecca Grainger[1], PhD

[1]Education Unit, University of Otago Wellington, Wellington, New Zealand

[2]University of Otago Wellington, Wellington, New Zealand

[3]Education Unit, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

[4]Hutt Valley District Health Board, Lower Hutt, New Zealand

Corresponding Author:
Tehmina Gladman, PhD
Education Unit
University of Otago Wellington
PO Box 7343
Newtown
Wellington, 6242
New Zealand
Phone: 64 04 918 6749
Email: tehmina.gladman@otago.ac.nz

Abstract

Background: To realize the potential for mobile learning in clinical skills acquisition, medical students and their teachers should be able to evaluate the value of an app to support student learning of clinical skills. To our knowledge, there is currently no rubric for evaluation of quality or value that is specific for apps to support medical student learning. Such a rubric might assist students to be more confident in using apps to support their learning.

Objective: The objective of this study was to develop an instrument that can be used by health professional educators to rate the value of a mobile app to support health professional student learning.

Methods: Using the literature, we developed a list of potential criteria for the evaluation of educational app value, which were then refined with a student group using a modified nominal group technique. The refined list was organized into themes, and the initial rubric, Mobile App Rubric for Learning (MARuL, version 1), was developed. iOS and Android app stores were searched for clinical skills apps that met our inclusion criteria. After the 2 reviewers were trained and the item descriptions were refined (version 2), a random sample of 10 included apps, 5 for each mobile operating system, was reviewed. Interitem and interrater analyses and discussions with the reviewers resulted in refinement of MARuL to version 3. The reviewers completed a review of 41 clinical skills mobile apps, and a second round of interitem and interrater reliability testing was performed, leading to version 4 of the MARuL.

Results: Students identified 28 items (from an initial set of 144 possible items) during the nominal group phase, and these were then grouped into 4 themes: teaching and learning, user centered, professional, and usability. Testing and refinement with reviewers reduced the list to 26 items. Internal consistency for MARuL was excellent (α=.96), and the interrater reliability as measured by the intraclass correlation coefficient (ICC) was good (ICC=0.66).

Conclusions: MARuL offers a fast and user-friendly method for teachers to select valuable apps to enhance student learning.

(JMIR Mhealth Uhealth 2020;8(7):e18015) doi:10.2196/18015

KEYWORDS

questionnaire design; medical education; health occupations students; just-in-time learning; self-directed learning; mobile phone; rubric; mobile learning; mobile apps; mhealth; digital learning

Introduction

Background

Smartphones and tablets have made mobile learning an important component of education, enabling learning anywhere, at any time, using mobile apps [1]. Most education apps have specific functions or aims, such as providing resources for reference while learning or in practice, learning activities and games, or organizing activities related to learning. Apps can be found by keyword search in the app stores (eg, Google Play Store, iOS App Store), through recommendations, or within app store–determined categories [2].

In medical education, the use of apps for reference and learning on the go (*just in time*) for ongoing professional development is widespread [1,3]. However, previous work suggests that medical students may still prefer textbooks and lectures as a learning resource [4,5] and predominantly use apps for reference and revision [1,3,4,6]. Although the determinants of preference for textbooks have not been explored, there may be barriers to finding apps that are both relevant and valuable for student learning [5]. Within app stores, star ratings and reviews are the main indicators of app quality or value; however, these can be subjective and only relevant if the reviewer has needs and expectations similar to the potential app user [7,8]. Furthermore, a search of medical apps reveals a very large number of potential apps to choose from and many that have not received enough reviews to get a rating. Owing to the number of options to choose from, and the lack of good quality guidance to inform their choices, students may not be accessing and using apps that could support their learning. Providing a standard way of evaluating apps for medical education may improve students' ability to find or identify apps to support their learning.

Evaluating Apps

To avoid the app overload caused by the increasing number of apps in the app stores, potential app users are recommended to use the literature to identify valuable apps [2,9,10]. There is evolving literature aiming to identify apps for health conditions; to support self-management, education, or behavior change [11,12]; and to evaluate app quality [10,13]. To date, similar literature in education has focused on apps for use by teachers working with students with learning disabilities [14] or preschool student learning [15]. In addition to providing curated lists of apps, this literature often evaluates app quality with instruments that include some generic aspects of proposed quality; for example, a component to evaluate the suitability of design and aesthetics. However, these rubrics also have components or domains specific to the type or main function of an app. For instance, the evaluation rubric for health care smartphone apps by Jin and Kim [11] emphasizes the input of medical experts and the developers' citing of authoritative sources. In comparison, the rubric by Lee and Kim [12] for evaluating educational apps emphasizes the *teaching and learning* component of an app. When developing a rubric to evaluate apps, a balance between general and specific criteria is needed to ensure that the rubric is both reliable and objective [10]. To our knowledge, there is currently no rubric for evaluation of quality or value that is specific for apps to support medical student learning. Such a rubric might assist students to be more confident in using apps to support their learning.

App Evaluation Rubrics

Most multidimensional app evaluation rubrics include both user-centered features (eg, engagement) and technology-centered features (eg, functionality) [10]. Although the model for education app evaluation by Lee and Kim [12] emphasizes teaching and learning, it also recognized the need for a technologically stable app, with "quick interactions (fast loading times) and error-free stability" [12]. It was concluded that factors must be considered together because they are interdependent in nature rather than independent. Stoyanov et al [10] also emphasized the importance of considering multiple features of an app to increase the objectivity and reliability of an app evaluation scale.

An additional factor that is relevant for health-related mobile apps is the credibility of the app content. Having apps with evidence-based approaches to health should minimize the risk of harm and promote safe application of knowledge by app users [16-18]. In the medical education setting, for example, when learning clinical skills, it is critical to learn skills that are relevant and accurate in the local health care environment. Apps that focus on clinical skills—that is, apps supporting learning of history taking, examination, communication, or procedures—provide opportunities for just-in-time learning in clinical settings where textbooks are not available, making apps supporting acquisition of clinical skills a good test case for the development of an app evaluation rubric.

Purpose and Research Question

To realize the potential for mobile learning in clinical skills acquisition, medical students and their teachers should be able to evaluate the value of an app for supporting the learning of clinical skills. Our aim was to develop a rubric that can be used by staff to evaluate the value of apps for learning the clinical skills of history taking, examination, communication, or procedures so that recommendations can be made to students. The initial research question was as follows: "What key domains of value need to be included in a reliable measure that teachers can use to rate mobile apps for just-in-time learning?"

Methods

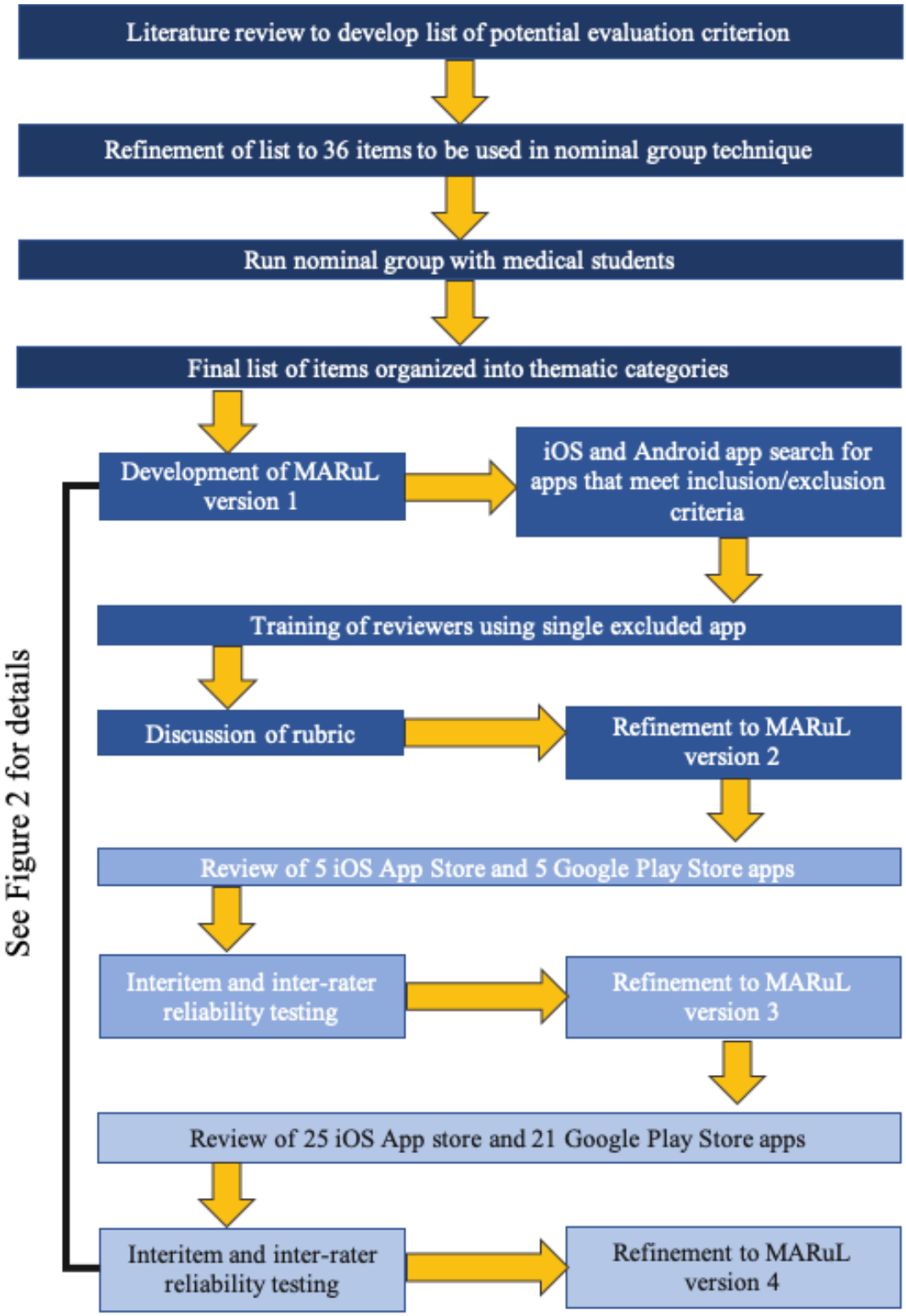
Overview

The development of evaluative rubrics for apps uses some or all of the following steps. A literature search is completed to identify previous literature with quality evaluation criteria, followed by a discussion among topic experts to group criteria and develop scale items [10]. A survey or focus group is conducted to gain expert opinion on the relevance of items generated from the literature search [12]. Users and experts evaluate apps with the draft rubric to test its reliability and validity [11,12], with items in the rubric refined accordingly. Many rubrics are structured using overarching factors with relevant subterms to clarify and give examples [10,11,19].

We undertook a literature review to identify potential rubric items, which were evaluated by a group of medical students

using a modified nominal group technique. The performance of the preliminary rubric was then evaluated by clinical and education experts for use with clinical skills apps, which were identified using a comprehensive search strategy in Google Play and iTunes app stores. The rubric was then refined through expert feedback and statistical analysis using the classical test theory (Figure 1).

Figure 1. Process for developing the Mobile App Rubric for Learning. MARuL: Mobile App Rubric for Learning.



Definitions

We defined *value* as referring to an object that is perceived to have utility in meeting short- or long-term goals [20]. *Just-in-time learning* is defined as a method of learning that is driven by the learner, when and where they require it [21]. With this definition, the learner anticipates learning and performance requirements rather than responding to them [21]. Just-in-time learning occurs temporally close to a clinical learning encounter, for example, with a mobile device [22,23]. *Clinical skills* apps

are those that include opportunities to learn and improve history taking, examination, communication, or procedures [24]. A *rubric* is a scoring method that uses evaluative criteria with quality definitions and a clear scoring strategy to determine the quality of the global concept being scored [25].

Development of Potential Rubric Items

A literature review was undertaken to identify criteria previously used to evaluate the quality of mobile apps for education. An expert librarian conducted a literature search on Scopus, MEDLINE, and Google Scholar (between August 14 and 21, 2018) using search terms related to the evaluation of educational mobile apps using evidence-based methods (Textboxes 1-4). After duplicates were removed, the records were exported into Rayyan QCRI (Qatar Computing Research Institute), a cloud-based method for completing systematic reviews of the literature [26], and 4 educational experts (TG, SR, SG, and RG) independently reviewed the titles and abstracts of the records to identify articles that potentially met the inclusion criteria. The inclusion criteria used were as follows: clear descriptions of the concepts, list, or rubric for the evaluation of app quality; reliability or validity testing of the method of evaluation; and the evaluation was for educational mobile apps, which we defined as apps whose primary purpose was to support the education of any population. Articles were excluded if they were literature reviews or they described a framework for evaluation without describing a specific measure. Articles were included if 3 or more reviewers agreed on their relevance.

Textbox 1. Device related literature search terms.

- app
- smart*
- phone*
- cellular
- smartphone
- smart phone
- mobile
- tablet

Textbox 2. Action related literature search terms.

- evaluat*
- assess*
- apprais*
- measure*
- validat*
- test*
- determine*
- measur*

Textbox 3. Measure related literature search terms.

- criteria
- checklist*
- rubric*
- framework*
- quality
- useful*

Textbox 4. Teaching and learning related literature search terms.

- educat*
- learn*
- teach*
- student*
- education*
- learning*
- educate

The full texts of all identified relevant articles were reviewed in detail by a reviewer (GT), and the concepts or terms that measure or assess app quality were extracted along with definitions. The extracted terms and definitions were then grouped and organized by GT, a medical student representing the end user for the rubric, and TG, an academic with expertise in electronic learning (e-learning) and educational psychology. Overarching concepts (called *categories*) were identified and related terms were grouped together within the concepts. For instance, *use without an internet connection* was grouped with *platform*, *syncing*, *updates*, and *compatibility* under the concept of *technical specifications*. Similar terms were merged.

Nominal Group Technique

A convenience sample of 10 medical students, recruited using targeted invitations, participated in a nominal group held at the University of Otago Wellington, School of Medicine, on December 6, 2018. Ethical approval was obtained from the Otago Human Ethics Committee, reference number D18/337, and written consent was obtained from the students.

In the nominal group technique, group interaction is facilitated by the leader and verbal interaction is restricted to a discussion between the leader and participants with no discussion between participants [27]. The technique is particularly useful in ensuring that all members of a group are heard. This makes it an ideal method in a group where there are varied levels of experience with the topic under discussion. Ranking of ideas occurs using votes or a Likert-type scale [27]. A voting system with a predetermined maximum of 3 rounds of voting was used after the initial ranking of terms. Limited voting was chosen to reduce the possibility of participant fatigue [28].

Students were asked to review the refined list of terms/concepts for app evaluation (with definitions) and indicate their top 20 terms on a ranking sheet without discussion. They were also asked to add any important missing terms or concepts. Each student then in turn stated a term on their list not previously offered by a student, which was recorded by the leader. This continued until all terms from each student's top 20 had been recorded.

Students were then given an opportunity for discussion before the first round of voting. Students voted on item inclusion in the final list using the options *keep*, *unsure*, and *discard*. Overall, 7 or more votes for the same option were considered a majority. Items that did not receive a majority vote were recorded and discussed before the next voting round. In the second round, students were encouraged to vote *keep* or *discard* but still had the option of *unsure*. In the second and third rounds, having 6 or more votes was considered a majority. In the third round, only options *keep* and *discard* were allowed. Results from each round were recorded by a group facilitator, whereas the leader facilitated group discussions and answered questions.

Development of Rubric

The terms chosen by the nominal group made up the preliminary rubric, which we named *Mobile App Rubric for Learning* (MARuL). Two authors (TG and GT) grouped the terms into themes separately, and then came together to discuss final category names and grouping of terms, with subterms used to develop descriptors for the scale.

App Search

The search of the iOS and Google Play stores for clinical skills apps was undertaken from January 15 to February 1, 2019. The search was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines for systematic reviews, with search terms confirmed by discussion (GT, TG, and RG) after preliminary searches (Textbox 5) [29,30]. The inclusion criteria were as follows: available in English; includes at least one of the keywords in the title or description; includes an interactive element requiring some form of user input; target audience includes medical students; and support for iOS 10 or later/Android version 7 or later. The exclusion criteria were as follows: priced over NZ $10.00 (US $6.30) for a one-off price or monthly subscription (based on expected student willingness to pay); reference-only apps (passive with no student input); designed for staff-only use in formative or summative assessment contexts; were a complement to other software (not standalone; requires a log-in/sign up).

Textbox 5. App search terms.

- clinical skills
- OSCE
- objective structured clinical examination
- medical history taking
- clinical history taking
- patient history
- medical examination
- physical examination
- clinical examination
- clinical exam
- physical exam
- medical exam
- planning and explaining
- patient education

A data screening and extraction spreadsheet was developed and refined by 2 authors (GT and TG) using Airtable [31]. The app search of the iOS and Google Play Store was conducted independently by 2 authors (GT and TG). Apps were screened based on their title and the description in the app store. App name, developer, operating system, reviewer, and inclusion decision were recorded for all screened apps. For the excluded apps, the first identified exclusion criterion was recorded. GT and TG then reviewed all the apps with discrepant decisions and reached a consensus through discussion.

Testing and Refinement of the Rubric

To initially test the reliability of the MARuL, we developed a protocol based on previous research [10,30] and trialed this with a randomly chosen subset of 10 apps included from the search. The 10 randomly selected apps, 5 Android apps and 5 iOS apps, were downloaded between May 20 and 22, 2019. The iPhones used were iPhone 6 and 6 plus running iOS 12.3.1. The Android phones were Samsung Galaxy J1Ace running Android 5.1.1. Following the initial testing, we refined the rubric and completed reliability testing with the remaining included apps identified in our search of the app stores. Trialing of the rubric with clinical skills apps was completed independently by 2 authors: (1) a hospital-based junior doctor (3 years postgraduate) training in internal medicine (JM) and (2) the e-learning facilitator for the clinical years at a different campus of our medical school (SG). The first reviewer was able to review the apps from the point of view of a near-peer teacher of medical students, whereas the second reviewer reviewed from the point of view of a learning expert with over 10 years of experience in medical education.

For training, the reviewers first used the MARuL to evaluate a previously excluded app, Clinical Skills by George Sim on iOS. The app was excluded because of a lack of interactive elements. The 2 reviewers each downloaded the app on an iPhone running iOS 12.2, tested the app features for 10 min and then independently evaluated the app using the MARuL. Following their review, they met with TG via videoconferencing to discuss their scoring of the app and ensure their understanding of the rubric items and process.

Once the reviewers (JM and SG) had a clear understanding of the items and process for reviewing the apps, they each downloaded the same 5 randomly selected apps for both mobile operating systems. They independently spent a minimum of 10 min using each app before evaluating the app with the MARuL. MARuL ratings for each app and time taken to complete MARuL rating were collected using Qualtrics (Qualtrics, Provo, UT) [32,], downloaded to an Excel spreadsheet, and analyzed using RStudio and appropriate packages [33-36].

Internal consistency and interrater reliability were calculated, and then a discussion was held via videoconferencing to review discrepancies in scoring on the rubric and identify any refinements required to the items of the MARuL. The MARuL was then revised to rewrite some of the descriptors and remove an item that was considered redundant by the reviewers. The remaining apps were then independently trialed and evaluated using the revised MARuL.

We calculated the internal consistency of the categories and the overall value measure using Cronbach alpha. Cronbach alpha measures how interrelated a set of items in a scale are, with scores ranging from 0 to 1, and higher scores indicating a stronger interitem relationship [37]. Interrater reliability of the categories and overall value measure was calculated using the intraclass correlation coefficient (ICC). The ICC measures how much of the difference between sets of scores is because of measurement error, and it ranges from 0 to 1, with higher scores indicating stronger interrater reliability [38].

Results

Rubric Items

The literature search yielded 193 unique articles. After reviewing the titles and abstracts, 134 articles were eliminated. Furthermore, 8 articles were removed following full-text review.

From the remaining 51 articles, 144 quality criteria were extracted, including main and descriptive subterms, 69 of which were main terms. After the consensus discussion and deletion of overlapping items and organization, a list of 36 main terms from 46 articles remained. (Multimedia Appendix 1 has the full list of terms and references).

Nominal Group

The nominal group had 10 students (6 female and 4 male) who had completed 2 years (n=2), 3 years (n=1), or 4 years (n=7) of medical school. Among them, 6 students identified as New Zealand European/Pākehā, 3 as Māori, and 1 as Sri Lankan. The nominal group ranking produced a list of 35 of the 36 terms; the only term excluded from voting was "product description." Following the first vote, 18 items were kept and 17 were to be voted on again. The second vote on the 17 terms resulted in 10 being kept, 3 discarded, and 4 to be voted on again. After the third vote, 1 term was discarded and 3 received an equal number of keep and discard votes. These terms were "sharing," "motivation," and "self-directedness." The final list of terms voted on by medical students included 28 items. Table 1 shows the initial set of items and the outcome of the initial ranking and 3 voting rounds.

Table 1. Outcome of nominal group votes.

Term	Initial ranking	First round of vote	Second round of voting	Third round of voting
Satisfaction	Yes	Keep	__[a]	—
Ease of use	Yes	Keep	—	—
Perceived usefulness	Yes	Keep	—	—
Information quality	Yes	Keep	—	—
Functionality	Yes	Keep	—	—
Engagement	Yes	Keep	—	—
In line with professional standards	Yes	Keep	—	—
Relevance to course	Yes	Keep	—	—
Credibility of developers	Yes	Keep	—	—
Privacy of information	Yes	No decision	Discard	—
Cost	Yes	Keep	—	—
Advantage of using app	Yes	Keep	—	—
Efficiency	Yes	Keep	—	—
Instructional features	Yes	No decision	Keep	—
Capacity to generate learning	Yes	Keep	—	—
Aesthetics	Yes	No decision	Keep	—
Quantity of information	Yes	No decision	Keep	—
User ratings	Yes	No decision	Discard	—
Intention to reuse	Yes	Keep	—	—
Technical specifications	Yes	Keep	—	—
Feedback	Yes	No decision	Keep	—
Pedagogy	Yes	Keep	—	—
Perceived enjoyment	Yes	No decision	No decision	Discard
Perceived importance	Yes	No decision	Keep	—
Subjective quality	Yes	No decision	Keep	—
Sharing	Yes	No decision	No decision	No decision-Discard
Motivation	Yes	No decision	No decision	No decision-Discard
Transparent	Yes	No decision	Keep	—
User experience	Yes	Keep	—	—
Purpose	Yes	No decision	Keep	—
Self-directedness	Yes	No decision	No decision	No decision-Discard
Playfulness	Yes	No decision	Discard	—
Lack of ads	Yes	Keep	—	—
Differentiation	Yes	No decision	Keep	—
User interactivity	Yes	No decision	Keep	—
Product description	No	—	—	—

[a]The decision taken at each round of voting is shown. The voting round where a Keep or Discard decision is made ends the decision making for that item.

Development of Rubric

The 28 items on the list determined by the student nominal group were grouped by 2 authors (GT and TG) into 4 themes based on their similar definitions and the aspects of value they appeared to measure. The categories formed were user-centered measures (n=7), teaching and learning measures (n=9), professional measures (n=4), and usability measures (n=8).

Each category consists of a set of items that are described by posing questions for the user to consider. The questions were developed using definitions for the terms (written by GT) and the authors' perspective on what was most important with regard to that measure. After consulting the literature, a 5-point Likert-type scale was chosen as the rating tool (0=does not fulfill the item requirements, 1=poorly fulfills requirements, 2=somewhat fulfills requirements, 3=mostly fulfills requirements, and 4=fully meets requirements) with the descriptors for each point on the scale written to answer the item questions. The rubric scale descriptors were developed by 2 authors (TG and GT) with reference to the literature. Some items were adapted from other rating scales such as the Mobile App Rating Scale (MARS) [10], and these are clearly acknowledged on the final rubric (Multimedia Appendix 2). Scores for each item on the initial rubric are added to give a total score out of 112. Scores are used to classify apps as not at all valuable (<50), potentially valuable (51-69), and probably valuable (>69).

App Search

A total of 1291 iOS apps and 4190 Android apps were screened by title and description in the Apple and Google Play Stores, respectively. In iOS, 81 apps from 14 search terms (Textbox 5) were identified before removal of duplicates. After removal of duplicates, 35 unique apps were included for evaluation. For Android apps, the search terms identified 106 apps, of which 29 were unique. Apps found by only 1 researcher were not included because of concerns about the consistency of access to the app using typical search criteria. This gave a total of 64 apps with which to test the MARuL rubric.

Reliability Analysis

Initial testing and review of the rubric using the 10 selected apps occurred between June 18 and 22, 2019. One of the Android apps was unable to be tested by one of the reviewers as it did not run after installation on the device provided for testing and was excluded from further analysis.

Cronbach alpha for the overall value was excellent (α=.95), and for each of the categories was acceptable to excellent (α=.78-.96). The ICC for overall value was good (ICC=0.81) and moderate to good for each of the categories (ICC=0.71-0.85). Pearson correlations showed moderate-to-strong correlations between the categories (r=0.49-0.91; Table 2). Through analysis and discussion with the reviewers, a further refinement of the descriptors of 5 of the items was completed and 1 item (transparency) was removed from the professional category. In addition, 2 further items (cost and advertisements) in the usability category were considered for removal because of poor statistical performance, but the reviewers felt they should remain as they were likely to be of more immediate concern both to students and to those individuals considering the value of apps for learning. Figure 2 shows the process of development of the MARuL rubric from the initial rubric to the final version.

Version 3 of the MARuL rubric, tested with the remaining 54 apps, consisted of 27 items across 4 categories: user-centered measures (n=7), teaching and learning measures (n=9), professional measures (n=3), and usability measures (n=8). The 54 remaining apps were downloaded and tested between July 29 and October 8, 2019. Of the 54 apps for which testing was attempted, only 46 (25 iOS and 21 Android) could be tested completely for review. As the search took place, 7 apps were removed from their respective app stores and 1 app had updated to require sign up, one of the exclusion criteria, to be used. However, as 41 apps was the minimum sample size needed to determine interrater reliability with 90% assurance, analysis was continued [37].

The mean time to complete the MARuL rubric for each app was 8 min (SD 0.69). The 2 reviewers showed completion times ranging from 1 to 25 min (JM) and 1 to 33 min (SG). The median review time for each reviewer was 6.6 min (JM) and 4.1 min (SG). The discussion with the 2 reviewers indicated that in a small number of cases, there were challenges in evaluating the app, including difficulties with installation or using the app, which accounted for the longer evaluation times.

Cronbach alpha for the overall value was excellent at .96. The categories showed good-to-excellent internal consistency (α=.70-.96). Pearson correlations between the categories were moderate to strong (r range 0.58-0.90). Interrater reliability was also fair to good, with an overall value ICC (2-way) of 0.66 and categories with ICC ranging from 0.45 to 0.75.

From the internal consistency results, it was determined that the item for *cost* was performing in the reverse of its expected direction and did not add useful information to the usability category. It was removed from the rubric, leaving the usability category with 7 items, and the analyses were rerun. With cost removed, the alpha for the usability category improved from .70 to .82 (see Table 3 for full details of all categories).

Table 2. Reliability statistics for the initial version of the Mobile App Rubric for Learning after review of 10 apps.

Rubric categories	Intraclass correlation coefficient score[a]	Cronbach α[b]	Pearson r[c]			
			Teaching and learning	User centered	Professional	Usability
Teaching and learning	0.85	.89	1.00	0.91	0.83	0.72
User centered	0.78	.96	N/A[d]	1.00	0.72	0.71
Professional	0.71	.87	N/A	N/A	1.00	0.49
Usability	0.71	.78	N/A	N/A	N/A	1.00

[a]Column 1 presents the interrater reliability scores for each category.

[b]Column 2 presents the interitem consistency for each category.

[c]Columns 3 to 6 present the correlations between categories presented in the top right half of the table only.

[d]N/A: not applicable.

Figure 2. Rubric development process. MARuL: Mobile App Rubric for Learning.

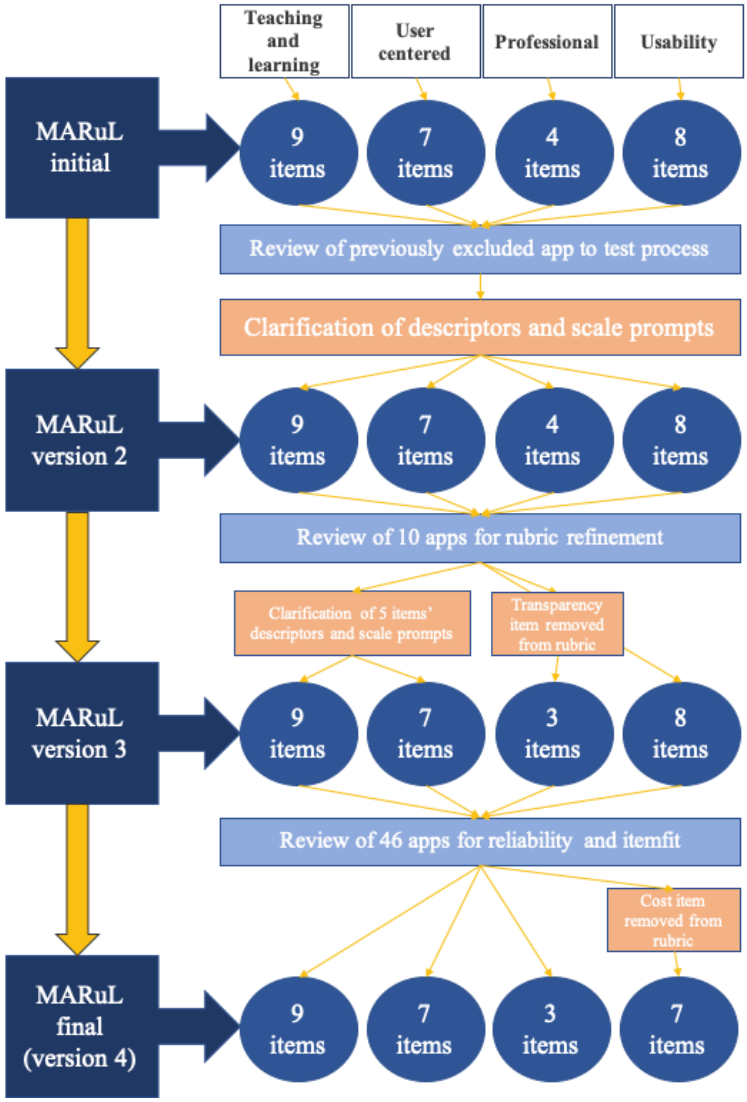


Table 3. Item statistics by category for final version of the Mobile App Rubric for Learning.

Category and item	Cronbach α	Values, mean (SD)
Teaching and learning	**.91**	**12.82 (6.93)**
Purpose	.75	2.33 (1.35)
Pedagogy	.83	1.97 (1.27)
Generates learning	.90	1.61 (1.05)
Quantity of information	.85	1.80 (1.39)
Relevance to study	.88	1.90 (1.04)
Instructional features	.76	1.15 (1.21)
User interactivity	.57	1.29 (0.95)
Feedback	.40	0.76 (0.88)
Efficiency	.93	1.42 (1.16)
User centered	**.96**	**10.89 (7.17)**
Subjective quality	.93	1.24 (1.13)
Satisfaction	.95	1.51 (1.23)
Perceived usefulness	.92	1.65 (1.15)
Perceived importance	.90	1.47 (1.03)
User Experience	.81	2.18 (1.00)
Intention to reuse	.94	1.33 (1.22)
Engagement	.91	1.52 (1.12)
Professional	**.74**	**5.94 (3.40)**
In line with standards	.79	2.46 (1.02)
Credibility	.82	2.94 (1.60)
Information quality	.83	1.43 (1.54)
Usability	**.82**	**16.47 (4.49)**
Aesthetics	.83	2.47 (0.93)
Functionality	.70	3.01 (0.94)
Differentiation	.76	1.40 (0.87)
Ease of use	.70	2.92 (0.80)
Advertisements	.36	3.73 (0.86)
Technical specifications	.61	1.23 (0.99)
Advantage of app	.87	1.71 (1.07)

The final version of the MARuL rubric (version 4) is provided in Multimedia Appendix 2.

Discussion

Principal Findings

The rigorous development of the MARuL has provided a robust and reliable instrument that can be used by medical students and their teachers to evaluate the value of apps to support just-in-time medical student learning. Potential rubric items were identified from a literature search; and medical students, the end users, identified relevant items via a structured nominal group technique. The preliminary instrument was refined, directed by analysis of internal consistency and interrater reliability, and the final MARuL instrument showed acceptable reliability and usability. Although rubrics for the evaluation of education and health apps are common, they tend to be generic [11,12]. We have developed a specific instrument for evaluating the value of medical education apps for learning.

We took a multidimensional approach to developing the MARuL in line with guidance from the literature to date [10,12]. Where appropriate, we adapted items, with acknowledgment, from extant instruments like the MARS. The MARS is a literature-informed, rigorously developed instrument that is widely used in the evaluation of health apps [10]. Our rubric measures the overall value of an app and represents the overall and category scores. This allows evaluation of how the app performs in different domains of value. To our knowledge, the use of multiple items to assess *professional measures* is unique to this rubric; medicine emphasizes the importance of a credible

and reliable source of information when informing student learning or patient education. The other 3 categories—*teaching and learning*, *user centered*, and *usability*—are common among evaluative measures [11,12]. These were included in the rubric because students agreed that they address aspects of apps that contribute to valuable student learning. During testing for the internal consistency of each category of the MARuL, cost was observed to be negatively related to the other items in the usability category. That is, apps with a high cost received a low score for the item, as students were less likely to pay for more expensive apps, but a high cost was related to high scores for other usability items. Cost remains an important consideration when choosing an app for student learning, as a lower cost is desirable, for individual students or institutional purchase, but a balance must be maintained between cost and other aspects of usability. Therefore, cost is captured in the basic information about each app (Multimedia Appendix 2), which is also the approach taken in the MARS [10].

The strength of this design is our focus on the end users, medical students, as the main source of input in the development of this rubric. This contrasts with most measures that rely on experts in education and technology to develop rubrics for evaluating apps [10,19]. Although special expertise is important when developing an accurate and reliable rubric, we believe the user voice is as important, if not more so, to ensure that the rubric is fit for purpose. Using a modified nominal group technique and student participants, we were able to confirm that the current literature surrounding evaluation of apps corresponded to student ideas about what made apps valuable to them in their learning. It also gives the user ownership of the means to evaluate the value of the technology they will be accessing.

Just-in-time learning is a common practice used by medical students in both the clinical environment and study situations [39]. Just-in-time learning especially relates to clinical skills, as medical students are constantly refining their history taking and physical examination skills with peers, in simulated clinical encounters, and with real patients. At an average of 8 min to complete after trialing an app, the MARuL rubric is easy to use and provides a fast evaluation of apps for learning.

Limitations

Although care was taken in the development and implementation of this project, there were some potential limitations. One researcher (GT) reviewed the final set of articles generated from the literature search to develop the list of potential evaluation criteria. As many articles were included, this was a large task, and it is possible that potential evaluation terms were overlooked and not extracted. Having two researchers consult the literature would have minimized this possibility; however, the number of terms extracted and their overlap, plus the nominal group's lack of additions during the initial ranking of terms, gives us confidence that the evaluation terms extracted from the literature search was comprehensive.

The recruitment of students to the nominal group was by convenience sampling. Research team members asked students to participate in the group. Every effort was made to recruit a variety of students, but the timing of the nominal group in the summer holidays limited the number of students available to take part. A related potential limitation is that students in the nominal group may not have felt confident in rejecting many of the criteria found during the literature review because of concerns around power and hierarchy. However, this was mitigated by having a medical student researcher facilitate the nominal group with another research team member, with no direct effect on student assessment, acting in a support role.

Finally, there are potential limitations because of the smaller than anticipated number of apps reviewed to complete the reliability testing. However, although eight apps identified in the original search were no longer available at the time of rubric testing, the total number reviewed was still acceptable for reliability analysis.

Next Steps

Our next steps for this research include further refinement of the rubric and construct validity testing to compare the MARuL with measures that evaluate health-related apps and education-oriented apps to measure convergent and discriminant validity. We have also adapted the language to create a student version of the rubric and plan to use it to look at the relationship between student evaluation of app value and teacher evaluation of app value. Finally, we believe that with minor changes to the language of some of the items, the rubric can be used with other types of health professional learning apps, for example, apps that focus on student learning of orthopedic skills.

Conclusions

The MARuL is a quick and user-friendly method that teachers can use to evaluate the value of an app for just-in-time learning. Through the inclusion of both experts and student stakeholders in the development process, it should be a robust method for teachers to use when deciding whether to download an app to recommend to students for just-in-time clinical skills learning.

Acknowledgments

This research was supported by a grant from the Otago Medical School Medical Education Research Fund. The authors would like to thank Michael Fouchelle, University of Otago Wellington Medical and Health Sciences Library, for his assistance with the initial literature search. They would like to thank Dr Hemakumar Devan, University of Otago Department of Physiotherapy, for acting as an internal reviewer during the preparation of this manuscript. Aspects of this research were presented at the 2019 Association for Medical Educators in Europe conference.

Authors' Contributions

TG developed the initial research question and methodology, designed the rubric, completed the app store searches, analyzed the data, and wrote the drafts of the manuscript. GT assisted with the development of the methodology, reviewed the literature, facilitated the nominal group, designed the rubric, completed the app store searches, and edited the manuscript. SG assisted with the development of the methodology, reviewed the apps using the MARuL, assisted in refinement of the MARuL, and edited the manuscript. JM reviewed the apps using the MARuL and assisted in refinement of the MARuL. SR and RG assisted with the development of the methodology and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Terms used to evaluate the value of apps and/or mobile technology.
[PDF File (Adobe PDF File), 95 KB - mhealth_v8i7e18015_app1.pdf]

Multimedia Appendix 2
Final version of the Mobile App Rubric for Learning (MARuL).
[PDF File (Adobe PDF File), 122 KB - mhealth_v8i7e18015_app2.pdf]

References

1. Payne KF, Wharrad H, Watts K. Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
2. Aungst TD, Clauson KA, Misra S, Lewis TL, Husain I. How to identify, assess and utilise mobile medical applications in clinical practice. Int J Clin Pract 2014 Feb;68(2):155-162. [doi: 10.1111/ijcp.12375] [Medline: 24460614]
3. Wallace S, Clark M, White J. 'It's on my iPhone': attitudes to the use of mobile computing devices in medical education, a mixed-methods study. BMJ Open 2012;2(4):- [FREE Full text] [doi: 10.1136/bmjopen-2012-001099] [Medline: 22923627]
4. Wynter L, Burgess A, Kalman E, Heron JE, Bleasel J. Medical students: what educational resources are they using? BMC Med Educ 2019 Jan 25;19(1):36 [FREE Full text] [doi: 10.1186/s12909-019-1462-9] [Medline: 30683084]
5. Koehler N, Yao K, Vujovic O, McMenamin C. Medical students' use of and attitudes towards medical applications. J Mob Technol Med 2012 Dec 22;1(4):16-21. [doi: 10.7309/jmtm.73]
6. Boruff JT, Storie D. Mobile devices in medicine: a survey of how medical students, residents, and faculty use smartphones and other mobile devices to find information. J Med Libr Assoc 2014 Jan;102(1):22-30 [FREE Full text] [doi: 10.3163/1536-5050.102.1.006] [Medline: 24415916]
7. iMedicalApps - Reviews of Medical Apps & Healthcare Technology. 2019. URL: https://www.imedicalapps.com/ [accessed 2018-07-13]
8. Dolan B. Top Five Medical Apps at Harvard Medical School. Mobi Health News. 2011. URL: https://www.mobihealthnews.com/10745/top-five-medical-apps-at-harvard-medical-school [accessed 2018-07-13]
9. Scott IA, Scuffham P, Gupta D, Harch TM, Borchi J, Richards B. Going digital: a narrative overview of the effects, quality and utility of mobile apps in chronic disease self-management. Aust Health Rev 2020 Feb;44(1):62-82. [doi: 10.1071/AH18064] [Medline: 30419185]
10. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
11. Jin M, Kim J. Development and evaluation of an evaluation tool for healthcare smartphone applications. Telemed J E Health 2015 Oct;21(10):831-837. [doi: 10.1089/tmj.2014.0151] [Medline: 26431261]
12. Lee J, Kim S. Validation of a tool evaluating educational apps for smart education. J Educ Comput Res 2015 Mar 26;52(3):435-450. [doi: 10.1177/0735633115571923]
13. 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 [FREE Full text] [doi: 10.1016/j.jbi.2018.05.017] [Medline: 29852318]
14. Ok MW, Kim MK, Kang EY, Bryant BR. How to find good apps: an evaluation rubric for instructional apps for teaching students with learning disabilities. Interv Sch Clin 2015 Jun 19;51(4):244-252. [doi: 10.1177/1053451215589179]
15. Papadakis S, Kalogiannakis M, Zaranis N. Designing and creating an educational app rubric for preschool teachers. Educ Inf Technol 2017 Jan 30;22(6):3147-3165. [doi: 10.1007/s10639-017-9579-0]
16. Hurst EJ. Evaluating health and wellness mobile applications. J Hosp Librariansh 2018 Jul 17;18(3):266-278. [doi: 10.1080/15323269.2018.1472928]

17. Böhme C, von Osthoff MB, Frey K, Hübner J. Development of a rating tool for mobile cancer apps: information analysis and formal and content-related evaluation of selected cancer apps. J Cancer Educ 2019 Feb;34(1):105-110. [doi: 10.1007/s13187-017-1273-9] [Medline: 28819928]
18. Fiore P. How to evaluate mobile health applications: a scoping review. Stud Health Technol Inform 2017;234:109-114. [Medline: 28186025]
19. Chener TS, Lee C, Fegely A, Santaniello LA. A detailed rubric for assessing the quality of teacher resource apps. J Inf Technol Educ Innov Pract 2016;15:117-143. [doi: 10.28945/3527]
20. Eccles JS, Wigfield A. In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. Pers Soc Psychol Bull 2016 Jul 2;21(3):215-225. [doi: 10.1177/0146167295213003]
21. Brandenburg DC, Ellinger AD. The future: just-in-time learning expectations and potential implications for human resource development. Adv Dev Hum Resour 2016 Jul 21;5(3):308-320. [doi: 10.1177/1523422303254629]
22. Rothman BS, Gupta RK, McEvoy MD. Mobile technology in the perioperative arena: rapid evolution and future disruption. Anesth Analg 2017 Mar;124(3):807-818. [doi: 10.1213/ANE.0000000000001858] [Medline: 28151816]
23. Jamu JT, Lowi-Jones H, Mitchell C. Just in time? Using QR codes for multi-professional learning in clinical practice. Nurse Educ Pract 2016 Jul;19:107-112. [doi: 10.1016/j.nepr.2016.03.007] [Medline: 27428702]
24. HARDEN RM. AMEE guide no 14: outcome-based education: part 1-an introduction to outcome-based education. Med Teach 1999;21(1):7-14. [doi: 10.1080/01421599979969]
25. Popham J. What's wrong--and what's right--with rubrics. Educ Lead 1997;55(2):72-75 [FREE Full text]
26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
27. O'Neil MJ, Jackson L. Nominal group technique: a process for initiating curriculum development in higher education. Stud High Educ 2006 Aug 5;8(2):129-138. [doi: 10.1080/03075078312331378994]
28. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as delphi and nominal group in medical education research. Med Teach 2017 Jan;39(1):14-19. [doi: 10.1080/0142159X.2017.1245856] [Medline: 27841062]
29. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
30. 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]
31. Airtable. 2019. URL: https://www.airtable.com/ [accessed 2019-01-10]
32. Qualtrics. 2019. URL: https://www.qualtrics.com/ [accessed 2019-03-01]
33. Rstudio: Open Source & Professional Software for Data.. 2019. URL: http://www.rstudio.com/ [accessed 2019-09-11]
34. Gamer M, Lemon J, Fellows I, Singh P. irr: Various Coefficients of Interrater Reliability and Agreement. The Comprehensive R Archive Network. 2019. URL: https://CRAN.R-project.org/package=irr [accessed 2019-09-11]
35. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. The Comprehensive R Archive Network. 2018. URL: https://CRAN.R-project.org/package=psych [accessed 2019-09-11]
36. Wickham H. Reshaping data with the reshape package. J Stat Soft 2007;21(12):-. [doi: 10.18637/jss.v021.i12]
37. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
38. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012 Dec 20;31(29):3972-3981. [doi: 10.1002/sim.5466] [Medline: 22764084]
39. Davies BS, Rafique J, Vincent TR, Fairclough J, Packer MH, Vincent R, et al. Mobile medical education (MoMEd) - how mobile information resources contribute to learning for undergraduate clinical students - a mixed methods study. BMC Med Educ 2012 Jan 12;12:1 [FREE Full text] [doi: 10.1186/1472-6920-12-1] [Medline: 22240206]

Abbreviations

e-learning: electronic learning
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
MARuL: Mobile App Rubric for Learning

Edited by G Eysenbach; submitted 28.01.20; peer-reviewed by N Farič, R Grad; comments to author 25.03.20; revised version received 29.04.20; accepted 30.04.20; published 31.07.20.

Please cite as:
Gladman T, Tylee G, Gallagher S, Mair J, Rennie SC, Grainger R
A Tool for Rating the Value of Health Education Mobile Apps to Enhance Student Learning (MARuL): Development and Usability Study
JMIR Mhealth Uhealth 2020;8(7):e18015
URL: https://mhealth.jmir.org/2020/7/e18015
doi:10.2196/18015
PMID:32735228

Original Paper

Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study

Sang Hoon Chae[1], MD, PhD; Yushin Kim[2], PhD; Kyoung-Soub Lee[3], MS; Hyung-Soon Park[1,3], PhD

[1]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[2]Major of Sports Health Rehabilitation, Cheongju University, Cheongju, Republic of Korea

[3]Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

Corresponding Author:
Hyung-Soon Park, PhD
Department of Mechanical Engineering
Korea Advanced Institute of Science and Technology
291 Daehak-ro, Yuseong-gu
Daejeon, 34141
Republic of Korea
Phone: 82 42 350 3038
Email: hyungspark@kaist.ac.kr

Abstract

Background: Recent advancements in wearable sensor technology have shown the feasibility of remote physical therapy at home. In particular, the current COVID-19 pandemic has revealed the need and opportunity of internet-based wearable technology in future health care systems. Previous research has shown the feasibility of human activity recognition technologies for monitoring rehabilitation activities in home environments; however, few comprehensive studies ranging from development to clinical evaluation exist.

Objective: This study aimed to (1) develop a home-based rehabilitation (HBR) system that can recognize and record the type and frequency of rehabilitation exercises conducted by the user using a smartwatch and smartphone app equipped with a machine learning (ML) algorithm and (2) evaluate the efficacy of the home-based rehabilitation system through a prospective comparative study with chronic stroke survivors.

Methods: The HBR system involves an off-the-shelf smartwatch, a smartphone, and custom-developed apps. A convolutional neural network was used to train the ML algorithm for detecting home exercises. To determine the most accurate way for detecting the type of home exercise, we compared accuracy results with the data sets of personal or total data and accelerometer, gyroscope, or accelerometer combined with gyroscope data. From March 2018 to February 2019, we conducted a clinical study with two groups of stroke survivors. In total, 17 and 6 participants were enrolled for statistical analysis in the HBR group and control group, respectively. To measure clinical outcomes, we performed the Wolf Motor Function Test (WMFT), Fugl-Meyer Assessment of Upper Extremity, grip power test, Beck Depression Inventory, and range of motion (ROM) assessment of the shoulder joint at 0, 6, and 12 months, and at a follow-up assessment 6 weeks after retrieving the HBR system.

Results: The ML model created with personal data involving accelerometer combined with gyroscope data (5590/5601, 99.80%) was the most accurate compared with accelerometer (5496/5601, 98.13%) or gyroscope data (5381/5601, 96.07%). In the comparative study, the drop-out rates in the control and HBR groups were 40% (4/10) and 22% (5/22) at 12 weeks and 100% (10/10) and 45% (10/22) at 18 weeks, respectively. The HBR group (n=17) showed a significant improvement in the mean WMFT score (P=.02) and ROM of flexion (P=.004) and internal rotation (P=.001). The control group (n=6) showed a significant change only in shoulder internal rotation (P=.03).

Conclusions: This study found that a home care system using a commercial smartwatch and ML model can facilitate participation in home training and improve the functional score of the WMFT and shoulder ROM of flexion and internal rotation in the treatment of patients with chronic stroke. This strategy can possibly be a cost-effective tool for the home care treatment of stroke survivors in the future.

Trial Registration: Clinical Research Information Service KCT0004818; https://tinyurl.com/y92w978t

(JMIR Mhealth Uhealth 2020;8(7):e17216) doi:10.2196/17216

KEYWORDS

home-based rehabilitation; artificial intelligence; machine learning; wearable device; smartwatch; chronic stroke

Introduction

Stroke is a major cause of disability in adults. About 13.7 million cases of stroke occur each year globally, but half of the patients are unable to restore enough upper extremity function required for daily living [1,2]. The rehabilitation required after stroke has been limited to the first 3 to 6 months of hospitalization following the stroke [3]. For the best recovery following stroke and prevention of recurrence, stroke survivors need ongoing home rehabilitation [4-7]. Previous literature has proven that continued home rehabilitation can activate neuroplasticity in chronic poststroke patients and result in greatly enhanced clinical outcomes [8-10]. In addition, the need for a high-quality home health care system is drawing greater attention with the recent COVID-19 pandemic. The major barriers in delivering high-quality home rehabilitation services are high cost and labor intensiveness [11,12]. Therefore, socioeconomically deprived people are less likely to receive high-quality rehabilitation care and more likely to experience recurrence and poor quality of life [13,14]. The burdensome labor of home care also puts the care giver and receiver at risk for poor mental health and depression [15,16].

To overcome the barriers for home rehabilitation, potential technology-enabled solutions have been suggested. For example, there are two kinds of technology used as solutions (vision-based solution and wearable sensor–based solution). The vision-based approach (eg, interactive television or Kinect) could be easier to use since it does not require any wearing of devices [17-21]. A vision-based system can only be used within a limited range of space, whereas wearable systems can be used anywhere, which would be advantageous for promoting the frequency of use [22,23].

To promote the frequency of use, we developed an upper limb home-based rehabilitation (HBR) system using wearable sensors embedded in a commercial smartwatch. A machine learning (ML) algorithm implemented by a convolutional neural network (CNN) was used to recognize four kinds of home exercise activities. While participants perform these home exercises, the HBR system makes it possible to share their home exercise data with therapists at remote locations. It helps therapists to encourage and communicate with chronic stroke survivors.

We conducted a prospective comparative study to evaluate the effectiveness of our HBR system. As the long-term goal of this study, we intended to investigate the benefits of using artificial intelligence–based HBR compared with those of conventional therapy. Herein, we compared the clinical outcomes of an experimental (HBR) group using the HBR system with those of a control group performing conventional home exercises. We hypothesized that the HBR group would show enhanced clinical outcomes compared with the control group [24]. This paper elaborates on the technological advancements pertaining to the detection of home exercise activities using a smartwatch (the ML model) and the results from a clinical trial.

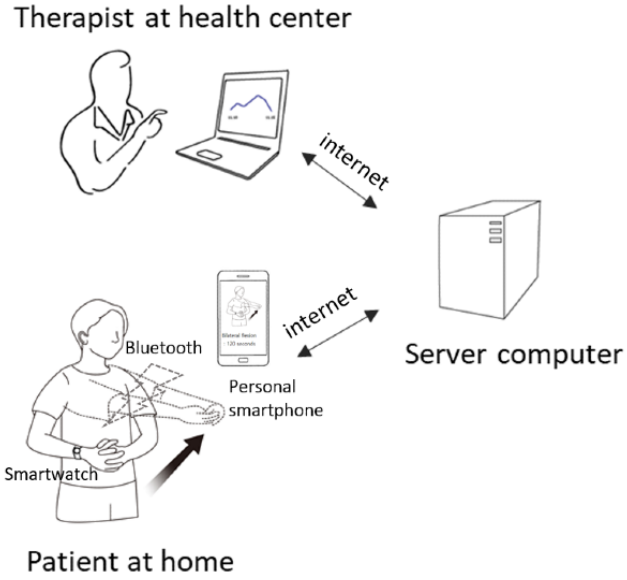
Methods

Development of an HBR System

Overview of an HBR System

We implemented an HBR system that can connect patients and therapists at a distance. Figure 1 presents an overview of our HBR system. To make the interface simple and user-friendly, we used a commercial smartwatch (watch style W270, LG,) that can be connected to a personal smartphone after installing a custom-programmed app. In our system, the smartwatch, which includes an inertial measurement unit (IMU) sensor, sent sensor data to the smartphone via Bluetooth communication while patients were doing exercise. The personal smartphone served as a platform for receiving sensor data, classifying the data, and transmitting the results to a server computer via the internet (Multimedia Appendix 1). The apps for the smartwatch and smartphone were developed using the Android Software Development Kit (Android Studio 2.3, Google).

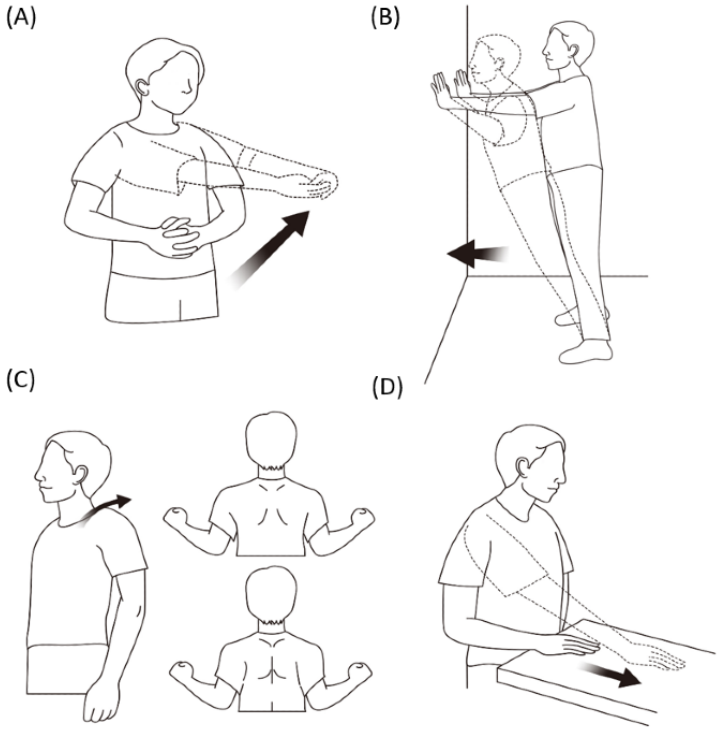
Figure 1. Overview of the home-based rehabilitation system.



Selection of Home Rehabilitation Exercise Tasks

We selected four exercise tasks based on bilateral movement therapy, which is called bilateral arm training rehabilitation. Previous literature has shown that bilateral arm training can induce reorganization in contralateral motor networks by interhemispheric crosstalk and evoke functional recovery of the upper extremities in chronic stroke survivors [25,26]. As shown in Figure 2, the following exercises were selected: (1) bilateral shoulder flexion with both hands interlocked; (2) wall push exercise; (3) active scapular exercise; and (4) towel slide exercise.

Figure 2. Home rehabilitation exercises for the upper limbs. (A) Bilateral shoulder flexion; (B) Wall push exercise; (C) Active scapular exercise; (D) Towel slide exercise.



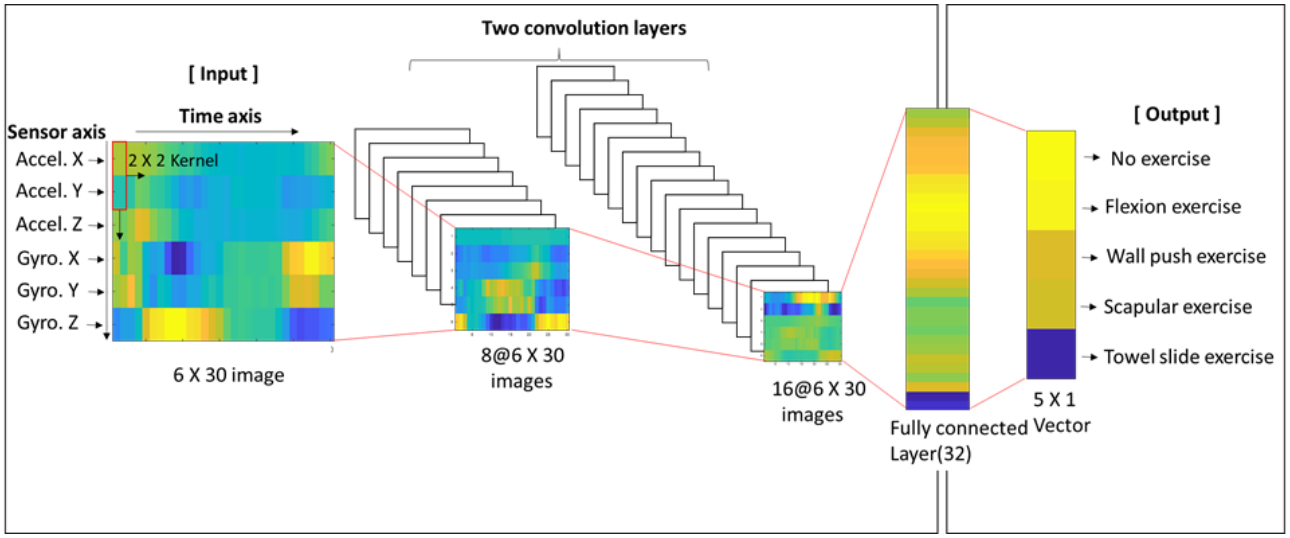
Machine Learning Algorithm for Home Exercise Recognition

There are various kinds of deep learning algorithms for human activity recognition [27,28]. Among them, we selected the CNN as it has been reported to be highly accurate in human activity recognition and simpler than other algorithms since it does not need feature extraction [29,30]. We made a program for building the ML model with Python script (Python 3.5, Python Software) and CNNs in the TensorFlow platform (Tensorflow 1.7.0, Google).

The kinematic data structure from the IMU sensor consists of three-axis (x, y, and z) accelerometer and gyroscope data. When

a patient is exercising while wearing the smartwatch, the accelerometer and gyroscope of the IMU sensor measure acceleration and velocity during the exercise. Since all sensor data are in a time series sampled at 10 Hz, the entire data can be represented by a two-dimensional matrix with a time axis (horizontal) and a sensor axis (vertical) as shown in Figure 3. We used the sliding time window method and applied a 3-second time window according to the experimental results that compared the performance of the ML algorithm at various time windows [31].

Figure 3. Baseline convolutional neural network architecture.



The training data essential for implementing the ML algorithm were obtained on the first day of the meeting. As it was difficult to meet patients with chronic stroke due to the difficulty of moving, we gathered the data on the same day just after explaining about the four kinds of home exercises. Participants were asked to repeat them 15 times in two sessions wearing a smartwatch.

Figure 3 reflects our baseline CNN architecture. Two convolution layers, which have 8 and 16 feature maps, are followed by a fully connected layer that has 32 nodes. Rectified units are employed as activation functions, and SoftMax functions are used for evaluating the final five output node values.

We experimented with the following two types of ML models: a ML model built with the personal data set and another ML model built with the total data set. The personal data set was composed of the exercise data of the user, whereas the total data set consisted of all participants' exercise data including the user. In order to evaluate the accuracy of each model, we applied a cross-validation test.

Cross-Validation Test for Accuracy Comparison

A five-fold cross-validation test was performed to test the accuracy of the ML model in recognizing exercise tasks. We divided the data into one test data set and four training data sets. The training data sets were used to build the ML model, and the test data set was used to determine the accuracy of the ML model built. Thus, we compared the accuracy of the model created by personal data versus total data. Additionally, we compared the accuracy between models based on each sensor data (accelerometer only, gyroscope only, and accelerometer and gyroscope combined) to determine which sensor data are most accurate for exercise prediction. Accuracy was calculated by using the following formula:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Development of Mobile Apps

We implemented the following three different android apps: (1) smartwatch app; (2) smartphone app for patients; and (3) smartphone app for physiotherapists (Android Studio 2.3, Google). The smartwatch app is designed to transmit sensor data to a smartphone as soon as the exercise button on the smartphone is pressed and to stop transmission when the use of the app on the smartphone has ended. There is no start or stop button on the smartwatch. We made the smartphone app automatically close as soon as the android app of the smartphone shuts down. The smartphone app for patients acts as a platform for starting the smartwatch, detecting home rehabilitation, and transmitting exercise time data to the server computer. The personalized ML model embedded in the smartphone app recognizes the type of exercise that the participant is doing. After recognition of the exercise, the smartphone calculates and transmits the exercise time via the internet. It also shows the personal rehabilitation time of the previous 3 days on pressing a button in the app. Lastly, the smartphone app for physiotherapists provides the physiotherapists with the rehabilitation status of all enrolled patients for the past 1 month for convenient statistical evaluation.

Clinical Trial: Prospective Comparative Study

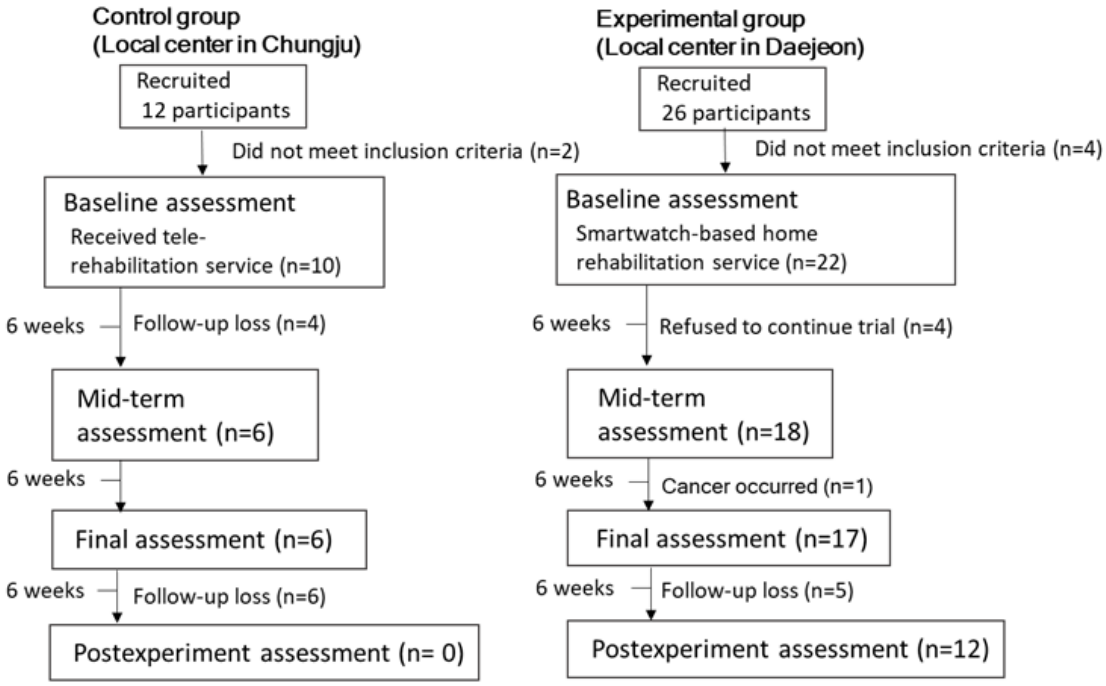
Experiment Design

We performed a clinical trial in two local health care centers located in two cities in South Korea (Cheongju [control group] and Daejeon [HBR group]). In the two centers, we recruited 12 and 26 patients with chronic stroke, respectively. The inclusion criteria were as follows: (1) age 40 to 70 years, (2) mild to

moderate neurologic deficit with hemiplegia, (3) more than 6 months after the onset of stroke, (4) 24 points or more in the Korean version of the Mini-Mental State Examination (K-MMSE), and (5) ability to understand the procedures and communicate with the supervisor. The exclusion criteria were as follows: (1) arthritis of the glenohumeral joint, (2) rotator cuff tear, (3) cervical root syndrome, (4) subluxation of the shoulder joint, (5) reluctance to follow the home exercise regimens of this study, and (6) no smartphone with Android OS.

Figure 4 shows the time flow of the study. According to the criteria, we excluded two patients in the control group owing to shoulder pain during exercise. In the HBR group, four patients were excluded (one patient had rotator cuff repair surgery previously, one had shoulder subluxation, and two had shoulder pain during exercise). After excluding those patients, 10 and 22 patients were initially enrolled in the control and HBR groups, respectively. While performing the home rehabilitation program, four patients in the control group dropped out. Drop-out was determined by a therapist who undertook the task of home-based rehabilitation for participants who did not respond to phone calls. In the HBR group, four patients gave up using our HBR system, as they were unfamiliar with the information technology devices and experienced difficulties in their usage. One patient missed the rehabilitation owing to deterioration caused by other underlying diseases. Finally, six patients in the control group and 17 patients in the HBR group completed the protocol at 12 weeks. To determine the changes in clinical scores, we tried to conduct an assessment at 6 weeks after the final assessment (18 weeks). Participants in the control group did not respond to our call, whereas participants in the HBR group cooperated with the requirement for assessment.

Figure 4. Time flow of the study.



All patients in the control group received personal education about the four exercise tasks for 30 minutes at the beginning of study enrollment. In the control group, the participants received a printed handout to remind them about how to perform the four exercise tasks. In contrast, participants in the HBR group received the same education and were given a smartwatch, and the HBR apps were installed on their own smartphones on the first day of the meeting. During the education, we acquired learning data for the ML algorithm in the HBR group. The physiotherapist taught each individual patient how to perform the four exercise tasks, and the training data were labeled manually while the patient was practicing each of the four tasks. In other words, with the smartwatch worn, participants were asked to repeat each home exercise 15 times in two sessions. In total, data were collected for 120 exercise attempts.

In both groups, weekly calls were made by the same therapists [17]. To avoid bias from the examiner, two therapists equally divided the participants present in both groups and managed them. They encouraged the participants to perform home exercise and answered any questions regarding how to perform the home exercise from participants in both groups. Since the control group did not use any sensor at home, one additional question was asked regarding how much time the control participants spent on home exercise.

In the HBR group, the participants were able to obtain their own home exercise results, and it was possible for the therapist to access the data of all the participants. Thereafter, the physiotherapist communicated with the participants and encouraged them based on the home exercise data collected by the HBR system (Multimedia Appendix 2).

All participants were asked to come to the local health care centers for outcome evaluation, and two physical therapists at each center conducted the clinical assessments at 0 weeks (baseline), 6 weeks (mid-term), and 12 weeks (final). In addition, we conducted one more assessment at 18 weeks, which was a 6-week follow-up after the completion of the home rehabilitation

program, to examine the change after our home rehabilitation program. For functional scoring, the Fugl-Meyer Assessment of Upper Extremity (FMA-UE) and Wolf Motor Function Test (WMFT) were used. We also evaluated psychologic depression using Beck Depression Inventory (BDI), grip power using a dynamometer (Patterson Medical), and shoulder range of motion (ROM) angle using a goniometer.

Statistical Analysis

We used descriptive statistics to characterize the demographics and analyzed the difference between the control and HBR groups at baseline using the Mann-Whitney *U* test. We compared the clinical results of functional recovery (WMFT and FMA-UE), grip power, BDI, and ROM using the Friedman test. As a post-hoc analysis, the Wilcoxon signed-rank test with Bonferroni correction was used. SPSS software was utilized for all statistical analyses (SPSS statistics 25, IBM Corp).

The sample size was determined according to a previous study (using virtual reality home training) related to this study [32]. Based on the result of a mean ROM increase of 41.67° (SD 22.29°), we calculated the sample size using G*Power software (two-tailed; α error, .05; power [1–β error], 0.8; effect size, 1.869; loss rate, 10%). This power analysis showed a sample size of 6.

Results

Accuracy Results of the HBR System

It was impossible to detect the accuracy of the home exercise activity logs as the ground truth of home exercise motion of the participants was unknown owing to matters pertaining to personal privacy. Instead, we attempted to ascertain what types of sensor data can detect home exercise activities most accurately via a cross-validation test.

The results shown in Table 1 represent the accurate values that were calculated by the cross-validation test with different ML models depending on various input data and sensor data. With regard to the input data, the ML model trained by accelerometer combined with gyroscope data had the best accuracy compared with other models. In particular, the ML model developed using personal data (99.9%) was more accurate than the model developed using total data (95.8%), although the amount of personal data was much smaller than that of total data.

Multimedia Appendix 3 shows the results of the cross-validation test.

Table 1. Accuracy of the convolutional neural network model according to exercise.

Exercise	Personal data (%)			Total data (%)		
	A[a]	G[b]	A+G[c]	A	G	A+G
No exercise	100 (1224/1224)[d]	95.8 (1172/1224)	100 (1224/1224)	97.4 (1192/1224)	97.8 (1197/1224)	100 (1224/1224)
Bilateral flexion	98.5 (1103/1120)	97.8 (1095/1120)	99.0 (1109/1120)	96.1 (1076/1120)	98.1 (1099/1120)	97.2 (1089/1120)
Wall push	99.0 (1014/1024)	93.7 (959/1024)	100 (1024/1024)	93.8 (960/1024)	86.5 (886/1024)	93.2 (954/1024)
Active scapula	93.0 (1030/1108)	97.6 (1081/1108)	100 (1108/1108)	92.2 (1022/1108)	87.0 (964/1108)	93.0 (1030/1108)
Towel slide	100 (1125/1125)	95.5 (1074/1125)	100 (1125/1125)	94.2 (1060/1125)	88.7 (998/1125)	95.5 (1074/1125)
Total	98.1 (5496/5601)	96.0 (5381/5601)	99.9 (5590/5601)	94.8 (5310/5601)	91.8 (5144/5601)	95.8 (5371/5601)

[a]Accelerometer data.

[b]Gyroscope data.

[c]Accelerometer combined with gyroscope data.

[d]Indicates the value of correct samples divided by total samples.

Results of the Clinical Trial

The study was approved by the Institutional Review Board (IRB no.: IRB-17-299). Informed consent was obtained from all participants. This study was supported by the Korea Advanced Institute of Science and Technology-funded Global Singularity Research Program. Patients were recruited from March 2018 to September 2018, and home exercise data were collected until February 2019. As of March 2019, we enrolled 23 stroke survivors for data analysis. Drop-out rates in the control and HBR groups were 40% (4/10) and 22% (5/22) at 12 weeks and 100% (10/10) and 45% (10/22) at 18 weeks, respectively.

Table 2 presents the demographics and baseline assessment findings. There were no relevant differences between the two groups.

To evaluate exercise compliance at home, in the control group, a telephone survey was the only approach to determine the home exercise activities of participants. Thus, we called them and asked how much time they exercised at home and encouraged

exercise. We found that they performed home exercise for about 13.6 (SD 4.85) min/day. However, the numbers obtained for the control group might not be accurate owing to the limitations of a verbal survey. In contrast, in the HBR group, the home exercise results of all the participants were provided by the smartphone app. Thus, we encouraged participants to perform home exercise based on the data.

Table 2. Patient demographics and baseline assessment in the control and home-based rehabilitation groups.

Characteristic	Control group (n=6), mean (SD)	HBR[a] group (n=17), mean (SD)	*P* value[b]
Age (years)	64.5 (9.6)	58.3 (9.3)	.25
Functional assessment test			
WMFT[c]	38.8 (25.6)	39.7 (22.2)	.91
FMA-UE[d]	29.0 (14.2)	36.6 (18.6)	.35
Grip power (kg)	11.7 (11.6)	13.3 (12.7)	.75
BDI[e]	24.2 (11.2)	17.88 (14.7)	.28
Shoulder ROM[f]			
Flexion	82.0 (59.07)	74.5 (45.3)	.91
Extension	40.8 (19.6)	28.7 (21.0)	.11
Internal rotation	50.8 (31.2)	50.43 (24.5)	.51
External rotation	23.4 (28.1)	16.84 (17.69)	.97

[a]HBR: home-based rehabilitation.

[b]*P* values were calculated with the Mann-Whitney *U* test.

[c]WMFT: Wolf Motor Function Test.

[d]FMA-UE: Fugl-Meyer Assessment of Upper Extremity.

[e]BDI: Beck Depression Inventory.

[f]ROM: range of motion.

On average, participants in the HBR group performed bilateral flexion exercise for 7.27 (SD 10.1) min/day, wall push exercise for 3.76 (SD 9.01) min/day, active scapula exercise for 4.82 (SD 9.62) min/day, and towel slide for 6.70 (SD 11.87) min/day. In total, home exercise was performed for an average of 22.57 (SD 37.69) min/day.

Table 3 presents the clinical results at the baseline, mid-term (6 weeks), and final assessments (12 weeks). In total, 23 individuals with chronic stroke completed this research (control: 6; HBR: 17). In the HBR group, the WMFT, BDI, and shoulder ROM of flexion and internal rotation showed relevant progression (Multimedia Appendix 4). However, FMA-UE showed no significant difference (*P*=.46). In the control group, there was no significant difference, except for the ROM of internal rotation (*P*=.03). In both groups, there was no relevant difference in the grip power test.

We tried to determine the change in clinical results after the completion of the home rehabilitation program, which had a duration of 12 weeks. Thus, we compared the clinical outcomes at the final assessment (12 weeks) with that at 6 weeks after removing the HBR system (18 weeks). However, there was no relevant difference (Multimedia Appendix 5).

Table 3. Clinical results in the control and home-based rehabilitation groups during the experiment.

Characteristic	Control group (n=6)				HBR[a] group (n=17)			
	0 weeks, mean (SD)	6 weeks, mean (SD)	12 weeks, mean (SD)	*P* value[b]	0 weeks, mean (SD)	6 weeks, mean (SD)	12 weeks, mean (SD)	*P* value[b]
Functional assessment test								
WMFT[c]	38.8 (25.6)	40.3 (25.7)	42.2 (22.8)	.69	39.7 (22.7)	40.5 (23.6)	42.5 (23.7)	.02
FMA-UE[d]	29.0 (14.2)	30.0 (14.2)	28.5 (16.1	.72	36.6 (18.7)	37.5 (18.4)	38.5 (18.3)	.46
Grip power (kg)	11.7 (11.6)	11.0 (10.4)	10.9 (10.3)	.47	13.3 (12.7)	12.9 (12.0)	14.8 (12.1)	.34
BDI[e]	24.2 (11.2)	10.0 (8.6)	8.8 (7.2)	.11	17.9 (14.7)	10.0 (8.8)	8.0 (9.9)	.06
Shoulder ROM[f]								
Flexion	82.0 (59.1)	90.6 (65.3)	87.5 (61.0)	.21	74.5 (45.3)	93.9 (52.3)	94.7 (48.9)	<.001
Extension	40.8 (19.6)	29.5 (18.9)	32.8 (20.4)	.38	28.7 (21.0)	31.5 (16.2)	34.7 (19.9)	.16
Internal rotation	50.8 (31.2)	48.5 (28.6)	57.3 (32.0)	.03	50.4 (24.5)	70.3 (28.3)	63.5 (26.9)	.001
External rotation	23.4 (28.1)	23.6 (30.0)	26.6 (27.7)	.76	16.8 (17.7)	15.4 (18.1)	16.9 (18.4)	.20

[a]HBR: home-based rehabilitation.

[b]Overall *P* values were calculated with the Friedman test.

[c]WMFT: Wolf Motor Function Test.

[d]FMA-UE: Fugl-Meyer Assessment of Upper Extremity.

[e]BDI: Beck Depression Inventory.

[f]ROM: range of motion.

Discussion

Principal Findings

In this study, we performed a comprehensive assessment based on a ML algorithm and wearable device. We developed an HBR system using a commercial smartwatch with the ML model and evaluated the effectiveness of the HBR system via a clinical trial. The ML model based on a CNN algorithm showed good to excellent accuracy ranging from 86.5% to 100%, and the clinical trial showed a relevant increase in ROM and the WMFT function score.

While developing the HBR system using a commercial smartwatch, determining the types of sensors that provide maximal accuracy was an important issue. According to previous research that used an IMU sensor for activity recognition, an accelerometer is the most accurate sensor for activity recognition [29,30,33-36]. Based on the results of the cross-validation test in our study, the accelerometer signal combined with gyroscope findings provided the most accurate results. This is consistent with the result in the study by Hyunh et al [37], which attempted to detect falls by using an IMU sensor at the chest. It was reported that adding a gyroscope can reduce the false-positive rate and increase specificity from 82.72% to 96.20%. However, in our study, the difference in the accuracy of the results obtained when using an accelerometer and an accelerometer combined with a gyroscope was relatively small (1.1%-1.8%), and in the case of active scapular exercises, the accuracy of the gyroscope was even higher than that of other approaches. Thus, we believe that the choice of the most accurate sensor may depend on the type of exercise and the location of the sensor. Our research, which involved the detection of repetitive and slow home exercise tasks by a smartwatch, showed that the combination of an accelerometer and a gyroscope provided the most accurate signal. However, considering that the improvement in accuracy with the addition of a gyroscope was relatively small and that the addition required doubled computation and battery loading for the gyroscope, we believe that an accelerometer-only signal could be an alternative choice.

Since the learning data set is a decisive factor in optimizing the ML algorithm, we compared the accuracy of the ML model built with personal data and that built with total data. The ML model based on total data was built with data from all participants and the ML model based on personal data was implemented by using the participants' own data. Although the amount of total data was larger than that of personal data, each exercise motion in the total data set represents a mix of different motions of all participants. Therefore, through this comparison,

we attempted to determine whether the quantity or quality of data is important. According to the results, we found that quality was more important. The ML model built with only personal data (99.9%), which represented the quality of data, was more accurate than the ML model built with total data (95.4%), which represented the quantity of data. This means that data personalization was more important than the total amount of data, especially for chronic stroke patients who had various disabilities and individual motion characteristics. We think that the different exercise motions of other patients contaminated the data consistency and had a bad influence on the ML model [38].

With regard to the clinical trial, the HBR group showed significant functional recovery (mean difference=2.8, *P*=.02) in the WMFT. However, FMA-UE did not show significant results (mean difference=1.9, *P*=.46). We think this is related to the different traits of both functional assessment methods. FMA-UE (total 66 scores) is an assessment tool for identifying motor impairment, and it involves reflex activity (6), flexor synergy (12), extensor synergy (6), combining synergy (6), movement out of synergy (6), wrist (10), hand (14), and coordination/speed (6) on an ordinal scale from 0 to 2 (0, none; 1, partial; 2, complete). In contrast, the WMFT (total 75 scores) is a test for assessing functional performance, providing insight into joint-specific and integrative limb movements graded from 0 to 5, with 15 function-based tasks [39]. According to the study performed by Wolf et al [40], the WMFT is more sensitive than FMA-UE for assessing functional improvement in less affected stroke patients. Thus, the different results of the two functional tests indicate that the home rehabilitation exercise for 12 weeks had beneficial effects in functional recovery, but it was not enough to change synergic movement or hand and wrist function.

In terms of shoulder ROM, we found a significant increase in shoulder flexion (*P*=.02) and internal rotation ROM (*P*=.001) in the HBR group by the Friedman test. According to the post-hoc analysis involving the Wilcoxon signed-rank test with Bonferroni correction, significant increases in the first 6 weeks of home exercise were noted for shoulder joint ROM with flexion (*P*=.004) and internal rotation (*P*=.001). However, there was no change in external rotation and extension ROM. Regarding the reasons for ROM increase, we think it is associated with the exercise protocol of our study. Among the four kinds of home exercises in our study protocol, bilateral flexion, wall push, and towel slide required wide movements of shoulder flexion and internal rotation. The exercise time records from our HBR system support this since patients performed the shoulder bilateral flexion exercise for the longest time when compared with other exercises. The shoulder extension ROM exercise was not included in our home exercise protocol. Although shoulder external rotation is required to perform the active scapular exercise, the external rotation ROM was not increased. This result is associated with the fact that chronic stroke patients usually have internally rotated joint contractures associated with impaired motor synergy [41,42]. We consider that the absence of a relevant change in the hand grip test was also related with the fact that our home exercise protocol required wide movement of the shoulder joint, but less motion of the wrist and hand joints.

The benefit of the HBR system was not only clinical improvement but also a decreased drop-out rate, which might encourage the application of wearable systems for HBR. We found that the drop-out rate was lower in the HBR group than in the control group at 12 weeks (5/22, 22% vs 4/10, 40%) and 18 weeks (10/22, 45% vs 10/10, 100%). After this study was completed, we interviewed two participants who dropped out to find out why they decided not to continue the study. They said that they became less interested in the conventional home rehabilitation program because the weekly phone calls did not help bring about any visible improvement and bothered them. We think the HBR system has a good influence on the motivation for home exercise and the relationship with physical therapists. According to the self-determination theory, which refers to each person's ability to make choices and manage their own life [24], people need to experience a sense of belonging and attachment with others, which is called "connection or relatedness." In addition, people need to feel in control of their own behaviors and goals, which is called "autonomy." Our HBR system would assist patients to record an exercise time (autonomy) and to communicate with a clinician (connection or relatedness).

Regarding the depression index, previous randomized controlled trials reported that home rehabilitation can reduce the incidence of depression [43,44]. However, BDI in our study did not show a relevant difference. It only showed a trend of positive effects (*P*=.06). The finding might be significant with a greater number of participants or a longer period because our protocol was relatively shorter than the period in the literature [43,44].

Lastly, there was no relevant difference in the HBR group between 12 weeks and 18 weeks (6 weeks after the final assessment without the HBR system). However, we believe that the HBR system is more effective when used consistently in home care because most of the clinical outcomes at 18 weeks showed a decreasing trend compared with 12 weeks.

Comparison With Prior Work

Previously, Wade emphasized that enabling self-directed practice is critical for stroke rehabilitation [45]. Regarding the strategy of self-directed practice, it has been shown that verbal encouragement does not have an impact on increasing rehabilitation activity after stroke [46]. Therefore, various methods of self-management training for upper limb rehabilitation have been suggested, including robot-assisted therapy. For example, Markopoulos et al [47] and Holden et al [48] developed a watch-like device. They used a visual feedback system as a self-management tool, but there was no remote supervision with the therapist. The mobile health system proposed by Dobkin [17] uses a similar strategy as that of our approach, which is called the rehabilitation internet-of-things (RIoT) device. However, we applied the ML model for home exercise detection and used a commercial smartwatch to simplify the user device interface, which has been regarded by previous researchers as the most important factor for use in clinical practice [17,49].

With regard to robot-assisted therapy, Lo et al [50] reported that robot-assisted therapy showed no relevant difference at 12 weeks and only showed improvement over 36 weeks when compared with typical care. Additionally, it costed US $15,562 for the 36-week program [50]. In contrast, our HBR system increased flexion ROM at 6 weeks and showed improvement of the WMFT score at 12 weeks. Considering the treatment cost of robot-assisted therapy, our HBR system strategy could be a better treatment modality with similar clinical improvement.

Limitations

There are several limitations in this study. First, the total number of patients who completed our program was relatively small to derive statistically strong evidence, particularly in the control group. Further work with a larger sample size would be helpful for more confirmative conclusions. Second, there was a discrepancy in the number of participants in the control and HBR groups. Only six participants from the control group were enrolled in the data analysis process. However, while carrying out our research, the loss of participants was inevitable in the control group because they were tired of receiving calls regarding management without any benefit, indicating the limitation of a conventional method. Third, there could have been loss of time measurement in the HBR group since some patients stated that they sometimes performed home exercise without the smartwatch owing to the inconvenience of wearing the smartwatch. Therefore, we think the exercise time recorded in the database was an underestimation of the real home exercise time. Fourth, the actual accuracy of exercise detection at home was not assessed. Although some researchers have attempted to address the privacy preservation of sensitive personal data based on a deep learning algorithm [51], we did not implement this approach and only calculated the accuracy based on a five-fold cross-validation test. Therefore, the actual accuracy, which is the correct prediction rate of exercise detection at home, could not be assessed because all patients wanted to protect their privacy. Fifth, there could have been selection bias associated with the positions of local health centers at different locations. Although we cannot quantify the difference, we think the bias was not relevant enough because both centers are closely located (50 km away) and the socioeconomic status is similar.

Conclusions

This study found that a home care system using a commercial smartwatch and ML model can facilitate participation in home training and improve the functional score of the WMFT and shoulder ROM of flexion and internal rotation in the treatment of patients with chronic stroke. We recommend our HBR system strategy as an innovative and cost-effective home care treatment modality.

Acknowledgments

This paper is based on research that has been conducted as part of the Korea Advanced Institute of Science and Technology-funded Global Singularity Research Program (2020).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Home-based rehabilitation system application video.
[MP4 File (MP4 Video), 21490 KB - mhealth_v8i7e17216_app1.mp4]

Multimedia Appendix 2
App view of home exercise results for the user and supervisor.
[PDF File (Adobe PDF File), 89 KB - mhealth_v8i7e17216_app2.pdf]

Multimedia Appendix 3
Accuracy calculation by the cross-validation test.
[PDF File (Adobe PDF File), 56 KB - mhealth_v8i7e17216_app3.pdf]

Multimedia Appendix 4
Clinical results with the home-based rehabilitation system.
[MP4 File (MP4 Video), 61489 KB - mhealth_v8i7e17216_app4.mp4]

Multimedia Appendix 5
Comparison between 12 weeks and 18 weeks in the home-based rehabilitation group.
[PDF File (Adobe PDF File), 12 KB - mhealth_v8i7e17216_app5.pdf]

Multimedia Appendix 6

CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 637 KB - mhealth_v8i7e17216_app6.pdf]

References

1. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJ. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. Stroke 2003 Sep;34(9):2181-2186. [doi: 10.1161/01.STR.0000087172.16305.CD] [Medline: 12907818]
2. GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):439-458 [FREE Full text] [doi: 10.1016/S1474-4422(19)30034-1] [Medline: 30871944]
3. Swayne O, Rothwell J, Ward N, Greenwood R. Stages of motor output reorganization after hemispheric stroke suggested by longitudinal studies of cortical physiology. Cereb Cortex 2008 Aug;18(8):1909-1922 [FREE Full text] [doi: 10.1093/cercor/bhm218] [Medline: 18234688]
4. Fisher RJ, Gaynor C, Kerr M, Langhorne P, Anderson C, Bautz-Holter E, et al. A Consensus on Stroke. Stroke 2011 May;42(5):1392-1397. [doi: 10.1161/strokeaha.110.606285]
5. Thorsén AM, Holmqvist LW, de Pedro-Cuesta J, von Koch L. A randomized controlled trial of early supported discharge and continued rehabilitation at home after stroke: five-year follow-up of patient outcome. Stroke 2005 Feb;36(2):297-303. [doi: 10.1161/01.STR.0000152288.42701.a6] [Medline: 15618441]
6. Legg L, Langhorne P, Outpatient Service Trialists. Rehabilitation therapy services for stroke patients living at home: systematic review of randomised trials. Lancet 2004 Jan 31;363(9406):352-356. [doi: 10.1016/S0140-6736(04)15434-2] [Medline: 15070563]
7. Johansson T, Wild C. Telerehabilitation in stroke care--a systematic review. J Telemed Telecare 2011;17(1):1-6. [doi: 10.1258/jtt.2010.100105] [Medline: 21097560]
8. Hsieh Y, Chang K, Hung J, Wu C, Fu M, Chen C. Effects of Home-Based Versus Clinic-Based Rehabilitation Combining Mirror Therapy and Task-Specific Training for Patients With Stroke: A Randomized Crossover Trial. Arch Phys Med Rehabil 2018 Dec;99(12):2399-2407. [doi: 10.1016/j.apmr.2018.03.017] [Medline: 29702070]
9. Hiengkaew V, Vongsirinavarat M. Home-Based Physical Therapy for Individuals With Stroke in Thailand. Home Health Care Management & Practice 2016 Jul 09;28(4):209-215. [doi: 10.1177/1084822316630278]
10. Wilkins KB, Owen M, Ingo C, Carmona C, Dewald JPA, Yao J. Neural Plasticity in Moderate to Severe Chronic Stroke Following a Device-Assisted Task-Specific Arm/Hand Intervention. Front Neurol 2017;8:284 [FREE Full text] [doi: 10.3389/fneur.2017.00284] [Medline: 28659863]
11. Rajsic S, Gothe H, Borba HH, Sroczynski G, Vujicic J, Toell T, et al. Economic burden of stroke: a systematic review on post-stroke care. Eur J Health Econ 2019 Feb;20(1):107-134. [doi: 10.1007/s10198-018-0984-0] [Medline: 29909569]
12. Visser-Meily A, Post M, van de Port I, van Heugten C, van den Bos T. Psychosocial functioning of spouses in the chronic phase after stroke: improvement or deterioration between 1 and 3 years after stroke? Patient Educ Couns 2008 Oct;73(1):153-158. [doi: 10.1016/j.pec.2008.03.011] [Medline: 18450411]
13. Marshall IJ, Wang Y, Crichton S, McKevitt C, Rudd AG, Wolfe CD. The effects of socioeconomic status on stroke risk and outcomes. Lancet Neurol 2015 Dec;14(12):1206-1218. [doi: 10.1016/S1474-4422(15)00200-8] [Medline: 26581971]
14. Mohan KM, Wolfe CD, Rudd AG, Heuschmann PU, Kolominsky-Rabas PL, Grieve AP. Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. Stroke 2011 May;42(5):1489-1494. [doi: 10.1161/STROKEAHA.110.602615] [Medline: 21454819]
15. Cameron JI, Cheung AM, Streiner DL, Coyte PC, Stewart DE. Stroke survivor depressive symptoms are associated with family caregiver depression during the first 2 years poststroke. Stroke 2011 Feb;42(2):302-306. [doi: 10.1161/STROKEAHA.110.597963] [Medline: 21164124]
16. Denno MS, Gillard PJ, Graham GD, DiBonaventura MD, Goren A, Varon SF, et al. Anxiety and depression associated with caregiver burden in caregivers of stroke survivors with spasticity. Arch Phys Med Rehabil 2013 Sep;94(9):1731-1736. [doi: 10.1016/j.apmr.2013.03.014] [Medline: 23548544]
17. Dobkin BH. A Rehabilitation-Internet-of-Things in the Home to Augment Motor Skills and Exercise Training. Neurorehabil Neural Repair 2017 Mar;31(3):217-227 [FREE Full text] [doi: 10.1177/1545968316680490] [Medline: 27885161]
18. 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]
19. Maglaveras N, Chouvarda I, Koutkias V, Lekka I, Tsakali M, Tsetoglou S, et al. Citizen centered health and lifestyle management via interactive TV: The PANACEIA-ITV health system. AMIA Annu Symp Proc 2003:415-419 [FREE Full text] [Medline: 14728206]
20. Miyoshi H, Kimura Y, Tamura T, Sekine M, Okabe I, Hara K. Smart Living - Home Rehabilitation Training System Using an Interactive Television. 2015 Presented at: 6th European Conference of the International Federation for Medical and Biological Engineering; September 7-11, 2014; Dubrovnik, Croatia p. 661-665 URL: https://link.springer.com/chapter/10.1007%2F978-3-319-11128-5_165#citeas

21. Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. J Neuroeng Rehabil 2014 Jul 03;11:108 [FREE Full text] [doi: 10.1186/1743-0003-11-108] [Medline: 24996956]
22. Krishnan NC, Cook DJ. Activity Recognition on Streaming Sensor Data. Pervasive Mob Comput 2014 Feb 01;10(Pt B):138-154 [FREE Full text] [doi: 10.1016/j.pmcj.2012.07.003] [Medline: 24729780]
23. Reeder B, David A. Health at hand: A systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Dec;63:269-276 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
24. Patrick H, Williams GC. Self-determination theory: its application to health behavior and complementarity with motivational interviewing. Int J Behav Nutr Phys Act 2012 Mar 02;9:18 [FREE Full text] [doi: 10.1186/1479-5868-9-18] [Medline: 22385676]
25. Cauraugh JH, Summers JJ. Neural plasticity and bilateral movements: A rehabilitation approach for chronic stroke. Prog Neurobiol 2005 Apr;75(5):309-320. [doi: 10.1016/j.pneurobio.2005.04.001] [Medline: 15885874]
26. Wenderoth N. Neural Networks Involved in Cyclical Interlimb Coordination as Revealed by Medical Imaging Techniques. In: Neuro-Behavioral Determinants of Interlimb Coordination. Boston, MA: Springer; 2004.
27. Singh D. Human Activity Recognition Using Recurrent Neural Networks. In: Machine Learning and Knowledge Extraction. 2017 Presented at: International Cross-Domain Conference for Machine Learning and Knowledge Extraction; 29 August-1 September; Reggio, Italy p. 267-274 URL: https://link.springer.com/chapter/10.1007/978-3-319-66808-6_18#citeas
28. Ordóñez FJ, Roggen D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors (Basel) 2016 Jan 18;16(1) [FREE Full text] [doi: 10.3390/s16010115] [Medline: 26797612]
29. Chen Y, Xue Y. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. 2015 Presented at: 2015 IEEE International Conference on Systems, Man, and Cybernetics; October 9-12, 2015; Kowloon, China p. 1488-1492.
30. Weiss G, Timko J, Gallagher C, Yoneda K, Schreiber A. Smartwatch-based activity recognition: A machine learning approach. 2016 Presented at: 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); February 24-27, 2016; Las Vegas, NV, USA.
31. Lee KS, Chae S, Park HS. Optimal Time-Window Derivation for Human-Activity Recognition Based on Convolutional Neural Networks of Repeated Rehabilitation Motions. 2019 Presented at: 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR); June 24-28, 2019; Toronto, ON, Canada.
32. Sin H, Lee G. Additional virtual reality training using Xbox Kinect in stroke survivors with hemiplegia. Am J Phys Med Rehabil 2013 Oct;92(10):871-880. [doi: 10.1097/PHM.0b013e3182a38e40] [Medline: 24051993]
33. Mortazavi B, Pourhomayoun M, Alsheikh G, Alshurafa N, Lee S, Sarrafzadeh M. Determining the single best axis for exercise repetition recognition and counting on smartwatches. 2014 Presented at: 2014 11th International Conference on Wearable and Implantable Body Sensor Networks; June 16-19, 2014; Zurich, Switzerland p. 33-38.
34. Margarito J, Helaoui R, Bianchi A, Sartor F, Bonomi A. User-Independent Recognition of Sports Activities From a Single Wrist-Worn Accelerometer: A Template-Matching-Based Approach. IEEE Trans Biomed Eng 2016 Apr;63(4):788-796. [doi: 10.1109/TBME.2015.2471094] [Medline: 26302509]
35. Ramezani R, Zhang W, Xie Z, Shen J, Elashoff D, Roberts P, et al. A Combination of Indoor Localization and Wearable Sensor-Based Physical Activity Recognition to Assess Older Patients Undergoing Subacute Rehabilitation: Baseline Study Results. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e14090 [FREE Full text] [doi: 10.2196/14090] [Medline: 31293244]
36. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil 2017 Mar 11;14(1):20 [FREE Full text] [doi: 10.1186/s12984-017-0229-y] [Medline: 28284228]
37. Huynh QT, Nguyen UD, Irazabal LB, Ghassemian N, Tran BQ. Optimization of an Accelerometer and Gyroscope-Based Fall Detection Algorithm. Journal of Sensors 2015;2015:1-8. [doi: 10.1155/2015/452078]
38. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. J Med Internet Res 2016 Feb 26;18(2):e41 [FREE Full text] [doi: 10.2196/jmir.4738] [Medline: 26920122]
39. Chen P, Kwong PW, Lai CK, Ng SS. Comparison of bilateral and unilateral upper limb training in people with stroke: A systematic review and meta-analysis. PLoS One 2019;14(5):e0216357 [FREE Full text] [doi: 10.1371/journal.pone.0216357] [Medline: 31120910]
40. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. Stroke 2001 Jul;32(7):1635-1639. [doi: 10.1161/01.str.32.7.1635] [Medline: 11441212]
41. Sackley C, Brittle N, Patel S, Ellins J, Scott M, Wright C, et al. The prevalence of joint contractures, pressure sores, painful shoulder, other pain, falls, and depression in the year after a severely disabling stroke. Stroke 2008 Dec;39(12):3329-3334. [doi: 10.1161/STROKEAHA.108.518563] [Medline: 18787199]
42. Ada L, Goddard E, McCully J, Stavrinos T, Bampton J. Thirty minutes of positioning reduces the development of shoulder external rotation contracture after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2005 Feb;86(2):230-234. [doi: 10.1016/j.apmr.2004.02.031] [Medline: 15706548]

43. Chaiyawat P, Kulkantrakorn K. Randomized controlled trial of home rehabilitation for patients with ischemic stroke: impact upon disability and elderly depression. Psychogeriatrics 2012 Sep;12(3):193-199 [FREE Full text] [doi: 10.1111/j.1479-8301.2012.00412.x] [Medline: 22994618]
44. Lai S, Studenski S, Richards L, Perera S, Reker D, Rigler S, et al. Therapeutic exercise and depressive symptoms after stroke. J Am Geriatr Soc 2006 Feb;54(2):240-247. [doi: 10.1111/j.1532-5415.2006.00573.x] [Medline: 16460374]
45. Wade D. Rehabilitation - a new approach. Part four: a new paradigm, and its implications. Clin Rehabil 2016 Feb;30(2):109-118. [doi: 10.1177/0269215515601177] [Medline: 26715679]
46. Boysen G, Krarup L, Zeng X, Oskedra A, Kõrv J, Andersen G, ExStroke Pilot Trial Group. ExStroke Pilot Trial of the effect of repeated instructions to improve physical activity after ischaemic stroke: a multinational randomised controlled clinical trial. BMJ 2009 Jul 22;339:b2810 [FREE Full text] [doi: 10.1136/bmj.b2810] [Medline: 19900934]
47. Markopoulos P, Timmermans A, Beursgens L, van Donselaar R, Seelen H. Us'em: the user-centered design of a device for motivating stroke patients to use their impaired arm-hand in daily life activities. Conf Proc IEEE Eng Med Biol Soc 2011;2011:5182-5187. [doi: 10.1109/IEMBS.2011.6091283] [Medline: 22255506]
48. Holden A, McNaney R, Balaam M, Thompson R, Hammerla N, Ploetz T. CueS: cueing for upper limb rehabilitation in stroke. 2015 Presented at: British HCI 2015: 2015 British Human Computer Interaction Conference; July 2015; Lincoln Lincolnshire, United Kingdom.
49. Singh D. Users' Perceptions and Attitudes Towards Smart Home Technologies. 2018 Presented at: 16th International Conference, ICOST 2018; July 10-12, 2018; Singapore.
50. Lo AC, Guarino PD, Richards LG, Haselkorn JK, Wittenberg GF, Federman DG, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. N Engl J Med 2010 May 13;362(19):1772-1783 [FREE Full text] [doi: 10.1056/NEJMoa0911341] [Medline: 20400552]
51. Psychoula I, Merdivan E, Singh D, Chen L, Chen F, Hanke S. A Deep Learning Approach for Privacy Preservation in Assisted Living. 2018 Presented at: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 19-23, 2018; Athens, Greece p. 710-715.

Abbreviations

BDI: Beck Depression Inventory
CNN: convolutional neural network
FMA-UE: Fugl-Meyer Assessment of Upper Extremity
HBR: home-based rehabilitation
IMU: inertial measurement unit
ML: machine learning
ROM: range of motion
WMFT: Wolf Motor Function Test

Edited by G Eysenbach; submitted 04.12.19; peer-reviewed by A Holzinger, KL Ong, C Lai; comments to author 28.01.20; revised version received 22.03.20; accepted 14.05.20; published 09.07.20.

Please cite as:
Chae SH, Kim Y, Lee KS, Park HS
Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study
JMIR Mhealth Uhealth 2020;8(7):e17216
URL: http://mhealth.jmir.org/2020/7/e17216/
doi:10.2196/17216
PMID:32480361

Original Paper

In-Home Rehabilitation Using a Smartphone App Coupled With 3D Printed Functional Objects: Single-Subject Design Study

Jeanne Langan[1*], PhD, PT; Sutanuka Bhattacharjya[2*], PhD, OTR; Heamchand Subryan[3], MArch, MFA; Wenyao Xu[4], PhD; Baicheng Chen[4]; Zhengxiong Li[4], MSc; Lora Cavuoto[5], PhD

[1]Department of Rehabilitation Science, University at Buffalo, Buffalo, NY, United States

[2]Department of Occupational Therapy, Georgia State University, Atlanta, GA, United States

[3]Center for Inclusive Design and Environmental Access, University at Buffalo, Buffalo, NY, United States

[4]Computer Science and Engineering Department, University at Buffalo, Buffalo, NY, United States

[5]Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY, United States

[*]these authors contributed equally

Corresponding Author:
Jeanne Langan, PhD, PT
Department of Rehabilitation Science
University at Buffalo
522 Kimball Tower
3435 Main St
Buffalo, NY, 14214
United States
Phone: 1 7168292905
Email: jlangan@buffalo.edu

Abstract

Background: Stroke is a major cause of long-term disability. While there is potential for improvements long after stroke onset, there is little to support functional recovery across the lifespan. mHealth solutions can help fill this gap. mRehab was designed to guide individuals with stroke through a home program and provide performance feedback.

Objective: To examine if individuals with chronic stroke can use mRehab at home to improve upper limb mobility. The secondary objective was to examine if changes in limb mobility transferred to standardized clinical assessments.

Methods: mRehab consists of a smartphone coupled with 3D printed household items: mug, bowl, key, and doorknob. The smartphone custom app guides task-oriented activities and measures both time to complete an activity and quality of movement (smoothness/accuracy). It also provides performance-based feedback to aid the user in self-monitoring their performance. Task-oriented activities were categorized as (1) object transportation, (2) prehensile grip with supination/pronation, (3) fractionated finger movement, and (4) walking with object. A total of 18 individuals with stroke enrolled in the single-subject experimental design study consisting of pretesting, a 6-week mRehab home program, and posttesting. Pre- and posttesting included both in-laboratory clinical assessments and in-home mRehab recorded samples of task performance. During the home program, mRehab recorded performance data. A System Usability Scale assessed user's perception of mRehab.

Results: A total of 16 participants completed the study and their data are presented in the results. The average days of exercise for each mRehab activity ranged from 15.93 to 21.19 days. This level of adherence was sufficient for improvements in time (t_{15}=2.555, *P*=.02) and smoothness (t_{15}=3.483, *P*=.003) in object transportation. Clinical assessments indicated improvements in functional performance (t_{15}=2.675, *P*=.02) and hand dexterity (t_{15}=2.629, *P*=.02). Participant's perception of mRehab was positive.

Conclusions: Despite heterogeneity in participants' use of mRehab, there were improvements in upper limb mobility. Smartphone-based portable technology can support home rehabilitation programs in chronic conditions such as stroke. The ability to record performance data from home rehabilitation offers new insights into the impact of home programs on outcomes.

Trial Registration: ClinicalTrials.gov NCT04363944; https://clinicaltrials.gov/ct2/show/NCT04363944

(JMIR Mhealth Uhealth 2020;8(7):e19582) doi:10.2196/19582

KEYWORDS

stroke; rehabilitation; smart technology; 3D printing; usability

Introduction

Background

Stroke is a major cause of disability, leading to restriction of occupational performance for stroke survivors [1,2]. It is estimated that 30%-60% of stroke survivors continue to have residual limitations in upper extremity movements after traditional rehabilitation services [3]. At the end of rehabilitation services, survivors are commonly given a written home exercise program to guide recovery in chronic stages of stroke [4]. Shortcomings of the written home exercise program include complaints of being unengaging and patients not continuing the program [4]. Knowing that upper limb motor deficits can reduce quality of life [5], it is important to support survivors to recover as much function as possible. Upper limb recovery after stroke is identified as a research priority by survivors of stroke, caregivers, and health professionals [6].

Research demonstrates that individuals with chronic stroke are capable of making gains in performance with continued practice. The research so far has focused on interventions led by therapists [7,8]. It is improbable that direct oversight by a therapist is a feasible solution for long-term recovery. For chronic conditions such as stroke, better supporting the individual's ability to self-manage their long-term recovery could offer a more sustainable approach. Use of mHealth (ie, mobile technology to manage health) offers the opportunity for individuals to engage in rehabilitative activities while monitoring their performance and managing their health behaviors [9,10]. mHealth apps can assist users in meeting basic needs, thereby giving a sense of autonomy and competence [11]. In addition, participants have reported that it is enjoyable to use apps [12]. Smart devices are equipped with interactive components (eg, sensors, cameras, speakers, and vibrators) capable of measuring human movement and providing feedback [13]. Readily available smartphone technology can be the basis of a home rehabilitation system.

There has been an increase in app development for stroke rehabilitation. A review of apps designed for stroke survivors or their caregivers found that 62% of apps addressed language or communication [14]. Other apps addressed stroke risk calculation, identifying acute stroke, atrial fibrillation, direction to emergency room or nearest certified stroke center, visual attention therapy, and a mere 4% addressed physical rehabilitation [14]. Importantly, apps for rehabilitation did not focus on upper limb function [14]. Use of technology to guide and measure performance in task-specific training of the upper extremity after stroke has primarily included clinical or laboratory-based interventions [15,16]. Task-specific programs are function based, with practice of tasks relevant to activities of daily life, and have been shown to be efficacious [17,18]. Use of instrumented objects in a laboratory setting has resulted in patients reporting they enjoyed the experience [15]. There has been less research on the use of portable technology for upper limb rehabilitation in a home setting for individuals with chronic arm/hand deficits after stroke.

Previous Work

mRehab (mobile Rehab) was created to better support in-home upper limb rehabilitation programs (Figure 1) [13]. It incorporates a task-oriented approach and immediate performance-based feedback. Exercise programs that include feedback have resulted in better outcomes compared with programs without feedback [19,20]. mRehab consists of 3D printed household objects (a mug, bowl, key, and doorknob) integrated with a smartphone and an app. The app guides participants through practice of activities of daily living, for example, sipping from a mug. It can also consistently measure time to complete an activity and quality of movement (smoothness/accuracy) during the performance of activities of daily living. The system is described in more detail in previous articles that have evaluated it in primarily laboratory-based settings [13,21].

Figure 1. In-home use of mRehab: (A) selecting an activity in mRehab; (B) turning key activity; and (C) vertical mug transfer activity.

There is little information on in-home use of technology for rehabilitation in chronic stroke. While technology-based systems designed for rehabilitation have been developed, they have typically been examined in laboratory or clinical settings [22,23]. The results of this study will provide much needed evidence of the ability of individuals with chronic stroke to use technology in a home-based program with oversight only upon request. This mimics clinical practice, in which patients are discharged from rehabilitation with a home program and then need to self-manage their recovery. We examine the individual's adherence to exercise and if they required support with the technology. The impact of the home-based mRehab program on functional mobility was also examined. While individuals with chronic stroke were selected for the first examination of mRehab in a home-based setting, the system has the potential to be used by individuals that have arm/hand deficits due to other underlying pathology.

Methods

Participants

The study was approved by the University at Buffalo Institutional Review Board and all participants provided written informed consent. A total of 18 participants were recruited from the local community. Participants were included if they were (1) at least 18 years of age and living in the community, (2) were 6 or more months after stroke, and (3) had a minimum score of 124 on the Mattis Dementia Rating Scale (MDRS) [24]. Participants were excluded if they met any of the following criteria: (1) acute or chronic pain that would interfere with participation, (2) severely limited range of motion of the upper limb, (3) absent or severely impaired proprioception of the upper limb, (4) musculoskeletal or circulatory conditions affecting the upper limb, (5) spasticity graded as 3 or greater for upper extremity movement on the Modified Ashworth Scale (MAS), or (6) botulinum toxin injections for spasticity management within 3 months of starting the study. These inclusion and exclusion criteria were established to select participants that were likely to have the cognitive and physical capacity to use mRehab.

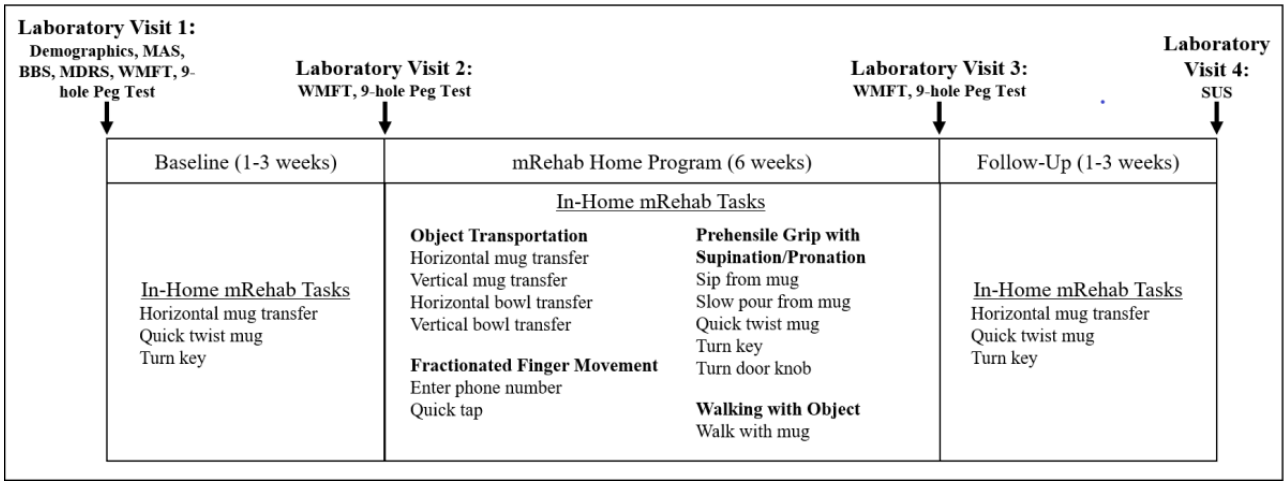
Design

A single-subject experimental design with multiple baselines was used. A strength of the single-subject study design is that participants serve as their own control. There is variability in the degree of arm/hand deficits for survivors of stroke, making it challenging to establish an equivalent control group. The single-subject design offers an alternative approach that is commonly used in assessing populations with stroke [25,26]. Each participant had a varying length of the baseline and follow-up periods to establish that the intervention, rather than time, was the primary reason for any observed change in performance.

Procedure

Over a 10-week period, participants completed baseline measurements, a 6-week mRehab home program, and follow-up measurements. Baseline measurements consisted of both in-laboratory and in-home measurements (Figure 2). Participants attended 2 laboratory visits prior to starting the home program. During the first laboratory visit, participants completed a demographic questionnaire and clinical assessments. The Berg Balance Scale (BBS) [27] was used to determine if participants had sufficient balance (score greater than 42) to participate in the walk with mug activity. The Wolf Motor Function Test (WMFT) [28] and Nine-Hole Peg Test [29] were the clinical outcome measures. An occupational therapist demonstrated the mRehab system to the participant. The participant learned to operate the smartphone, mRehab app, and mRehab restricted mode. The mRehab restricted mode was designed to sample baseline in-home performance of 3 representative mRehab activities: horizontal mug transfer, quick twist of the mug, and turn key. Only 3 repetitions of each baseline activity could be performed in a session, for a maximum of 3 sessions per baseline week. Repetitions were limited to avoid improving performance during the baseline period. The app did not give feedback during the restricted mode. At the second laboratory visit, participants completed the same clinical assessments (Figure 2) and learned the remaining 9 mRehab activities for the home program. Participants were instructed to contact the research team if they had questions or concerns. Participant's contacts to the research team were recorded.

Figure 2. mRehab study timeline. BBS: Berg Balance Scale; MAS: Modified Ashworth Scale; MDRS: Mattis Dementia Rating Scale; SUS: Systems Usability Scale; WMFT: Wolf Motor Function Test.



During the home program, participants could select from all 12 mRehab activities (Figure 2). It was suggested that participants complete 10 repetitions of each activity daily, 5 days per week for 6 weeks. The mRehab app recorded and provided feedback (both visual and auditory) on the user's performance (repetitions, time to complete, and smoothness/accuracy) at the end of each activity.

The follow-up phase was similar to the baseline phase (Figure 2). The third laboratory visit examined changes in performance of the clinical assessments immediately after the completion of the mRehab home program. Between the third and fourth laboratory visit, participants again used mRehab in the restricted mode without feedback to sample in-home performance. At the fourth laboratory visit, participants returned the 3D printed objects and completed the Systems Usability Scale (SUS) [30]. SUS allows for subjective assessment of perceived usability. Participants responded to 10 questions using a Likert 1-5 scale (1=strongly disagree and 5=strongly agree). Percentile rank out of a possible 100 was calculated.

Statistical Analyses

mRehab Data

Performance was examined on each mRehab task and on composites of similar activities: (1) object transportation (horizontal and vertical mug and bowl transfers); (2) prehensile grip with supination/pronation (sip from mug, slowly pour water from mug, quick twist of the mug, turn key, and turn door knob); (3) fractionated finger movement (entering phone number and quick tap); and (4) walking with object activity (here, walk with mug was left out of the analyses because not all participants could perform this activity). The composite score represented the average time to complete the activities in a category.

mRehab data were examined at both individual and group levels. As is typical with single-subject designs, visual inspection of individual's data was used as the first stage of analysis of adherence and performance change during the intervention. Quantitative changes in motor performance based on mRehab data were examined using paired *t* tests for the following comparisons: (1) baseline compared with follow-up using the average time to complete and smoothness of each activity, (2) first compared with last training day in the 6-week home program using the average time to complete and smoothness of each activity, and (3) first compared with last training day using composite scores. To examine an a priori hypothesis that the amount of training impacts outcome, correlations between the total number of repetitions performed for each activity during the 6-week home program and changes in performance for each activity were examined.

Clinical Assessments

Changes in clinical assessments were examined using paired *t* tests. The average time for tasks in the WMFT was used in the analyses [31]. To meet the normality assumption, the log transform of the average WMFT score was used for the statistical analysis. For both the WMFT and the Nine-Hole Peg Test, the scores from the first and second in-laboratory visits were averaged (scores were not different [$P>.05$]) to account for variability in the performance of individuals with stroke. This averaged preintervention score was compared with the third in-laboratory visit to assess the immediate change in performance following use of mRehab. Because some of the mRehab activities resemble tasks in the WMFT, we examined changes for each task in the WMFT to assess if we were in essence training for the clinical test. Significance for all tests was set at $P<.05$.

Usability

Perceived usability of mRehab was examined. The range and the average of the SUS percentile rank scores were reported. The full assessment of usability including qualitative assessments is beyond the scope of this paper and will be reported in another paper.

Results

In-Home Use of mRehab

A total of 18 participants with stroke were recruited from the community (Table 1); 2 participants did not complete the study: 1 participant reported being unable to use mRehab without caregiver assistance and did not wish to continue the home program, whereas the other completed only the first in-laboratory visit and decided he did not have sufficient time in his schedule to complete the full study. The performance data of the remaining 16 participants are reported in the results.

During in-home use of mRehab, 10 participants contacted the research team reporting difficulties with the system. Home visits were made to 7 participants. The most common reason for a home visit was to replace the 3D printed door knob or key or both. The construction of these objects was modified during the study to improve durability. Changing the direction of fill in the 3D printing process improved the product. A full report on usability will be discussed in detail in another paper.

Table 1. Participant demographics.

Participant code	Age	Gender	Years after stroke	Paretic side	Reported dominant arm	Reported dominant arm prior to stroke	MDRS[a]	WMFT[b] (s)[c]
s01[d]	57	F	2	R	L	R	132	43.10
s02	54	F	7	L	R	L	144	4.84
s03	68	M	4	R	L	R	142	13.78
s04	61	F	12	R	L	R	140	10.12
s05	78	F	1	L	R	R	140	4.75
s06	66	M	14	L	R	L	140	44.77
s07	73	M	1	L	R	L	139	39.37
s08	61	M	0.5	L	R	R	142	1.85
s09	62	F	2	R	L	R	124	23.96
s10	67	M	1	R	L	R	130	8.65
s11	76	M	6	R	L	R	133	2.25
s12	43	M	5	R	R	R	143	2.95
s13	76	M	4	R	R	L	144	2.06
s14	39	F	4	R	L	R	143	2.43
s15	78	M	3	L	R	R	134	34.00
s16[d]	56	M	6	L	R	L	143	80.55[e]
s17	72	M	11	R	L	R	142	81.12
s18	37	M	1	CL	R	R	141	4.95

[a]MDRS: Mattis Dementia Rating Scale.

[b]WMFT: Wolf Motor Function Test.

[c]Average visits 1 and 2.

[d]Indicates participant did not complete the study.

[e]Participant only completed visit 1.

mRehab In-Home Recorded Data

Visual analyses of individual data show differences between participants in adherence and performance of the mRehab activities. As an example, Figure 3 shows individual data sets for the time to complete the activity horizontal mug transfer during the baseline, 6-week program, and follow-up. Participants demonstrated variance in the number of days exercised, the rate of performance change, and the stability of performance. The majority of participants reduced their time to complete the horizontal mug transfer by the end of the 6-week program.

Quantitative changes in mRehab performance were examined. Baseline and follow-up data were compared for horizontal mug transfer and key turn. The quick twist of the mug data were excluded from analysis because only few participants could twist quickly enough for the sensor to capture the movement. Participants demonstrated a decrease in time from baseline to follow-up in the horizontal mug transfer (t_{15}=2.14, *P*=.05; Figure 4), and the decrease for key turn (t_{15}=1.86, *P*=.08) approached the commonly accepted α .05 level (Figure 5). Comparing performance on the first day of training with the last training day in the 6-week program (Table 2), there was a trend across activities for improved scores in both time and smoothness at the last session. All object transportation activities and quick tap, a fractionated finger movement activity, reduced in time to complete. On the last day of the program, vertical and horizontal mug transfer were performed more smoothly and quick tap more accurately. No correlation examining the number of repetitions completed during the home program and changes in performance of an activity (time to complete or smoothness) resulted in *P* values <.05. Comparing composite scores from the first and last day of training, object transportation improved in time (t_{15}=2.555, *P*=.02; Figure 5). Neither the composite score for prehensile grip with supination/pronation nor fractionated finger movement demonstrated significant improvement in time (Figure 5).

Figure 3. Data sets from 16 participants showing their performance on horizontal mug transfer during baseline, 6-week home program, and follow-up. The number of days they did this activity is on the x axis. Asterisk (*) indicates the average time to complete horizontal mug transfer on that day was above 10 seconds.

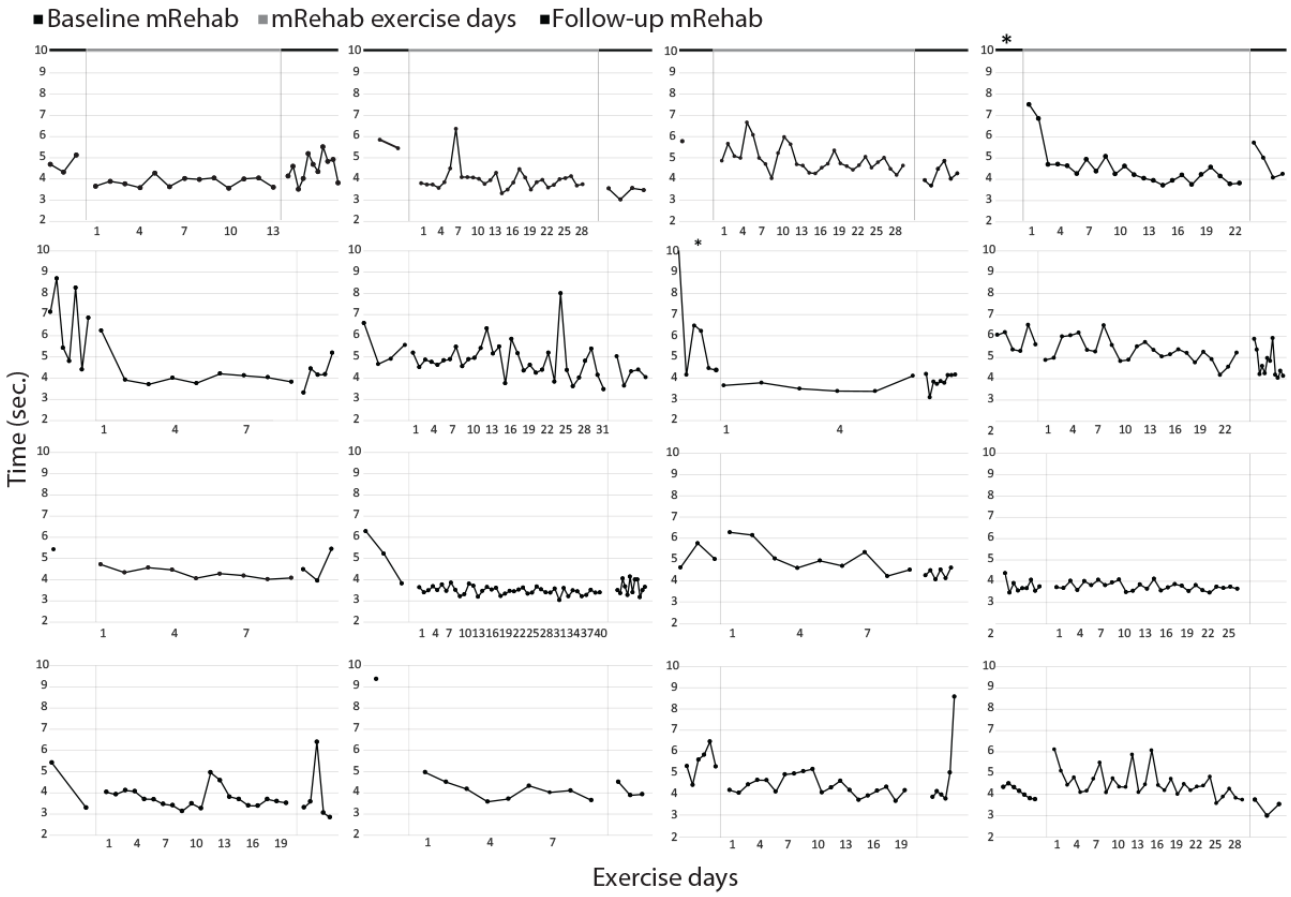


Figure 4. Pre- and postintervention average time to complete horizontal mug transfer and turning a key. * indicates a *P* value <.05.

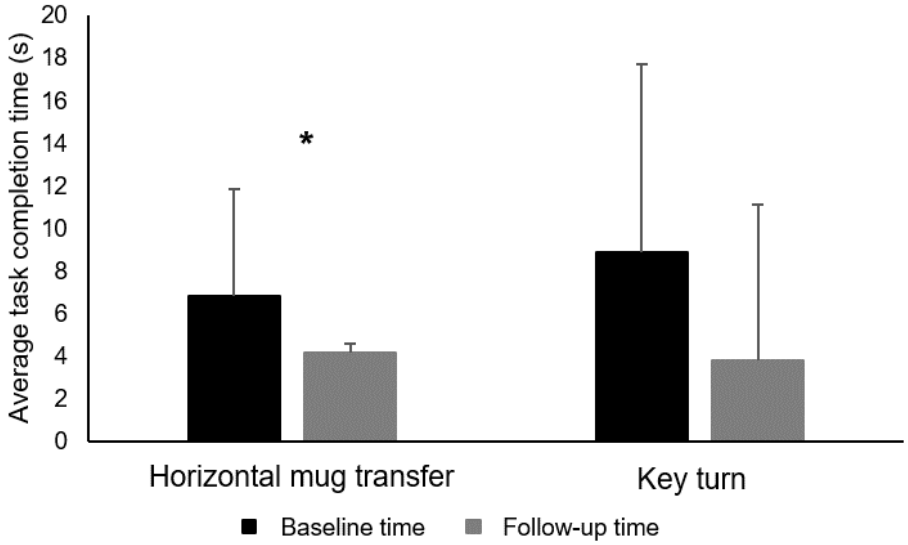


Figure 5. Start and end of intervention average time to complete task for composite task groupings. Error bars represent standard deviations. * indicates a *P* value <.05.

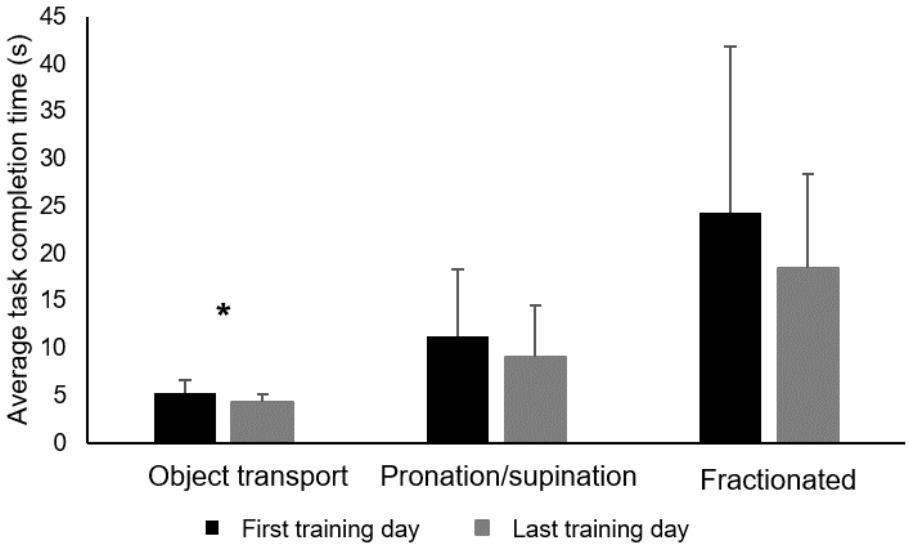


Table 2. Group data performance changes on each activity in mRehab.

Activity	N	Number of exercise days, mean (SD)	Total number of repetitions, mean (SD)	First session time, mean (SD)	Last session time, mean (SD)	*P* value (for last session time)	First session smoothness, mean (SD)	Last session smoothness, mean (SD)	*P* value (for last session smoothness)
Object transportation									
Vertical bowl transfer	16	20.75 (10.60)	282.06 (210.00)	5.04 (0.91)	4.03 (0.46)	<.001	511.99 (297.33)	410.66 (128.81)	—[a]
Horizontal bowl transfer	16	20.88 (10.64)	289.44 (223.34)	5.03 (1.32)	4.01 (0.60)	.001	478.53 (315.49)	375.39 (137.23)	—
Vertical mug transfer	16	21.06 (10.64)	283.69 (227.94)	5.22 (1.22)	4.19 (0.55)	.003	699.24 (321.74)	492.70 (226.60)	.02
Horizontal mug transfer	16	21.19 (10.62)	292.31 (224.02)	4.96 (1.22)	4.00 (0.58)	.009	585.83 (233.81)	431.23 (159.97)	.04
Prehensile grip with supination/pronation									
Sip from mug	16	20.00 (10.81)	202.19 (151.48)	9.51 (4.71)	9.62 (1.39)	—	394.17 (239.46)	331.11 (61.77)	—
Slow pour with mug	15	20.06 (10.24)	148.94 (137.23)	25.07 (8.58)	20.69 (4.08)	.06	807.39 (323.21)	655.29 (249.62)	.048
Quick twist mug	14	18.36 (12.53)	228.29 (195.81)	4.30 (5.24)	3.34 (5.28)	—	747.08 (1394.73)	532.92 (1125.29)	.07
Turning a key	14	15.93 (11.01)	177.20 (141.26)	6.03 (5.48)	4.48 (7.32)	—	41.17 (33.28)	42.53 (65.15)	—
Turning a doorknob	15	18.81 (11.00)	199.13 (139.04)	6.65 (10.75)	2.21 (1.53)	—	33.58 (32.05)	31.76[b] (38.02)	—
Fractionated finger movement									
Enter phone number	16	19.13 (10.00)	153.13 (131.05)	30.86 (26.28)	22.35 (14.05)	—	NA[c]	NA	—
Quick tap	16	19.38 (10.33)	177.20 (141.26)	17.05 (9.42)	14.75 (7.54)	.005	NA	NA	—
Walking with object									
Walk with mug	11	16.87 (12.48)	183.07 (178.70)	NA	NA	—	967.42 (99.42)	997.98 (77.81)	—

[a]Not applicable.

[b]Removed data 2SD outside of the mean from 1 participant that experienced a broken door knob.

[c]Not assessed due to lower N.

Clinical Assessments

Analyses of the WMFT using the average time to complete a task and the natural log transform both resulted in rejecting the null hypothesis. Only the average time to complete a task was reported. Participants improved performance from baseline to follow-up testing on both the Nine-Hole Peg Test (t_{15}=2.629, *P*=.02) and the WMFT (t_{15}=2.675, *P*=.02). We explored how each task of the WMFT changed from baseline to follow-up. Tasks in which performance improved (*P*<.05) included moving the hand from table to box (front), reaching and retrieving a 1-lb weight, and folding a towel. Tasks in which the decrease in time neared the α .05 level include moving the weighted arm from the table to the box (*P*=.07) and turning a key in a lock (*P*=.08).

Usability

Examining usability of mRehab, the percentile rank on the SUS ranged from 60 to 97. The average score was 81.7.

Discussion

Principal Findings

This study is novel in using scalable components, smartphones and 3D printed items, to create a portable rehabilitation system. Furthermore, extended in-home use of a system by end users without regular oversight is uncommon in research. Approximately 89% of participants (16/18) completed the 6-week mRehab home program. This demonstrates that participants can use mRehab in-home, with technical support provided as needed, to enhance upper limb function. This is encouraging as both individuals with stroke and their caregivers report feeling that more rehabilitation would be beneficial [32,33]. The combination of the in-home mRehab data set and laboratory-based clinical assessments provides insight into adherence, task-specific training, and generalized performance gains with mRehab home-based rehabilitation.

For an exercise program to be effective there needs to be some degree of practice. The dosage of practice necessary to make gains is not well understood. A Cochrane review suggests that 30-60 minutes of rehabilitation per day, 5-7 days per week is effective [34]. Another review presented evidence that high-intensity and high repetition task-oriented and task-specific training is effective [35]. In this study, the average number of exercise days and repetitions was roughly half of the recommendation. The self-selected dosage was sufficient for improved performance in both mRehab and clinical measurements. We anticipated that individuals that practiced more would have larger improvements in the practiced mRehab activities. However, the data did not confirm this. It is possible that a larger cohort would have demonstrated a positive correlation. It is also possible that multiple mechanisms contributed to improvements with limited practice. Neurophysiology studies show that neuroplastic changes occur with learning a new skill and not after rote repetitive movement [36-38]. If participants can identify when repetitions become *rote repetitive movement*, they may reduce the dosage and more efficiently complete their home program. Task-specific training has shown to be more efficient compared with other nonspecific training approaches [30,39]. Besides, the addition of feedback has been shown to be effective [40].

There is an effort to better define and measure rehabilitation interventions to more fully understand what influences outcomes [41]. The data recorded by mRehab combined with performance changes on clinical assessments provide an opportunity to consider how exercise programs may impact performance on clinical outcome measures. We considered potential connections comparing tasks that decreased in time in both mRehab and WMFT. Time taken to complete all mRehab object transportation activities decreased. These activities would require adequate scapular, shoulder, elbow, wrist, and hand mobility and stability. Likewise, folding a towel in the WMFT would involve similar mobility and stability. In the WMFT, participants' scores improved for retrieving a 1-lb weight. While mRehab did not include progressive resistive training, repetitive task training has been shown to improve strength after stroke [42]. There were improvements on the Nine-Hole Peg Test, even though manipulation of objects using a pincer grasp was not part of the mRehab program. Participants did, however, demonstrate an improvement in performance of quick tap which requires fractionated finger movement. Taken together, it suggests that movement components trained within mRehab activities translate to other functional tasks. Having large-scale documentation of home exercise can lead to a better understanding of what form of exercise is most impactful on function.

A recent survey showed that clinicians perceive mRehab interventions as being important for supporting the function of patients at home and in the community, and improving adherence to home programs [43]. Therefore, it is necessary to perform research that examines how programs can be delivered at home. Other systems designed to improve upper limb function including custom hand–wrist orthosis and electrical stimulation, both designed to assist movement, or biofeedback to augment feedback during motor-based games have more commonly been assessed in clinical settings with the support of clinicians [44]. Tablet-based apps created to improve dexterity in the general population have been examined in individuals with stroke in a clinical setting, demonstrating that most individuals with stroke could use the system [12]. A gaming system designed for stroke used in a home-based setting, but included regular visits with clinicians, found similar results to this study, improvement in pre- to posttesting, but the correlation with practice was unremarkable. The percentage of days the participants used the gaming system ranged from 54% to 100% [25]. Overall, results from mHealth apps/systems appear promising, but much more research is needed to provide clinicians with the information they need to inform their decision making for mHealth home programs.

Limitations

While performance and usability of mRehab were assessed in laboratory prior to this study [11,45], the extended in-home use revealed flaws in the system. About 63% (10/16) of participants called to receive technical support and about 44% (7/16) of participants received home visits for assistance. The technical difficulties could have limited performance changes. In mRehab, the prehensile grip with supination/pronation activities did not demonstrate reduced times. It is possible that the difficulties with breakage of 3D door knobs and keys impeded performance improvements during training. Extended use of mRehab also demonstrated that the system did not work well for all individuals in the study. While inclusion and exclusion criteria were designed to select individuals that were a good fit with this intervention, 1 person did not complete the study. It is challenging to determine what combination of assessments will best predict adherence to mHealth-based programs. Further investigation is necessary to assure home programs are tailored to the individual's abilities.

Interestingly, when participants rated the usability at the end of the study, the average usability score for mRehab was 81.7. Modest technology assistance may have impacted the usability ratings. The use of technology in home programs is low [46]. It is possible that clinician's decision to use written home programs rather than technology is to eliminate the need for technology assistance. Not only is research necessary to create technology for rehabilitation, but also we must better understand how technology needs to be introduced and supported for successful use in self-managing long-term recovery.

Conclusion

The use of technology to improve home-programs and long-term recovery is promising. It can benefit both individuals with stroke in improving function and the field of rehabilitation in better understanding long-term recovery.

Acknowledgments

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number R21 HD092243-01.

Authors' Contributions

JL and SB oversaw recruitment, training and clinical assessments, and wrote the first draft of the manuscript. WX, BC, and ZL oversaw app function and mRehab data collection. LC and HS oversaw usability of mRehab objects. LC conducted statistics. All authors provided suggestions or revisions to the manuscript.

Conflicts of Interest

None declared.

References

1. Shah S, Vanclay F, Cooper B. Efficiency, effectiveness, and duration of stroke rehabilitation. Stroke 1990 Feb;21(2):241-246. [doi: 10.1161/01.STR.21.2.241]
2. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. The Lancet 2011 May;377(9778):1693-1702. [doi: 10.1016/S0140-6736(11)60325-5]
3. Masiero S, Armani M, Rosati G. Upper-limb robot-assisted therapy in rehabilitation of acute stroke patients: Focused review and results of new randomized controlled trial. JRRD 2011;48(4):355-366. [doi: 10.1682/jrrd.2010.04.0063]
4. Miller KK, Porter RE, DeBaun-Sprague E, Van Puymbroeck M, Schmid AA. Exercise after Stroke: Patient Adherence and Beliefs after Discharge from Rehabilitation. Topics in Stroke Rehabilitation 2016 Jun 23;24(2):142-148. [doi: 10.1080/10749357.2016.1200292]
5. Fryer C, Luker J, McDonnell M, Hillier S. Self management programmes for quality of life in people with stroke. Cochrane Database Syst Rev 2016 Aug 22;2016(8):CD010442 [FREE Full text] [doi: 10.1002/14651858.CD010442.pub2] [Medline: 27545611]
6. Pollock A, St George B, Fenton M, Firkins L. Top 10 research priorities relating to life after stroke--consensus from stroke survivors, caregivers, and health professionals. Int J Stroke 2014 Apr 11;9(3):313-320. [doi: 10.1111/j.1747-4949.2012.00942.x] [Medline: 23227818]
7. Kwakkel G, Veerbeek JM, van Wegen EEH, Wolf SL. Constraint-induced movement therapy after stroke. The Lancet Neurology 2015 Feb;14(2):224-234 [FREE Full text] [doi: 10.1016/S1474-4422(14)70160-7]
8. Agostini M, Moja L, Banzi R, Pistotti V, Tonin P, Venneri A, et al. Telerehabilitation and recovery of motor function: a systematic review and meta-analysis. J Telemed Telecare 2015 Feb 22;21(4):202-213. [doi: 10.1177/1357633x15572201]
9. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
10. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
11. Li J, Zhang C, Li X, Zhang C. Patients' emotional bonding with MHealth apps: An attachment perspective on patients' use of MHealth applications. International Journal of Information Management 2020 Apr;51:102054. [doi: 10.1016/j.ijinfomgt.2019.102054]
12. Kizony R, Zeilig G, Dudkiewicz I, Schejter-Margalit T, Rand D. Tablet Apps and Dexterity. Journal of Neurologic Physical Therapy 2016;40(1):31-39. [doi: 10.1097/npt.0000000000000110]
13. Bhattacharjya S, Stafford MC, Cavuoto LA, Yang Z, Song C, Subryan H, et al. Harnessing smartphone technology and three dimensional printing to create a mobile rehabilitation system, mRehab: assessment of usability and consistency in measurement. J Neuroeng Rehabil 2019 Oct 29;16(1):127 [FREE Full text] [doi: 10.1186/s12984-019-0592-y] [Medline: 31665036]
14. Nussbaum R, Kelly C, Quinby E, Mac A, Parmanto B, Dicianno BE. Systematic Review of Mobile Health Applications in Rehabilitation. Arch Phys Med Rehabil 2019 Jan;100(1):115-127. [doi: 10.1016/j.apmr.2018.07.439] [Medline: 30171827]
15. Szturm T, Peters JF, Otto C, Kapadia N, Desai A. Task-specific rehabilitation of finger-hand function using interactive computer gaming. Arch Phys Med Rehabil 2008 Nov;89(11):2213-2217. [doi: 10.1016/j.apmr.2008.04.021] [Medline: 18996252]
16. Klamroth-Marganska V, Blanco J, Campen K, Curt A, Dietz V, Ettlin T, et al. Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial. The Lancet Neurology 2014 Feb;13(2):159-166. [doi: 10.1016/s1474-4422(13)70305-3]
17. Pollock A, Farmer S, Brady M, Langhorne P, Mead G, Mehrholz J, et al. Interventions for improving upper limb function after stroke. Cochrane Database Syst Rev 2014 Nov 12;11(11):CD010820 [FREE Full text] [doi: 10.1002/14651858.CD010820.pub2] [Medline: 25387001]

18. Hubbard IJ, Parsons MW, Neilson C, Carey LM. Task-specific training: evidence for and translation to clinical practice. Occup Ther Int 2009 Sep;16(3-4):175-189 [FREE Full text] [doi: 10.1002/oti.275] [Medline: 19504501]
19. Dobkin BH, Plummer-D'Amato P, Elashoff R, Lee J, SIRROWS Group. International randomized clinical trial, stroke inpatient rehabilitation with reinforcement of walking speed (SIRROWS), improves outcomes. Neurorehabil Neural Repair 2010;24(3):235-242 [FREE Full text] [doi: 10.1177/1545968309357558] [Medline: 20164411]
20. Cirstea C, Ptito A, Levin M. Feedback and Cognition in Arm Motor Skill Reacquisition After Stroke. Stroke 2006 May;37(5):1237-1242. [doi: 10.1161/01.str.0000217417.89347.63]
21. Cavuoto LA, Subryan H, Stafford M, Yang Z, Bhattacharjya S, Xu W, et al. Understanding User Requirements for the Design of a Home-Based Stroke Rehabilitation System. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2018 Oct Presented at: 2018 Human Factors and Ergonomics Society (HFES) International Annual Meeting; 2018; Philadelphia, PA p. 1037-1041. [doi: 10.1177/1541931218621239]
22. Bobin M, Bimbard F, Boukallel M, Anastassova M, Ammi M. SpECTRUM: Smart ECosystem for sTRoke patient s Upper limbs Monitoring. Smart Health 2019 Aug;13:100066. [doi: 10.1016/j.smhl.2019.01.001]
23. Borghese NA, Pirovano M, Lanzi PL, Wüest S, de Bruin ED. Computational Intelligence and Game Design for Effective At-Home Stroke Rehabilitation. Games Health J 2013 Apr;2(2):81-88 [FREE Full text] [doi: 10.1089/g4h.2012.0073] [Medline: 24761321]
24. Llebaria G, Pagonabarraga J, Kulisevsky J, García-Sánchez C, Pascual-Sedano B, Gironell A, et al. Cut-off score of the Mattis Dementia Rating Scale for screening dementia in Parkinson's disease. Mov Disord 2008 Aug 15;23(11):1546-1550. [doi: 10.1002/mds.22173] [Medline: 18546326]
25. Slijper A, Svensson KE, Backlund P, Engström H, Sunnerhagen K. Computer game-based upper extremity training in the home environment in stroke persons: a single subject design. J Neuroeng Rehabil 2014 Mar 13;11(1):35 [FREE Full text] [doi: 10.1186/1743-0003-11-35] [Medline: 24625289]
26. Moss A, Nicholas M. Language Rehabilitation in Chronic Aphasia and Time Postonset. Stroke 2006 Dec;37(12):3043-3051. [doi: 10.1161/01.str.0000249427.74970.15]
27. Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. Phys Ther 1997 Aug 01;77(8):812-819. [doi: 10.1093/ptj/77.8.812] [Medline: 9256869]
28. Wolf S, Lecraw D, Barton L, Jann B. Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. Experimental Neurology 1989 May;104(2):125-132. [doi: 10.1016/s0014-4886(89)80005-6]
29. Oxford Grice K, Vogel KA, Le V, Mitchell A, Muniz S, Vollmer MA. Adult norms for a commercially available Nine Hole Peg Test for finger dexterity. Am J Occup Ther 2003 Sep 01;57(5):570-573. [doi: 10.5014/ajot.57.5.570] [Medline: 14527120]
30. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS). In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2013 Sep 30 Presented at: 2018 Human Factors and Ergonomics Society (HFES) International Annual Meeting; 2018 October; Philadelphia, PA p. 192-196. [doi: 10.1177/1541931213571043]
31. Wolf SL, Thompson PA, Morris DM, Rose DK, Winstein CJ, Taub E, et al. The EXCITE trial: attributes of the Wolf Motor Function Test in patients with subacute stroke. Neurorehabil Neural Repair 2005 Sep 30;19(3):194-205. [doi: 10.1177/1545968305276663] [Medline: 16093410]
32. Tistad M, Tham K, von Koch L, Ytterberg C. Unfulfilled rehabilitation needs and dissatisfaction with care 12 months after a stroke: an explorative observational study. BMC Neurol 2012 Jun 18;12(1):40 [FREE Full text] [doi: 10.1186/1471-2377-12-40] [Medline: 22708545]
33. Duxbury S, Depaul V, Alderson M, Moreland J, Wilkins S. Individuals with stroke reporting unmet need for occupational therapy following discharge from hospital. Occup Ther Health Care 2012 Jan 23;26(1):16-32. [doi: 10.3109/07380577.2011.621514] [Medline: 23899105]
34. Pollock A, Baer G, Campbell P, Choo PL, Forster A, Morris J, et al. Physical rehabilitation approaches for the recovery of function and mobility following stroke. Cochrane Database Syst Rev 2014 Apr 22(4):CD001920 [FREE Full text] [doi: 10.1002/14651858.CD001920.pub3] [Medline: 24756870]
35. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. PLoS One 2014 Feb;9(2):e87987 [FREE Full text] [doi: 10.1371/journal.pone.0087987] [Medline: 24505342]
36. Plautz EJ, Milliken GW, Nudo RJ. Effects of repetitive motor training on movement representations in adult squirrel monkeys: role of use versus learning. Neurobiol Learn Mem 2000 Jul;74(1):27-55. [doi: 10.1006/nlme.1999.3934] [Medline: 10873519]
37. Nudo RJ, Plautz EJ, Frost SB. Role of adaptive plasticity in recovery of function after damage to motor cortex. Muscle Nerve 2001 Aug;24(8):1000-1019. [doi: 10.1002/mus.1104] [Medline: 11439375]
38. Remple MS, Bruneau RM, VandenBerg PM, Goertzen C, Kleim JA. Sensitivity of cortical movement representations to motor experience: evidence that skill learning but not strength training induces cortical reorganization. Behavioural Brain Research 2001 Sep;123(2):133-141. [doi: 10.1016/s0166-4328(01)00199-1]

39. Page SJ. Intensity Versus Task-Specificity After Stroke. American Journal of Physical Medicine & Rehabilitation 2003;82(9):730-732. [doi: 10.1097/01.phm.0000078226.36000.a5]
40. Novak I. Effective home programme intervention for adults: a systematic review. Clin Rehabil 2011 Dec 10;25(12):1066-1085. [doi: 10.1177/0269215511410727] [Medline: 21831927]
41. Dijkers MP, Ferraro MK, Hart T, Packel A, Whyte J, Zanca JM. Toward a rehabilitation treatment taxonomy: summary of work in progress. Phys Ther 2014 Mar 01;94(3):319-321. [doi: 10.2522/ptj.20130999] [Medline: 24481596]
42. de Sousa DG, Harvey LA, Dorsch S, Glinsky JV. Interventions involving repetitive practice improve strength after stroke: a systematic review. J Physiother 2018 Oct;64(4):210-221 [FREE Full text] [doi: 10.1016/j.jphys.2018.08.004] [Medline: 30245180]
43. Morris J, Jones M, Thompson N, Wallace T, DeRuyter F. Clinician Perspectives on mRehab Interventions and Technologies for People with Disabilities in the United States: A National Survey. Int J Environ Res Public Health 2019 Oct 31;16(21):4220 [FREE Full text] [doi: 10.3390/ijerph16214220] [Medline: 31683536]
44. Graham J. Technology for rehabilitating the upper limb after stroke: the hand. British Journal of Neuroscience Nursing 2014 Apr;10(Sup2):19-24. [doi: 10.12968/bjnn.2014.10.sup2.19]
45. Bhattacharjya S. Assessing the Functionality of a Portable Rehabilitation System Used for Self-Management of Upper Extremity Function by Individuals With Stroke. Am J Occup Ther 2018 Jul 01;72(4_Supplement_1):7211505134p1. [doi: 10.5014/ajot.2018.72s1-po6031]
46. Langan J, Subryan H, Nwogu I, Cavuoto L. Reported use of technology in stroke rehabilitation by physical and occupational therapists. Disabil Rehabil Assist Technol 2018 Oct 16;13(7):641-647. [doi: 10.1080/17483107.2017.1362043] [Medline: 28812386]

Abbreviations

BBS: Berg Balance Scale
MAS: Modified Ashworth Scale
MDRS: Mattis Dementia Rating Scale
mRehab: mobile Rehab
SUS: Systems Usability Scale
WMFT: Wolf Motor Function Test

Edited by G Eysenbach; submitted 28.04.20; peer-reviewed by L Tedesco Triccas, I Schiering; comments to author 20.05.20; revised version received 04.06.20; accepted 25.06.20; published 22.07.20.

Please cite as:
Langan J, Bhattacharjya S, Subryan H, Xu W, Chen B, Li Z, Cavuoto L
In-Home Rehabilitation Using a Smartphone App Coupled With 3D Printed Functional Objects: Single-Subject Design Study
JMIR Mhealth Uhealth 2020;8(7):e19582
URL: http://mhealth.jmir.org/2020/7/e19582/
doi:10.2196/19582
PMID:32706702

Original Paper

Haptic Nudges Increase Affected Upper Limb Movement During Inpatient Stroke Rehabilitation: Multiple-Period Randomized Crossover Study

Nada Elizabeth June Signal[1], BHSc, MHSc, PhD; Ruth McLaren[1], BHSc, MHPrac; Usman Rashid[1], BEng, MSc, PhD; Alain Vandal[2], MA, PhD; Marcus King[3], BE; Faisal Almesfer[4], BE, ME; Jeanette Henderson[5], BHSc; Denise Taylor[1], BPhys, MSc, PhD

[1]Health and Rehabilitation Research Institute, Auckland University of Technology, Auckland, New Zealand

[2]Department of Statistics, University of Auckland, Auckland, New Zealand

[3]Callaghan Innovation, Christchurch, New Zealand

[4]Exsurgo Rehabilitation, Auckland, New Zealand

[5]Assessment, Treatment and Rehabilitation Department, Waitakere Hospital, Waitemata District Health Board, Auckland, New Zealand

Corresponding Author:
Nada Elizabeth June Signal, BHSc, MHSc, PhD
Health and Rehabilitation Research Institute
Auckland University of Technology
90 Akoranga Drive
Northcote
Auckland, 0627
New Zealand
Phone: 64 21929144
Email: nada.signal@aut.ac.nz

Abstract

Background: As many as 80% of stroke survivors experience upper limb (UL) disability. The strong relationships between disability, lost productivity, and ongoing health care costs mean reducing disability after stroke is critical at both individual and society levels. Unfortunately, the amount of UL-focused rehabilitation received by people with stroke is extremely low. Activity monitoring and promotion using wearable devices offer a potential technology-based solution to address this gap. Commonly, wearable devices are used to deliver a haptic nudge to the wearer with the aim of promoting a particular behavior. However, little is known about the effectiveness of haptic nudging in promoting behaviors in patient populations.

Objective: This study aimed to estimate the effect of haptic nudging delivered via a wrist-worn wearable device on UL movement in people with UL disability following stroke undertaking inpatient rehabilitation.

Methods: A multiple-period randomized crossover design was used to measure the association of UL movement with the occurrence of haptic nudge reminders to move the affected UL in 20 people with stroke undertaking inpatient rehabilitation. UL movement was observed and classified using movement taxonomy across 72 one-minute observation periods from 7:00 AM to 7:00 PM on a single weekday. On 36 occasions, a haptic nudge to move the affected UL was provided just before the observation period. On the other 36 occasions, no haptic nudge was given. The timing of the haptic nudge was randomized across the observation period for each participant. Statistical analysis was performed using mixed logistic regression. The effect of a haptic nudge was evaluated from the intention-to-treat dataset as the ratio of the odds of affected UL movement during the observation period following a “Planned Nudge” to the odds of affected limb movement during the observation period following “No Nudge.”

Results: The primary intention-to-treat analysis showed the odds ratio (OR) of affected UL movement following a haptic nudge was 1.44 (95% CI 1.28-1.63, *P*<.001). The secondary analysis revealed an increased odds of affected UL movement following a Planned Nudge was predominantly due to increased odds of spontaneous affected UL movement (OR 2.03, 95% CI 1.65-2.51, *P*<.001) rather than affected UL movement in conjunction with unaffected UL movement (OR 1.13, 95% CI 0.99-1.29, *P*=.07).

Conclusions: Haptic nudging delivered via a wrist-worn wearable device increases affected UL movement in people with UL disability following stroke undertaking inpatient rehabilitation. The promoted movement appears to be specific to the instructions given.

Trial Registration: Australia New Zealand Clinical Trials Registry 12616000654459; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370687&isReview=true

(JMIR Mhealth Uhealth 2020;8(7):e17036) doi:10.2196/17036

KEYWORDS

stroke; rehabilitation; physical activity; movement; disability; technology; upper limb; wearable; haptic; nudge

Introduction

Although the incidence of stroke has reduced, its burden continues to grow as more people are surviving after stroke and living with disability [1]. Direct and indirect health care costs following stroke are strongly correlated with stroke disability, with greater disability associated with greater costs [2,3]. Around 80% of stroke survivors experience upper limb (UL) disability, with only 5%-20% achieving full recovery of UL function [4-6]. UL disability has subsequent impacts on independence in activities of daily living, discharge destination, return to work, quality of life, and mood [7-10].

Effective rehabilitation involves high-dose, intensive, task-specific activity [11]. Meta-analyses of randomized controlled trials suggest there is a dose-response relationship, with higher doses of rehabilitation resulting in better outcomes [12-14]. However, studies describing usual stroke care illustrate that the dose of UL rehabilitation received by people with stroke is extremely low, with as little as 4-6 minutes in physiotherapy sessions and 11-17 minutes in occupational therapy sessions [15]. Movement of the affected UL outside formal therapy sessions during inpatient rehabilitation is also low [16]. Consequently, affected UL movement dose, both within formal therapy sessions and across the rehabilitation day, is currently insufficient to reduce UL disability following stroke.

Rehabilitation technologies have been proposed as potential solutions to the limited dose of rehabilitation this population receives [17]. A number of technological solutions have been developed for use in UL stroke rehabilitation, including virtual reality, gaming, and robotics [17-19]. Despite indications of effectiveness [20,21], the use of rehabilitation technologies is not yet pervasive in stroke care [18]. Therapists have identified several barriers to adopting rehabilitation technologies, including concerns about patient safety, whether the technology effectively addresses a clinical need, and the feasibility of technologies from time, space, and cost perspectives [22,23]. The poor uptake of rehabilitation technologies is inconsistent with research involving people with stroke that indicates rehabilitation technologies can support engagement and interest in performing repetitive rehabilitation activities and offer a means of social support [19,24,25]. Activity monitoring and promotion using wearable devices is a potential low-cost and feasible rehabilitation technology. Research investigating the effect of wearable devices on outcomes following stroke is in its infancy. Preliminary indications suggest that wearable devices may increase the amount and intensity of physical activity undertaken during rehabilitation [26-28] and potentially contribute to improved functional outcomes [29]. However, the effect of wearable devices on UL rehabilitation and outcomes has been less well studied [30,31], with much of the research to date focusing on the accuracy and validity of accelerometry measurement of real-world UL movement [32-34].

Wearable devices that deliver haptic nudges have been used to promote physical activity in both healthy and patient populations [35-39]. Commonly, a haptic nudge reminder is delivered via a small motor embedded inside a wearable device. Wearers are encouraged to respond to a haptic reminder by performing a particular behavior. For example, a haptic nudge might be used to remind the wearer to stand up and move after an extended period of sitting or to undertake rehabilitation exercises. However, despite the pervasiveness of haptic nudging in consumer wearable devices, there remains much to learn about the effectiveness of haptic nudging in promoting behaviors in patient populations. Haptic nudges have been used to effectively promote behaviors in people with autism spectrum disorder [40] and traumatic brain injury engaged in a rehabilitation task [41]. Research also suggests that wearable devices are feasible and well tolerated in people with stroke [30,42,43], with preliminary data indicating haptic nudging via wearable devices may promote affected UL movement [31]. The aim of this study was to estimate the effect of haptic nudging delivered via a wrist-worn wearable device on UL movement in people with UL disability following stroke undertaking inpatient rehabilitation.

Methods

Study Design

A multiple-period randomized crossover design was used to measure the association of UL movement with the occurrence of haptic nudge reminders to move the affected UL in 20 people with stroke undertaking inpatient rehabilitation. UL movement was observed and classified using movement taxonomy across 72 one-minute observation periods from 7:00 am to 7:00 pm on a single weekday. On 36 occasions, a haptic nudge (Nudge) to move the affected UL was provided just before the observation period, and on the other 36 occasions, no haptic nudge was given (No Nudge). The timing of the haptic nudge was randomized. Approval for this study was obtained from the New Zealand Health and Disability Ethics Committee (16/NTA/74).

Participants

All people with stroke admitted to the rehabilitation service from July 2018 through December 2018 were considered for inclusion in this study. Participants were included if they had a confirmed diagnosis of stroke based on the Oxford classification system [44], presented with UL deficit as a result of stroke as determined by their rehabilitation therapist, were deemed medically stable and fit for rehabilitation by their medical consultant, and provided written informed consent.

Participants were excluded if they had cognitive, behavioral, or communication impairments that, in the opinion of the research team (RM, DT, NS), would limit their ability to participate in the research (for example, if the person was unable to follow a 2-step verbal command or recall the details of the research study); were within 3 days of planned discharge from inpatient rehabilitation; or reported shoulder pain.

Procedure

Potential participants were identified and referred to the research team by their rehabilitation therapist. They were then informed about the study and screened against the inclusion and exclusion criteria by a trained research assistant. Eligible participants provided written informed consent and identified a mutually agreeable day for data collection. Demographic, clinical, and medical information was gathered from the medical record of consenting participants by the rehabilitation therapist. Collected data included age, sex, ethnicity, date of stroke, type of stroke, side of body most affected, dominant hand prior to stroke, date of admission to the rehabilitation ward, estimated date of discharge, comorbidities, and medications.

On the day of data collection, the participant was fitted with a BuzzNudge wearable device on the wrist of the affected UL. The BuzzNudge is a Bluetooth-enabled wearable device with a 2.3 V coin vibration motor (Precision Microdrives Ltd, Model 310-103), which provided 3 consecutive vibratory stimuli of 0.3 seconds duration at 150 Hz within 1.5 seconds, representing a similar magnitude of stimulus to the vibration of a phone. The researcher explained the value of moving the affected UL after stroke and instructed the participant to “move, try and move, or visualize moving their (affected) arm” following a nudge. The researcher emphasized that the participant should do whatever movement they felt they could manage. If sensation was impaired in the affected UL such that the participant could not feel the haptic nudge, the device was worn on the less affected UL, but the participant was still instructed to use the haptic nudge as a reminder to move the affected UL.

During data collection, participants were followed discreetly out of their field of view around the rehabilitation ward, therapy areas, and hospital facilities by a trained researcher (where feasible). The researcher manually recorded UL movement for 1 minute every 10 minutes [45]. Each minute of observation was broken into 6 epochs of 10 seconds using a silent interval timer. UL movement was classified according to a previously defined taxonomy: (1) unilateral affected UL movement; (2) unilateral unaffected UL movement; (3) bimanual movement, where movement of both ULs was observed to achieve a common task or purpose; (4) bilateral limb movement, where movement of both ULs was observed to achieve independent or unrelated tasks; and (5) no movement [16]. When patients were not able to be directly observed (ie, because curtains were drawn or when in showers or toilets), activity was recorded after conferring with the participant, staff, or family members, as appropriate. In circumstances where the activity could not be estimated (eg, during 4 randomly scheduled observer breaks), activity was coded as unobserved [16,46].

Haptic nudge reminders were triggered by the researcher via Bluetooth immediately before movement observation according to a planned randomization schedule. For half of the observation periods, a haptic nudge was to be provided, and for half, a haptic nudge was not to be provided. Further details regarding the randomization schedule are presented in Multimedia Appendix 1. Haptic nudges were not triggered if the participant was not visible, the participant was asleep, or a nudge was considered inappropriate (eg, if the participant was drinking a hot beverage or undertaking an assessment procedure). Any scheduled nudge that was not given was recorded as a “Missed Nudge.”

Statistical Analysis

Coded data were entered into a Microsoft Excel spreadsheet. Descriptive analysis was used to examine the amount and type of UL movement. If a participant withdrew from the study, their movement observations were coded as missing values, and their scheduled nudges were coded as Missed Nudges. Statistical analysis was performed using mixed logistic regression. Total affected UL movement was collated based on all observations in which the affected UL was moved: total affected UL movement = unilateral affected UL movement + bilateral limb movement + bimanual movement.

In the primary analysis using the intention-to-treat dataset, nudges were represented as a fixed effect factor with two levels: Planned Nudge (ie, Nudge + Missed Nudge) and No Nudge, meaning the analysis considered whether a nudge was planned or not, rather than delivered. The effect of Planned Nudges was evaluated as the ratio of the odds of affected UL movement during the observation period following a Planned Nudge to the odds of affected UL movement during the observation period following No Nudge. More formally, the primary null hypothesis tested with the model was: H_0: $OR_{\text{Planned Nudge/No Nudge}} = 1$.

The sensitivity of the effect of the Planned Nudge to missing values was tested with the pooled effect of 10 worst-case random imputations at the level of the participant with the worst outcomes.

In the secondary analysis using the intention-to-treat dataset, two additional models were used to evaluate the effect of a Planned Nudge on unilateral affected UL movement and the sum of bilateral limb movement + bimanual movement, respectively. A post-hoc exploratory analysis with an instrumental variable approach was used to evaluate the local average treatment effect (also known as complier average causal effect) of the haptic nudge reminder. This analysis considered the effect of the haptic nudge when delivered (Nudged) compared with no haptic nudge (Not Nudged) irrespective of the schedule on total affected UL movement. To explain the variation in UL movement across the day, all models fitted the data with smooth natural splines that had 1 degree of freedom per hour. To account for correlated repeated measures, the models included hierarchical random effects per participant and per hour within participant. Statistical analyses were performed using R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) with *lme4* version 1.1-21 [47], *splines* version 3.5.1 [48], and *emmeans* version 1.3.4 [39]. The threshold of statistical significance was set at .05. A detailed statistical analysis report containing code snippets and additional graphical representations of data is available in Multimedia Appendix 1.

Results

Participant Characteristics

In total, 20 people consented to participate in this study (Table 1). Participants' median age was 76 years (IQR 68-83 years), and the median time since stroke was 23.5 days (IQR 8.25-38.25 years); 9 participants had left hemiparesis, 10 had right hemiparesis, and 1 participant had bilateral symptoms with the left UL more affected than the right. Five participants had total anterior circulation syndrome, 10 had partial anterior circulation syndrome, 4 had lacunar circulation syndrome, and 1 had posterior circulation syndrome. Of the participating patients, 4 had hemorrhagic stroke, and 16 had ischemic stroke. Participant 2 was withdrawn from the study when 6 nudges in a row were unable to be delivered due to a technical error. Participant 6 asked to withdraw 20 minutes into data collection due to experiencing anxiety associated with wearing the device.

Table 1. Participant demographics.

Participant	Age range (years)	Gender	Stroke classification	Days since stroke	AUL[a]	AUL = dominant UL[b]	Device worn on AUL
1	70-79	Male	LACS-I[c]	9	Left	No	Yes
2[d]	80-89	Female	TACS-I[e]	39	Left	No	Yes
3	70-79	Female	TACS-I	59	Left	No	Yes
4	40-49	Female	LACS-H[f]	8	Right	Yes	Yes
5	60-69	Female	PACS-I[g]	5	Right	Yes	Yes
6[d]	80-89	Male	PACS-I	34	Left	No	Yes
7	70-79	Male	TACS-I	27	Right	Yes	Yes
8	80-89	Female	PACS-I	7	Left	No	Yes
9	80-89	Male	TACS-I	67	Right	Yes	Yes
10	60-69	Male	PACS-I	36	Right	Yes	Yes
11	50-59	Female	TACS-I	25	Left	Yes	Yes
12	70-79	Male	LACS-H	33	Left	No	No
13	80-89	Male	PACS-I	12	Left	No	Yes
14	60-69	Male	LACS-I	3	Right	Yes	No
15	70-79	Male	PACS-H[h]	40	Right	Yes	Yes
16	80-89	Female	PACS-I	22	Right	Yes	Yes
17	60-69	Female	PACS-I	6	Left	No	Yes
18	60-69	Male	POCS-I[i]	10	Bilateral	Yes	Yes
19	80-89	Male	PACS-I	9	Right	Yes	Yes
20	80-89	Male	PACS-H	160	Right	Yes	Yes

[a]AUL: affected upper limb.

[b]UL: upper limb.

[c]LACS-I: lacunar circulation syndrome ischemic.

[d]Participant withdrawn.

[e]TACS-I: total anterior circulation syndrome ischemic.

[f]LACS-H: lacunar circulation syndrome hemorrhagic.

[g]PACS-I: partial anterior circulation syndrome ischemic.

[h]PACS-H: partial anterior circulation syndrome hemorrhagic.

[i]POCS-I: posterior circulation syndrome ischemic.

Data Completeness

In total, 7517 of a possible 8640 observations of movement in 10-second time intervals were recorded across the 20 participants (median 414 observations/participant, IQR 402-420 observations/participant, range 12-432 observations/participant), representing data completeness of 87.0%. Data loss was due to the 2 participants who withdrew (769/8640 observations, 8.9%) and the remaining participants being away from the service for appointments or involved in private personal hygiene activities (354/8640 observations, 4.1%).

Of the 720 Planned Nudges, 32.6% (235 Nudges) were not delivered (Missed Nudge), with 8.7% (63/720 Nudges) ascribed

to the 2 participants that withdrew. For the remaining Missed Nudges, participants were sleeping for 12.6% (90/720) of the Planned Nudges; Nudges could not be directly observed for 6.7% (48/720) of the Planned Nudges; a nudge was deemed inappropriate for 2.8% (20/720) of the Planned Nudges; a technical error prevented nudging for 1.4% (10/720) of the Planned Nudges; and the reason was not stated for 0.4% (3/720) of the Planned Nudges.

UL Movement

During observations without a nudge scheduled (No Nudge), the affected UL moved 19.2% of the time; 15.6% of movement occurred in conjunction with the unaffected UL, and only 3.6% of the time the movement was of the affected UL by itself. The unaffected UL moved 39.2% of the time, with half of this movement (23.6%) being movement of the unaffected UL by itself. Participants used one or both ULs for 42.8% of the observation time.

Haptic Nudge Effect

The results of the statistical analyses are presented in Table 2. The treatment effect of the intervention is represented by the odds ratio (OR) for Planned Nudge versus No Nudge. This OR indicated that the odds of moving the affected UL either independently (unilateral affected UL movement) or in concert with the unaffected limb (bimanual movement or bilateral limb movement) was 1.44 times greater following a Planned Nudge than following No Nudge. The proportions estimated by the model for the affected UL movement (unilateral affected UL movement + bilateral limb movement + bimanual movement) recorded during the observation periods following a Planned Nudge and No Nudge were 26.7% (95% CI 15.4%-42.2%) and 20.2% (95% CI 11.2%-33.6%), respectively. Therefore, the average absolute increase in the proportion of affected UL movement with the intervention was 6.5% (95% CI 4.2%-8.6%), representing an increase of 32.2% in average activity. The sensitivity analysis showed that the effect of the Planned Nudge on unilateral affected UL movement + bilateral limb movement + bimanual movement was robust to missing values (P<.001). The proportion of observation periods with affected UL movement following Planned Nudges and No Nudges by participants is represented in Table 3.

The secondary analysis revealed that the odds of moving the affected UL independently (unilateral affected UL movement) was 2.03 times greater following a Planned Nudge than following No Nudge. However, the OR for either bilateral or bimanual movement (bilateral limb movement + bimanual movement) was only 1.13 times greater. The exploratory analysis revealed that the OR for the effect of the haptic nudge when delivered (Nudged) compared with no haptic nudge (Not Nudged) irrespective of the schedule was 1.64.

Table 2. Odds of an affected upper limb (UL) movement recorded during the observation periods.

Estimate	Odds ratio	95% CI	Standard error	Z value	*P* value
Planned nudge/no nudge					
Primary analysis: AU[a]+BiL[b]+BiM[c]	1.44	1.28-1.63	0.09	5.86	<.001
Sensitivity analysis: AU+BiL+BiM	1.30	1.16-1.46	0.08	4.4	<.001
Secondary analysis: AU	2.03	1.65-2.51	0.22	6.65	<.001
Secondary analysis: BiL+BiM	1.13	0.99-1.29	0.08	1.80	.07
Nudged/not nudged					
Exploratory analysis: AU+BiL+BiM	1.64	1.42-1.89	0.12	6.77	<.001

[a]AU: unilateral affected upper limb movement.

[b]BiL: bilateral limb movement.

[c]BiM: bimanual movement.

Table 3. Proportion of observations with affected upper limb movment (unilateral affected upper limb movement + bilateral limb movement + bimanual movement) following No Nudge and Planned Nudge.

Participant	No Nudge (%)	Planned Nudge (%)
1	7.41	11.11
2[a]	1.85	0.00
3	26.39	30.09
4	3.24	13.89
5	46.30	52.31
6[a]	3.70	0.00
7	18.06	12.96
8	35.19	38.89
9	1.39	5.56
10	16.67	9.26
11	5.56	7.87
12	6.48	0.93
13	30.09	37.96
14	15.74	42.13
15	20.83	39.35
16	32.87	43.98
17	23.61	37.96
18	38.89	24.54
19	25.93	43.06
20	23.61	16.67

[a]Participant withdrawn.

Discussion

Principal Findings

This multiple-period randomized crossover study explored the effect of haptic nudging on UL movement in people with stroke. Haptic nudging increased the likelihood that a person with stroke moved their affected UL by 1.44 times (P<.001) during the subacute rehabilitation phase. Haptic nudging resulted in a relative increase in the proportion of affected UL movement of 32.2%. The actual amount and type of UL movement observed without nudging in our study was comparable with our previously published research [16] and other observational studies [33,46]. This strengthens the assertion that haptic nudges influence the amount of affected UL movement in people with stroke. This study is the first to specifically investigate whether haptic nudges delivered by a wrist-worn wearable device influence the amount of UL movement undertaken during rehabilitation following stroke. Given the limited amount of spontaneous UL movement following stroke [16,46,49] and the challenges associated with increasing the dose of UL rehabilitation [15,50], these research findings indicate that haptic nudging represents a potentially powerful stroke rehabilitation tool that could be easily implemented in clinical practice.

Our secondary analysis clarified the type of UL movement promoted by haptic nudging. Participants were 2.03 times more likely to move their affected UL in isolation (unilateral affected UL movement) following a haptic nudge compared with moving the affected UL in concert with the unaffected UL (bimanual movement + bilateral limb movement), which was just 1.13 times more likely. Given that participants were instructed to move their affected UL rather than both limbs together, the instructions related to haptic nudging may be important in determining exactly which movements are promoted. This specificity in effect has been noted in other studies in which people with stroke altered their behavior in response to feedback. Dobkin and Plummer-D'Amato [51] gave daily feedback in relation to gait speed to people with stroke undertaking rehabilitation. They reported the feedback group had significantly increased gait speed but did not change walking endurance or independence compared with the control group (usual care). Our study contributes to the growing body of research suggesting that drawing attention to specific aspects of movement and physical activity throughout the rehabilitation day can influence patient behavior during stroke rehabilitation [26-31].

The greater odds of moving the affected UL in the minute following a planned nudge resulted in a 32.2% increase in the average amount of movement. Another small-scale proof-of-concept study involving people with stroke (n=7) [30] indicated an increase of 19.7% in the average amount of affected UL movement in the hour following a haptic nudge reminder

to undertake exercises. In that study, participants were instructed to perform up to 80 repetitions of task-specific training in response to a haptic nudge, but only received a median of 4 nudges across the rehabilitation day. In contrast, participants in our study received an average of 27 haptic nudges and were instructed to move, try to move, or think about moving their affected UL following a nudge. It is not yet known how the frequency of nudges, burden of the required behavioral response, and capacity to integrate that response into everyday activities influence the magnitude and duration of the haptic nudge effect in people with stroke.

Results for individual participants illustrated that there was considerable variability in response to haptic nudging. For example, 13 participants increased the amount of movement of their affected UL in response to nudging, with 8 exhibiting large relative increases. In contrast, 5 participants had a reduction in the amount of affected UL movement in response to haptic nudging. There appeared to be no relationship between individual response and participant age, stroke severity, hemiplegic side, or whether the hemiplegic side was the dominant hand. Although the researchers checked that the participant understood the instruction to move the affected UL in response to haptic nudging at the beginning of data collection, it remains unclear whether cognitive, communication, perceptual, or sensory deficits influenced participants' ability to attend and respond to the nudge. Other nudging modalities (eg, auditory tones, lights, and text messaging) may be effective and more appropriate than haptic vibration for some people with stroke [52]. In addition, we relied on observation of movement, and it was not possible to determine whether participants who were more severely affected were attempting to or thinking about moving their affected UL. This could be addressed in future research by using alternative data collection methods such as ecological momentary assessment [53] or electroencephalography to determine movement intention [54]. One participant who had moved less in response to haptic nudging had a cerebellar stroke that influenced UL movement bilaterally; advocating increased movement of the more affected limb might have been inappropriate in that case. The relationship between haptic nudge efficacy and clinical and demographic factors requires further investigation to ensure that this type of technology is used appropriately.

The magnitude of effect in response to haptic nudging might have been underestimated in our study given that we included observations in which planned nudges were not delivered (Missed Nudges), for example, when a participant was asleep or not visible to the researcher. This assertion was supported by the exploratory analysis that revealed the OR for the effect of the scheduled nudge when actually delivered compared with no haptic nudge was 1.64. It is also noteworthy that our participants were advised of the value of moving the affected UL after stroke on a single occasion. The effect of haptic nudging may be enhanced by providing regular positively framed information on the consequences of nudged behavior (eg, "more movement promotes recovery"), encouraging explicit action planning (eg, "I will move my affected arm by…when I feel the nudge"), repeated practice of the desired behavioral response to the haptic nudge, and tracking and reporting the desired behavior [35,55,56]. In commercial wearable devices, haptic nudging is commonly coupled with other behavior change and persuasive strategies including education, gamification, social support via social network services, and reward systems [56,57]. It is likely that the magnitude of effect of a comprehensive rehabilitation wearable technology that incorporates haptic nudging with other behavioral change and persuasive strategies would have a larger effect than the effect of haptic nudging alone, as estimated in this study [35,58,59].

This study sought to investigate the effect of haptic nudging on UL movement across an inpatient rehabilitation day; we did not explore the effect on UL movement over a longer timeframe or in a community setting. It is possible that people with stroke habituate or become less responsive to haptic nudging with everyday use. Conversely, they may learn to respond to haptic nudging more effectively over time. Understanding the effect over time is important, as increasing the dose of upper limb movement to a therapeutic level through continued engagement over a matter of weeks to months is likely required to promote functional gains. In the subacute phase following stroke, adherence to wearable devices has been reported as high [31,42], although use appears to dwindle over time [42]. This is consistent with studies in healthy community-dwelling people, where half to two-thirds of purchasers continued to use wearable devices 6 months after purchase [35,36]. In healthy populations, uptake and ongoing use of such devices are influenced by personal characteristics, including age, computer self-efficacy, ease of use, usual levels of physical activity, internalization of intention to change, and personality [36,60]. Although wearable devices have been found to be acceptable to people with stroke [61], one of our participants withdrew at the beginning of data collection because wearing the device made him feel anxious. Previous research indicates that the use of wearable technologies may increase anxiety in clinical populations [62,63]. When developing wearable devices to promote physical activity and movement in people with stroke it may be important to consider the personal and clinical characteristics of the intended users, when and where in the continuum of care the device will be used and for how long, and how users' engagement and adherence can be supported.

Limitations

A key limitation of this study was that the researcher observing and recording movement was also responsible for triggering the nudge and therefore not blinded to the intervention. While the randomization schedule was designed to address lag, where the effect of a nudge influences subsequent movement observation periods (refer to Multimedia Appendix 1), the duration of the nudge effect was unknown and might have influenced subsequent observations. While participants were blinded to the study hypothesis, it is possible that the research protocol, particularly being observed by a researcher, may have influenced the likelihood they moved in response to the haptic nudge. Documentation of the number of potential participants screened and the reasons for exclusion from the study would have helped to interpret the external validity of the study findings. A more detailed evaluation of the included participants' sensorimotor, perceptual, cognitive, and communication impairments along with measurement of their UL functional

abilities may have allowed for a more nuanced interpretation of the effect of haptic nudging in people with different clinical presentations of stroke.

Conclusions

Haptic nudging increased the likelihood that a person with stroke moved their affected UL by 1.44 times. This equated to an increase of 32% in the average amount of affected UL movement. Participants were twice (OR 2.03) as likely to move their affected UL in isolation (unilateral affected UL movement) in response to haptic nudging, compared with movement in conjunction with the unaffected UL (OR 1.13), indicating that the effect of haptic nudging was specific to the behavioral instructions given. Given the limited amount of spontaneous UL movement following stroke and the challenges associated with increasing the dose of UL rehabilitation, haptic nudging as part of a comprehensive wearable device aimed at increasing the dose of UL movement represents a potentially powerful stroke rehabilitation tool.

Acknowledgments

This research was funded by a grant from the New Zealand Medical Technologies Centre of Research Excellence. The authors are grateful to the Waitemata District Health Board and the staff of the rehabilitation service at Waitakere Hospital who supported this research. Simon Tse, Rohil Chauhan, Shikha Chaudhary, Abdallah Abou El Ela, and Nicole Wallen contributed to data collection.

Authors' Contributions

NS, DT, and MK conceptualized the study and acquired funding. DT, AV, RM, and NS designed the methodology, and RM, DT, and NS performed project administration. RM and JH collected the data, and MK and FA were responsible for the technology. Data analysis was performed by UR and AV, and UR created the visualizations. Supervision was performed by NS and DT. The original draft was written by NS, RM, and UR, and all authors reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Statistical analysis for the haptic nudge study.
[PDF File (Adobe PDF File), 738 KB - mhealth_v8i7e17036_app1.pdf]

References

1. Yu A, Kapral MK. More people are surviving after acute stroke. BMJ 2019 May 22;365:l2150 [FREE Full text] [doi: 10.1136/bmj.l2150] [Medline: 31122927]
2. Dewilde S, Annemans L, Peeters A, Hemelsoet D, Vandermeeren Y, Desfontaines P, et al. Modified Rankin scale as a determinant of direct medical costs after stroke. Int J Stroke 2017 Jun 06;12(4):392-400. [doi: 10.1177/1747493017691984] [Medline: 28164742]
3. Mittmann N, Seung SJ, Hill MD, Phillips SJ, Hachinski V, Coté R, et al. Impact of Disability Status on Ischemic Stroke Costs in Canada in the First Year. Can. J. Neurol. Sci 2014 Dec 02;39(6):793-800. [doi: 10.1017/s0317167100015638]
4. Kwakkel G, van Peppen R, Wagenaar RC, Wood Dauphinee S, Richards C, Ashburn A, et al. Effects of augmented exercise therapy time after stroke: a meta-analysis. Stroke 2004 Nov;35(11):2529-2539. [doi: 10.1161/01.STR.0000143153.76460.7d] [Medline: 15472114]
5. Houwink A, Nijland RH, Geurts AC, Kwakkel G. Functional recovery of the paretic upper limb after stroke: who regains hand capacity? Arch Phys Med Rehabil 2013 May;94(5):839-844. [doi: 10.1016/j.apmr.2012.11.031] [Medline: 23201317]
6. Hendricks HT, van Limbeek J, Geurts AC, Zwarts MJ. Motor recovery after stroke: A systematic review of the literature. Archives of Physical Medicine and Rehabilitation 2002 Nov;83(11):1629-1637. [doi: 10.1053/apmr.2002.35473]
7. Morris JH, van Wijck F, Joice S, Donaghy M. Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. Disabil Rehabil 2013 Feb;35(4):291-299. [doi: 10.3109/09638288.2012.691942] [Medline: 22691224]
8. Wang YC, Kapellusch J, Garg A. Important factors influencing the return to work after stroke. Work 2014;47(4):553-559. [doi: 10.3233/WOR-131627] [Medline: 23531585]
9. Frank M, Conzelmann M, Engelter S. Prediction of Discharge Destination after Neurological Rehabilitation in Stroke Patients. Eur Neurol 2010;63(4):227-233. [doi: 10.1159/000279491]
10. Faria-Fortini I, Michaelsen SM, Cassiano JG, Teixeira-Salmela LF. Upper Extremity Function in Stroke Subjects: Relationships between the International Classification of Functioning, Disability, and Health Domains. Journal of Hand Therapy 2011 Jul;24(3):257-265. [doi: 10.1016/j.jht.2011.01.002]

11. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. PLoS One 2014 Feb;9(2):e87987 [FREE Full text] [doi: 10.1371/journal.pone.0087987] [Medline: 24505342]
12. Schneider EJ, Lannin NA, Ada L, Schmidt J. Increasing the amount of usual rehabilitation improves activity after stroke: a systematic review. J Physiother 2016 Oct;62(4):182-187 [FREE Full text] [doi: 10.1016/j.jphys.2016.08.006] [Medline: 27637769]
13. Lohse KR, Lang CE, Boyd LA. Is More Better? Using Metadata to Explore Dose–Response Relationships in Stroke Rehabilitation. Stroke 2014 Jul;45(7):2053-2058. [doi: 10.1161/strokeaha.114.004695]
14. Wattchow KA, McDonnell MN, Hillier SL. Rehabilitation Interventions for Upper Limb Function in the First Four Weeks Following Stroke: A Systematic Review and Meta-Analysis of the Evidence. Archives of Physical Medicine and Rehabilitation 2018 Feb;99(2):367-382. [doi: 10.1016/j.apmr.2017.06.014]
15. Hayward KS, Brauer SG. Dose of arm activity training during acute and subacute rehabilitation post stroke: a systematic review of the literature. Clin Rehabil 2015 Dec;29(12):1234-1243. [doi: 10.1177/0269215514565395] [Medline: 25568073]
16. McLaren R, Signal N, Lord S, Taylor S, Henderson J, Taylor D. The volume and timing of upper limb movement in acute stroke rehabilitation: still room for improvement. Disability and Rehabilitation 2019 Apr 05:1-6. [doi: 10.1080/09638288.2019.1590471]
17. Crosbie JH, McNeill MDJ, Burke J, McDonough S. Utilising technology for rehabilitation of the upper limb following stroke: the Ulster experience. Physical Therapy Reviews 2013 Jul 19;14(5):336-347. [doi: 10.1179/108331909x12540993897892]
18. Langan J, Subryan H, Nwogu I, Cavuoto L. Reported use of technology in stroke rehabilitation by physical and occupational therapists. Disability and Rehabilitation: Assistive Technology 2017 Aug 16;13(7):641-647. [doi: 10.1080/17483107.2017.1362043] [Medline: 28812386]
19. Lam MY, Tatla SK, Lohse KR, Shirzad N, Hoens AM, Miller KJ, et al. Perceptions of Technology and Its Use for Therapeutic Application for Individuals With Hemiparesis: Findings From Adult and Pediatric Focus Groups. JMIR Rehabil Assist Technol 2015 Feb 10;2(1):e1 [FREE Full text] [doi: 10.2196/rehab.3484] [Medline: 28582236]
20. Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. J NeuroEngineering Rehabil 2018 Mar 27;15(1). [doi: 10.1186/s12984-018-0370-2]
21. Veerbeek JM, Langbroek-Amersfoort AC, van Wegen EEH, Meskers CGM, Kwakkel G. Effects of Robot-Assisted Therapy for the Upper Limb After Stroke. Neurorehabil Neural Repair 2016 Sep 24;31(2):107-121. [doi: 10.1177/1545968316666957]
22. Chen CC, Bode RK. Factors Influencing Therapists' Decision-Making in the Acceptance of New Technology Devices in Stroke Rehabilitation. American Journal of Physical Medicine & Rehabilitation 2011;90(5):415-425. [doi: 10.1097/phm.0b013e318214f5d8]
23. Scott K, Signal N, Taylor D, Kayes N. Clinician's perspectives on what influences uptake of new technologies into rehabilitation practice. In: Design4Health. 2018 Presented at: 5th International Conference on Design4Health; 4th-6th September 2018; Sheffield, England URL: https://research.shu.ac.uk/design4health/wp-content/uploads/2018/09/664.pdf
24. Lemke M, Rodríguez Ramírez E, Robinson B, Signal N. Motivators and barriers to using information and communication technology in everyday life following stroke: a qualitative and video observation study. Disabil Rehabil 2019 Jan 27:1-9. [doi: 10.1080/09638288.2018.1543460] [Medline: 30686063]
25. Seo NJ, Arun KJ, Hur P, Crocher V, Motawar B, Lakshminarayanan K. Usability evaluation of low-cost virtual reality hand and arm rehabilitation games. J Rehabil Res Dev 2016;53(3):321-334 [FREE Full text] [doi: 10.1682/JRRD.2015.03.0045] [Medline: 27271199]
26. Kanai M, Izawa KP, Kobayashi M, Onishi A, Kubo H, Nozoe M, et al. Effect of accelerometer-based feedback on physical activity in hospitalized patients with ischemic stroke: a randomized controlled trial. Clin Rehabil 2018 Feb 05;32(8):1047-1056. [doi: 10.1177/0269215518755841]
27. Grau-Pellicer M, Lalanza J, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Topics in Stroke Rehabilitation 2019 Dec 02;27(5):354-368. [doi: 10.1080/10749357.2019.1691816]
28. Atkins A, Cannell J, Barr C. Pedometers alone do not increase mobility in inpatient rehabilitation: a randomized controlled trial. Clin Rehabil 2019 Apr 08;33(8):1382-1390. [doi: 10.1177/0269215519838312]
29. Hassett L, van den Berg M, Lindley RI, Crotty M, McCluskey A, van der Ploeg HP, et al. Digitally enabled aged care and neurological rehabilitation to enhance outcomes with Activity and MObility UsiNg Technology (AMOUNT) in Australia: A randomised controlled trial. PLoS Med 2020 Feb 18;17(2):e1003029. [doi: 10.1371/journal.pmed.1003029]
30. Da-Silva RH, van Wijck F, Shaw L, Rodgers H, Balaam M, Brkic L, et al. Prompting arm activity after stroke: A clinical proof of concept study of wrist-worn accelerometers with a vibrating alert function. J Rehabil Assist Technol Eng 2018;5:2055668318761524 [FREE Full text] [doi: 10.1177/2055668318761524] [Medline: 31191927]
31. Da-Silva RH, Moore SA, Rodgers H, Shaw L, Sutcliffe L, van Wijck F, et al. Wristband Accelerometers to motiVate arm Exercises after Stroke (WAVES): a pilot randomized controlled trial. Clin Rehabil 2019 Mar 07;33(8):1391-1403. [doi: 10.1177/0269215519834720]

32. Wei WXJ, Fong KNK, Chung RCK, Myint JMWW, Cheung HKY, Chow ESL. Utility of a unilateral accelerometer for monitoring upper extremity use in subacute stroke patients after discharge from hospital. Assistive Technology 2018 Jan 25;31(4):193-198. [doi: 10.1080/10400435.2017.1414085]
33. Chin LF, Hayward KS, Soh AJA, Tan CM, Wong CJR, Loh JW, et al. An accelerometry and observational study to quantify upper limb use after stroke during inpatient rehabilitation. Physiother Res Int 2019 May 16;24(4). [doi: 10.1002/pri.1784]
34. Parker J, Powell L, Heller B, Schweiss E, Mawson S. Upper limb activity in chronic post-stroke survivors: A comparison of accelerometry data with the Action Research Arm Test (ARAT). Annals of Physical and Rehabilitation Medicine 2018 Jul;61:e187. [doi: 10.1016/j.rehab.2018.05.430]
35. 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]
36. Rupp MA, Michaelis JR, McConnell DS, Smither JA. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. Applied Ergonomics 2018 Jul;70:77-87. [doi: 10.1016/j.apergo.2018.02.005]
37. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
38. Orme MW, Weedon AE, Saukko PM, Esliger DW, Morgan MD, Steiner MC, et al. Findings of the Chronic Obstructive Pulmonary Disease-Sitting and Exacerbations Trial (COPD-SEAT) in Reducing Sedentary Time Using Wearable and Mobile Technologies With Educational Support: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e84. [doi: 10.2196/mhealth.9398]
39. Sloan RA, Kim Y, Sahasranaman A, Müller-Riemenschneider F, Biddle SJH, Finkelstein EA. The influence of a consumer-wearable activity tracker on sedentary time and prolonged sedentary bouts: secondary analysis of a randomized controlled trial. BMC Res Notes 2018 Mar 22;11(1). [doi: 10.1186/s13104-018-3306-9]
40. Finn L, Ramasamy R, Dukes C, Scott J. Using WatchMinder to Increase the On-Task Behavior of Students with Autism Spectrum Disorder. J Autism Dev Disord 2015 May;45(5):1408-1418. [doi: 10.1007/s10803-014-2300-x] [Medline: 25377769]
41. Dvorkin AY, Ramaiya M, Larson EB, Zollman FS, Hsu N, Pacini S, et al. A "virtually minimal" visuo-haptic training of attention in severe traumatic brain injury. J NeuroEngineering Rehabil 2013;10(1):92. [doi: 10.1186/1743-0003-10-92]
42. Ezeugwu VE, Manns PJ. The Feasibility and Longitudinal Effects of a Home-Based Sedentary Behavior Change Intervention After Stroke. Archives of Physical Medicine and Rehabilitation 2018 Dec;99(12):2540-2547. [doi: 10.1016/j.apmr.2018.06.014]
43. Fong KNK, Yang NYH, Chan MKL, Chan DYL, Lau AFC, Chan DYW, et al. Combined effects of sensory cueing and limb activation on unilateral neglect in subacute left hemiplegic stroke patients: a randomized controlled pilot study. Clin Rehabil 2013 Jul;27(7):628-637. [doi: 10.1177/0269215512471959] [Medline: 23405025]
44. Bamford J, Sandercock P, Dennis M, Warlow C, Burn J. Classification and natural history of clinically identifiable subtypes of cerebral infarction. The Lancet 1991 Jun;337(8756):1521-1526. [doi: 10.1016/0140-6736(91)93206-o]
45. Bernhardt J, Chan J, Nicola I, Collier JM. Little therapy, little physical activity: rehabilitation within the first 14 days of organized stroke unit care. J Rehabil Med 2007 Jan;39(1):43-48 [FREE Full text] [doi: 10.2340/16501977-0013] [Medline: 17225037]
46. Bernhardt J, Chitravas N, Meslo IL, Thrift AG, Indredavik B. Not all stroke units are the same: a comparison of physical activity patterns in Melbourne, Australia, and Trondheim, Norway. Stroke 2008 Jul;39(7):2059-2065. [doi: 10.1161/STROKEAHA.107.507160] [Medline: 18451351]
47. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J. Stat. Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
48. R Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
49. Hayward KS, Eng JJ, Boyd LA, Lakhani B, Bernhardt J, Lang CE. Exploring the Role of Accelerometers in the Measurement of Real World Upper-Limb Use After Stroke. Brain Impairment 2015 Nov 10;17(1):16-33. [doi: 10.1017/brimp.2015.21]
50. Bayley MT, Hurdowar A, Richards CL, Korner-Bitensky N, Wood-Dauphinee S, Eng JJ, et al. Barriers to implementation of stroke rehabilitation evidence: findings from a multi-site pilot project. Disability and Rehabilitation 2012 May 28;34(19):1633-1638. [doi: 10.3109/09638288.2012.656790]
51. Dobkin BH, Plummer-D'Amato P, Elashoff R, Lee J, SIRROWS Group. International randomized clinical trial, stroke inpatient rehabilitation with reinforcement of walking speed (SIRROWS), improves outcomes. Neurorehabil Neural Repair 2010;24(3):235-242 [FREE Full text] [doi: 10.1177/1545968309357558] [Medline: 20164411]
52. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil 2017 Dec 11;14(1):20 [FREE Full text] [doi: 10.1186/s12984-017-0229-y] [Medline: 28284228]
53. Dunton GF. Ecological Momentary Assessment in Physical Activity Research. Exerc Sport Sci Rev 2017 Dec;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]

54. Jochumsen M, Rovsing C, Rovsing H, Cremoux S, Signal N, Allen K, et al. Quantification of Movement-Related EEG Correlates Associated with Motor Training: A Study on Movement-Related Cortical Potentials and Sensorimotor Rhythms. Front Hum Neurosci 2017;11:604 [FREE Full text] [doi: 10.3389/fnhum.2017.00604] [Medline: 29375337]
55. 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]
56. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive Technology in Mobile Applications Promoting Physical Activity: a Systematic Review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
57. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior Change Techniques Present in Wearable Activity Trackers: A Critical Analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
58. Abedtash H, Holden R. Systematic review of the effectiveness of health-related behavioral interventions using portable activity sensing devices (PASDs). J Am Med Inform Assoc 2017 Sep 01;24(5):1002-1013 [FREE Full text] [doi: 10.1093/jamia/ocx006] [Medline: 28339620]
59. 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]
60. Micallef N, Baillie L, Uzor S. Time to Exercise! An aide memoire stroke app for post stroke arm rehabilitation. In: Proceedings of 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2016 Presented at: 8th International Conference on Human-Computer Interaction with Mobile Devices and Services; 2016; Florence, Italy p. 6-9.
61. Held JPO, Luft AR, Veerbeek JM. Encouragement-Induced Real-World Upper Limb Use after Stroke by a Tracking and Feedback Device: A Study Protocol for a Multi-Center, Assessor-Blinded, Randomized Controlled Trial. Front. Neurol 2018 Jan 25;9. [doi: 10.3389/fneur.2018.00013]
62. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit Use and Activity Levels Throughout a Physical Activity Intervention: Exploratory Analysis from a Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Feb 05;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
63. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]

Abbreviations

AU: unilateral affected upper limb movement.
AUL: affected upper limb.
BiL: bilateral limb movement.
BiM: bimanual movement.
LACS-H: lacunar circulation syndrome hemorrhagic.
LACS-I: lacunar circulation syndrome ischemic.
OR: odds ratio.
PACS-H: partial anterior circulation syndrome hemorrhagic.
PACS-I: partial anterior circulation syndrome ischemic.
POCS-I: posterior circulation syndrome ischemic.
TACS-I: total anterior circulation syndrome ischemic.
UL: upper limb.

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by M Whelan, K Ng, N Cecilia; comments to author 10.01.20; revised version received 15.03.20; accepted 13.05.20; published 29.07.20.

Please cite as:
Signal NEJ, McLaren R, Rashid U, Vandal A, King M, Almesfer F, Henderson J, Taylor D
Haptic Nudges Increase Affected Upper Limb Movement During Inpatient Stroke Rehabilitation: Multiple-Period Randomized Crossover Study
JMIR Mhealth Uhealth 2020;8(7):e17036
URL: https://mhealth.jmir.org/2020/7/e17036
doi:10.2196/17036
PMID:32723718

Original Paper

Patient Experiences of Rehabilitation and the Potential for an mHealth System with Biofeedback After Breast Cancer Surgery: Qualitative Study

Louise Brennan[1,2,3], BSc; Threase Kessie[4], PhD, BA; Brian Caulfield[2,3], BSc, MSc, PhD

[1]Physiotherapy Department, Beacon Hospital, Dublin, Ireland

[2]Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland

[3]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

[4]Maynooth University Innovation Lab, Maynooth University, Kildare, Ireland

Corresponding Author:
Louise Brennan, BSc
Physiotherapy Department
Beacon Hospital
Sandyford Business Park
Dublin
Ireland
Phone: 353 851437364
Email: louise.brennan@ucdconnect.ie

Abstract

Background: Physiotherapy-led home rehabilitation after breast cancer surgery can protect against the development of upper limb dysfunction and other disabling consequences of surgery. A variety of barriers can limit physical rehabilitation outcomes, and patients may benefit from more support during this time. Mobile health (mHealth) systems can assist patients during rehabilitation by providing exercise support, biofeedback, and information. Before designing mHealth systems for a specific population, developers must first engage with users to understand their experiences and needs.

Objective: The aims of this study were to explore patients' rehabilitation experiences and unmet needs during home rehabilitation after breast cancer surgery and to understand their experiences of mHealth technology and the requirements they desire from an mHealth system.

Methods: This was the first stage of a user-centered design process for an mHealth system. We interviewed 10 breast cancer survivors under the two main topics of "Rehabilitation" and "Technology" and performed a thematic analysis on the interview data.

Results: Discussions regarding rehabilitation focused on the acute and long-term consequences of surgery; unmet needs and lack of support; self-driven rehabilitation; and visions for high-quality rehabilitation. Regarding technology, participants reported a lack of mHealth options for this clinical context and using non-cancer–specific applications and wearables. Participants requested an mHealth tool from a reliable source that provides exercise support.

Conclusions: There are unmet needs surrounding access to physiotherapy, information, and support during home rehabilitation after breast cancer surgery that could be addressed with an mHealth system. Breast cancer survivors are open to using an mHealth system and require that it comes from a reliable source and focuses on supporting exercise performance.

(JMIR Mhealth Uhealth 2020;8(7):e19721) doi:10.2196/19721

KEYWORDS

breast cancer; physiotherapy; rehabilitation; mHealth; biofeedback; user-centred design; cancer

Introduction

Breast cancer accounts for 1 in 4 cancer diagnoses in women in Europe [1]. A pattern of decreased mortality rates and increased incidence means that there is an ever-increasing number of women living with and beyond a breast cancer diagnosis [2]. The adverse side effects of breast cancer treatment can last for years after treatment finishes, limiting survivors'

quality of life and participation in activities of daily living [3,4]. Addressing the consequences of treatment in this growing population through rehabilitation is an essential component of the holistic management of this disease and an important goal for researchers and clinicians [5,6].

Upper limb dysfunction is a prevalent, persistent, and disabling consequence of breast cancer treatment, which is reported by up to 62% of women 6 years after treatment [3,7,8]. Symptoms such as shoulder girdle pain, weakness, reduced range of movement (ROM), and lymphedema are specifically associated with surgery for breast cancer [8-10]. Physiotherapy-led rehabilitation is recommended after breast cancer surgery to prevent development of upper limb dysfunction, to assist a return to full function, and to promote self-management and maintenance of healthy lifestyle behaviors [11,12]. A key component of physiotherapy rehabilitation is the home exercise program (HEP), which is usually prescribed to patients on the first day after surgery. Once discharged from hospital, the patient is expected to continue doing this rehabilitation exercise program daily at home without supervision from a physiotherapist, a process referred to as "home rehabilitation." Performing exercise programs unsupervised at home can be challenging, and in the general population adherence rates to home exercise are as low as 65% [13]. In a breast cancer population there are additional barriers to rehabilitation to consider, such as the incapacitating side effects of cancer treatment, hospital admissions, and difficulty accessing physiotherapy in hospital or on discharge [14,15]. A large US-based survey by Reigle and Zhang [16] found that 61% of breast cancer survivors were instructed to perform upper body exercises after their surgery and only 28% received these instructions from a physiotherapist. Accessing a physiotherapist with experience in managing breast cancer–related upper limb impairments can be difficult, due to the low numbers of physiotherapists within this specialty [17]. Provision of a preoperative review and postoperative follow-up care for up to 1 year after surgery with long-term ongoing surveillance is recommended to optimize physical and functional well-being after breast cancer surgery [5,18,19]. Although access to physiotherapy services for breast cancer surgery varies between and within nations, it rarely meets these recommendations [20-22]. Considering the above barriers to rehabilitation, many women may benefit from more support after surgery to optimize their rehabilitation and limit the development of upper limb dysfunction.

The Global Observatory for electronic health (eHealth) defines mobile health (mHealth) as "medical and public health practice supported by mobile devices…and other wireless devices" [23]. MHealth apps can effectively support patients after surgery [24-27]. There is a promising role for mHealth apps and wearable sensors to enhance breast cancer rehabilitation by acting as motivational tools, to measure exercise interventions, and to support self-management [28-30]. Interventions that combine an app with a wearable external sensor can collect biomechanical or physiological data and feed this back to the user, allowing them to alter their exercise performance if needed; this process is called biofeedback [31,32]. Biofeedback during exercise is educational and motivational, can improve rehabilitation quality and safety, and can facilitate patient engagement [33-35]. Biofeedback mHealth systems can enable the collection of objective, longitudinal, and real-world data related to exercise performance in the home environment [30,36]. MHealth and biofeedback technologies have had an impact in several areas of breast cancer care, such as symptom management [37], screening [38], and general survivorship care [39]. However, these technologies are still in the very early stages of development and adoption in breast cancer rehabilitation. To date, innovations have mainly focused on physical activity interventions [40-42], lymphedema management [28], psychological well-being [43], and general survivorship care [29,44]. Very few interventions concentrate on physiotherapy rehabilitation exercises. There is a need to address this gap by developing an mHealth system that contains biofeedback features to support patients during physiotherapy rehabilitation in breast cancer.

To design an effective mHealth system, developers must first understand the experiences and unmet needs of the target population [45,46]. Unmet needs occur with a high incidence across the entire spectrum of breast cancer care [47,48]. Although some studies have investigated the unmet needs of patients within specific aspects of care, such as information needs [49,50] or psychosocial needs [51], to date, little is known about unmet needs during physiotherapy rehabilitation, despite its importance in reducing the adverse side effects of surgery. It is also essential in the early stages of digital technology development to understand the perspectives and experiences of the user in regard to the proposed technology [52,53]. Several studies have investigated breast cancer survivors' attitudes to mHealth for general physical activity [54,55]; however, there is a need to understand their attitudes toward, experiences of, and requirements from mHealth technology for physiotherapy-led breast cancer home rehabilitation.

The aim of this study was therefore to explore the experiences and unmet needs of women during home rehabilitation following surgery for breast cancer and to gather survivors' perspectives on and requirements from mHealth technology for postoperative breast cancer rehabilitation.

Methods

Study Design

A qualitative research study design using semistructured interviews with breast cancer survivors was undertaken to explore their experiences and needs regarding rehabilitation and technology. This methodology was chosen to allow a detailed description of experiences and opinions, which can produce rich, detailed, and complex data, even when collected from a small number of respondents [56,57]. The study was conducted and reported according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist for qualitative studies [58]. Ethics approval for this study was granted by University College Dublin Human Research Ethics Committee. All participants provided written informed consent to take part in this study.

Participants and Recruitment

Participants for this study were women with a history of breast cancer living in the wider Dublin area, who were recruited from a sports club whose members are all survivors of breast cancer. This sample was chosen to ensure the data collected represented the general perspectives of breast cancer survivors and was not influenced by where they received cancer treatment. It was important to capture how consequences of surgery and postoperative rehabilitation may have impacted participants in the medium- and long-term [5]. Therefore, participants were individuals who had completed their treatment and could reflect on their experiences across and beyond the entire treatment pathway. Participants were required to be 18 years of age or older and to have had surgery for breast cancer within the last 5 years. This time limit served to ensure that current rehabilitation practices were represented, and to improve reliability of experience recollection. The lead researcher was introduced to the club members by a physiotherapist who is professionally associated with the club. Individuals interested in participating contacted the lead researcher, at which point they were screened for inclusion suitability. A sample size of 10 participants was chosen. This is in line with recommendations by Guest et al [59], who found that data saturation occurred within 12 interviews and meta-themes could be detected at 6 interviews. Other qualitative literature with breast cancer survivors used a similar number of participants [55,60,61].

Data Collection

A literature review of unmet needs and technology in breast cancer rehabilitation, as discussed in the introduction to this study, informed the development of a flexible, semistructured interview guide, which is presented in Textbox 1. The interviews, which were audio-recorded, contained two main topics with two distinct aims: the first topic (questions 1.1-1.7) aimed to explore rehabilitation experiences and the second topic (questions 2.1-2.3) aimed to explore perspectives on, and opportunities for, mHealth during rehabilitation. The interview guide was developed in consultation with a postdoctoral researcher with expertise in qualitative data collection and analysis. Each participant had a one-on-one semistructured interview with the lead researcher (LB), who is a female physiotherapist and PhD researcher. Interviews lasted approximately 45 minutes and took place in a private room in the sports club, the participant's home, or a similar private venue between April and May 2019.

Textbox 1. Semistructured interview guide used in this study.

Topic 1: Rehabilitation experiences

- 1.1 Please talk me through your process of rehabilitation, from the time of the surgery until now.
- 1.2 Can you describe your awareness of how the operation would impact you, both short and long term?
- 1.3 Tell me about a time throughout the rehabilitation process that you found difficult.
- 1.4 What were your expectations of the physiotherapy care you would receive? How did they match the reality?
- 1.5 How well-equipped did you feel to do the postoperative rehabilitation at home? How confident did you feel doing the exercises?
- 1.6 Were you given exercise instructions and information? How?
- 1.7 Is there anything that might have helped you during rehabilitation after the surgery?

Topic 2: Digital technologies in breast cancer rehabilitation

- 2.1 Did you use any modern technologies to help you during your rehabilitation?
- 2.2 Do you use any technologies for your general health and well-being?
- 2.3 When discussing your experiences of physiotherapy, you mentioned [insert problem]. If there was a piece of technology to help you with this issue, what would it do?

Data Analysis

The interviews were transcribed verbatim and anonymized transcripts were imported into the qualitative data analysis management software NVivo (QSR International). The lead researcher applied a thematic analysis with a semantic, mixed inductive and deductive approach to analyze the data, following the process outlined by Braun and Clarke [56]. After a data familiarization stage, the data was categorized into codes, which contained conceptually similar ideas or actions. Applying an iterative process, the relationships between codes were analyzed and themes were formed by grouping related codes together. A second researcher who specializes in qualitative data analysis (TK) applied the same inductive thematic analysis process to review and code extracts from 7 transcripts. The two researchers compared codes and themes, and resolved any differences in coding through discussion. Data saturation was determined when no new themes and relationships among the interview data were found [62]. The lead researcher then performed a final revision of the thematic analysis to ensure that the themes accurately reflected the content of the data set and to screen for internal and external heterogeneity of themes [63]. Final themes were agreed, defined, and named by the two researchers. Illustrative quotes for each code and theme were identified. Participants did not receive a copy of the transcript or provide feedback on the findings.

Results

Participant and Surgery Characteristics

In total, 10 women participated in the semistructured interviews. Participant characteristics are summarized in Table 1. Information on the participants' surgeries and physiotherapy received are described in Table 2. Postoperative physiotherapy services mainly consisted of a single postoperative assessment, which for the purposes of this study will be defined as a single visit by the physiotherapist to review the patient, teach postoperative exercises, and assist in mobility or other needs. Those who had physiotherapy arranged as a follow-up (n=2) or a later referral (n=2) reported the experience to be very helpful and educational, with supportive and highly skilled physiotherapists.

Table 1. Participant characteristics.

Participant characteristic	Participants, n
Female	10
Age (years)	
35-44	1
45-54	6
55-64	2
65-74	1
Ethnicity	
White (Irish)	10
Highest level of education	
Trade/technical/vocational training	1
Diploma	1
Some university	1
Bachelor's degree	3
Master's degree	3
Doctorate	1
Employment status	
Unable to work	2
Employed	5
Self-employed	1
Temporary retirement	1
Retired	1
Approximate annual household income (€)	
30,000-50,000	3
50,000-75,000	4
75,000-100,000	2
>100,000	1
Marital status	
Single	1
Married or domestic partnership	7
Separated or divorced	2

Table 2. Characteristics of breast cancer surgery for each participant[a].

Type of breast surgery	Physiotherapy received after surgery	Extent of axillary surgery	Time since surgery (years)	Public or private health care
Participant 1				
Bilateral mastectomy	Postoperative assessment	Axillary clearance	4	Public
Axillary clearance				
Participant 2				
Lumpectomy	Postoperative assessment	Axillary clearance	2.4	Private
Lumpectomy and axillary clearance				
Participant 3				
Axillary clearance	Unknown[b]	Axillary clearance	1.2	Public patient in private hospital
Mastectomy	Unknown			
Reconstruction (type unknown)	Seen daily in hospital			
Participant 4				
SLNB[c] and lumpectomy	No	Axillary clearance	4	Private
Axillary clearance	Participant unsure			
Mastectomy	Postoperative assessment			
DIEP[d] flap reconstruction	Postoperative assessment and out-patient follow-up; later referred by oncologist			
Participant 5				
Bilateral mastectomy	Postoperative assessment and later self-referral	Axillary clearance	3.4	Public patient in private hospital
Participant 6				
SLNB	Postoperative assessment	SLNB	2.5	Private
Lumpectomy				
Participant 7				
Mastectomy with reconstruction	Postoperative assessment	Unknown	4	Public
Participant 8				
Mastectomy with DIEP flap reconstruction	Postoperative assessment	Unknown	2.2	Private
Participant 9				
Mastectomy with implant reconstruction	Postoperative assessment	Unknown	1.5	Private
Participant 10				
SLNB	Postoperative assessment and out-patient follow-up	Axillary clearance	1	Private
Mastectomy and axillary clearance				

[a]To preserve anonymity, participants are not listed in order and participant numbers do not correspond to those used throughout the manuscript.

[b]"Unknown" refers to missing data that was not gathered during the interviews [64].

[c]SNLB: sentinel lymph node biopsy.

[d]DIEP: deep inferior epigastric perforators.

Topic One: Rehabilitation Experiences

Under the topic of "rehabilitation experiences," four main themes were identified: acute and long-term consequences of surgery; unmet needs and lack of support; self-driven rehabilitation; and visions for high quality rehabilitation.

Theme 1: Acute and Long-Term Consequences of Surgery

Participants experienced short-term and long-term consequences of the surgery, which impacted their function and quality of life. Long-term upper limb dysfunction was described by Participant (P) 6: "This arm is definitely still a bit weaker. If I go to pick things up, picking up groceries or whatever, I don't

have the same strength in this arm." As well as physical impairments of stiffness and weakness, participants experienced feelings of overwhelm, fear, and anxiety. The prospect of developing side effects of surgery could be worrying; P8 stated, "I was terrified of lymphedema." Most participants were surprised by how they felt emotionally after the surgery, and report not being prepared for what to expect:

> *I had no idea that there would be such life changing things going on in my body. You kind of think, you'll have your surgery…and then life will go back to normal, but it doesn't.* [P6]

Secondary musculoskeletal complications developed for several participants, such as frozen shoulder, bursitis, and postural problems. P2 reported how she could not move her upper limb into the required position to receive radiotherapy:

> *The first time around, I developed a frozen shoulder. This was very evident when I was going for my radiotherapy, I couldn't do it.*

Theme 2: Unmet Needs and Lack of Support

Participants were not always aware of the importance of physiotherapy and rehabilitation and several felt that there was a lack of emphasis on rehabilitation throughout their hospital stay:

> *I mean I was told, but it wasn't heavily enforced and it wasn't really stressed. I don't feel that it was high priority.* [P5]

As demonstrated in Table 2, most participants only had one assessment with the physiotherapist, with no follow-up appointment. P2 explained:

> *Your physiotherapist comes to you in the ward, goes through the exercises with you, but other than that, there isn't any real physio support, you do the exercises yourself.*

This was felt to be insufficient, especially if concerns arose after this assessment, as P6 expressed: "I didn't really feel like I had a physiotherapist that I could ring up and talk to." There were also challenges with how the HEP was provided. The sheet of paper containing descriptions of the exercises was carefully minded by some participants, but others, including P4, didn't take the same approach: "you don't know where you have left it half the time." Several participants felt that the physiotherapy assessment took place too soon after the surgery. P10 explained: "I was probably still on morphine, so I probably wasn't paying enough attention."

As a result of the above unmet needs, P1 reported feeling uncertain about how to perform the exercises:

> *I didn't even know whether I was doing it right or not, you know, sitting on the bed supposed to be doing this or this, or whatever, and I didn't bother then.* [P1]

Additionally, there were difficulties with the postoperative self-management advice that they had received from the hospital. The information was "very unstructured and informal" [P5] and two participants felt it was too focused on what not to do: "don't do this, don't do this; it wasn't reassuring to me in any sense" [P4]. Upon discharge from hospital, treatment side effects restricted participants' abilities to do the HEP; they reported feeling "quite weakened" from chemotherapy [P8], that radiotherapy "affected my skin and it made the stretching very sore" [P10] and that fatigue was "the biggest impediment to doing (the HEP)" [P4]. Financial barriers prevented two participants from arranging private physiotherapy appointments. Time and scheduling were additional barriers to home rehabilitation, as was a fear of movement:

> *I think a woman is naturally terrified of it and afraid that those stretch things are going to, you know, hurt, so, I think there is just a barrier.* [P8]

At a time when professional advice and support was much needed, participants felt alone and unsupported:

> *Certainly, post-surgery you feel quite isolated, whereas, in chemo you are meeting oncology nurses all the time, but post-surgery…you are on your own.* [P8]

Having little contact with health care professionals was especially difficult when participants experienced unexpected side effects, as P8 reported: "I found it quite shocking and terrifying…essentially I had nobody to turn to."

Theme 3: Self-Driven Rehabilitation

A desire and motivation to recover was expressed by all participants. Despite the unmet rehabilitation needs and barriers to home rehabilitation, 6 participants expressed they had had a strong internal motivation which led them to persevere with the HEP. P4 asserted: "if you say to me that this is what you need to do to get better, I'll do it." Another motivating factor was personal experience of upper limb dysfunction:

> *There is no question about it. I felt if I didn't stretch my arm…it would be very stiff and, and very, very sore.* [P6]

When participants did have follow-up physiotherapy appointments, this sense of accountability was an external motivating factor. Participants were proactive in fulfilling their own rehabilitation needs; for example, they sourced information, support and advice as required:

> *It's hard to do nothing, therefore whatever is at hand, you will look up, research and investigate or speak to…people like to feel like they are being pro-active about their own recovery.* [P5]

There was a strong role for peer support, and speaking to other individuals with breast cancer was an effective way to gain knowledge, advice and assistance:

> *It was very rarely involved interacting with professionals, it was almost exclusively peer to peer, asking other women, what to do about lymphoedema, what to do about, when to wear a sleeve?* [P5]

The holistic benefits of exercise were mentioned by all participants. Some participants tried new forms of exercise to address surgery-related upper body dysfunction, including yoga, Pilates, and going to the gym:

> *My posture was horrendous, I was just completely rounded, you know, to hide… So now, through the Pilates…my posture is greatly improved.* [P2]

Overall, 3 participants reported that they did not engage in exercise prior to surgery, but continued because they saw the benefits of the particular activity. P1 stated that, "There is a whole new me, from being a couch potato all my life, to doing this." Additionally, 8 participants discussed the social and psychological benefits of exercise, which ranged from having improved confidence and feeling calmer, to having the opportunity to get fresh air and make new friends. Regarding the sports club, participants spoke of finding "a great camaraderie" [P3] and of "meeting people who have been through something similar, you can all relate" [P4]. P10 expressed gratitude for the club and her experience as a member of that community:

> *I also think that was the silver lining to the whole experience 'cos now I have a whole life around that, and a whole lot of new friends, so that has been brilliant.*

Theme 4: Visions for High Quality Rehabilitation

Discussions of good experiences with health care professionals and recommendations for service improvements provided insights into patient preferences for breast cancer rehabilitation. Participants appreciated proactive care and patient education. P6 described a valuable interaction with her physiotherapist: "she wasn't forceful but she made it clear that if you don't do your exercises you will be left with less movement." All participants recommended more access to physiotherapy services, especially follow-up assessments after discharge from hospital:

> *I would have liked to, maybe, have gone through the exercises that I was trying to do, check that I was doing them properly.* [P3]

Several participants recommended providing patients with more information about what to expect after the surgery, regarding diet, exercise, and breast prostheses. One participant suggested this could be done through a physiotherapist-led preoperative information session, while another requested a postoperative support group:

> *A support group for people, for the after-effects of surgery and everything else, which would include the physiotherapist plus maybe other things as well.* [P6]

Topic Two: Technology

The experiences of the participants with mHealth technology and the features that participants would like to see in an mHealth tool for breast cancer rehabilitation were the two distinct themes from this portion of the interviews.

Theme 1: Experience with Digital Technology for Health

The digital technologies used by participants for health and rehabilitation are detailed in Table 3. None of the participants used a breast cancer–specific mHealth tool; instead, participants applied technology for the general population to meet their cancer rehabilitation needs. Only 1 participant did not use any digital technology for health, stating that, "I think here we are seriously looking at an old age thing" [P3].

Technology played two main roles: first, it acted alongside health care professionals, for example to help patients understand information provided by the hospital: "[The terminology] was all new language to me, so when I came home I did google all that" [P1]. Second, it compensated for unmet needs in health care provision, for example to provide exercise advice: "I would do Pilates, I would watch an app to do that" [P6]. Overall, 6 participants used activity monitors, and they perceived them as good motivators to be physically active, providing a sense of achievement when participants reached their goals. Some disadvantages of mHealth were expressed, such as feeling overwhelmed by the volume and content of information: "you can really live in cancer-land" [P6]. Concerns about data privacy were raised: "What happens if [your personal information] somehow gets into the wrong hands?" [P7]. Additionally, it was acknowledged that technology cannot replace the role of the human in health care:

> *As much as you want apps or websites, you also need the human bit, you need them all.* [P10]

Table 3. Participants' use of digital technology and mHealth.

Technology characteristic	Participants, n
Smartphone owner	
Yes	10
Uses smartphone apps or wearables in daily life	
Yes	10
Used mHealth tool designed for postoperative rehabilitation	
No	10
Used or uses digital technology for general cancer rehabilitation issues[a]	
Activity monitor: wearable	5
Activity monitor: smartphone	4
Guided exercise app	2
Mindfulness app	2
Cancer-themed podcast	1
Online breast cancer magazine	1
Nothing	1
Used or uses digital technology for health information seeking	
Yes	5
Yes, occasionally	2
Not sure	2
No	1

[a]Some participants used more than one digital technology for rehabilitation issues.

Theme 2: Requirements for an mHealth System

All participants stated that they would have liked to have tried out an mHealth system during home rehabilitation if such a system had been available. Throughout the discussion, participants suggested features they would like to see in such a system (Table 4). These features largely focused on exercise information and guidance. In total, 4 participants suggested having a visual aid in the form of a physiotherapist or a biofeedback avatar performing the exercises alongside them. They felt that an avatar should resemble someone who could have had breast cancer treatment, but not someone currently in treatment:

> *I don't think you would need anybody bald, you wouldn't want, necessarily, that reminder. Maybe not somebody with gorgeous locks, short hair would be fine.* [P5]

A desire for both audio and visual exercise biofeedback was expressed by P9: "If there is going to be an audio, having something visual as well, it's there if I wanted to check, 'Yeah, I am doing this correctly.'" It was important to many that the system should be provided from the hospital: "if it's through the hospital, then, you know, it's got that legitimacy" [P5]. A record of exercise performance was a popular feature, not only because it would provide feedback to patients, but because it would remotely provide feedback to the physiotherapist as well:

> *I would like to know that…somebody had taken an interest and a follow-up in what I am doing…and then they would be able to intervene if they saw I wasn't doing it properly.* [P1]

Regarding the style of the system, there was a strong preference for a "less is more" approach, with concise information that is specific to the patient's own situation:

> *Your concentration is shot after cancer when you are having treatment…You need to ask a question and get an answer.* [P2]

There was a preference toward content that is positive and life-affirming, delivered with an encouraging and reassuring tone:

> *I think maybe some encouragement because it's such a difficult time, not just 'do this,' but 'how are you finding it?' You know, 'how's your day?'* [P5]

Table 4. Desired features in an mHealth home rehabilitation system: participant suggestions and frequency of suggestion.

Feature	Suggestion frequency, n
Information on exercises	8
Record of exercise sessions	6
Reminders	5
Motivational messaging	4
Exercise videos	4
Personalization of content	3
Contact with physiotherapist	3
Interact with other patients	2
Postoperative information	2
Goal setting	2
Repetition counter	2
Exercise progression generator	1

Discussion

Principal Findings

This study presented detailed insights into the physical rehabilitation experiences and unmet needs of 10 women after surgery for breast cancer. It has highlighted the barriers to home rehabilitation and presents participants' visions for high-quality rehabilitation. Through this, the role of an mHealth system with biofeedback features to reduce barriers and improve quality of care can be seen. Additionally, this study explored patient preferences and requirements for such a system. Under the topic of "rehabilitation," themes covering the experiences of acute and long-term consequences of surgery; participants' unmet needs and feelings of lack of support; the motivation and need to take the lead in one's own rehabilitation; and participants' visions for high-quality rehabilitation were identified. Under the "technology" topic, the sample's experience with digital technology and mHealth were outlined, and key requirements for an mHealth home rehabilitation support system were identified.

Comparison with Prior Work

The findings regarding the acute and long-term consequences of breast cancer surgery were wholly in line with previous research. Throughout the "rehabilitation" section of the interview, participants reported a desire for increased access to physiotherapy services, both before and after surgery, and at later stages of treatment. Alongside clinical practice guidelines, other research with patient populations also supports these recommendations for routine preoperative and follow-up physiotherapy services [18,65,66]. Many participants had extensive breast surgery and axillary clearance (Table 2), which are risk factors for the development of lymphedema [67] and arm and shoulder impairments [68], indicating that this sample had substantial rehabilitation needs from the time of surgery onward. All participants experienced postoperative upper limb dysfunction, which for some resolved after the acute postoperative period, but, as reported by Ewertz et al [7], these symptoms could last long after treatment completion. As also described by Easley et al [69] in their analysis of interviews with young breast cancer survivors, at the time of treatment many participants were not made aware of rehabilitation services or the importance of rehabilitation. Easley et al stated that their participants also experienced a lack of support and had "no place to turn for help." In work by Harder et al [70] and Lee et al [71], participants often felt that the perioperative information they received was insufficient and poorly timed; participants in this study shared this perspective. The recurrence of these themes throughout literature that spans a decade suggests that improvements in accessing rehabilitation services are not happening quickly enough, and that innovative methods of improving access to care are needed.

Implications of Findings

The unmet needs described by participants in this study include a need to be educated in advance about the impact of the surgery and the importance of rehabilitation; a need for more physiotherapy services; a need for timely and clearly presented information regarding postoperative care; a need for more support on discharge and the ability to contact the appropriate health care professional when concerns arise. These unmet needs highlight areas in which improvements to care could be made, either through traditional service development methods or through the use of technology. An mHealth system, such as an app, could assist in these improvements by being a portable source of postoperative support by providing information, prescribed exercises, and a method of remote contact with a physiotherapist. Improved clinical outcomes have been demonstrated by combining a standard postoperative HEP with an mHealth app and biofeedback technology [35,72]. Additionally, Lambert et al [73] report that people with musculoskeletal conditions self-report adhering to their HEP better with an app than with a paper handout.

An important finding of this study was the participants' high levels of motivation to recover after surgery. Although adherence to HEPs has been found to be low in many populations [13,74,75], it is not necessarily the case in this population. Essery et al [76] report that self-motivation is a

positive predictor of adherence to home rehabilitation. Although many participants reported doing their exercises "religiously," there was uncertainty regarding the exercises and barriers to performing the HEP. This may have led participants to perform poor-quality rehabilitation or to stop doing the HEP prematurely. Improving motivation to exercise should therefore not be a primary goal of mHealth in this context; instead, the focus should be on improving the quality of rehabilitation.

No participants used mHealth technology for their postoperative rehabilitation. This reflects the lack of digital technology specifically designed for this period of breast cancer care and a lack of awareness of available apps, rather than a lack of desire to use mHealth, as all participants stated they would have used a suitable mHealth tool if it was available. Older age was a limiting factor for 1 participant for use of digital technology in daily life, but it was not a barrier to her desire to use mHealth. This is consistent with research by Abelson et al [77] which found that older adults are as willing as younger adults to engage with mHealth. Technologies with the capacity for objective, detailed in-clinic assessment are emerging, such as the Kinect (Microsoft Corp) reachable workspace system, as applied in a breast cancer population by Uhm et al [78], and a smartphone-based fatigue test developed by Cuesta-Vargas et al [79]. As a key stakeholder in these technologies, the openness of patients to the use of effective, helpful technologies is essential. After the acute stage of recovery, participants often used non–cancer-specific consumer technologies for physical activity and mental well-being. However, these may not meet the needs of individuals with breast cancer, as they do not consider fatigue, upper limb dysfunction, and other disease-specific considerations [80]. Zhou et al [81] developed a program using WeChat to provide physical, psychological, and social guidance after breast cancer surgery. Although this did not support the user through biofeedback features, it provided an opportunity to communicate remotely with a nurse or doctor from the breast surgery team at a specific time of day.

Overall, participants saw value in an mHealth tool to support patients during home rehabilitation after breast cancer surgery. They requested a system which focuses on providing exercise information, feedback on progress, and guidance in an audiovisual manner. This is echoed in a 2019 exploration of breast cancer survivors' preferences for mHealth physical activity interventions by Phillips et al [80], who found that exercise progress feedback and exercise scheduling were highly desirable features. Several of the suggested features for an mHealth system can be categorized as biofeedback features (record of exercise sessions, repetition counter, motivational messaging), and these can be enabled through the use of a wearable external sensor. This sensor data could also be applied to optimize additional feedback-related features (contact with physiotherapist, interact with other patients, goal setting, exercise progression generator). External sensors are increasingly being used in cancer care for both research and clinical purposes [30,82]. They can be facilitators to exercise, as found by Kokts-Porietis et al [55] who used wrist-worn activity monitors in a physical activity intervention with breast cancer survivors.

Finally, it was important to participants that an mHealth system would be provided by the hospital or a health care provider, and that it would include a remote connection with the physiotherapist. Chandra et al [83] found that physiotherapy patients mainly felt "motivation by support when their physiotherapist could view their exercise record, but were also unsure if physiotherapists would actually view it. Physiotherapists were cautiously positive about remote communications, but had reservations regarding increased workload. This feature is clearly valuable to the patient, but should be developed closely with the physiotherapist to ensure it is appropriate for all stakeholders. Additionally, participants were keen to retain some face-to-face contact with the physiotherapist, and therefore an mHealth system in this context should be provided by the physiotherapist as an adjunct to regular care.

Study Limitations

Breast cancer is a disease which affects 1 in 8 women across their lifetime, and therefore the experiences of women with breast cancer are incredibly varied and impossible to capture in a single study. The themes presented strongly across the participants and therefore appear to be highly representative of the population [62]. Specific limitations of our sample are that they were self-selecting and were actively engaged in a sports team after their treatment. Although participants were not all habitual exercisers prior to their diagnosis, they may have inherent positive attitudes toward exercise performance. Due to limited human resources, a full analysis of the transcripts was performed by only one author. A second reviewer analyzed sections of the transcripts to validate the lead researcher's analysis, and all authors reviewed and discussed the themes, codes, and quotes.

Conclusions and Future Work

An mHealth intervention has the potential to support patients during rehabilitation after breast cancer surgery. Understanding unmet needs of users and their perspectives on mHealth at an early phase of development enables the co-design of a user-friendly and relevant system. A thematic analysis of interviews with breast cancer survivors reflecting on their rehabilitation experiences demonstrates that, while participants were motivated to exercise, a lack of support on discharge and a lack of emphasis on rehabilitation led to uncertainty during rehabilitation and feelings of being alone and unsupported. Participants stated they would have welcomed an mHealth support tool that was reliable, provided by the hospital, and focused on rehabilitation support. The results of this analysis support a drive for improvements in breast cancer rehabilitation, which can be led by technological and nontechnological means. The impact of rehabilitation promotion and facilitation of access to postoperative physiotherapy services on rehabilitation outcomes should be investigated. These results provide motivation and rationale to develop an mHealth system with biofeedback that strives to improve outcomes for these often-undertreated consequences of breast cancer treatment.

Acknowledgments

LB is part of Cancer: Activating Technology for Connected Health (CATCH) ITN and is supported by a grant from the European Union's Horizon 2020 research and innovation program under a Marie Sklodowska-Curie grant (grant number 722012). The authors would like to thank all the participants for sharing their experiences and Ms Ailish Daly for her continued support throughout this work.

Authors' Contributions

LB conceived and designed the study, conducted the data collection, led the data analysis and interpretation, and drafted the manuscript. TK assisted with qualitative data analysis and interpretation, and with manuscript review. BC participated in study design and manuscript review. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. National Cancer Registry Ireland. Survival Statistics. National Cancer Registry Ireland. 2020. URL: https://www.ncri.ie/data/survival-statistics [accessed 2020-07-26]
3. Hayes SC, Johansson K, Stout NL, Prosnitz R, Armer JM, Gabram S, et al. Upper-body morbidity after breast cancer: incidence and evidence for evaluation, prevention, and management within a prospective surveillance model of care. Cancer 2012 Apr 15;118(8 Suppl):2237-2249 [FREE Full text] [doi: 10.1002/cncr.27467] [Medline: 22488698]
4. Brach M, Cieza A, Stucki G, Füssl M, Cole A, Ellerin B, et al. ICF Core Sets for breast cancer. J Rehabil Med 2004 Jul(44 Suppl):121-127 [FREE Full text] [doi: 10.1080/16501960410016811] [Medline: 15370759]
5. Stout NL, Binkley JM, Schmitz KH, Andrews K, Hayes SC, Campbell KL, et al. A prospective surveillance model for rehabilitation for women with breast cancer. Cancer 2012 Apr 15;118(8 Suppl):2191-2200 [FREE Full text] [doi: 10.1002/cncr.27476] [Medline: 22488693]
6. Macmillian Cancer Support. Throwing Light on the Consequences of Cancer and its Treatment. Macmillan Cancer Support. 2013. URL: https://www.macmillan.org.uk/documents/aboutus/research/researchandevaluationreports/throwinglightontheconsequencesofcanceranditstreatment.pdf [accessed 2020-07-26]
7. Ewertz M, Jensen AB. Late effects of breast cancer treatment and potentials for rehabilitation. Acta Oncologica 2011 Jan 13;50(2):187-193. [doi: 10.3109/0284186x.2010.533190]
8. Hidding JT, Beurskens CHG, van der Wees PJ, van Laarhoven HWM, Nijhuis-van der Sanden MWG. Treatment related impairments in arm and shoulder in patients with breast cancer: a systematic review. PLoS One 2014;9(5):e96748 [FREE Full text] [doi: 10.1371/journal.pone.0096748] [Medline: 24816774]
9. Lee TS, Kilbreath SL, Refshauge KM, Herbert RD, Beith JM. Prognosis of the upper limb following surgery and radiation for breast cancer. Breast Cancer Res Treat 2008 Jul;110(1):19-37. [doi: 10.1007/s10549-007-9710-9] [Medline: 17899373]
10. Stubblefield MD, Keole N. Upper body pain and functional disorders in patients with breast cancer. Phys Med Rehabil 2014 Feb 19;6(2):170-183. [doi: 10.1016/j.pmrj.2013.08.605] [Medline: 24360839]
11. National Institute for Healthcare and Excellence. Early and locally advanced breast cancer: diagnosis and management, NICE guideline. nice.org.uk. 2018 Jul. URL: https://www.nice.org.uk/guidance/ng101/resources/early-and-locally-advanced-breast-cancer-diagnosis-and-management-pdf-66141532913605 [accessed 2020-07-26]
12. McNeely M, Campbell K, Ospina M, Bh R, Dabbs K, Tp K, et al. Exercise interventions for upper-limb dysfunction due to breast cancer treatment (review). Cochrane Database of Systematic Reviews 2010. [doi: 10.1002/14651858.cd005211]
13. Bassett SF. The assessment of patient adherence to physiotherapy rehabilitation. New Zeal J Physiother 2003;31(2):66. [doi: 10.4172/2165-7025.1000e124]
14. Blaney J, Lowe-strong A, Rankin J, Campbell A, Allen J. The cancer rehabilitation journey: barriers to and facilitators of cancer-related fatigue. Physical Therapy 2010 Aug;90(8):1147. [doi: 10.2522/ptj.20090278]
15. Husebø AML, Karlsen B, Allan H, Søreide JA, Bru E. Factors perceived to influence exercise adherence in women with breast cancer participating in an exercise programme during adjuvant chemotherapy: a focus group study. J Clin Nurs 2015 Feb 03;24(3-4):500-510. [doi: 10.1111/jocn.12633] [Medline: 24890796]
16. Reigle BS, Zhang B. Women's Rehabilitation Experiences Following Breast Cancer Surgery. Rehabil Nurs 2018;43(4):195-200. [doi: 10.1097/rnj.0000000000000168] [Medline: 29957696]

17. Kenyon K, Hebron C, Vuoskoski P, McCrum C. Physiotherapists' experiences of managing upper limb movement impairments due to breast cancer treatment. Physiother Theory Pract 2020 Jan 07;36(1):71-84. [doi: 10.1080/09593985.2018.1480077] [Medline: 29877744]
18. Harris SR, Schmitz KH, Campbell KL, McNeely ML. Clinical practice guidelines for breast cancer rehabilitation: syntheses of guideline recommendations and qualitative appraisals. Cancer 2012 Apr 15;118(8 Suppl):2312-2324 [FREE Full text] [doi: 10.1002/cncr.27461] [Medline: 22488705]
19. Stout Gergich NL, Pfalzer LA, McGarvey C, Springer B, Gerber LH, Soballe P. Preoperative assessment enables the early diagnosis and successful treatment of lymphedema. Cancer 2008 Jun 15;112(12):2809-2819 [FREE Full text] [doi: 10.1002/cncr.23494] [Medline: 18428212]
20. Stubblefield MD. The Underutilization of Rehabilitation to Treat Physical Impairments in Breast Cancer Survivors. PM R 2017 Sep 21;9(9S2):S317-S323. [doi: 10.1016/j.pmrj.2017.05.010] [Medline: 28942906]
21. Mulcahy S, Prendergast J, Foley G, O'Hare A, Murphy E, Guinan EM, et al. Exercise Rehabilitation Services Provided by Physiotherapy Departments in Cancer Care in Ireland. Ir Med J 2018 Oct 11;111(9):818. [Medline: 30556666]
22. Canestraro A, Nakhle A, Stack M, Strong K, Wright A, Beauchamp M, et al. Oncology Rehabilitation Provision and Practice Patterns across Canada. Physiother Can 2013 Jan;65(1):94-102 [FREE Full text] [doi: 10.3138/ptc.2011-53] [Medline: 24381389]
23. WHO Global Conservatory for eHealth. New horizons for health through mobile technologies. Geneva: World Health Organisation; 2011.
24. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-Term Outcomes of Digital Versus Conventional Home-Based Rehabilitation After Total Knee Arthroplasty: Prospective, Parallel-Group Feasibility Study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]
25. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-Week Digital Care Program for Chronic Knee Pain on Pain, Mobility, and Surgery Risk: Randomized Controlled Trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
26. Dunphy E, Hamilton FL, Spasić I, Button K. Acceptability of a digital health intervention alongside physiotherapy to support patients following anterior cruciate ligament reconstruction. BMC Musculoskelet Disord 2017 Nov 21;18(1):471 [FREE Full text] [doi: 10.1186/s12891-017-1846-0] [Medline: 29162071]
27. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018 May;24(4):303-316. [doi: 10.1177/1357633X17697501] [Medline: 28350282]
28. Fu MR, Axelrod D, Guth AA, Rampertaap K, El-Shammaa N, Hiotis K, et al. mHealth self-care interventions: managing symptoms following breast cancer treatment. mHealth 2016 Jul 22;2:28-28 [FREE Full text] [doi: 10.21037/mhealth.2016.07.03] [Medline: 27493951]
29. Baseman J, Revere D, Baldwin L. A Mobile Breast Cancer Survivorship Care App: Pilot Study. JMIR Cancer 2017 Sep 26;3(2):e14 [FREE Full text] [doi: 10.2196/cancer.8192] [Medline: 28951383]
30. Beg MS, Gupta A, Stewart T, Rethorst CD. Promise of Wearable Physical Activity Monitors in Oncology Practice. J Oncol Pract 2017 Feb;13(2):82-89 [FREE Full text] [doi: 10.1200/JOP.2016.016857] [Medline: 28387544]
31. Giggins OM, Persson UM, Caulfield B. Biofeedback in rehabilitation. J Neuroeng Rehabil 2013 Jun 18;10:60 [FREE Full text] [doi: 10.1186/1743-0003-10-60] [Medline: 23777436]
32. Wang Q, Markopoulos P, Yu B, Chen W, Timmermans A. Interactive wearable systems for upper body rehabilitation: a systematic review. J Neuroeng Rehabil 2017 Dec 11;14(1):20 [FREE Full text] [doi: 10.1186/s12984-017-0229-y] [Medline: 28284228]
33. Lauber B, Keller M. Improving motor performance: selected aspects of augmented feedback in exercise and health. Eur J Sport Sci 2014;14(1):36-43. [doi: 10.1080/17461391.2012.725104] [Medline: 24533493]
34. Riel H, Matthews M, Vicenzino B, Bandholm T, Thorborg K, Rathleff MS. Feedback leads to better exercise quality in adolescents with patellofemoral pain. Med Sci Sports Exerc ? 2018;50(1):35. [doi: 10.1249/MSS.0000000000001412]
35. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Home-based Rehabilitation With A Novel Digital Biofeedback System versus Conventional In-person Rehabilitation after Total Knee Replacement: a feasibility study. Sci Rep 2018 Jul 26;8(1):11299 [FREE Full text] [doi: 10.1038/s41598-018-29668-0] [Medline: 30050087]
36. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital Care for Chronic Musculoskeletal Pain: 10,000 Participant Longitudinal Cohort Study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
37. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
38. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile Phone Multilevel and Multimedia Messaging Intervention for Breast Cancer Screening: Pilot Randomized Controlled Trial. JMIR mHealth uHealth 2017 Nov 07;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]

39. Tiong SS, Koh E, Delaney G, Lau A, Adams D, Bell V, et al. An e-health strategy to facilitate care of breast cancer survivors: A pilot study. Asia Pac J Clin Oncol 2016 Jun;12(2):181-187. [doi: 10.1111/ajco.12475] [Medline: 26935343]
40. Schaffer K, Panneerselvam N, Loh KP, Herrmann R, Kleckner IR, Dunne RF, et al. Systematic Review of Randomized Controlled Trials of Exercise Interventions Using Digital Activity Trackers in Patients With Cancer. J Natl Compr Canc Netw 2019 Jan 18;17(1):57-63 [FREE Full text] [doi: 10.6004/jnccn.2018.7082] [Medline: 30659130]
41. Lee H, Uhm KE, Cheong IY, Yoo JS, Chung SH, Park YH, et al. Patient Satisfaction with Mobile Health (mHealth) Application for Exercise Intervention in Breast Cancer Survivors. J Med Syst 2018 Nov 06;42(12):254. [doi: 10.1007/s10916-018-1096-1] [Medline: 30402781]
42. Dorri S, Asadi F, Olfatbakhsh A, Kazemi A. A Systematic Review of Electronic Health (eHealth) interventions to improve physical activity in patients with breast cancer. Breast Cancer 2020 Jan 12;27(1):25-46. [doi: 10.1007/s12282-019-00982-3] [Medline: 31187411]
43. Lally RM, Kupzyk KA, Bellavia G, Hydeman J, Gallo S, Helgeson VS, et al. CaringGuidance after breast cancer diagnosis eHealth psychoeducational intervention to reduce early post-diagnosis distress. Support Care Cancer 2020 May;28(5):2163-2174 [FREE Full text] [doi: 10.1007/s00520-019-05028-0] [Medline: 31414245]
44. Lozano-Lozano M, Martín-Martín L, Galiano-Castillo N, Álvarez-Salvago F, Cantarero-Villanueva I, Fernández-Lao C, et al. Integral strategy to supportive care in breast cancer survivors through occupational therapy and a m-health system: design of a randomized clinical trial. BMC Med Inform Decis Mak 2016 Nov 25;16(1):150 [FREE Full text] [doi: 10.1186/s12911-016-0394-0] [Medline: 27887610]
45. De Vito Dabbs A, Myers BA, McCurry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
46. Coleman R, Clarkson J, Dong H, Cassim J. Design for Inclusivity: A Practical Guide to Accessible, Innovative and User-Centred Design. Oxon, UK: Routledge; 2016.
47. von Heymann-Horan AB, Dalton SO, Dziekanska A, Christensen J, Andersen I, Mertz BG, et al. Unmet needs of women with breast cancer during and after primary treatment: A prospective study in Denmark. Acta Oncologica 2013 Jan 16;52(2):382-390. [doi: 10.3109/0284186x.2012.746468]
48. Capelan M, Battisti NML, McLoughlin A, Maidens V, Snuggs N, Slyk P, et al. The prevalence of unmet needs in 625 women living beyond a diagnosis of early breast cancer. Br J Cancer 2017 Oct 10;117(8):1113-1120 [FREE Full text] [doi: 10.1038/bjc.2017.283] [Medline: 28859057]
49. Halbach SM, Ernstmann N, Kowalski C, Pfaff H, Pförtner TK, Wesselmann S, et al. Unmet information needs and limited health literacy in newly diagnosed breast cancer patients over the course of cancer treatment. Patient Educ Couns 2016 Sep;99(9):1511-1518. [doi: 10.1016/j.pec.2016.06.028] [Medline: 27378079]
50. Sheehy EM, Lehane E, Quinn E, Livingstone V, Redmond HP, Corrigan MA. Information Needs of Patients With Breast Cancer at Years One, Three, and Five After Diagnosis. Clin Breast Cancer 2018 Dec;18(6):e1269-e1275. [doi: 10.1016/j.clbc.2018.06.007] [Medline: 30153977]
51. Martínez Arroyo O, Andreu Vaíllo Y, Martínez López P, Galdón Garrido MJ. Emotional distress and unmet supportive care needs in survivors of breast cancer beyond the end of primary treatment. Support Care Cancer 2019 Mar 9;27(3):1049-1057. [doi: 10.1007/s00520-018-4394-8] [Medline: 30094729]
52. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
53. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Key principles for user-centred systems design. Behaviour & Information Technology 2003 Nov;22(6):397-409. [doi: 10.1080/01449290310001624329]
54. Phillips SM, Conroy DE, Keadle SK, Pellegrini CA, Lloyd GR, Penedo FJ, et al. Breast cancer survivors' preferences for technology-supported exercise interventions. Support Care Cancer 2017 Dec;25(10):3243-3252 [FREE Full text] [doi: 10.1007/s00520-017-3735-3] [Medline: 28470368]
55. Kokts-Porietis RL, Stone CR, Friedenreich CM, Froese A, McDonough M, McNeil J. Breast cancer survivors' perspectives on a home-based physical activity intervention utilizing wearable technology. Support Care Cancer 2019 Aug;27(8):2885-2892. [doi: 10.1007/s00520-018-4581-7] [Medline: 30554276]
56. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
57. Crouch M, McKenzie H. The logic of small samples in interview-based qualitative research. Social Science Information 2016 Jun 29;45(4):483-499. [doi: 10.1177/0539018406069584]
58. 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]
59. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903] [Medline: 19638638]

60. Brunet J, Taran S, Burke S, Sabiston CM. A qualitative exploration of barriers and motivators to physical activity participation in women treated for breast cancer. Disabil Rehabil 2013 Jun 17;35(24):2038-2045 [FREE Full text] [doi: 10.3109/09638288.2013.802378] [Medline: 23772995]
61. Larsson IL, Jönsson C, Olsson AC, Gard G, Johansson K. Women's experience of physical activity following breast cancer treatment. Scand J Caring Sci 2008 Sep;22(3):422-429. [doi: 10.1111/j.1471-6712.2007.00546.x] [Medline: 18466196]
62. Fusch P, Ness L. Are we there yet? Data saturation in qualitative research. Qual Rep 2015;20(9):1409-1416.
63. Patton M. Qualitative Research and Evaluation Methods. In: 3rd ed. Thousand Oaks, CAage Publications. ISBN. Thousand Oaks, CA: Sage Publications; 2001.
64. Singh S, Richards L. Missing data: finding "central" themes in qualitative research. Qualitative Research 2003;3(1):17.
65. Binkley JM, Harris SR, Levangie PK, Pearl M, Guglielmino J, Kraus V, et al. Patient perspectives on breast cancer treatment side effects and the prospective surveillance model for physical rehabilitation for women with breast cancer. Cancer 2012 Apr 15;118(8 Suppl):2207-2216 [FREE Full text] [doi: 10.1002/cncr.27469] [Medline: 22488695]
66. Pidlyskyj K, Roddam H, Rawlinson G, Selfe J. Exploring aspects of physiotherapy care valued by breast cancer patients. Physiotherapy 2014 Jun;100(2):156-161. [doi: 10.1016/j.physio.2014.03.006] [Medline: 24780632]
67. Hayes SC, Janda M, Cornish B, Battistutta D, Newman B. Lymphedema After Breast Cancer: Incidence, Risk Factors, and Effect on Upper Body Function. JCO 2008 Jul 20;26(21):3536-3542. [doi: 10.1200/jco.2007.14.4899]
68. Nesvold I, Dahl AA, Løkkevik E, Marit Mengshoel A, Fosså SD. Arm and shoulder morbidity in breast cancer patients after breast-conserving therapy versus mastectomy. Acta Oncol 2008 Jul 08;47(5):835-842. [doi: 10.1080/02841860801961257] [Medline: 18568481]
69. Easley J, Miedema B. Rehabilitation after breast cancer: recommendations from young survivors. Rehabil Nurs 2012 Jun 20;37(4):163-170. [doi: 10.1002/rnj.21] [Medline: 22744987]
70. 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]
71. Lee TS, Kilbreath SL, Sullivan G, Refshauge KM, Beith JM. Patient Perceptions of Arm Care and Exercise Advice After Breast Cancer Surgery. Oncology Nursing Forum 2009 Dec 31;37(1):85-91. [doi: 10.1188/10.onf.85-91]
72. Blanquero J, Cortés-Vega MD, García-Frasquet MA, Sánchez-Laulhé PR, Nieto Díaz de Los Bernardos MI, Suero-Pineda A. Exercises using a touchscreen tablet application improved functional ability more than an exercise program prescribed on paper in people after surgical carpal tunnel release: a randomised trial. J Physiother 2019 Apr;65(2):81-87 [FREE Full text] [doi: 10.1016/j.jphys.2019.02.008] [Medline: 30926400]
73. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
74. van der Wardt V, Hancox J, Gondek D, Logan P, Nair RD, Pollock K, et al. Adherence support strategies for exercise interventions in people with mild cognitive impairment and dementia: A systematic review. Prev Med Rep 2017 Sep;7:38-45 [FREE Full text] [doi: 10.1016/j.pmedr.2017.05.007] [Medline: 28593121]
75. Campbell R, Evans M, Tucker M, Quilty B, Dieppe P, Donovan JL. Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. J Epidemiol Community Health 2001 Feb;55(2):132-138 [FREE Full text] [doi: 10.1136/jech.55.2.132] [Medline: 11154253]
76. Essery R, Geraghty AWA, Kirby S, Yardley L. Predictors of adherence to home-based physical therapies: a systematic review. Disabil Rehabil 2017 Mar 21;39(6):519-534. [doi: 10.3109/09638288.2016.1153160] [Medline: 27097761]
77. Abelson JS, Symer M, Peters A, Charlson M, Yeo H. Mobile health apps and recovery after surgery: What are patients willing to do? Am J Surg 2017 Oct;214(4):616-622. [doi: 10.1016/j.amjsurg.2017.06.009] [Medline: 28666581]
78. Uhm KE, Lee S, Kurillo G, Han JJ, Yang J, Yoo YB, et al. Usefulness of Kinect sensor-based reachable workspace system for assessing upper extremity dysfunction in breast cancer patients. Support Care Cancer 2020 Feb 29;28(2):779-786. [doi: 10.1007/s00520-019-04874-2] [Medline: 31144172]
79. Cuesta-Vargas A, Pajares B, Trinidad-Fernandez M, Alba E, Roldan-Jiménez C. Inertial Sensors Embedded in Smartphones as a Tool for Fatigue Assessment Based on Acceleration in Survivors of Breast Cancer. Phys Ther 2020 Mar 10;100(3):447-456. [doi: 10.1093/ptj/pzz173] [Medline: 32031221]
80. Phillips SM, Courneya KS, Welch WA, Gavin KL, Cottrell A, Nielsen A, et al. Breast cancer survivors' preferences for mHealth physical activity interventions: findings from a mixed methods study. J Cancer Surviv 2019 Apr;13(2):292-305. [doi: 10.1007/s11764-019-00751-3] [Medline: 30912011]
81. Zhou K, Wang W, Zhao W, Li L, Zhang M, Guo P, et al. Benefits of a WeChat-based multimodal nursing program on early rehabilitation in postoperative women with breast cancer: A clinical randomized controlled trial. Int J Nurs Stud 2020 Jun;106:103565. [doi: 10.1016/j.ijnurstu.2020.103565] [Medline: 32272281]
82. Gresham G, Schrack J, Gresham LM, Shinde AM, Hendifar AE, Tuli R, et al. Wearable activity monitors in oncology trials: Current use of an emerging technology. Contemp Clin Trials 2018 Jan;64:13-21 [FREE Full text] [doi: 10.1016/j.cct.2017.11.002] [Medline: 29129704]

83. Chandra H, Oakley I, Silva H. Designing to support prescribed home exercises: understanding the needs of physiotherapy patients. In: Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design - NordiCHI '12. 2012 Oct Presented at: Proc 7th Nord Conf Human-Computer Interact Mak Sense Through Des - Nord 12 Copenhagen; October 14-17; Copenhagen. [doi: 10.1145/2399016.2399108]

Abbreviations

eHealth: electronic health
HEP: home exercise program
mHealth: mobile health
P: participant
ROM: range of motion

Edited by G Eysenbach; submitted 29.04.20; peer-reviewed by H Harder, R Ferrer-Peña, M Lozano-Lozano; comments to author 21.05.20; revised version received 26.05.20; accepted 14.06.20; published 29.07.20.

Please cite as:
Brennan L, Kessie T, Caulfield B
Patient Experiences of Rehabilitation and the Potential for an mHealth System with Biofeedback After Breast Cancer Surgery: Qualitative Study
JMIR Mhealth Uhealth 2020;8(7):e19721
URL: http://mhealth.jmir.org/2020/7/e19721/
doi:10.2196/19721
PMID:32687476

Original Paper

The Mobile Health App Trustworthiness Checklist: Usability Assessment

Afua van Haasteren[1], MSc, PhD; Effy Vayena[1], MSc, PhD; John Powell[2], PhD

[1]Health Ethics and Policy Lab - Department of Health Sciences and Technology, Swiss Federal Institute of Technology, Zurich, Switzerland

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

Corresponding Author:
John Powell, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Woodstock Rd
Oxford, OX2 6GG
United Kingdom
Phone: 44 (0)1865 617768
Email: john.powell@phc.ox.ac.uk

Abstract

Background: The mobile health (mHealth) app trustworthiness (mHAT) checklist was created to identify end users' opinions on the characteristics of trustworthy mHealth apps and to communicate this information to app developers. To ensure that the checklist is suited for all relevant stakeholders, it is necessary to validate its contents.

Objective: The purpose of this study was to assess the feasibility of the mHAT checklist by modifying its contents according to ratings and suggestions from stakeholders familiar with the process of developing, managing, or curating mHealth apps.

Methods: A 44-item online survey was administered to relevant stakeholders. The survey was largely comprised of the mHAT checklist items, which respondents rated on a 5-point Likert scale, ranging from *completely disagree* (1) to *completely agree* (5).

Results: In total, seven professional backgrounds were represented in the survey: administrators (n=6), health professionals (n=7), information technology personnel (n=6), managers (n=2), marketing personnel (n=3), researchers (n=5), and user experience researchers (n=8). Aside from one checklist item—"the app can inform end users about errors in measurements"—the combined positive ratings (ie, *completely agree* and *agree*) of the checklist items overwhelmingly exceeded the combined negative ratings (ie, *completely disagree* and *disagree*). Meanwhile, two additional items were included in the checklist: (1) business or funding model of the app and (2) details on app uninstallation statistics.

Conclusions: Our results indicate that the mHAT checklist is a valuable resource for a broad range of stakeholders to develop trustworthy mHealth apps. Future studies should examine if the checklist works best for certain mHealth apps or in specific settings.

(JMIR Mhealth Uhealth 2020;8(7):e16844) doi:10.2196/16844

KEYWORDS

checklist; trustworthiness; trust; mobile health apps; validation; survey

Introduction

From self-diagnosis to wellness, mobile health (mHealth) apps have evolved as a conduit for individuals to play more pronounced roles in their own health care [1,2]. Sustaining the uptake of mHealth apps is particularly necessary to realize their potential in health systems. Among others, mHealth apps are perceived as a vehicle to enhance patient-provider communication, boost patient attempts to self-monitor health conditions, as well as to advance patient empowerment [3]. Despite these positive outcomes, several studies have found significant flaws in the privacy, security, and safety claims of several mHealth apps on the market [4,5].

There is evidence to suggest that end users abandon or reject mHealth apps that they perceive as untrustworthy [6-8]. So far, regulatory bodies such as the US Food and Drug Administration have been more concerned about the safety of mHealth apps that purport to be medical devices or their accompanying

add-ons [9]. The wellness app space lacks quality assurance frameworks and clear guidance from respective authorities [10]. In light of the relaxed regulatory approach to mHealth apps, evaluation tools have been proposed to assist end users in determining which apps are safe or secure and, thus, which can be trusted [11,12]. Very few of these tools target app developers, although they are vital stakeholders in ensuring that mHealth apps are trustworthy from the outset [13]. To fill this gap, we created the mHealth app trustworthiness (mHAT) checklist via a focus group study with end users of health apps. The mHAT checklist is displayed in Multimedia Appendix 1 and details of the study can be found elsewhere [14].

The purpose of this study was to validate the mHAT checklist by modifying its contents according to ratings and suggestions from stakeholders familiar with the process of developing, managing, or curating mHealth apps to ensure that its contents are applicable. Since the checklist goes beyond the technical aspects of app development, procuring the suggestions of different stakeholders will ensure that it is suited for anyone likely to be involved in app development. The value of validating the checklist was reinforced by a preliminary feedback exercise during the development of the mHAT checklist. Among the six experts with information technology (IT) backgrounds that provided feedback, there was a consensus that the checklist must be validated among a wider range of stakeholders.

Methods

Study Design

We designed a cross-sectional online survey to validate the mHAT checklist. Ethical approval for this study was granted by the ETH Zurich (Swiss Federal Institute of Technology) Ethics Commission (EK 2019-N-20). To ensure that this study poses little to no risks to the respondents, we did not require any personal details [15]. Taking part in the survey implied consent. The survey was administered through the SurveyMonkey online survey tool (SurveyMonkey Inc) between March 14 and July 10, 2019.

Survey Development

The first question in the 44-item open survey screened for eligible respondents to ensure that only individuals with experience in developing apps advanced on to the next page. Another question requested participants' areas of expertise and was accompanied by four professions—administrators; marketing personnel; software designers, programmers, or developers; and user experience (UX) researchers—as well as a field for free-text answers. To provide an overview of the types of apps created by participants, the next question asked for the function of an app respondents have created within the last 5 years. The remaining 41 survey items consisted of the mHAT checklist items and required answers on a 5-point Likert scale, ranging from *completely disagree* (1) to *completely agree* (5) [16].

Aside from the screening question, which blocked noneligible participants from proceeding further, no other special functions, such as skip logic, were applied to the survey. The number of questions on each page of the survey differed according to the items in each segment of the checklist. Consequently, page 1 contained the 11 questions from the informational content section, whereas page 2 was comprised of the seven questions under *organizational attributes*. SurveyMonkey estimated 10 minutes as the average completion time.

Participant Recruitment

Our target population was a convenience sample of adults over the age of 18 years who are knowledgeable about the processes involved in developing mobile apps. Among others, administrators (ie, business and systems); marketing personnel; software designers, programmers, and developers; and UX researchers from all geographical locations were eligible for this study. In line with the principles for calculating sample sizes for surveys, the minimum desired sample size was 30 [17]. Participants were recruited by propagating a link of the survey hosted on the SurveyMonkey platform (1) to relevant individuals via email and (2) on social media (ie, Twitter). In line with minimizing access to participants' personal information, participants' Internet Protocol (IP) addresses were not tracked, and their computers were also not assigned unique identification numbers.

Data Analysis

In accordance with recommended practice, the data from the Likert scales were analyzed with descriptive statistics, such as ranks, modes, frequencies, and ranges [16,18]. The frequencies and percentages of the survey items were presented as *positive*, *neutral*, and *negative* using Microsoft Excel. Positive ratings refer to the sum of *completely agree* and *agree*, whereas negative ratings are the sum of *completely disagree* and *disagree*. Meanwhile, the neutral values present the raw ratings of each item on the Likert scale. Only the checklist items with positive ratings (ie, *completely agree* and *agree*) exceeding negative ratings (ie, *completely disagree* and *disagree*) were retained. The data derived from the open-ended questions were analyzed by content analysis in NVivo 12 (QSR International) [19]. Novel suggestions of items to include in the checklist were catalogued accordingly.

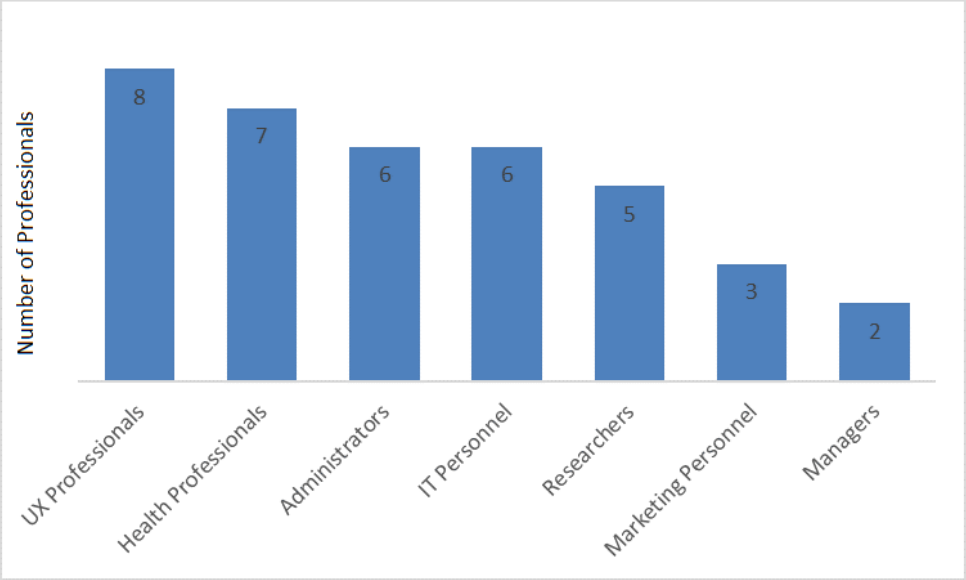
Results

Participant Characteristics

Of the 144 individuals that responded to calls to participate in this survey, 49 (34.0%) with some professional experience in developing mobile apps were eligible to participate. Out of these 49 participants, 23 (47%) completed the survey in its entirety, while 26 (53%) others completed it partially.

Overall, 22 respondents indicated their professional expertise among the four fixed-choice professions provided in the survey: 6 (27%) administrators; 3 (14%) marketing personnel; 5 (23%) software designers, programmers, or developers; and 8 (36%) UX researchers. Another 15 free-text responses showed additional professional backgrounds: 7 (47%) health professionals, 5 (33%) researchers, 2 (13%) managers, and 1 (7%) IT professional. Figure 1 presents the frequencies of the professions represented in this survey.

Figure 1. Professional expertise of respondents (n=37). IT: information technology; UX: user experience.



The respondents had collectively created apps with 33 functions within the last 5 years. By aggregating similar apps, 20 out of 33 (61%) sought to assist end users to self-manage or monitor chronic diseases, such as diabetes and mental health, while 8 (24%) apps intended to promote healthy lifestyles through diet and or exercise. The complete list of app functions can be found in Multimedia Appendix 2.

Survey Responses According to mHAT Categories

In Table 1, we rank the *positive* (ie, sum of *completely agree* and *agree*), *neutral*, and *negative* (ie, sum of *completely disagree* and *disagree*) ratings according to categories of the checklist. Aside from one checklist item—"the app can inform end users about errors in measurements"—the combined positive ratings (ie, *completely agree* and *agree*) of the checklist items overwhelmingly exceeded the combined negative ratings (ie, *completely disagree* and *disagree*). The individual ratings for *completely disagree*, *disagree*, *neutral*, *agree*, and *completely agree* can be retrieved from Multimedia Appendix 3.

Table 1. Rankings of respondent ratings of checklist items.

Item	Negative, n (%)	Neutral, n (%)	Positive, n (%)
Informational content (n=29)			
The privacy policies accompanying the app can be concise, clear, and easy to understand	1 (3)	6 (21)	22 (76)
The terms of service accompanying the app can be concise and easy to read	1 (3)	7 (24)	21 (72)
The app can be programmed such that it does not require too many end-user personal data	2 (7)	6 (21)	21 (72)
The app can provide accurate measurements	3 (10)	5 (17)	21 (72)
The app can be accompanied by clear end-user safety guidelines	3 (10)	6 (21)	20 (69)
The app can be created with evidence from robust research	4 (14)	5 (17)	20 (69)
The app can inform end users about errors in measurements	11 (38)	8 (28)	10 (34)
The app can recommend regular updates to:			
Fix bugs inherent within the app	5 (17)	4 (14)	20 (69)
Amend app contents based on improved research	4 (14)	7 (24)	18 (62)
The information on the app can be certified by an:			
In-house team	4 (14)	6 (21)	19 (66)
External third-party team	5 (17)	6 (21)	18 (62)
The app can ensure that personalized data for end users is precise	1 (3)	12 (41)	16 (55)
The research-backed evidence used to create the app can be easy to locate and understand	6 (21)	8 (28)	15 (52)
The app can highlight potential risks or side effects resulting from its use	9 (31)	8 (28)	12 (41)
Organizational attributes (n=28)			
My company can be transparent about our data-handling history and data breaches	2 (7)	4 (14)	22 (79)
My company can demonstrate that it values data-protection regulations	2 (7)	4 (14)	22 (78)
My company can employ skilled personnel within the app development domain to perform all tasks relating to the app	4 (14)	4 (14)	20 (71)
My company can adopt clear policies on how to handle end-user data	3 (11)	6 (21)	19 (68)
Our app can be affiliated with a nongovernmental organization or a reputable government agency	5 (18)	8 (29)	15 (54)
My company has developed similar apps in the past	8 (29)	5 (18)	15 (54)
My company has other reputable products or services to associate the app with	7 (25)	7 (25)	14 (50)
Societal influences (n=25)			
End users can readily suggest the app to others	2 (8)	4 (16)	19 (76)
The app store can display how often the app has been downloaded	3 (12)	5 (20)	17 (68)
The app can display the positive reviews that it receives	5 (20)	6 (24)	14 (56)
To ensure that end users locate the app, it can be made to appear:			
In the top results of search engines	6 (24)	10 (40)	9 (36)
As a featured app in the app store	4 (16)	10 (40)	11 (44)
The app can accompany a wearable device	5 (20)	10 (40)	10 (40)
Technology-related features (n=24)			
The app can be easy to use and have a friendly end-user interface	1 (4)	4 (17)	19 (79)
The app can be made aesthetically appealing	1 (4)	5 (21)	18 (75)
The app can be programmed to send out a reasonable number of notifications	1 (4)	5 (21)	18 (75)
The data generated from the app can be anonymized to make individuals unidentifiable	3 (13)	3 (13)	18 (75)
The app can be easily accessed by the end users it aims to target	1 (4)	5 (21)	18 (75)
The data generated from the app can be secured by end-to-end encryption	0 (0)	7 (29)	17 (71)

Item	Negative, n (%)	Neutral, n (%)	Positive, n (%)
The data generated from the app can be:			
Stored locally on the device	3 (13)	5 (21)	16 (67)
Encrypted	1 (4)	7 (29)	16 (67)
The app features can be customized by end users	4 (16)	5 (20)	15 (62)
Privacy can be a core consideration throughout the life cycle of the app	4 (17)	7 (29)	13 (54)
End users can easily access all of their data (eg, address and billing information)	4 (17)	7 (29)	13 (54)
User control (n=23)			
The app can allow end users to easily delete their data	1 (4)	6 (26)	16 (70)
The app can seek explicit end-user permission before sharing data with third parties	3 (13)	6 (26)	14 (61)
The app can allow end users to opt in or decide which data can be stored or processed	2 (9)	7 (30)	14 (61)
The app can give end users the freedom to control how their data are used	2 (9)	8 (35)	13 (57)
The app can allow end users to restrict data sharing to third parties such as social networking sites	1 (4)	9 (39)	13 (57)
The app can designate end users as proprietors (ie, owners) of their data	7 (30)	7 (30)	9 (39)

Suggestions for Additional Survey Items

There were two new suggestions of items to include in the checklist: (1) transparency about business models and funding streams as well as (2) statistics on app uninstallations. These items were catalogued under the *organizational attributes (reputation)* and *informational content (transparency)* categories of the checklist, respectively. In the final version of the mHAT checklist found in Multimedia Appendix 4, the new items are indicated with an asterisk.

Discussion

Principal Findings

We have conducted a stakeholder survey to validate the mHAT checklist, a practical tool for app developers to create trustworthy mHealth apps. The checklist items can be considered feasible in many different settings, since the majority of the items were rated positively by the disparate stakeholders—UX researchers, administrators, IT personnel, health care professionals, researchers, managers, and marketing personnel—who reviewed its contents.

Throughout the survey, it appeared that those items that could apply to disparate mHealth apps were rated higher than those items that may be relevant to certain apps but not others. For example, “the app can inform end users about errors in measurements” was the most negatively rated item, whereas “the app can be easy to use and have a friendly end-user interface” was the most positively rated. One plausible explanation for this difference in ratings may be due to the aspiration to design any app, regardless of function, to be user friendly for its end users. The extent to which an app can have no measurement errors, however, will differ from the function of the app (eg, diabetes versus mental health).

The most negatively rated items in the survey—“the app can inform end users about error in measurements” and “the app can highlight potential risks or side effects resulting from its use”—demonstrated that app developers place marginal value in informing end users about the errors, risks, or side effects that may arise from their apps. Regardless of the reasons that contributed to these ratings, such information is vital to upholding end users’ trust in mHealth apps. Since plenty of evidence suggests that some apps on the market perpetuate inaccurate advice that may threaten patient safety, it is imperative that app developers are transparent about these issues [4,20]. Without a clear list of app risks or errors, end-user concerns about the trustworthiness of mHealth apps are bound to continue.

Since the checklist item “the app can inform end users about error in measurements” received more negative than positive ratings, it was excluded from the checklist. Meanwhile, respondents suggested including two new items: (1) a disclosure of the business model and (2) uninstallation statistics. Indeed, when app developers obscure their business models, they are perceived negatively by end users [6]. Including the uninstallation statistics of an app, however, is interesting for two reasons. On the one hand, it could signal transparency on the part of app developers and thus boost end users’ trust. On the other hand, it may deter end users from downloading the app altogether, since unpopular apps are unlikely to be downloaded in the first place [21].

Strengths and Limitations

To the best of our knowledge, the mHAT checklist is the only empirically validated checklist that presents the attributes of trustworthy mHealth apps. This validated checklist is a robust tool for assisting app developers to create trustworthy apps. Our study has limitations. Despite multiple attempts to obtain a minimum sample size of 30 participants, challenges in recruiting participants meant that fewer participants could take part in the survey. Nonetheless, 23 participants answered all the survey items affording informative statistical analysis. The convenience sample recruited for this study is another limitation. Since these samples are usually unrepresentative of an entire population, our study may have missed out on capturing the opinions of

some relevant stakeholders [22,23]. Further, ineligible individuals may have participated in the survey for two reasons: (1) it was an open survey and (2) to minimize access to respondents' personal information, their computers were not assigned unique identification numbers. Nonetheless, there was no incentive to participate, since no compensation was awarded.

Implications for Future Research

The mHAT checklist informs the conversation on the expectations of trustworthy mHealth apps. Future studies should assess whether the checklist is suitable for all mHealth apps or whether it is better suited for certain apps than others. Through additional studies, the contents of the checklist can be improved to retain useful items and exclude redundant ones. More enquiries about the underlying reasons why app developers see little value in informing end users about the risks, side effects, or errors of their apps is also warranted.

Conclusions

This study presents a validated mHAT checklist: a useful guide for app developers to create trustworthy health apps. The 41-item checklist is comprised of five main categories—informational content, organizational attributes, societal influences, technology-related factors, and user control—and 11 subcategories—information accuracy, understandability, transparency, brand familiarity, reputation, recommendations, external factor, usability, privacy, autonomy, and empowerment (see Multimedia Appendix 4).

Acknowledgments

This project was funded by a Swiss National Science Foundation grant awarded to EV (PP000P3-157556; 4075210-167223). The authors would like to acknowledge Dr Constantinos Koshiaris, Chelu Mfalila, and Dr Chrysanthi Papoutsi of the Nuffield Department of Primary Care Health Sciences, University of Oxford, for their assistance with the survey.

Authors' Contributions

AvH designed and carried out the survey under the supervision of EV and JP.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The unvalidated mobile health app trustworthiness (mHAT) checklist.
[PDF File (Adobe PDF File), 74 KB - mhealth_v8i7e16844_app1.pdf]

Multimedia Appendix 2
Functions of apps developed by respondents.
[DOCX File , 19 KB - mhealth_v8i7e16844_app2.docx]

Multimedia Appendix 3
Frequencies and percent ratings of survey items.
[DOCX File , 30 KB - mhealth_v8i7e16844_app3.docx]

Multimedia Appendix 4
The updated mobile health app trustworthiness (mHAT) checklist. New items to the checklist are indicated with an asterisk.
[DOCX File , 151 KB - mhealth_v8i7e16844_app4.docx]

References

1. Lupton D, Jutel A. 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015 May;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]
2. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
3. East ML, Havard BC. Mental health mobile apps: From infusion to diffusion in the mental health social system. JMIR Ment Health 2015;2(1):e10 [FREE Full text] [doi: 10.2196/mental.3954] [Medline: 26543907]
4. Wicks P, Chiauzzi E. 'Trust but verify': Five approaches to ensure safe medical apps. BMC Med 2015 Sep 25;13:205 [FREE Full text] [doi: 10.1186/s12916-015-0451-z] [Medline: 26404791]
5. Plante TB, O'Kelly AC, Urrea B, MacFarlane ZT, Blumenthal RS, Charleston J, et al. User experience of instant blood pressure: Exploring reasons for the popularity of an inaccurate mobile health app. NPJ Digit Med 2018;1:31 [FREE Full text] [doi: 10.1038/s41746-018-0039-z] [Medline: 31304313]

6. Khalid H, Shihab E, Nagappan M, Hassan A. What do mobile app users complain about? IEEE Softw 2015 May;32(3):70-77. [doi: 10.1109/MS.2014.50]
7. Murnane E, Huffaker D, Kossinets G. Mobile health apps: Adoption, adherence, and abandonment. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015) colocated with the 19th International Symposium on Wearable Computers (ISWC 2015). 2015 Presented at: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015) colocated with the 19th International Symposium on Wearable Computers (ISWC 2015); September 7-11, 2015; Osaka, Japan p. 261-264. [doi: 10.1145/2800835.2800943]
8. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: Qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
9. Powell AC, Landman AB, Bates DW. In search of a few good apps. JAMA 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
10. Ferretti A, Ronchi E, Vayena E. From principles to practice: Benchmarking government guidance on health apps. Lancet Digit Health 2019 Jun;1(2):e55-e57 [FREE Full text] [doi: 10.1016/s2589-7500(19)30027-5]
11. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: Mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
12. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
13. Albrecht U, Noll C, Von Jan U. App-synopsis: Self-assessment on trust or distrust of health-apps. Stud Health Technol Inform 2014;202:233-236. [Medline: 25000059]
14. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. Digit Health 2019;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]
15. What is personal data? Information Commissioner's Office. Wilmslow, UK: Information Commissioner's Office; 2019. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/ [accessed 2020-07-10]
16. Boone HN, Boone DA. Analyzing Likert data. J Ext 2012 Apr;50(2):1-5 [FREE Full text]
17. Sue VM, Ritter LA. Conducting Online Surveys. 2nd edition. Thousand Oaks, CA: SAGE Publications; 2012.
18. Allen IE, Seaman CA. Likert scales and data analyses. Quality Progress. 2007 Jul. URL: http://rube.asq.org/quality-progress/2007/07/statistics/likert-scales-and-data-analyses.html [accessed 2019-07-10]
19. Taylor-Powell E, Renner M. Analyzing Qualitative Data. Madison, WI: University of Wisconsin—Extension, Cooperative Extension; 2003. URL: https://cdn.shopify.com/s/files/1/0145/8808/4272/files/G3658-12.pdf [accessed 2020-07-01]
20. 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]
21. Ickin S, Petersen K, Gonzalez-Huerta J. Why do users install and delete apps? A survey study. In: Proceedings of the 8th International Conference on Software Business (ICSOB 2017). Cham, Switzerland: Springer; 2017 Presented at: 8th International Conference on Software Business (ICSOB 2017); June 12-13, 2017; Essen, Germany p. 186-191 URL: http://www.gonzalez-huerta.net/wp-content/uploads/2017/06/ICSOB17.pdf [doi: 10.1007/978-3-319-69191-6_13]
22. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1-4 [FREE Full text] [doi: 10.11648/j.ajtas.20160501.11]
23. Hill RK. What sample size is "enough" in internet survey research. Interpersonal Comput Technol 1998 Jul;6(3-4):1-10 [FREE Full text]

Abbreviations

ETH Zurich: Swiss Federal Institute of Technology
IP: Internet Protocol
IT: information technology
mHAT: mobile health app trustworthiness
mHealth: mobile health
UX: user experience

Edited by G Eysenbach; submitted 30.10.19; peer-reviewed by A LeFevre, T Plante, R Yin; comments to author 12.12.19; revised version received 30.04.20; accepted 03.06.20; published 21.07.20.

Please cite as:
van Haasteren A, Vayena E, Powell J
The Mobile Health App Trustworthiness Checklist: Usability Assessment
JMIR Mhealth Uhealth 2020;8(7):e16844
URL: http://mhealth.jmir.org/2020/7/e16844/
doi:10.2196/16844
PMID:32706733

Original Paper

Untold Stories in User-Centered Design of Mobile Health: Practical Challenges and Strategies Learned From the Design and Evaluation of an App for Older Adults With Heart Failure

Victor Philip Cornet[1,2], MS; Tammy Toscos[2], PhD; Davide Bolchini[1], PhD; Romisa Rohani Ghahari[2], PhD; Ryan Ahmed[2], MS; Carly Daley[2,3], MS; Michael J Mirro[2,3,4], MD, FACC, FHRS, FAHA, FACP; Richard J Holden[4,5], PhD

[1]Department of Human-centered Computing, School of Informatics and Computing, IUPUI, Indianapolis, IN, United States

[2]Parkview Mirro Center for Research and Innovation, Parkview Health, Fort Wayne, IN, United States

[3]Department of BioHealth Informatics, School of Informatics and Computing, IUPUI, Indianapolis, IN, United States

[4]Department of Medicine, School of Medicine, Indiana University, Indianapolis, IN, United States

[5]Regenstrief Institute, Indianapolis, IN, United States

Corresponding Author:
Richard J Holden, PhD
Regenstrief Institute
1101 W 10th St #421
Indianapolis, IN, 46202
United States
Phone: 1 317 278 5323
Email: rjholden@iu.edu

Abstract

Background: User-centered design (UCD) is a powerful framework for creating useful, easy-to-use, and satisfying mobile health (mHealth) apps. However, the literature seldom reports the practical challenges of implementing UCD, particularly in the field of mHealth.

Objective: This study aims to characterize the practical challenges encountered and propose strategies when implementing UCD for mHealth.

Methods: Our multidisciplinary team implemented a UCD process to design and evaluate a mobile app for older adults with heart failure. During and after this process, we documented the challenges the team encountered and the strategies they used or considered using to address those challenges.

Results: We identified 12 challenges, 3 about UCD as a whole and 9 across the UCD stages of formative research, design, and evaluation. Challenges included the timing of stakeholder involvement, overcoming designers' assumptions, adapting methods to end users, and managing heterogeneity among stakeholders. To address these challenges, practical recommendations are provided to UCD researchers and practitioners.

Conclusions: UCD is a gold standard approach that is increasingly adopted for mHealth projects. Although UCD methods are well-described and easily accessible, practical challenges and strategies for implementing them are underreported. To improve the implementation of UCD for mHealth, we must tell and learn from these traditionally untold stories.

(JMIR Mhealth Uhealth 2020;8(7):e17703) doi:10.2196/17703

KEYWORDS

user-centered design; research methods; mobile health; digital health; mobile apps; usability; technology; evaluation; human-computer interaction; mobile phone

Introduction

The user-centered design (UCD), or human-centered design, is a framework for iteratively researching, designing, and evaluating services and systems by involving end users and other stakeholders throughout a project life cycle. [1-3]. Mobile health (mHealth) projects benefit from UCD by using input from patients, informal caregivers, clinicians, and other stakeholders during the project life cycle to create better designs and iteratively improve interventions, thus enhancing their

usability, acceptance, and potential success when implemented [4-6]. Increasingly, UCD has been recommended and adopted in mHealth projects to great success [7], with many examples of mHealth for people living with HIV [5,8], chronic conditions [9-11], or mental illness [6,12,13].

UCD is supported by popular tools and methods, such as cognitive task analysis, workflow studies and journey mapping, participatory design, rapid prototyping, usability testing, and heuristic evaluation [14-17]. Textbooks, articles, and other resources offer easily accessible and detailed guidance on the general UCD process and specific UCD methods [18,19]. However, an informal review of the literature reveals little information about the practical implementation of UCD methods.

Practical challenges reported in studies largely in non–health care domains include ensuring participants' representativeness of the target population [20]; threats to innovation [21]; difficulty communicating with people from different backgrounds [22]; and organizational barriers, such as not having convenient access to participants [23].

Although studies applying UCD for mHealth are on the rise, very few mHealth studies report the challenges they face while planning and executing UCD activities, or they do so parenthetically. UCD challenges may be unique or amplified in the field of mHealth. For example, there are known difficulties in evaluating the effectiveness of mHealth solutions, in part because of the variable and multifactorial nature of health and illness trajectories [6]. mHealth projects often involve unique stakeholders, drawn from vulnerable patient populations (eg, older adults, patients with chronic conditions) [24,25] and busy clinician experts [6,26], or sometimes both. Moreover, on the one hand, there is sometimes a mismatch between the use of UCD methods (eg, rapid prototyping, user testing) and emerging technologies (eg, sensors) and, on the other hand, traditions (eg, clinical trials) and technological conservatism that characterize much of the health care sector [27-29].

Given the presence and importance of the practical challenges for implementing UCD, coupled with the increased use of UCD for mHealth, we argue for the need to explicitly describe those challenges. As mHealth technologies become more pervasive, navigating practical UCD challenges is essential for the development of "safe, sound, and desirable" [30] mHealth solutions that improve health outcomes while involving stakeholders in the design process [31,32]. We believe that identifying, reporting, and discussing the *untold stories* of actually implementing UCD for mHealth will help in overcoming the significant gaps between research and practice [33,34].

Methods

Overview

The objectives of this study were as follows:

- Characterize practical challenges encountered while implementing UCD to design an mHealth app for older adults with heart failure.
- Discuss strategies that we used or considered using to manage these challenges.

In presenting these challenges and strategies, we offer illustrations from our own experiences, particularly the Power to the Patient (P2P) project (R21 HS025232) and cite others that have been described in the literature in and outside the mHealth arena.

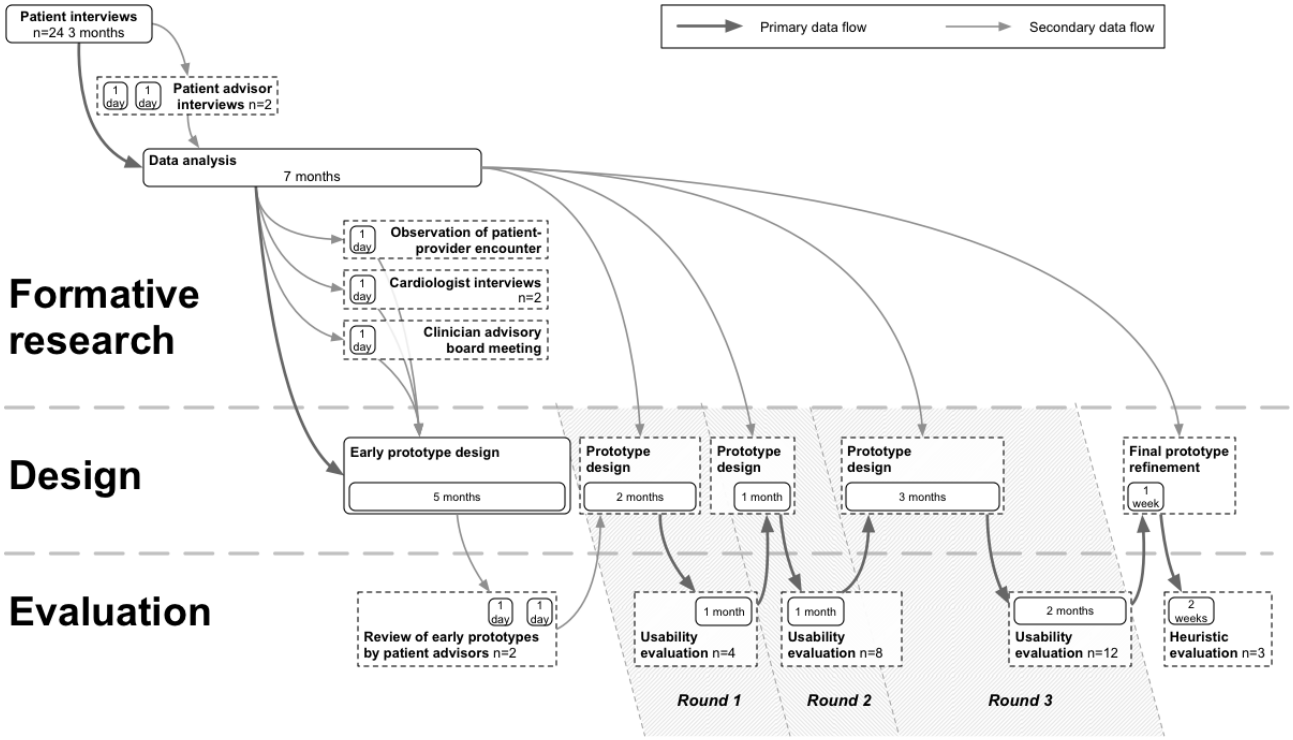
Power to the Patient Project

From 2017 to 2019, we performed a UCD study to design and evaluate information technology for older adults with heart failure. On the basis of previous literature, we knew that these patients had unmet needs and required additional support to monitor and manage symptoms and various related behaviors, including medication use, dietary and fluid restriction, and physical activity [35-37]. Our work focused on delivering information and decision support by leveraging a novel technological opportunity, namely, sharing with patients the data from their cardiac implanted electronic devices (CIEDs). Many patients with heart failure have CIEDs for the delivery of timely cardiac therapy and the capturing of data that can predict decompensation and other events leading to hospitalization and other downstream outcomes [38,39]. Patients seldom receive their CIED data [40,41], but technical and cultural changes are increasing the likelihood that they will in the future [41-43].

Figure 1 presents an overview of the project's timeline, and Textboxes 1 and 2 further elaborate on the methods used. Select methods and results from the project are available elsewhere [19,32,44-47].

The project began with a *problem analysis* or formative study of the domain as a precursor to design. This phase comprised interviews with 24 older adults with heart failure, half of whom had CIEDs, to learn about how they made decisions about their health. These interviews used the critical incident technique, a method that asks participants to recall and describe a specific event or scenario and uses probes to better understand the participants' thoughts and actions during the event or scenario [48,49]. We also examined participants' decision-making strategies in response to fictitious scenarios [16]. In these scenarios, individuals were presented with hypothetical situations related to data from a fictitious CIED and asked to think aloud as they made decisions about how to respond. Other formative study methods were a brief observation of a device clinic, meetings with 2 cardiologists, and sharing of findings and design work from 2 recent similar studies on CIED data sharing [43,50].

Figure 1. Timeline of the Power to the Patient project.



Textbox 1. Formative research methods used to establish the Power to the Patient's domain space.

Patient interviews

- Method: 70-min (1) critical incident interviews and (2) scenario-based cognitive interviews to understand the decision-making process of older adults with heart failure
- Participants: 24 English-speaking older adults (≥65 years) diagnosed with heart failure (New York Heart Association Class II-IV) and 14 accompanying support persons (family and friends). Patients were receiving care at Parkview Health (Fort Wayne, Indiana)
- Procedure: (1) Participants were asked to describe a recent minor adverse health event and were probed with questions about their thought, feelings, and actions. (2) Participants were presented with a picture of a fictitious device that could give them a CIED score representing their heart health; they were asked to describe what they would think and do depending on the score displayed on the device

Patient advisory meetings

- Method: One-on-one meetings with patient advisors soliciting feedback on (1) personas and use-case scenarios and (2) early design concepts and prototypes
- Participants: 2 older adults with heart failure from the community (Indianapolis, Indiana) who voluntarily assisted the study in an advisory capacity
- Procedure: (1) Advisors met with the research team to discuss the preliminary findings from the interviews and early persona development. They provided feedback on the findings, methods, and relevance of the work. (2) Advisors were presented with design alternatives of a Power to the Patient prototype, and then they interacted with it while thinking aloud using a computer and a mouse

Clinician advisory board meeting

- Method: Group dinner with clinician experts to elicit feedback on personas, use-case scenarios, and early concepts
- Participants: 7 Parkview Health clinicians (2 cardiologists, a device clinic supervisor, 2 technicians, a nurse, and the vice president of operations for the Parkview Heart Institute)
- Procedure: Personas and scenarios were presented, among other findings, and questions were asked of clinicians regarding the validity of the findings and related current protocols (some of which were subsequently collected)

Individual interviews with 2 cardiologists

Observation of clinical encounters with a patient in the device clinic

Textbox 2. Evaluation methods used during the Power to the Patient development.

Usability evaluations, round 1 (R1) and round 2 (R2)

- Method: 90-min task-driven evaluation sessions of Power to the Patient prototypes to assess usability (primarily) and acceptability (secondarily)
- Participants: 4 (R1) and 8 (R2) English-speaking older adults diagnosed with heart failure and 3 accompanying support persons (2 in R1 and 1 in R2)
- Procedure: Participants performed specific tasks in the prototype while thinking aloud. Testing occurred in a private room, with an interactive prototype made in Axure RP 9 running on a Samsung Galaxy S7 smartphone. Participants' manual interactions were video recorded. Pretest, participants completed a demographic survey (ie, age, gender, technology use, and education level) and the Newest Vital Sign health literacy screening (NVS) [51]; posttest, they completed the system usability scale [52], National Aeronautics and Space Administration Task Load Index (NASA-TLX) [51], and user acceptance survey. Participants were also interviewed about their understanding and projected use of Power to the Patient prototypes.

Usability evaluation, round 3 (R3)

- Method: 90-min scenario-driven evaluation sessions of Power to the Patient prototypes to assess acceptability (primarily) and usability (secondarily)
- Participants: 12 English-speaking older adults diagnosed with heart failure, with cardiac implanted electronic devices, and accompanied by 5 support persons
- Procedure: Participants simulated days 1 and 10 of longitudinal use of the Power to the Patient prototype while thinking aloud. They completed the same assessments as in earlier rounds and were interviewed about their understanding and projected use of the Power to the Patient prototype.

Heuristic evaluation

- Method: Heuristic evaluation questionnaire to assess usability of Power to the Patient prototypes
- Participants: 3 user-centered design experts external to the team
- Procedure: Participants explored the prototype based on 2 use cases. They then reported their observations for 9 heuristics and gave an overall rating for the usability of the prototype

Findings from the formative study were analyzed to develop personas, representing distinct ways patients made decisions, and use-case scenarios, representing decision-making situations in which hypothetical patients with CIEDs might find themselves. A model was also created depicting the flow of naturalistic decision making for heart failure self-care. These products were presented separately to 2 patient advisors and a panel of clinicians, who provided feedback on the realism and relevance of the personas and scenarios. The personas, use-case scenarios, and a review of the literature and market landscape (eg, app store reviews of similar mHealth products) were used to formulate requirements and early design concepts to be presented to patient advisors.

The design involved writing requirements and 5 months of iterative prototyping, concluding with an interactive prototype. Subsequently, we performed 3 rounds of formal laboratory-based usability testing with 24 participants, interleaved with periods of prototype redesign. Each round had a more complete prototype and an increasing number of participants (ie, n=4 in round 1, n=8 in round 2, and n=12 in round 3). The project concluded with a final refinement of the prototype and formal *heuristic* evaluations by 3 outside UCD experts.

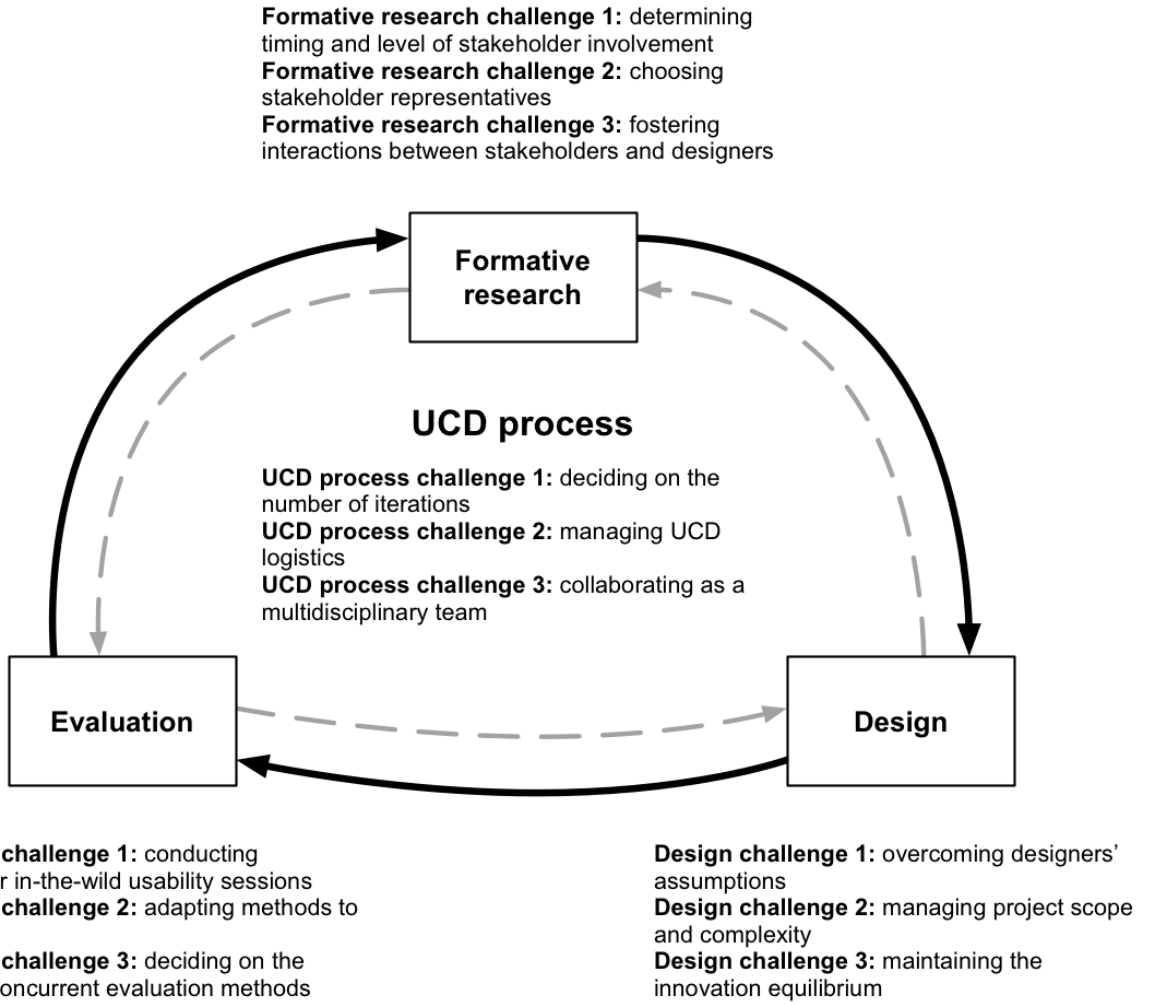
The P2P app designed in this study had 4 core patient-facing components: a heart health score calculated from CIED data (a fictitious concept inspired by existing research [53,54]), self-assessments on recommended heart failure self-care domains (eg, medication use, sodium-restricted diet), tips and strategies for better self-management, and logs of data captured by the app. We designed and tested different implementations of these core concepts, changing information architecture and amending feature sets as we received feedback from study participants. For example, in one iteration, assessments, tips, and strategies were organized using the concept of *plans* that users could select.

Results

Overview

The UCD process as a whole and the 3 major UCD phases of formative research, design, and evaluation present associated implementation challenges. Twelve of the most pervasive challenges are summarized in Figure 2. Next, we discuss each, with examples from our experience with P2P, supplemented by relevant literature.

Figure 2. UCD challenges encountered during Power to the Patient research, development, and evaluation. UCD: User-centered design.



Whole User-Centered Design Process

We identified 3 challenges related to the UCD process as a whole: deciding on the number of design-and-test iterations (UCD process challenge 1); managing logistics associated with UCD projects, such as preparing materials and recruitment costs (UCD process challenge 2); and collaborating as a multidisciplinary team by navigating misaligned goals and communication breakdowns (UCD process challenge 3). Recommendations for the UCD process are listed in Textbox 3.

Textbox 3. Recommendations for the user-centered design process.

1. Keep the number of iterations flexible, estimating the range based on available resources but adjusting as the project progresses.
2. Avoid endless iterations by having a clear criteria for when to end the design and testing activities.
3. Consult or collaborate with experienced practitioners to anticipate and manage logistic challenges.
4. Explicitly discuss and seek to align multidisciplinary team members' goals and preferences early on, managing these over time through open communication.
5. Involve a multilinguistic conductor to lead the team and coordinate members with diverse values, norms, practices, vocabularies, theories, and methods.

User-Centered Design Process Challenge 1: Deciding on the Number of Iterations

We planned the number of design-and-test iterations early in the project to satisfy the requirements of the funding agency and Institutional Review Board (IRB). Three iterations were originally estimated as feasible, given the timeline and amount of funding. Each iteration was assumed to require an average of at least 2 months of time and budget. As the project progressed, it became apparent that those assumptions were correct and that to adhere to budget and project timeline restrictions, we would be able to complete the 3 planned iterations. The number of participants in total and per iteration was based on several considerations. The total of 24 test participants were based on the project budget and timeline, as above, and the goal of having an average of 8 participants per

test, a typical upper range for formative testing. For the initial 2 iterations, our focus was on identifying overt usability issues, and based on Nielsen's recommendations, we planned to enroll about 4 to 8 participants (a few participants can uncover most usability issues [55]). In the last round, as our focus shifted toward evaluating acceptance and the extent to which the design was usable across a more diverse set of users, we recruited 12 participants.

Iterations in UCD benefit software development by reducing usability issues and improving features before it is too late or too expensive to make changes [56]. Iteration is generally recommended [56], but the number of iterations is not fixed and often depends on the project and how it progresses. However, deviating from a planned number of deviations can be costly or prohibited for other (eg, regulatory or management) reasons. Conversely, endless iteration is counterproductive and delays in-the-wild field testing and actual implementation.

UCD Process Challenge 2: Managing UCD Logistics

Many of our recurring challenges were related to logistics, such as meeting recruitment goals, preparing and managing materials (ie, instructions, presentations, and design documentation), and arranging a specific time for members and advisors to meet despite being geographically dispersed. Others have reported similar challenges in mHealth projects [57,58], but more often they are assumed and discussed informally between practitioners. This means that novice UCD practitioners tend to underestimate logistic challenges and have limited information on how to address them. UCD logistics challenges often mirror challenges to implementing patient-centered research in general, for example, the taxonomy of challenges by Holden et al [59], which includes patient identification and recruitment; privacy and confidentiality; conflicts with compensation; and logistical issues, such as travel, timing, and communication. Authors have described strategies to achieve buy-in, trust, transparency, accommodation (flexibility), openness, and anticipation [59,60], as well as checklists for implementing these general strategies [61].

Others have described the additional costs of recruiting end users for UCD projects [20,57]. It is even more difficult to recruit a representative sample in a technology project when technology ownership and proficiency are among the eligibility criteria [62] or the notion of technology leads individuals to decline participation [63]. Age- and illness-related physical and cognitive limitations may also exclude some individuals from mHealth studies [6,64]. In such cases, UCD projects must also consider whether to involve informal caregivers or others (eg, translators) who could assist patients during formative research or testing.

UCD Process Challenge 3: Collaborating as a Multidisciplinary Team

Our P2P project was the fruit of collaboration between a research university and a research center in a large health system, with support from patient and clinician advisors. Our team included a cardiologist coinvestigator who provided invaluable clinical information, feedback, and access to local and national clinical leaders. However, with disciplinary diversity comes disagreement, communication difficulty, and differences in assumptions, and although these are all desirable elements, they require efforts to manage, for example, by frequently asking team members to state their assumptions.

Multidisciplinary collaboration is often encouraged in UCD [6,65,66], including partnerships between designers and clinicians [67,68]. Those who have attempted such collaborations are aware of the methodological and cultural misalignment or divergent goals between Human-Computer Interaction technologists and clinicians [65,69]. Strategies to overcome these collaboration challenges include structured communication and the involvement of a *multilinguistic (symphonic) conductor*, a person who has learned "each team member's discipline- or profession-specific values, norms, practices, vocabularies, theories, and methods to coordinate and translate between dissimilar members" [70].

Formative Research

We encountered 3 types of formative research challenges in the P2P project: determining when and how to involve stakeholders (formative research challenge 1), recruiting participants and advisors who are representative of stakeholder groups (formative research challenge 2), and fostering meaningful interactions between stakeholders and designers despite personnel constraints (formative research challenge 3). Recommendations for formative research are listed in Textbox 4.

Textbox 4. Recommendations for formative research.

1. Involve stakeholders early.
2. Avoid over recruiting or collecting more data than can be expediently analyzed.
3. Use shorter-cycle iterative research sprints with smaller sample sizes instead [55].
4. Carefully balance expectations placed on stakeholders versus what they are able or willing to do.
5. Deprioritize but do not discard the ideas that stakeholders rejected early on before the ideas reached maturity.
6. When possible, generate involvement from multiple stakeholder groups.
7. Use diverse recruitment methods to ensure stakeholders are chosen for representativeness, not convenience.
8. Foster direct relations between designers and stakeholders.
9. Minimize avoidable personnel changes and practice cross-staffing across user-centered design phases.

Formative Research Challenge 1: Determining Timing and Level of Stakeholder Involvement

We involved patients and clinicians early in the P2P project in several ways, including informant interviews and feedback on analyses, requirements, and early design concepts. This helped learn lessons early, such as clinicians insisting that because of interindividual variability in physical activity, activity goals should be highly individualized, whereas other goals (eg, medication adherence) could be identical for all users. Early learning allowed earlier decisions about scope, facilitated evidence-based design choices, and prevented having to make costly future design revisions. In terms of the scope, early stakeholder involvement helped eliminate especially difficult or risky design concepts, for example, including medication titration advice.

Early involvement was not always simple. The number and depth of P2P interviews yielded more data than could be analyzed in the time allotted. As a result, several important findings from stakeholder interviews were not discovered until further in the design process, negating some of the benefits of early learning. Furthermore, having sought feedback early may have prematurely terminated concepts that were promising but premature at the time.

The forms of involvement vary from collecting extensive data to asking individuals to assess early products [65]. The level of involvement can also be adjusted between informing (as in interviews), advising (as in reviewing concepts), and doing (as in having stakeholders co-perform research or design work) [71]. However, more active or laborious stakeholder involvement risks asking individuals to do more than what is realistic, reasonable, or affordable [21,72]. This is often the case when individuals are asked to be co-designers without adequate training in design, compensation for their contribution, or understanding of the problem space. Although some involvement is essential to UCD, more is not always better [71].

Formative Research Challenge 2: Choosing Stakeholder Representatives

P2P was fortunate to obtain input from patients sampled from a pool of current patients; volunteer advisors who were willing to meet repeatedly with the design team; and various clinicians, some of whom also offered access to their clinic and protocols. Not every design team can easily access stakeholders for formative research in a timely manner, far less multiple stakeholder groups, especially when the stakeholders include busy professionals. Some researchers resort to gathering data from less representative convenience samples, including online services offering access to paid volunteers, such as Amazon's Mechanical Turk or the Qualtrics Panel [73-75].

Despite having adequate access, P2P was also limited in the variety and representativeness of stakeholders. The patients we interviewed in our formative research were all white, and two-third were male. Patient advisors were more likely to be educated and engaged in their health than peers, consistent with the general trend that patient advisors are rarely *ordinary people*. Clinicians in our study may have been more motivated than nonparticipants. Although stakeholder involvement is essential to UCD [1,76], it is predicated on stakeholders having unique knowledge or insights that designers do not have. However, stakeholders too have limited knowledge and represent primarily the communities to which they belong, meaning even with stakeholder involvement, there may exist multiple blind spots. When those who are involved differ from end users (eg, on race, education, or motivation), those blind spots may disadvantage underrepresented groups [63]. In practice, however, few design studies have the opportunity to conduct formative research with large samples representative of the population, whereas increasing sample size exacerbates formative research challenge 1 ("When and how much to involve stakeholders"), as discussed above.

Those who have worked extensively with patient advisory boards offer useful advice on assembling the right group of stakeholders, especially when they must work together, as on a panel. Suggestions include leadership commitment to listening to stakeholder suggestions, diverse recruitment (to avoid the abovementioned blindspots), careful selection of individuals who will work well with others, and adequate funding to compensate or otherwise support stakeholders [77].

Formative Research Challenge 3: Fostering Interactions Between Stakeholders and Designers

We attempted to promote direct interactions between designers and both patient and clinician stakeholders. Designers attended many of the formative research sessions or had direct access to the collected raw data. Furthermore, to ensure continuity, there was cross-staffing of formative research, design, and evaluation teams. One project member personally participated in almost every interview, feedback session, design meeting, and usability test. However, she was the only design team member who had interviewed patients and was therefore expected to be the *voice* of patient participants on the design team. Over time, turnover greatly reduced the number of team members who had been present from the beginning of the project and had therefore participated in any formative research activities.

Having designers interact directly with stakeholders, and especially end users, has been shown to yield better results [78] than hearing about the stakeholders and end users from another source [71]. Continual interaction with stakeholders during the UCD process helps designers gain a firsthand experience of the domain [78,79]. However, when projects progress sequentially from formative research with stakeholders to design and evaluation, turnover and staffing limitations may mean that those designing or testing the product may not have had such firsthand experience.

Design

We identified 3 challenges related to the design phase of UCD: overcoming designers' assumptions with empirical research findings (design challenge 1), managing project scope and complexity and avoiding *scope creep* (design challenge 2), and maintaining the *innovation equilibrium* by balancing new ideas with outside constraints (design challenge 3). Recommendations for design are listed in Textbox 5.

Textbox 5. Recommendations for design.

1. Engage stakeholders during design as ad hoc informants or co-designers to challenge incorrect assumptions.
2. Conduct iterative new rounds of data collection during the design phase as questions arise that are best answered by gathering evidence.
3. Seek simplicity and thus reducing complexity.
4. Monitor for scope creep [80] and overly complex designs, relative to what end users need.
5. Plan for feature deimplementation (ie, removing features from design), using techniques such as a formal *termination plan* [33].
6. Without stifling innovation, ensure stakeholders can rule out designs that are unsafe, unacceptable, infeasible, inconsistent with clinical reality, or otherwise impractical.
7. For innovative ideas transcending conventional practice, develop clear plans for how the design will fit in or overcome existing infrastructure constraints, regulations, preferences, and habits.

Design Challenge 1: Overcoming Designers' Assumptions

Designers naturally make decisions that are inspired by but immediately validated by end user evidence. Some of these decisions are based on assumptions that go unquestioned during design but are discovered to be incorrect during testing. This situation underscores the value of testing and the limitations of design. In the P2P project, for example, we incorporated rewards based on the literature. Nothing in the formative data contradicted the potential value of rewards, so it was not until testing that we learned that most participants thought the rewards were distracting. Some assumptions are also persistent and can be made despite disconfirming formative research findings. For example, members of the design team persistently believed that end users would have little technology experience, despite evidence to the contrary from formative research and usability testing.

Designer bias is difficult to overcome, even when UCD methods are used to collect contradictory evidence. The sequence of design following formative research means some assumptions are not tested or contradicted until the testing phase, by which time the assumptions may have greatly influenced the design. An alternative would be to conduct additional research to challenge the design team's assumptions during the design phase, but before formal testing [81]. Furthermore, designers should be judicious in the use of design techniques, such as personas, which can lead to oversimplification and encourage misleading assumptions about end users [82]. Other strategies to mitigate incorrect assumptions include conducting more frequent testing or including stakeholders on design teams to challenge assumptions during the design process [50,71].

Design Challenge 2: Managing Project Scope and Complexity

Similar to other design projects, P2P produced many ideas, which were easier to generate than to dismiss. As a result, we attempted to include in a single app a large variety of features. We also attempted to integrate these many features to produce a coherent product. Often, multiple features were being slowly designed in parallel, rather than perfecting 1 feature before moving on to the next. These conditions sometimes led to confusion about the purpose of the app. More features also meant less time and effort spent designing or testing each.

Complexity and scope need to be carefully managed to avoid natural tendencies to add (rather than subtract) from taking over. Additional research could be used to help prioritize features and determine which features are attractive to designers but not needed by end users [28]. When a project's scope is intentionally large, steps can be taken to create distinct modules (chunks of features) [83], which provide coherent structure and separation. If complexity is inevitable, the project team will need to plan for more extensive testing by conducting longer sessions or sessions with more users.

Design Challenge 3: Maintaining the Innovation Equilibrium

In our experience, designers, clinician stakeholders, and patient stakeholders were divided on what was possible for and needed from the product being designed. Generally, clinicians were more conservative, preferring to replicate existing practices and avoid less studied or riskier options. For instance, clinicians were more conservative than designers about how much unedited information and control over its interpretation to offer patients. Another point of contention was whether to integrate the product into other health information systems, including electronic medical records. Patients preferred integration, whereas designers were divided on leveraging those systems at the expense of their practical limitations and regulatory constraints. Innovation also conflicted with clinical reality, a case where a patient or designer might envision something that is not technically possible or clinically relevant [32]. For example, the design team assumed an ability to predict heart failure events through CIED data that were beyond publicly available scientific knowledge. Designers' innovative ideas could also be mismatched with what patient end users were used to and could comfortably perform. This may have been the case with patients' dislike of rewards or reluctance to rate their health using standard online rating conventions (eg, out of 5 stars). In general, end users tend to have more conventional preferences than designers [26]. In mHealth projects, patients may be unaware of or reluctant to suggest all the technological possibilities granted by smartphones [29], such as push notifications [84] or smartphone sensors [27].

In conversations with innovators, UCD professionals often hear the statement attributed to Henry Ford, "If I had asked people what they wanted, they would have said faster horses." The broader challenge is maintaining the *innovation equilibrium*: allowing innovators to innovate, while also allowing

stakeholders to influence or evaluate their design, especially when it comes to usability, safety, and privacy. The related challenge is to prevent innovation from creating products that commit what Cornet et al [32] call *type 2 design error*, which "occurs when designers do not accommodate the clinical reality, including biomedical knowledge, clinical workflows, and organizational requirements."

Evaluation

We identified 3 evaluation challenges: managing the tradeoff between laboratory and in-the-wild usability sessions (evaluation challenge 1); adapting standardized methods to the end user population, in our case, older adults (evaluation challenge 2); and deciding on the number of concurrent evaluation methods relative to the effort spent setting up sessions and analyzing data (evaluation challenge 3). Recommendations for evaluation are listed in Textbox 6.

Textbox 6. Recommendations for evaluation.

1. Use a laboratory setup for usability testing to improve efficiency and effectiveness.
2. Begin testing in the laboratory but transition to in-the-wild testing as time and budget allow.
3. Adapt methods to end user needs, when necessary, even if this means deviating from the standard.
4. Allow for flexibility and experimentation, at times sacrificing standardization.
5. Control the number of concurrent evaluation methods; use efficiency, pacing, and workload management strategies if multiple methods are implemented.

Evaluation Challenge 1: Conducting Laboratory or In-the-Wild Usability Sessions

P2P usability testing was conducted in a laboratory setting, albeit in meeting rooms without built-in usability or simulation equipment (eg, a control room, multicamera recording, eye tracking). Although this setting was adequate in most cases, it was at times inconvenient. The laboratory setting was more challenging because it required participants to test prototypes in a time and place dissimilar from the intended context of use. Participants spent 30 min using a prototype technology meant to be used for weeks, months, and years. They were then asked to project how they would use the technology in practice. In the third round of testing, scenarios were used to simulate several days in the prolonged use of the product to help participants project future acceptance and use. However, the cross-sectional and laboratory-based design of our testing limited our confidence in our findings regarding acceptance and future use, relative to findings of usability (eg, observed errors or subjective usability ratings).

Evaluating mHealth prototypes in a laboratory setting offers ideal conditions for detecting product software usability issues, such as navigation or layout issues. Such evaluation however lacks external validity in reproducing the context of the use of the mHealth product [85] and therefore fails to assess most issues related to product hardware usability, operating system usability, acceptance, and longer term outcomes (eg, changes in behavior or health) [6]. Laboratory evaluation is appropriate to quickly iterate on designs and address usability issues before in-the-wild testing to avoid fielding a poorly designed product. However, in-the-wild testing is expensive and time-intensive and may not be possible in every project.

Evaluation Challenge 2: Adapting Methods to End Users

We adapted the standard methods in several ways, including accommodating older adult participants. For example, we administered a simplified version of the System Usability Scale (SUS) self-report measure, which we developed specifically for older adults [9,86,87]. We built flexibility for breaks during testing, especially given the use of diuretic medications by patients with heart failure. We also discovered challenges that we had not anticipated, for example, a participant having difficulty completing computer-based surveys because of vision and motor impairment.

In the context of mHealth projects, adapting standardized usability evaluation methods to end users is often necessary to accommodate patient abilities and limitations. For example, most standardized usability scales have technical or difficult words [86]; thus, many studies edit these measures, for example, by changing the "cumbersome" in the SUS to "awkward" [88]. Although standardized methods ensure scientific reproducibility, rigidity in the UCD process can undermine the goal of iteratively improving a product, which often requires flexibility and experimentation [65]. If, for example, researchers discover that some older participants have difficulty using a touchscreen device, it may be worth adapting the protocol to permit the use of a mouse or stylus in subsequent testing.

Evaluation Challenge 3: Deciding on the Number of Concurrent Evaluation Methods

Our testing involved participant consenting, lengthy pre- and posttest surveys, posttest interviews, and task-based usability testing with think-aloud. In addition, each testing session required a pretest room and equipment setup and posttest aggregation of data from audio, video, computerized, and written recordings. At times, the multiplicity of methods in a single session resulted in testing sessions being cut short. Moreover, the amount of data collected during testing affected the speed at which the team could analyze usability test findings and prepare the next design for another round of testing.

Using multiple concurrent evaluation methods improves triangulation and therefore mHealth usability [89]. However, each method adds burden and affects the timeline. Thus, those implementing UCD should pursue strategies to reduce inefficiency (eg, use of a dedicated testing room to reduce setup labor), ensure pacing (eg, blocking off staff time for testing and

analysis), and reduce workload (eg, use of automated usability data collection or analysis) [90].

Discussion

Limitations

The challenges reported were based on our experience with a P2P project, supplemented by firsthand experience with multiple other mHealth projects and a review of the literature. However, the literature yielded few explicit depictions of challenges and less formal discussion of them. (This validated the goals of this paper.) Both our experiences and most of those described in the literature originated in academic environments, which have unique staffing, timing, and funding characteristics. The UCD implementation challenges and strategies encountered in the industry may be different, although an examination of gray literature (eg, blog posts and popular design books) shows some similarities. Finally, our recommendations are to be taken with caution, as they have not been formally validated across projects, project teams, or environments. The mHealth UCD community should actively debate these recommendations and produce new ones.

Conclusions

UCD implementation for mHealth apps can lead to highly usable and acceptable patient-centered and clinically valid solutions. Implementation is challenging, as the 12 practical challenges in this paper easily illustrate. However, these challenges can be overcome, and our recommendations may help others apply UCD to mHealth or similar arenas. Telling and learning from the typically *untold stories* will result in more efficient, effective, and sustainable mHealth design efforts, effectively bridging the gap between the science and practice of UCD and mHealth implementation. We call on our fellow researchers, designers, and UCD experts to document and share their own challenges and strategies toward improving the implementation of UCD.

Acknowledgments

This study was sponsored by the Agency for Healthcare Research and Quality (AHRQ) grant R21 HS025232 (Holden, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ. The authors acknowledge the participants for their contribution. The authors also thank their entire research team, their patient and clinical advisory board members, and Parkview Health staff for their contribution throughout the project. They also thank the two peer reviewers for their helpful comments on the manuscript.

Conflicts of Interest

MM has research funding from AHRQ and the following financial relationships with industry to disclose: research grants from Medtronic, Inc; consulting fees/honoraria with iRhythm Technologies, Inc; and Zoll Medical Corporation; and financial partnership with Medical Informatics Engineering, outside the submitted work.

References

1. Ergonomics of human-System Interaction — Part 210: Human-Centred Design for Interactive Systems. International Organization for Standardization. 2010. URL: https://www.iso.org/standard/52075.html [accessed 2020-06-10]
2. Gould J, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985;28(3):300-311 [FREE Full text] [doi: 10.1145/3166.3170]
3. Norman D. The Design of Everyday Things: Revised and Expanded Edition. New York, USA: Basic Books; 2013.
4. 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]
5. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
6. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mhealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
7. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
8. Wray T, Kahler C, Simpanen E, Operario D. User-centered, interaction design research approaches to inform the development of health risk behavior intervention technologies. Internet Interv 2019 Mar;15:1-9 [FREE Full text] [doi: 10.1016/j.invent.2018.10.002] [Medline: 30425932]
9. Srinivas P, Cornet VP, Holden RJ. Human factors analysis, design, and evaluation of engage, a consumer health IT application for geriatric heart failure self-care. Int J Hum Comput Interact 2017;33(4):298-312 [FREE Full text] [doi: 10.1080/10447318.2016.1265784] [Medline: 30429638]

10. Giunti G, Mylonopoulou V, Romero OR. More stamina, a gamified mhealth solution for persons with multiple sclerosis: research through design. JMIR Mhealth Uhealth 2018 Mar 2;6(3):e51 [FREE Full text] [doi: 10.2196/mhealth.9437] [Medline: 29500159]
11. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Lavigne AS, et al. A patient-centered mobile health system that supports asthma self-management (breathe): design, development, and utilization. JMIR Mhealth Uhealth 2019 Jan 28;7(1):e10956 [FREE Full text] [doi: 10.2196/10956] [Medline: 30688654]
12. Vilardaga R, Rizo J, Zeng E, Kientz J, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
13. Biagianti B, Hidalgo-Mazzei D, Meyer N. Developing digital interventions for people living with serious mental illness: perspectives from three mhealth studies. Evid Based Ment Health 2017 Nov;20(4):98-101 [FREE Full text] [doi: 10.1136/eb-2017-102765] [Medline: 29025862]
14. Nielsen J. Usability Engineering. Cambridge, MA: AP Professional; 1994.
15. Rubin J, Chisnell D. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. Indianapolis, IN: Wiley Publishing; 2008.
16. Crandall B, Klein GA, Klein G, Hoffman RR. Working Minds: A Practitioner's Guide to Cognitive Task Analysis. Cambridge, MA: MIT Press; 2006.
17. Simonsen J, Robertson T, editors. Routledge International Handbook of Participatory Design. New York, USA: Routledge; 2012.
18. Holden RJ, Voida S, Savoy A, Jones JF, Kulanthaivel A. Human factors engineering and human–computer interaction: supporting user performance and experience. In: Finnel JT, Dixon BE, editors. Clinical Informatics Study Guide: Text and Review. Cham, Switzerland: Springer International Publishing; 2016:287-307.
19. Cornet VP, Daley C, Cavalcanti LH, Parulekar A, Holden RJ. Design for self care. In: Sethumadhavan A, Sasangohar F, editors. Design for Health: Applications of Human Factors. Cambridge, MA: Academic Press; 2020:277-302.
20. Kujala S, Kauppinen M. Identifying and Selecting Users for User-Centered Design. In: Proceedings of the third Nordic Conference on Human-Computer Interaction. 2004 Presented at: NordiCHI'04; October 23-37, 2004; Tampere, Finland. [doi: 10.1145/1028014.1028060]
21. Heinbokel T, Sonnentag S, Frese M, Stolte W, Brodbeck F. Don't underestimate the problems of user centredness in software development projects - there are many!. Behav Inf Technol 1996 Jan;15(4):226-236 [FREE Full text] [doi: 10.1080/014492996120157]
22. Shah LM, Yang WE, Demo RC, Lee MA, Weng D, Shan R, et al. Technical guidance for clinicians interested in partnering with engineers in mobile health development and evaluation. JMIR Mhealth Uhealth 2019 May 15;7(5):e14124 [FREE Full text] [doi: 10.2196/14124] [Medline: 31094337]
23. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 9;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
24. 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]
25. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2019:139-151 [FREE Full text] [doi: 10.3233/978-1-61499-635-4-139]
26. van Velsen L, Wentzel J, van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
27. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
28. Mann DM, Quintiliani LM, Reddy S, Kitos NR, Weng M. Dietary approaches to stop hypertension: lessons learned from a case study on the development of an mHealth behavior change system. JMIR Mhealth Uhealth 2014 Oct 23;2(4):e41 [FREE Full text] [doi: 10.2196/mhealth.3307] [Medline: 25340979]
29. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
30. Albrecht U, von Jan U. Safe, sound and desirable: development of mHealth apps under the stress of rapid life cycles. Mhealth 2017;3:27 [FREE Full text] [doi: 10.21037/mhealth.2017.06.05] [Medline: 28828374]
31. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
32. Cornet VP, Daley C, Bolchini D, Toscos T, Mirro MJ, Holden RJ. Patient-centered design grounded in user and clinical realities: towards valid digital health. Proc Int Symp Hum Fact Erg Health Care 2019 Sep 15;8(1):100-104 [FREE Full text] [doi: 10.1177/2327857919081023]

33. Boustani MA, van der Marck MA, Adams N, Azar JM, Holden RJ, Vollmar HC, et al. Developing the agile implementation playbook for integrating evidence-based health care services into clinical practice. Acad Med 2019;94(4):556-561. [doi: 10.1097/ACM.0000000000002497]
34. Waterson P. Ergonomics and ergonomists: lessons for human factors and ergonomics practice from the past and present. In: Shorrock S, Williams C, editors. Human Factors and Ergonomics in Practice: Improving System Performance and Human Well-Being in the Real World. Boca Raton, FL: CRC Press; 2017:29-44.
35. Roger VL. Epidemiology of heart failure. Circ Res 2013 Aug 30;113(6):646-659 [FREE Full text] [doi: 10.1161/CIRCRESAHA.113.300268] [Medline: 23989710]
36. Lainscak M, Blue L, Clark A, Dahlström U, Dickstein K, Ekman I, et al. Self-care management of heart failure: practical recommendations from the patient care committee of the heart failure association of the European society of cardiology. Eur J Heart Fail 2011 Feb;13(2):115-126 [FREE Full text] [doi: 10.1093/eurjhf/hfq219] [Medline: 21148593]
37. Holden RJ, Schubert CC, Mickelson RS. The patient work system: an analysis of self-care performance barriers among elderly heart failure patients and their informal caregivers. Appl Ergon 2015 Mar;47:133-150 [FREE Full text] [doi: 10.1016/j.apergo.2014.09.009] [Medline: 25479983]
38. Klersy C, Boriani G, de Silvestri A, Mairesse G, Braunschweig F, Scotti V, Health Economics Committee of the European Heart Rhythm Association. Effect of telemonitoring of cardiac implantable electronic devices on healthcare utilization: a meta-analysis of randomized controlled trials in patients with heart failure. Eur J Heart Fail 2016 Feb;18(2):195-204 [FREE Full text] [doi: 10.1002/ejhf.470] [Medline: 26817628]
39. Hawkins NM, Virani SA, Sperrin M, Buchan IE, McMurray JJ, Krahn AD. Predicting heart failure decompensation using cardiac implantable electronic devices: a review of practices and challenges. Eur J Heart Fail 2016 Aug;18(8):977-986 [FREE Full text] [doi: 10.1002/ejhf.458] [Medline: 26663507]
40. Mirro M, Daley C, Wagner S, Ghahari RR, Drouin M, Toscos T. Delivering remote monitoring data to patients with implantable cardioverter-defibrillators: does medium matter? Pacing Clin Electrophysiol 2018 Nov;41(11):1526-1535. [doi: 10.1111/pace.13505] [Medline: 30225880]
41. Daley CN, Allmandinger A, Heral L, Toscos T, Plant R, Mirro M. Engagement of ICD patients: direct electronic messaging of remote monitoring data via a personal health record. EP Lab Digest 2015;15(5):- [FREE Full text]
42. Daley CN, Chen EM, Roebuck AE, Ghahari RR, Sami AF, Skaggs CG, et al. Providing patients with implantable cardiac device data through a personal health record: a qualitative study. Appl Clin Inform 2017 Oct;8(4):1106-1116 [FREE Full text] [doi: 10.4338/ACI-2017-06-RA-0090] [Medline: 29241248]
43. Rohani Ghahari R, Holden RJ, Flanagan M, Wagner S, Martin E, Ahmed R, et al. Using cardiac implantable electronic device data to facilitate health decision making: a design study. Int J Ind Ergon 2018 Mar;64:143-154 [FREE Full text] [doi: 10.1016/j.ergon.2017.11.002]
44. Daley CN, Bolchini D, Varrier A, Rao K, Joshi P, Blackburn J, et al. Naturalistic decision making by older adults with chronic heart failure: an exploratory study using the critical incident technique. Proc Hum Factors Ergon Soc Annu Meet 2018 Sep 27;62(1):568-572 [FREE Full text] [doi: 10.1177/1541931218621130]
45. Daley CN, Cornet VP, Patekar G, Kosarabe S, Bolchini D, Toscos T, et al. Uncertainty management among older adults with heart failure: responses to receiving implanted device data using a fictitious scenario interview method. Proc Int Sym Hum Fact Erg Health Care 2019 Sep 15;8(1):127-130 [FREE Full text] [doi: 10.1177/2327857919081030]
46. Holden RJ, Joshi P, Rao K, Varrier A, Daley CN, Bolchini D, et al. Modeling personas for older adults with heart failure. Proc Hum Factors Ergon Soc Annu Meet 2018 Sep 27;62(1):1072-1076 [FREE Full text] [doi: 10.1177/1541931218621246]
47. Holden RJ, Daley CN, Mickelson RS, Bolchini D, Toscos T, Cornet VP, et al. Patient decision-making personas: an application of a patient-centered cognitive task analysis (P-CTA). Appl Ergon 2020 Apr 15;87:103107 [FREE Full text] [doi: 10.1016/j.apergo.2020.103107] [Medline: 32310109]
48. Kain D. Owning significance: the critical incident technique in research. In: de Marrais K, Lapan SD, editors. Foundations for Research: Methods of Inquiry in Education and the Social Sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:69-85.
49. Flanagan JC. The critical incident technique. Psychol Bull 1954 Jul;51(4):327-358. [doi: 10.1037/h0061470] [Medline: 13177800]
50. Ahmed R, Toscos T, Ghahari RR, Holden RJ, Martin E, Wagner S, et al. Visualization of cardiac implantable electronic device data for older adults using participatory design. Appl Clin Inform 2019 Aug;10(4):707-718. [doi: 10.1055/s-0039-1695794] [Medline: 31533172]
51. Hart S, Staveland L. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol 1988;52:139-183 [FREE Full text] [doi: 10.1016/s0166-4115(08)62386-9]
52. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, England: Taylor & Francis; 1996:189-194.
53. Boehmer JP, Hariharan R, Devecchi FG, Smith AL, Molon G, Capucci A, et al. A multisensor algorithm predicts heart failure events in patients with implanted devices: results from the multisense study. JACC Heart Fail 2017 Mar;5(3):216-225 [FREE Full text] [doi: 10.1016/j.jchf.2016.12.011] [Medline: 28254128]

54. Gardner RS, Singh JP, Stancak B, Nair DG, Cao M, Schulze C, et al. Heartlogic multisensor algorithm identifies patients during periods of significantly increased risk of heart failure events: results from the multisense study. Circ Heart Fail 2018 Jul;11(7):e004669. [doi: 10.1161/CIRCHEARTFAILURE.117.004669] [Medline: 30002113]
55. Nielsen J, Landauer TK. A Mathematical Model of the Finding of Usability Problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Presented at: INTERCHI '93; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213 URL: https://dl.acm.org/doi/10.1145/169059.169166 [doi: 10.1145/169059.169166]
56. Nielsen J. Iterative user-interface design. Computer 1993 Nov;26(11):32-41 [FREE Full text] [doi: 10.1109/2.241424]
57. Baez M, Casati F. Agile Development for Vulnerable Populations: Lessons Learned and Recommendations. In: Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Society. 2018 Presented at: ICSE-SEIS'18; May 27-June 3, 2018; Gothenburg, Sweden. [doi: 10.1145/3183428.3183439]
58. Holden RJ, Bodke K, Tambe R, Comer RS, Clark DO, Boustani M. Rapid translational field research approach for eHealth R&D. Proc Int Sym Hum Fact Erg Health Care 2016 Jul 22;5(1):25-27 [FREE Full text] [doi: 10.1177/2327857916051003]
59. Holden RJ, Scott AM, Hoonakker PLT, Hundt AS, Carayon P. Data collection challenges in community settings: insights from two field studies of patients with chronic disease. Qual Life Res 2015 May;24(5):1043-1055 [FREE Full text] [doi: 10.1007/s11136-014-0780-y] [Medline: 25154464]
60. Holden RJ, Or CK, Alper SJ, Joy Rivera A, Karsh BT. A change management framework for macroergonomic field research. Appl Ergon 2008 Jul;39(4):459-474 [FREE Full text] [doi: 10.1016/j.apergo.2008.02.016] [Medline: 18417095]
61. Valdez RS, Holden RJ. Health care human factors/ergonomics fieldwork in home and community settings. Ergon Des 2016 Oct;24(4):4-9 [FREE Full text] [doi: 10.1177/1064804615622111] [Medline: 28781512]
62. Toscos T, Drouin M, Pater J, Flanagan M, Pfafman R, Mirro MJ. Selection biases in technology-based intervention research: patients' technology use relates to both demographic and health-related inequities. J Am Med Inform Assoc 2019 Aug 1;26(8-9):835-839. [doi: 10.1093/jamia/ocz058] [Medline: 31172193]
63. Holden RJ, Toscos T, Daley CN. Researcher reflections on human factors and health equity. In: Roscoe R, Chiou EK, Wooldridge AR, editors. Advancing Diversity, Inclusion, and Social Justice Through Human Systems Engineering. Boca Raton, FL: CRC Press; 2019:51-62.
64. Cosco TD, Firth J, Vahia I, Sixsmith A, Torous J. Mobilizing mHealth data collection in older adults: challenges and opportunities. JMIR Aging 2019 Mar 19;2(1):e10019 [FREE Full text] [doi: 10.2196/10019] [Medline: 31518253]
65. Pagliari C. Design and evaluation in ehealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
66. Poole ES. HCI and mobile health interventions: how human-computer interaction can contribute to successful mobile health interventions. Transl Behav Med 2013 Dec;3(4):402-405 [FREE Full text] [doi: 10.1007/s13142-013-0214-3] [Medline: 24294328]
67. Aidemark J, Askenäs L, Nygårdh A, Strömberg A. User involvement in the co-design of self-care support systems for heart failure patients. Procedia Comput Sci 2015;64:118-124 [FREE Full text] [doi: 10.1016/j.procs.2015.08.471]
68. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
69. Calvo RA, Dinakar K, Picard R, Christensen H, Torous J. Toward impactful collaborations on computing and mental health. J Med Internet Res 2018 Feb 9;20(2):e49 [FREE Full text] [doi: 10.2196/jmir.9021] [Medline: 29426812]
70. Holden RJ, Binkheder S, Patel J, Viernes SH. Best practices for health informatician involvement in interprofessional health care teams. Appl Clin Inform 2018 Jan;9(1):141-148 [FREE Full text] [doi: 10.1055/s-0038-1626724] [Medline: 29490407]
71. Kujala S. User involvement: a review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16 [FREE Full text] [doi: 10.1080/01449290301782]
72. Baek EO, Cagiltay K, Boling E, Frick T. User-centered design and development. In: Spector JM, Merrill MD, van Merrienboer J, Driscoll MP, editors. Handbook of Research on Educational Communications and Technology: A Project of the Association for Educational Communications and Technology. New York, USA: Taylor and Francis; 2008:660-668.
73. Kittur A, Chi EH, Suh BW. Crowdsourcing user studies with mechanical turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI'08; April 5-10, 2008; Florence, Italy URL: https://dl.acm.org/doi/10.1145/1357054.1357127 [doi: 10.1145/1357054.1357127]
74. Guillory J, Kim A, Murphy J, Bradfield B, Nonnemaker J, Hsieh Y. Comparing Twitter and online panels for survey recruitment of e-cigarette users and smokers. J Med Internet Res 2016 Nov 15;18(11):e288 [FREE Full text] [doi: 10.2196/jmir.6326] [Medline: 27847353]
75. Walter SL, Seibert SE, Goering D, O'Boyle EH. A tale of two sample sources: do results from online panel data and conventional data converge? J Bus Psychol 2018 Jul 21;34(4):425-452. [doi: 10.1007/s10869-018-9552-y]
76. Eshet E, Bouwman H. Context of use: the final frontier in the practice of user-centered design? Interact Comput 2016 Oct 13;29:368-390 [FREE Full text] [doi: 10.1093/iwc/iww030]
77. Sharma A, Angel L, Bui Q. Patient advisory councils: giving patients a seat at the table. Fam Pract Manag 2015;22(4):22-27 [FREE Full text] [Medline: 26176505]

78. Keil M, Carmel E. Customer-developer links in software development. Commun ACM 1995;38(5):33-44 [FREE Full text] [doi: 10.1145/203356.203363]
79. Béguin P. Design as a mutual learning process between users and designers. Interact Comput 2003 Oct;15(5):709-730 [FREE Full text] [doi: 10.1016/s0953-5438(03)00060-2]
80. Lorenzi NM, Riley RT. Managing change: an overview. J Am Med Inform Assoc 2000;7(2):116-124 [FREE Full text] [doi: 10.1136/jamia.2000.0070116] [Medline: 10730594]
81. Magrabi F, Ong M, Coiera E. Health IT for patient safety and improving the safety of health IT. Stud Health Technol Inform 2016;222:25-36. [Medline: 27198089]
82. Turner P, Turner S. Is stereotyping inevitable when designing with personas? Design Stud 2011 Jan;32(1):30-44 [FREE Full text] [doi: 10.1016/j.destud.2010.06.002]
83. Norman D. Living with Complexity. Cambridge, Massachusetts: The MIT Press; 2011.
84. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Danish HS, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
85. Cho H, Yen PY, Dowding D, Merrill JA, Schnall R. A multi-level usability evaluation of mobile health applications: a case study. J Biomed Inform 2018 Oct;86:79-89 [FREE Full text] [doi: 10.1016/j.jbi.2018.08.012] [Medline: 30145317]
86. Cornet VP, Daley CN, Srinivas P, Holden RJ. User-centered evaluations with older adults: testing the usability of a mobile health system for heart failure self-management. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep;61(1):6-10 [FREE Full text] [doi: 10.1177/1541931213601497] [Medline: 30930610]
87. Holden RJ, Campbell NL, Abebe E, Clark DO, Ferguson D, Bodke K, Brain Health Patient Safety Laboratory. Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults. Res Social Adm Pharm 2020 Jan;16(1):54-61. [doi: 10.1016/j.sapharm.2019.02.011] [Medline: 30853507]
88. Bangor A, Kortum P, Miller J. An empirical evaluation of the system usability scale. Int J Hum-Comput Interact 2008 Jul 30;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
89. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. J Biomed Inform 2016 Feb;59:115-129 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.008] [Medline: 26639894]
90. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CIED: cardiac implanted electronic device
mHealth: mobile health
P2P: Power to the Patient
SUS: system usability scale
UCD: user-centered design

Edited by G Eysenbach; submitted 06.01.20; peer-reviewed by A Khoja, K Yin; comments to author 10.04.20; revised version received 20.04.20; accepted 14.05.20; published 21.07.20.

Please cite as:
Cornet VP, Toscos T, Bolchini D, Rohani Ghahari R, Ahmed R, Daley C, Mirro MJ, Holden RJ
Untold Stories in User-Centered Design of Mobile Health: Practical Challenges and Strategies Learned From the Design and Evaluation of an App for Older Adults With Heart Failure
JMIR Mhealth Uhealth 2020;8(7):e17703
URL: http://mhealth.jmir.org/2020/7/e17703/
doi:10.2196/17703
PMID:32706745

Original Paper

Theme Trends and Knowledge Structure on Mobile Health Apps: Bibliometric Analysis

Cheng Peng[1], MD, PhD; Miao He[2], BSc; Sarah L Cutrona[3,4], MD, MPH; Catarina I Kiefe[3], MD, PhD; Feifan Liu[3], PhD; Zhongqing Wang[2,3], BSc, MBA

[1]Department of Ophthalmology, The Fourth Affiliated Hospital of China Medical University, Shenyang, China

[2]Department of Information Center, The First Hospital of China Medical University, Shenyang, China

[3]Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[4]Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

Corresponding Author:
Zhongqing Wang, BSc, MBA
Department of Information Center
The First Hospital of China Medical University
155 Nanjinbei Street
Shenyang
China
Phone: 86 15940082159
Email: wangzhongqing@cmu.edu.cn

Abstract

Background: Due to the widespread and unprecedented popularity of mobile phones, the use of digital medicine and mobile health apps has seen significant growth. Mobile health apps have tremendous potential for monitoring and treating diseases, improving patient care, and promoting health.

Objective: This paper aims to explore research trends, coauthorship networks, and the research hot spots of mobile health app research.

Methods: Publications related to mobile health apps were retrieved and extracted from the Web of Science database with no language restrictions. Bibliographic Item Co-Occurrence Matrix Builder was employed to extract bibliographic information (publication year and journal source) and perform a descriptive analysis. We then used the VOSviewer (Leiden University) tool to construct and visualize the co-occurrence networks of researchers, research institutions, countries/regions, citations, and keywords.

Results: We retrieved 2802 research papers on mobile health apps published from 2000 to 2019. The number of annual publications increased over the past 19 years. *JMIR mHealth and uHealth* (323/2802, 11.53%), *Journal of Medical Internet Research* (106/2802, 3.78%), and *JMIR Research Protocols* (82/2802, 2.93%) were the most common journals for these publications. The United States (1186/2802, 42.33%), England (235/2802, 8.39%), Australia (215/2802, 7.67%), and Canada (112/2802, 4.00%) were the most productive countries of origin. The University of California San Francisco, the University of Washington, and the University of Toronto were the most productive institutions. As for the authors' contributions, Schnall R, Kuhn E, Lopez-Coronado M, and Kim J were the most active researchers. The co-occurrence cluster analysis of the top 100 keywords forms 5 clusters: (1) the technology and system development of mobile health apps; (2) mobile health apps for mental health; (3) mobile health apps in telemedicine, chronic disease, and medication adherence management; (4) mobile health apps in health behavior and health promotion; and (5) mobile health apps in disease prevention via the internet.

Conclusions: We summarize the recent advances in mobile health app research and shed light on their research frontier, trends, and hot topics through bibliometric analysis and network visualization. These findings may provide valuable guidance on future research directions and perspectives in this rapidly developing field.

(JMIR Mhealth Uhealth 2020;8(7):e18212) doi:10.2196/18212

KEYWORDS

mobile app; mobile health; mhealth; digital health; digital medicine; bibliometrics; co-word analysis; mobile phone; VOSviewer

Introduction

Worldwide, the use of mobile phones has reached widespread popularity at an unprecedented rate [1]. There were more than 325,000 mobile health apps in 2017 [2]. According to Zion Market Research, the global mobile health app market will grow from $8.0 billion in 2018 to $111.1 billion by 2025 [3]. Recently, mobile health apps have seen significant growth. An increasing number of hospitals and health care institutions are using mobile health apps to monitor the development of diseases and improve health care outcomes [4]. Therefore, it is essential to understand the use, significance, trends, and research hot spots of mobile health apps in the health care domain.

Bibliometric analysis has been widely used in quantitative analysis of academic literature to describe the hot spots, trends, and contributions of scholars, journals, and countries/regions [5-8]. Co-word analysis, proposed in the late 1970s [9,10] as an important bibliometric technique, can identify the main themes, investigate hot spots, and detect knowledge in literature. Thus, bibliometrics can contribute to monitoring the development and patterns of effective publications [11]. In recent years, bibliometric analysis has been applied to biomedicine and health care [12-15]. In the current study, we used bibliometric quantitative analysis and network visualization to describe the research trends, research hot spots, emerging topics, and collaboration partners in the field of mobile health apps. Our study is the first one to quantitatively analyze the characteristics and hot topics of mobile health app research. Our study may provide valuable guidance on future research directions in this rapidly developing field.

Methods

Data Collection

Web of Science (WOS) is an extensive international database of academic information, including more than 9000 prestigious and high-impact research journals from all over the world. WOS contains various characteristics that can be used for bibliometric study, including title, author, institution, country/region, publication year, and keywords [9]. WOS has been recently receiving more attention as a reliable data source for bibliometric analysis in the biomedical domain, with applications in clinical and bench science research questions (eg, cardiovascular disease, diabetic kidney disease, and long noncoding RNA) [8,16,17]. On October 6, 2019, we conducted a publication search in WOS to find publications using the following search strategy: TS="mobile health app*" OR (TS="mobile app*" AND TS=("health*" OR "medic*" OR "clinic*" OR "hospital*")). Only full-length papers were included, and no language limitation was set. We validated the reliability of our search strategy by manually reviewing the retrieved publications [18]. All data from retrieved publications were collected and saved in TXT formats.

Data Analysis and Visualization Maps

We aimed to exploit bibliometric analysis to identify the knowledge structure, research frontiers, research hot spots, active authors, and other bibliometric information in the mobile health app area. Bibliometric analysis typically consists of the construction of bibliometric maps and the graphical representation of such maps [19]. Co-word analysis was used to calculate the frequency of co-occurrence of bibliographic information and perform hierarchical clustering based on the co-occurrence information [9,10]. Finally, the clusters were visualized graphically.

In this study, we have applied widely used bibliometric analysis tools on the WOS data. Bibliographic Item Co-Occurrence Matrix Builder version 2.0 [10] was used to extract and analyze bibliographic information on the publication years and the journal sources. VOSviewer (version 1.6.13; Leiden University) was used to extract bibliographic information on researchers, institutions, countries/regions, references, and keywords. VOSviewer uses the visualization of similarities mapping technique, which produces better structured maps than other popular multidimensional scaling techniques for bibliometrics [19]. Specifically, when constructing a map, VOSviewer takes as input a similarity matrix that is created using a similarity measure known as the association strength [20]. It calculates the similarity s_{ij} of two items i and j with the equation $s_{ij} = c_{ij} / (w_i w_j)$, where c_{ij} denotes the number of co-occurrences of items i and j, and where w_i and w_j denote the total number of occurrences of items i and j. Once the similarity matrix is created, VOSviewer maps all the items in a 2-dimensional map so that items with a high similarity will be located close to each other, while items with a low similarity will be located far from each other. Unlike other map-viewing programs, VOSviewer pays special attention to the graphical representation of bibliometric maps in an easy-to-interpret way [19].

Using network-mapping techniques, we created different bibliometric maps that included coauthorships of authors, institutions, and countries/regions; co-citations of references; and co-occurrence of keywords. Each node in a map is represented by a circle with a label. Larger circles indicate higher-frequency items. The color of each circle is determined by the clusters it belongs to. The thickness and length of links between nodes represent the association strength between corresponding nodes. A maximum of 500 lines was set to display the 500 strongest links between nodes.

Research Ethics

Data from bibliographic information were searched and downloaded from WOS. These were publicly available data. The extraction of these data did not involve interaction with human subjects or animals. Thus, there were no ethical issues involving the use of these data, and no approval from an ethics committee was required.

Results

Publication Outputs

Based on our search strategy, we identified and incorporated 2802 publications on mobile health apps from WOS. The number of annual publications on mobile health apps increased from 2 publications in 2000 to 692 publications in 2018 (2019 data are incomplete because they reflect only approximately 9 months of publications). Before 2013, the number of annual

publications did not exceed 100. However, the number of annual publications in 2014, 2015, 2016, 2017, and 2018 was 122, 263, 430, 507, and 692, respectively.

Distribution of Source Journals

Publications on mobile health app research were distributed across 1209 journals; 848 of these journals have published only 1 paper on mobile health apps. Table 1 lists the top 10 journals on this topic. *JMIR mHealth and uHealth* published the most papers (323/2802, 11.53%), followed by *Journal of Medical Internet Research* (106/2802, 3.78%), then *JMIR Research Protocols* (82/2802, 2.93%). The top 10 journals published 776 publications, accounting for 27.64% of all publications in this study.

Table 1. Top 10 journals publishing research on mobile health app research, 2000-2019.

Rank	Journal	Country	Categories	Publications, n	Percentage[a]
1	*JMIR mHealth and uHealth*	Canada	Medical informatics	323	11.53
2	*Journal of Medical Internet Research*	Canada	Medical informatics	106	3.78
3	*JMIR Research Protocols*	Canada	Medical informatics	82	2.93
4	*Plos One*	United States	Multidisciplinary sciences	47	1.68
5	*Journal of Medical Systems*	United States	Medical informatics	43	1.53
6	*BMC Medical Informatics and Decision Making*	England	Medical informatics	37	1.32
7	*International Journal of Medical Informatics*	Ireland	Medical informatics	37	1.32
8	*Telemedicine and e-Health*	United States	Health care sciences and services	37	1.32
9	*JMIR Mental Health*	Canada	Medical informatics and mental health	33	1.18
10	*BMJ Open*	England	Medicine, general and internal	31	1.11

[a]The total number of retrieved papers on mobile health apps from 2000 to 2019 (N=2802) was used as the denominator.

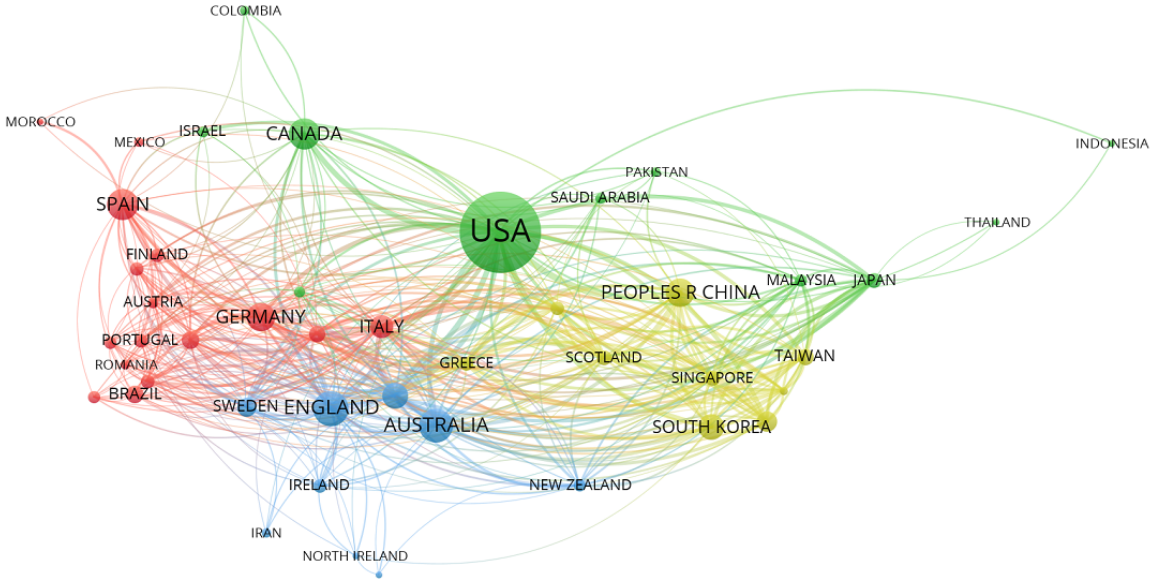
Distribution and Coauthorship of Countries/Regions

According to the search results, 2802 publications came from 104 countries/regions. Figure 1 shows the location of the top 30 countries/regions that were publishing mobile health app research. The United States has the largest number of publications (1186/2802, 42.33%) and England ranks second (235/2802, 8.39%), followed by Australia (215/2802, 7.67%) and Canada (112/2802, 4.00%).

Figure 1. Top 30 countries/regions publishing mobile health app research, 2000-2019.

As shown in Figure 2, the coauthorship analysis of countries/regions reflects the collaboration relationship between countries/regions in this field, as well as the degree of collaboration. The larger nodes represent more productive countries/regions in this field; the thickness and length of links between nodes represent the cooperative relationship between countries/regions. Figure 2 shows the 45 most productive countries/regions in this field from 4 collaboration clusters, which were distinguished by different colors.

Figure 2. The coauthorship network of countries/regions that contributed to mobile health app research, 2000-2019. Peoples R China: People's Republic of China. USA: United States of America.



Distribution and Coauthorship of Institutions

According to the search results, 3795 research institutions contributed to mobile health app research. Table 2 presents the top 10 most productive institutions in mobile health app research. The University of California San Francisco (67 publications) ranked first among all institutions identified, followed by the University of Washington (58 publications) and the University of Toronto (56 publications).

Table 2. Top 10 most productive institutions in mobile health app research, 2000-2019.

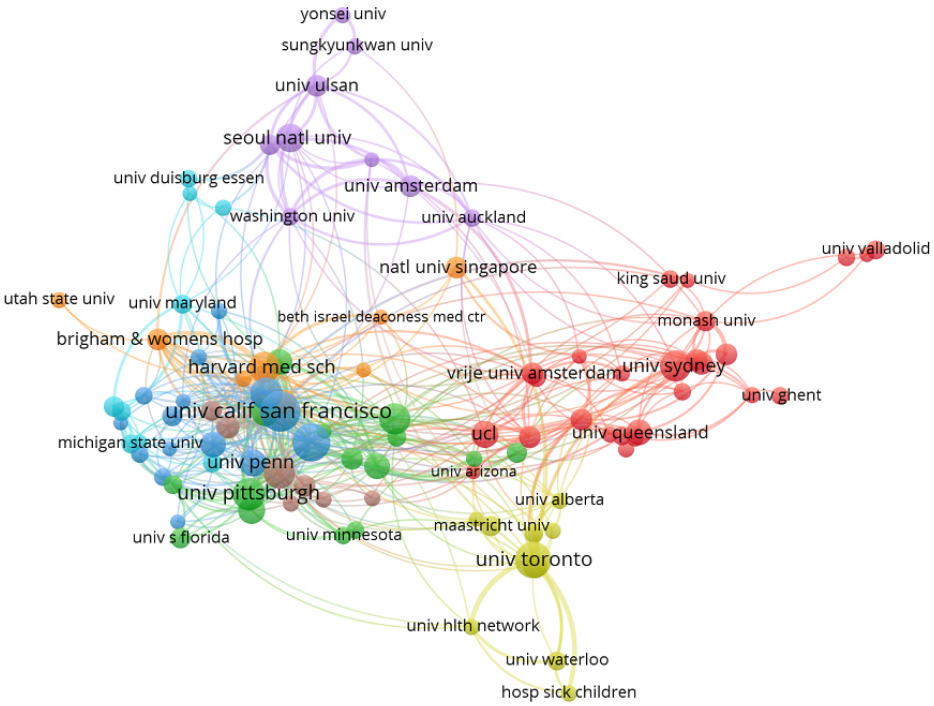
Rank	Institution	Country	Publications, n	Citations, n
1	Univ[a] of California San Francisco	United States	67	819
2	Univ of Washington	United States	58	511
3	Univ of Toronto	Canada	56	640
4	Stanford Univ	United States	46	432
5	Univ of Pittsburgh	United States	45	409
6	Harvard Medical School	United States	42	408
7	Columbia Univ	United States	39	387
8	Northwestern Univ	United States	39	296
9	Univ of Sydney	Australia	39	240
10	Seoul National Univ	South Korea	33	184

[a]Univ: university.

Coauthorship analysis was performed by VOSviewer to display the visualization network map of institutions in mobile health app research. The link between institutions is determined by the number of publications coauthored between them. The

coauthorship analysis of institutions shows that 99 institutions, each of which published at least 10 papers, formed 8 clusters. These clusters are shown in Figure 3 and distinguished by different colors.

Figure 3. The coauthorship network of institutions that contributed to mobile health app research, 2000-2019. Univ: university.



Distribution and Coauthorship of Authors

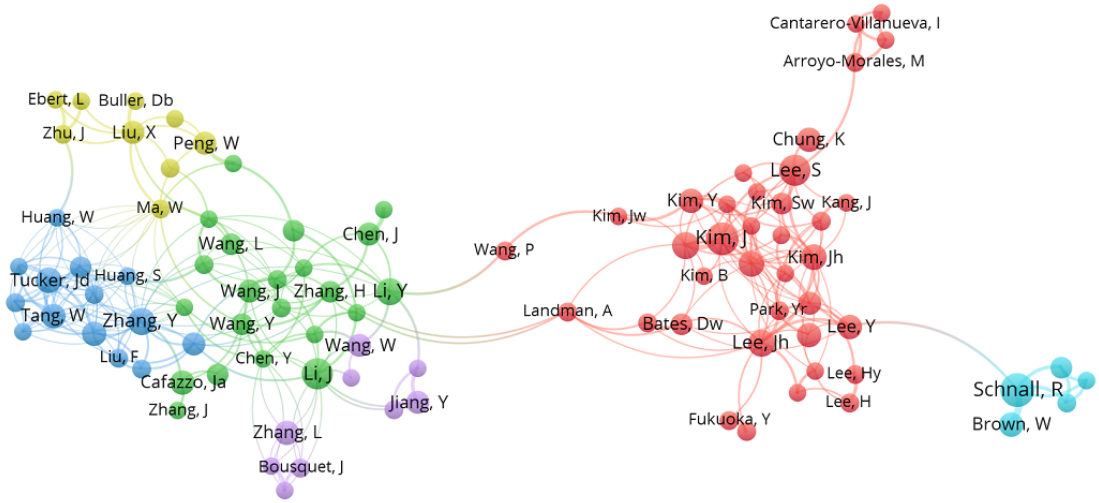
According to the search results, 2802 mobile health app publications were written by 13,040 authors, with an average of 5 authors per publication. Table 3 presents the top 10 most productive authors (all authors of each publication were ranked equally) in mobile health app research. Schnall R (15 publications) ranked first among all authors, followed by Kuhn E (14 publications), Lopez-Coronado M (14 publications), and Kim J (14 publications).

Table 3. Top 10 most productive authors in mobile health app research, 2000-2019.

Rank	Author	Publications, n	Citations, n
1	Schnall R	15	217
2	Kuhn E	14	223
3	Lopez-Coronado M	14	130
4	Kim J	14	43
5	Lee S	13	365
6	Li J	13	49
7	Torous J	12	271
8	Lee JH	10	107
9	Lee J	10	46
10	Zhang Y	10	36

Our coauthorship analysis of authors showed that 221 of 13,040 authors had published at least 4 papers, and the largest set of associated authors consisted of 95 authors in 6 clusters. The node label shows the author's name, and the node size represents the number of published publications. Links connecting 2 nodes represent coauthorship between the 2 authors, and thicker links represent more collaboration between the 2 authors, as shown in Figure 4.

Figure 4. The coauthorship network of authors who contributed to mobile health app research, 2000-2019.



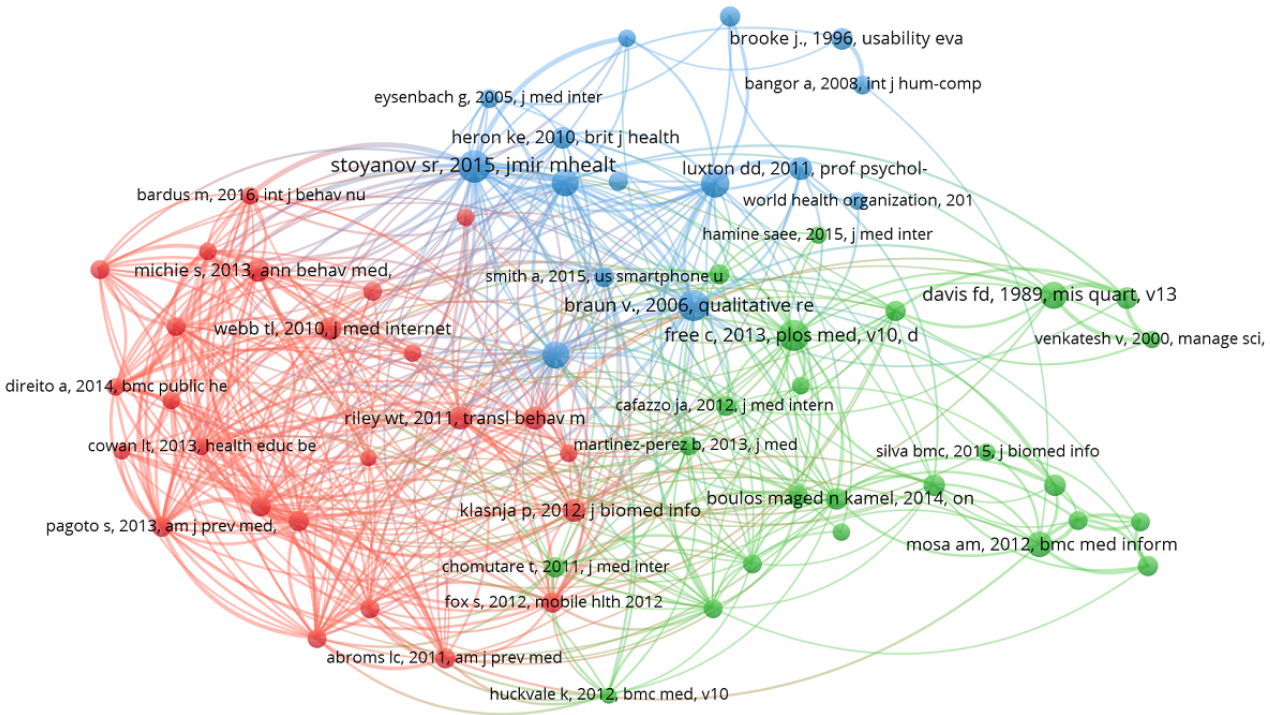
Reference Co-Citation Analysis

Through the co-citation analysis (examining references cited in publications), we explored the knowledge base for the mobile health apps field. We identified 2802 mobile health app publications, which cited 76,721 references, averaging 27 references per publication. The top 10 most frequently cited references are listed in Table 4. The publication that received the most citations, Stoyanov and colleagues' "Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps," was published in *JMIR mHealth and uHealth* in 2015 and received a total of 106 citations as of October 6, 2019. We chose the top 64 references, which were cited at least 30 times by the retrieved papers, to generate a visualization network map with VOSviewer of co-cited references in mobile health app research. This visualization network showed 3 main clusters marked in different colors, as shown in Figure 5.

Table 4. Top 10 cited references in mobile health app research, 2000-2019.

Rank	Author	Journal	Title	Citations, n
1	Stoyanov SR et al (2015)	*JMIR mHealth and uHealth*	Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps	106
2	Free C et al (2013)	*Public Library of Science Medicine*	The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis	101
3	Braun V and Clarke V (2006)	*Qualitative Research in Psychology*	Using Thematic Analysis in Psychology	98
4	Donker T et al (2013)	*Journal of Medical Internet Research*	Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review	85
5	Krebs P and Duncan DT(2015)	*JMIR mHealth and uHealth*	Health App Use Among US Mobile Phone Owners: A National Survey	79
6	Davis FD (1989)	*Management Information Systems Quarterly*	Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology	78
7	Dennison L et al (2013)	*Journal of Medical Internet Research*	Opportunities and Challenges for Smartphone Applications in Supporting Health Behavior Change: Qualitative Study	77
8	Mosa AM et al (2012)	*BMC Medical Informatics and Decision Making*	A Systematic Review of Healthcare Applications for Smartphones	71
9	Luxton DD et al (2011)	*Professional Psychology: Research and Practice*	mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare	58
10	Riley WT et al (2011)	*Translational Behavioral Medicine*	Health Behavior Models in the Age of Mobile Interventions	58

Figure 5. The co-citation network of references in mobile health app research, 2000-2019.

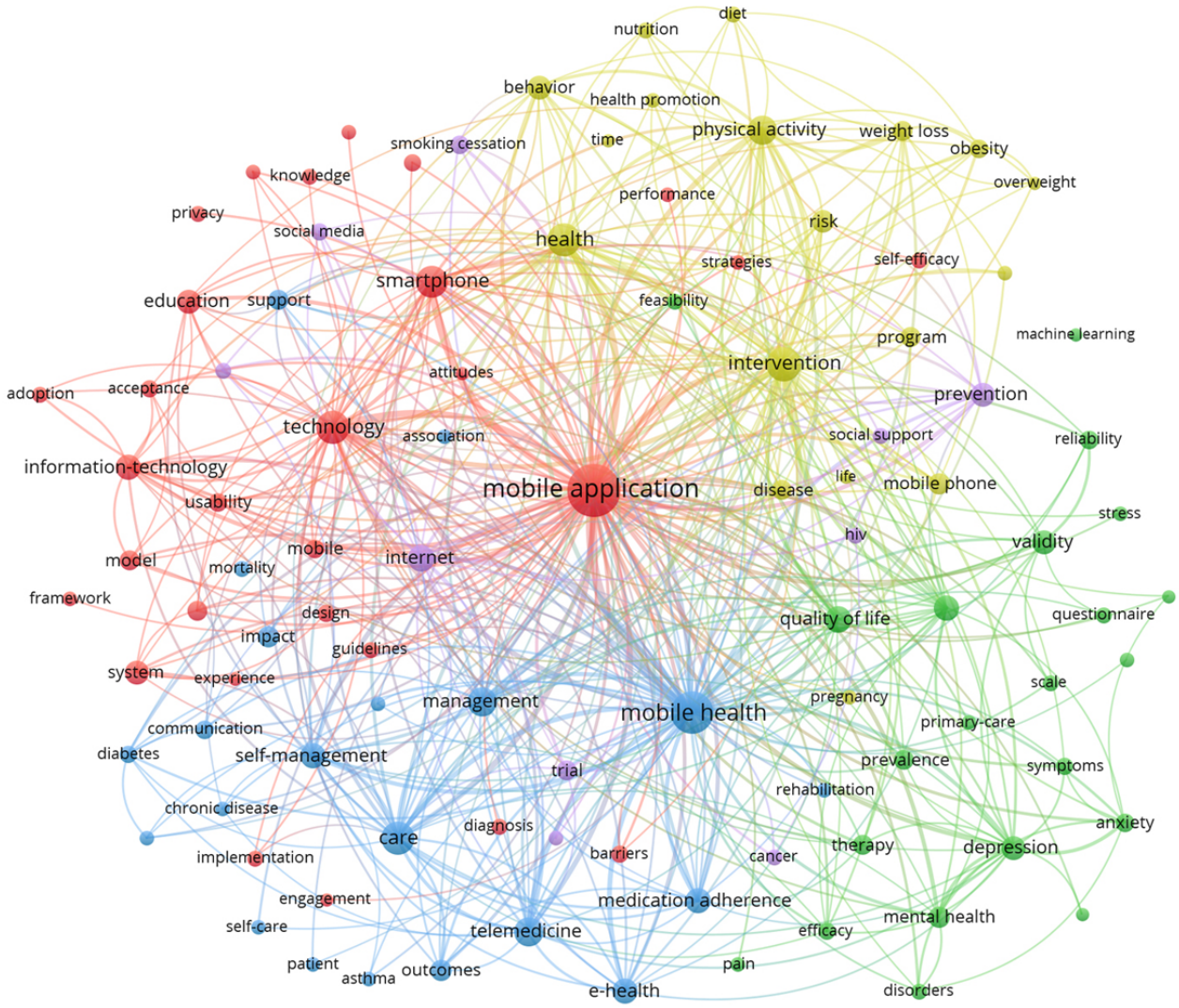


Co-Occurrence Analysis of Top 100 Keywords

Keywords cover the main topics of a publication and are well suited to be used for analyzing related research hot spots. The research hot spots of mobile health app research were identified through co-occurrence analysis of the top 100 keywords. We used VOSviewer to extract and cluster the top 100 keywords. Multimedia Appendix 1 shows the frequency and clustering of the top 100 keywords.

As shown in Figure 6, we used VOSviewer to build a visualization network map of the top 100 keywords in 5 clusters with their co-occurrence. The keywords mobile application (1124), mobile health (631), intervention (347), health (329), and technology (315) are located at the center of the visualization network map. The 5 clusters are represented by color: red (cluster 1), green (cluster 2), blue (cluster 3), yellow (cluster 4), and purple (cluster 5). The node label is the keyword, and the node size represents the frequency of the keywords. Links connecting 2 nodes represent a co-occurrence relationship between the keywords. For visualization purposes, the top 500 links among keywords were determined based on how frequently they co-occur in the same publication.

Figure 6. The co-occurrence network of the top 100 keywords in mobile health app research, 2000-2019.



Discussion

In the current study, we explored the bibliometric characteristics of mobile health app research and we identified research trends, research hot spots, and the knowledge base associated with mobile health apps through our co-word analysis of the top 100 keywords.

Global Trends in Mobile Health App Research

The change in the number of academic publications in a field is an important indicator of the evolving trends in this field. Mobile health app research included over 2800 publications around the world. The number of research papers published on mobile health apps annually has been increasing since 2000, with particularly notable gains in the past 5 years.

For journal sources, the top 3 journals publishing mobile health app research belong to the area of medical informatics, accounting for 18.24% (511/2802) of the total publications. Meanwhile, 848 journals had only 1 mobile health app–related publication, accounting for 30.26% (848/2802) of the total publications. Journals publishing mobile health app research were widely distributed across the general health domain, with higher concentration, as expected, in the field of medical informatics.

The Coauthorship Networks in Mobile Health App Research

Due to the accessibility of bibliometric analysis, coauthorship is frequently used as a proxy for research collaboration. A coauthorship network can reflect the collaborative relationship

among researchers and provide potential opportunities for other researchers to cooperate; examining this network can highlight potential opportunities for enhanced collaboration both within and outside of the existing network. The coauthorship network reflects author, institution, and country.

We found that the United States was the most significant contributor to mobile health app research. Of the publications identified, 65.06% (1823/2802) were published in the United States, England, Australia, and Canada, the current global leaders in mobile health app research. Figure 2 shows that the United States is the center of an international coauthorship network and cooperates with many countries/regions. The yellow coauthorship networks are mainly Asian countries/regions, such as China, Korea, and Singapore. The blue coauthorship networks are dominated by European countries such as Germany, Spain, and Italy. Our findings indicate that cooperation among countries/regions has certain regional characteristics. These groupings prompt speculation that cooperation on mobile health app research among countries/regions may be influenced by geographic proximity or by a shared language.

We found that 7 of the top 10 most productive institutions are from the United States, and the other 3 are from Canada, Australia, and South Korea. Figure 3 shows that the top 10 institutions are almost at the core of each coauthorship network. These most productive institutions and groups are leading the trends in mobile health app research. There is an excellent cooperative relationship between these institutions.

We identified 13,040 authors who have published research on mobile health apps. Of those authors, only 233 (1.79%) have published more than 4 papers in this emerging field, forming 5 relatively small coauthorship networks. We conclude from this finding that there are many researchers interested in pursuing mobile health app research, but collaboration between authors remains limited. Promoting collaboration between authors, institutions, and countries would expand the number of authors regularly publishing in this field and could contribute to more effective dissemination of innovative practices in mobile health app use.

Basic Knowledge and Hot Topics in Mobile Health Apps

More than three-quarters of the top 10 most frequently cited references in mobile health app research were published after 2011. This timeline is consistent with our understanding of mobile health app research as a rapidly emerging field of study. As shown in Figure 5, the most frequently cited references formed 3 clusters (shown in red, green, and blue). These clusters correspond to 3 basic groupings of research: (1) promotion of health behavior change (in red), (2) evaluation of quality of mobile health apps (in green), and (3) assessment of efficiency of mobile health apps (in blue).

Keywords are standardized terms used to ensure that publications are indexed uniformly by topic. Therefore, mapping the co-word network by analyzing the co-occurrence frequency of keywords from multiple publications is helpful to study the internal structure and the hot topics in the field of mobile health app research [21]. As shown in Figure 6, there were 5 clusters of mobile health app research that were formed by co-occurrence cluster analysis of the top 100 keywords. Combined with the characteristics of mobile health apps, the 5 clusters were analyzed as described below.

Cluster 1 (red cluster) mainly focuses on the technology and system development of mobile health apps and includes 29 high-frequency keywords, such as mobile app, technology, smartphone, system, model, usability, acceptance, design, devices, barriers, privacy, and attitudes. With the continuous development of smartphone and information technology, researchers need to update or develop new mobile health apps to meet the growing needs of patients and medical staff. The system development of mobile health apps mainly includes a user-computer interface, algorithms, privacy, design, and computer security, and it follows the principles of user-centered, convenient operation, safety, and stability [22,23]. As a new product, the effectiveness, quality, and accuracy of various mobile health apps used in health care need to be continuously evaluated through academic research [24,25].

Cluster 2 (green cluster) mainly focuses on mobile health apps used in mental health and includes 22 high-frequency keywords, such as quality of life, depression, validity, mental health, prevalence, therapy, anxiety, reliability, efficacy, disorders, questionnaire, stress, and cognitive behavioral therapy. It is reported that about 29% of humans suffer from mental illness in their lifetime, and more than 55% of these do not receive the treatment they need [26]. Mobile health apps can provide instant support, anonymity, customization, and low cost. These characteristics can potentially improve access to mental health services, thereby improving the equity of mental health resource allocation. Mobile health apps can be used as independent self-help mental health assessment tools and can be used to deliver online interventions aimed at diagnosis, treatment, and monitoring. Importantly, mobile health apps improve the accessibility of treatment through ecological momentary assessment to reduce the barriers to face-to-face help, especially in patients with depression, anxiety, stress, and other symptoms [27,28].

Cluster 3 (blue cluster) focuses on mobile health apps used as mobile health tools in telemedicine, chronic disease, and medication adherence management and includes 21 high-frequency keywords such as mobile health, care, telemedicine, self-management, eHealth, medication adherence, communication, diabetes, chronic disease, glycemic control, hypertension, and asthma. Telemedicine delivered using mobile health apps is an innovative model of health care, with significant potential to solve challenges in today's health care environment [29]. This approach can provide cost-effective solutions that bridge geographical and institutional barriers [30]. Mobile health app use for telemedicine is gaining in popularity in developing countries, where medical institutions are often remote and inaccessible [31,32]. Globally, chronic diseases currently account for 60% of the global disease burden [33]. Importantly, patients with chronic diseases are prone to secondary complications, which can be prevented by strengthening patient education and self-management. Mobile health apps can be used in a variety of environments, enhancing

their effectiveness as tools for self-management and monitoring [34]. They are widely used in the management of diabetes, hypertension, and asthma [35,36]. Mobile health apps can provide personalized medication adherence reminders and early warnings and improve medication adherence in patients [37,38].

Cluster 4 (yellow cluster) mainly focuses on mobile health apps used in health behavior and health promotion and includes 18 high-frequency keywords, such as intervention, health, physical activity, behavior, risk, weight loss, obesity, nutrition, diet, health promotion, and overweight. With the development of portable wearable devices and smart sensors, mobile health apps can provide self-tracking capabilities. People can track measures of interest such as weight, calories consumed, heart rate, respiratory rate, and exercise status, and can also record how they feel or how they are responding to treatment (eg, side effects). Tracking capabilities of this type can be used by people to promote the adoption of healthful behaviors, such as physical exercise, reasonable diet, and obesity prevention [39-41].

Cluster 5 (purple cluster) mainly focuses on mobile health apps used in disease prevention via the internet and includes 10 high-frequency keywords, such as internet, prevention, trial, smoking cessation, social media, cancer, and human immunodeficiency virus (HIV). The International Telecommunication Union estimates that 4.1 billion people were using the internet at the end of 2019 [42]. The internet can provide anonymity, low intervention costs, and the ability to fulfill effective solutions. Building on these characteristics, internet-based mobile health apps can promote disease prevention to solve health problems. Compared with traditional disease prevention, mobile health apps using social media technology can attract users in a more interactive way, provide convenient health education and rapid internet intervention, and achieve excellent results in HIV, smoking cessation, and cancer [43-45].

Limitations

Our study is, to our knowledge, the first bibliometric analysis of mobile health app–related publications. Still, there are some limitations to this study. First, there may be language bias because, although we did not place any limits on the language of publications in our study, most WOS publications are in English. Second, the quality of publications in WOS is not uniform. Conducting a weighted analysis of publications based on the assessment of quality was outside the scope of our study; therefore, it is possible that our analysis has given equal attention to publications of differing quality. Finally, the current data for analysis were only extracted from WOS, excluding data extracted from other search engines such as Scopus (Elsevier), PubMed, or Google Scholar (Google LLC). Thus, it is possible that publications appearing only through one of these other search engines have been missed. We plan to address this by exploring ways of combining different data sources in future work.

Conclusions

Through the bibliometric quantitative analysis and visualization network map of the data extracted from the WOS database, the current study reveals the research status, research trends, hot spots, and coauthorship network of mobile health app research. Mobile health app research is a new and promising field globally, with great potential for improving patient care and promoting health. By comprehensively summarizing the trends in mobile health app research, we expect this work may serve as a guide for facilitating future research directions to advance this field of research further.

Acknowledgments

This research was supported by the Humanities Social Science Foundation of Ministry of Education (Grant No. 17YJCZH184), the Philosophy and Social Sciences Promotion Project of China Medical University (Grant No. 111–3110118083), and the Health Big Data Research Project of China Medical University (Grant No. 111–HMB201901101).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Top 100 keywords and 5 clusters in mobile health app research, 2000-2019.
[DOCX File , 22 KB - mhealth_v8i7e18212_app1.docx]

References

1. 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]
2. Larson RS. A Path to Better-Quality mHealth Apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414. [doi: 10.2196/10414]
3. Zion Market Research. Global mHealth Apps Market Will Reach USD 111.1 Billion By 2025: Zion Market Research URL: https://www.globenewswire.com/news-release/2019/01/24/1704860/0/en/Global-mHealth-Apps-Market-Will-Reach-USD-111-1-Billion-By-2025-Zion-Market-Research.html [accessed 2019-01-24]
4. Martínez-Pérez B, de LTI, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]

5. Guler AT, Waaijer CJF, Palmblad M. Scientific workflows for bibliometrics. Scientometrics 2016 Feb 15;107(2):385-398. [doi: 10.1007/s11192-016-1885-6]
6. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and Visibility of "Digital Health" as a Keyword in Articles by JMIR Publications in the New Millennium: Bibliographic-Bibliometric Analysis. J Med Internet Res 2019 Dec 19;21(12):e10477 [FREE Full text] [doi: 10.2196/10477] [Medline: 31855190]
7. Taj F, Klein MCA, van Halteren A. Digital Health Behavior Change Technology: Bibliometric and Scoping Review of Two Decades of Research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311. [doi: 10.2196/13311]
8. Zou L, Sun L. Global diabetic kidney disease research from 2000 to 2017: A bibliometric analysis. Medicine (Baltimore) 2019 Mar;98(6):e14394 [FREE Full text] [doi: 10.1097/MD.0000000000014394] [Medline: 30732183]
9. Agarwal A, Durairajanayagam D, Tatagari S, Esteves S, Harlev A, Henkel R, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J Androl 2016;18(2):296. [doi: 10.4103/1008-682x.171582]
10. Li F, Li M, Guan P, Ma S, Cui L. Mapping Publication Trends and Identifying Hot Spots of Research on Internet Health Information Seeking Behavior: A Quantitative and Co-Word Biclustering Analysis. J Med Internet Res 2015 Mar 25;17(3):e81. [doi: 10.2196/jmir.3326]
11. Cozzens SE, Callon M, Law J, Rip A. Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World. Contemporary Sociology 1988 Nov;17(6):815. [doi: 10.2307/2073618]
12. Zhang C, Yu Q, Fan Q, Duan Z. Research collaboration in health management research communities. BMC Med Inform Decis Mak 2013 Apr 23;13(1). [doi: 10.1186/1472-6947-13-52]
13. Zhang J, Xie J, Hou W, Tu X, Xu J, Song F, et al. Mapping the Knowledge Structure of Research on Patient Adherence: Knowledge Domain Visualization Based Co-Word Analysis and Social Network Analysis. PLoS ONE 2012 Apr 5;7(4):e34497. [doi: 10.1371/journal.pone.0034497]
14. Wang Y, Wang Q, Wei X, Shao J, Zhao J, Zhang Z, et al. Global scientific trends on exosome research during 2007–2016: a bibliometric analysis. Oncotarget 2017 Apr 19;8(29). [doi: 10.18632/oncotarget.17223]
15. Zhao J, Yu G, Cai M, Lei X, Yang Y, Wang Q, et al. Bibliometric analysis of global scientific activity on umbilical cord mesenchymal stem cells: a swiftly expanding and shifting focus. Stem Cell Res Ther 2018 Feb 7;9(1). [doi: 10.1186/s13287-018-0785-5]
16. Miao Y, Xu S, Chen L, Liang G, Pu Y, Yin L. Trends of long noncoding RNA research from 2007 to 2016: a bibliometric analysis. Oncotarget 2017 Oct 10;8(47):83114-83127 [FREE Full text] [doi: 10.18632/oncotarget.20851] [Medline: 29137328]
17. Khan MS, Ullah W, Riaz IB, Bhulani N, Manning WJ, Tridandapani S, et al. Top 100 cited articles in cardiovascular magnetic resonance: a bibliometric analysis. J Cardiovasc Magn Reson 2016 Nov 21;18(1). [doi: 10.1186/s12968-016-0303-9]
18. Romero L, Portillo-Salido E. Trends in Sigma-1 Receptor Research: A 25-Year Bibliometric Analysis. Front. Pharmacol 2019 May 24;10. [doi: 10.3389/fphar.2019.00564]
19. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2009 Dec 31;84(2):523-538. [doi: 10.1007/s11192-009-0146-3]
20. Eck NJV, Waltman L. How to normalize cooccurrence data? An analysis of some well-known similarity measures. J. Am. Soc. Inf. Sci 2009 Aug;60(8):1635-1651. [doi: 10.1002/asi.21075]
21. Gan J, Cai Q, Galer P, Ma D, Chen X, Huang J, et al. Mapping the knowledge structure and trends of epilepsy genetics over the past decade. Medicine 2019;98(32):e16782. [doi: 10.1097/md.0000000000016782]
22. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
23. Martínez-Pérez B, de LTI, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
24. Singler K, Roth T, Beck S, Cunningham M, Gosch M. Development and initial evaluation of a point-of-care educational app on medical topics in orthogeriatrics. Arch Orthop Trauma Surg 2015 Dec 8;136(1):65-73. [doi: 10.1007/s00402-015-2366-8]
25. Vagge A, Wangtiraumnuay N, Pellegrini M, Scotto R, Iester M, Traverso CE. Evaluation of a Free Public Smartphone Application to Detect Leukocoria in High-Risk Children Aged 1 to 6 Years. J Pediatr Ophthalmol Strabismus 2019 Jul 01;56(4):229-232. [doi: 10.3928/01913913-20190516-01]
26. Anthes E. Mental health: There's an app for that. Nature 2016 Apr 07;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
27. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
28. Lemey C, Larsen ME, Devylder J, Courtet P, Billot R, Lenca P, et al. Clinicians' Concerns About Mobile Ecological Momentary Assessment Tools Designed for Emerging Psychiatric Problems: Prospective Acceptability Assessment of the MEmind App. J Med Internet Res 2019 Apr 25;21(4):e10111. [doi: 10.2196/10111]

29. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The Empirical Foundations of Telemedicine Interventions in Primary Care. Telemed J E Health 2016 May;22(5):342-375 [FREE Full text] [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
30. Silva BMC, Rodrigues JJPC, de LTDI, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
31. Déglise C, Suggs LS, Odermatt P. Short Message Service (SMS) Applications for Disease Prevention in Developing Countries. J Med Internet Res 2012 Jan 12;14(1):e3. [doi: 10.2196/jmir.1823]
32. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AHA, et al. Mobile Health (mHealth) Approaches and Lessons for Increased Performance and Retention of Community Health Workers in Low- and Middle-Income Countries: A Review. J Med Internet Res 2013 Jan 25;15(1):e17. [doi: 10.2196/jmir.2130] [Medline: 23353680]
33. World Health Organization. World Health Organization. Geneva, Switzerland: Noncommunicable Diseases and Mental Health, World Health Organization; 2002. Innovative care for chronic conditions: building blocks for action URL: https://www.who.int/chp/knowledge/publications/icccglobalreport.pdf?ua=1 [accessed 2020-05-30]
34. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: A Novel mHealth System for Supporting Self-Care in Management of Complex and Chronic Conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10 [FREE Full text] [doi: 10.2196/mhealth.2391] [Medline: 25100682]
35. Lum E, Jimenez G, Huang Z, Thai L, Semwal M, Boehm BO, et al. Decision Support and Alerts of Apps for Self-management of Blood Glucose for Type 2 Diabetes. JAMA 2019 Dec 16;321(15):1530-1532. [doi: 10.1001/jama.2019.1644] [Medline: 30990543]
36. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A Mobile Health Intervention for Self-Management and Lifestyle Change for Persons With Type 2 Diabetes, Part 2: One-Year Results From the Norwegian Randomized Controlled Trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
37. Zanetti-Yabur A, Rizzo A, Hayde N, Watkins AC, Rocca JP, Graham JA. Exploring the usage of a mobile phone application in transplanted patients to encourage medication compliance and education. The American Journal of Surgery 2017 Oct;214(4):743-747. [doi: 10.1016/j.amjsurg.2017.01.026]
38. Labovitz DL, Shafner L, Reyes GM, Virmani D, Hanina A. Using Artificial Intelligence to Reduce the Risk of Nonadherence in Patients on Anticoagulation Therapy. Stroke 2017 May;48(5):1416-1419. [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
39. Stavrositu C, Kim J. Self-Persuasion Through Mobile Applications: Exploring Different Routes to Health Behavioral Change. Cyberpsychology, Behavior, and Social Networking 2018 Aug;21(8):516-522. [doi: 10.1089/cyber.2018.0079]
40. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
41. Martin C, Gilmore L, Apolzan J, Myers C, Thomas D, Redman L. Smartloss: A Personalized Mobile Health Intervention for Weight Management and Health Promotion. JMIR Mhealth Uhealth 2016 Mar 16;4(1):e18 [FREE Full text] [doi: 10.2196/mhealth.5027] [Medline: 26983937]
42. ITUPublications. Measuring digital development: Facts & figures 2019 URL: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2020-05-31]
43. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown IW. Comparison of a User-Centered Design, Self-Management App to Existing mHealth Apps for Persons Living With HIV. JMIR Mhealth Uhealth 2015;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
44. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR Mhealth Uhealth 2016;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
45. Coughlin S, Thind H, Liu B, Champagne N, Jacobs M, Massey RI. Mobile Phone Apps for Preventing Cancer Through Educational and Behavioral Interventions: State of the Art and Remaining Challenges. JMIR Mhealth Uhealth 2016 May 30;4(2):e69. [doi: 10.2196/mhealth.5361]

Abbreviations

HIV: human immunodeficiency virus
WOS: Web of Science

Edited by G Eysenbach; submitted 11.02.20; peer-reviewed by C Kruse, I Moreira; comments to author 10.03.20; revised version received 20.04.20; accepted 26.04.20; published 27.07.20.

Please cite as:
Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z
Theme Trends and Knowledge Structure on Mobile Health Apps: Bibliometric Analysis
JMIR Mhealth Uhealth 2020;8(7):e18212
URL: https://mhealth.jmir.org/2020/7/e18212
doi:10.2196/18212
PMID:32716312

Original Paper

The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study

Pere Llorens-Vernet[1*], MSc; Jordi Miró[1*], PhD

Unit for the Study and Treatment of Pain-ALGOS, Research Center for Behavior Assessment, Universitat Rovira i Virgili, Tarragona, Spain
[*]all authors contributed equally

Corresponding Author:
Jordi Miró, PhD
Unit for the Study and Treatment of Pain-ALGOS
Research Center for Behavior Assessment
Universitat Rovira i Virgili
Carretera de Valls, s/n
Tarragona, 43007
Spain
Phone: 34 +34649298921
Email: jordi.miro@urv.cat

Abstract

Background: In recent years, there has been an exponential growth of mobile health (mHealth)–related apps. This has occurred in a somewhat unsupervised manner. Therefore, having a set of criteria that could be used by all stakeholders to guide the development process and the assessment of the quality of the apps is of most importance.

Objective: The aim of this paper is to study the validity of the Mobile App Development and Assessment Guide (MAG), a guide recently created to help stakeholders develop and assess mobile health apps.

Methods: To conduct a validation process of the MAG, we used the Delphi method to reach a consensus among participating stakeholders. We identified 158 potential participants: 45 patients as potential end users, 41 health care professionals, and 72 developers. We sent participants an online survey and asked them to rate how important they considered each item in the guide to be on a scale from 0 to 10. Two rounds were enough to reach consensus.

Results: In the first round, almost one-third (n=42) of those invited participated, and half of those (n=24) also participated in the second round. Most items in the guide were found to be important to a quality mHealth-related app; a total of 48 criteria were established as important. “Privacy,” “security,” and “usability” were the categories that included most of the important criteria.

Conclusions: The data supports the validity of the MAG. In addition, the findings identified the criteria that stakeholders consider to be most important. The MAG will help advance the field by providing developers, health care professionals, and end users with a valid guide so that they can develop and identify mHealth-related apps that are of quality.

(JMIR Mhealth Uhealth 2020;8(7):e17760) doi:10.2196/17760

KEYWORDS

assessment; Delphi method; MAG; mobile apps; mobile health; validity; guide

Introduction

Mobile apps are increasingly being used for health care [1-3]. The implementation of mobile devices such as phones, patient monitoring devices or personal digital assistants, and wireless devices has proven that they can be used for improving communication between patients and health professionals [4], and improving adherence to treatment [5]. Importantly, recent reports have suggested that smartphones have become the most popular technology among physicians [6,7]. In addition, there has been a sharp increase in the use of these technologies by the general population. For example, official estimates indicate that in 2019 a total of 65% of people had a smartphone, and by 2025, this figure will have increased to 80% [8].

However, this increase in use has occurred in a somewhat unsupervised manner; that is to say, it has not been regulated or supervised in any way. In addition, a large number of mobile health (mHealth) apps have been developed without any rigorous scientific basis [9,10] or having undergone any validation process, thus undermining the confidence of both patients and

health care professionals [11]. Moreover, information privacy practices are not transparent to users and, in many cases, are absent, opaque, or irrelevant [12]. Finally, there is mounting evidence to show that this lack of control and development without guidance is placing consumers at risk [13].

In an attempt to solve this problem, and guarantee the quality of existing and future health apps, various government-related initiatives have been taken at the regional level (eg, the proposal "AppSalut" [14,15] in Catalonia and the "AppSaludable Quality Seal" [16] in Andalusia, Spain), the national level (eg, "Good practice guidelines on health apps and smart devices [mobile health or mhealth]" [17] in France; "Health apps & co: safe digital care products with clearer regulations" [18] in Germany; "Medical devices: software applications [apps]" [19] in the United Kingdom; "Policy for Device Software Functions and Mobile Medical Applications" [20] in the United States; "Regulation of Software as a Medical Device" [21,22] in Australia), and the international level, such as the "Green Paper on mobile health" by the European Commission [23]. In general, these initiatives provide recommendations and regulations to establish how health apps should be and guarantee their quality. However, they show important differences on the key criteria. For example, "Appsalut" emphasizes usability issues [14,15], while "Regulation of Software as a Medical Device" emphasizes safety [21,22] as it equates health apps with medical devices. Clinicians and researchers have also attempted to provide specific solutions to this major problem [24]. For example, Stoyanov and colleagues [25] developed a scale to classify and rate the quality of mHealth apps. There have also been other attempts to provide alternatives for assessing mHealth apps (eg, [26,27]), each one of which suggests its own quality criteria. All these attempts have positive and negative characteristics. A major limitation common to many of these initiatives is that they have been created from one narrow perspective and focusing on, for example, a specific health problem or intervention such as emergency interventions [27] or a stakeholder such as adolescents [26]. In addition, some of them have been created from a specific perspective, for example, taking into account usability issues rather than safety. Thus, there is no common set of criteria that can be used by all stakeholders to guide the development process and the assessment of the apps' quality.

Recently, to help overcome these limitations, we conducted a study to develop such a guide: the Mobile App Development and Assessment Guide (MAG) [28]. We studied the guidelines, frameworks, and standards available in the field of health app development, with a particular focus on the world regions where the mHealth market was most significant, and pinpointed the criteria that could be recommended as a general standard. We suggested a guide containing 36 criteria that could be used by stakeholders. Our preliminary study showed that stakeholders found them to be important. They also found them easy to understand and use [28].

However, that study had some limitations. Most importantly, although the criteria identified underwent a preliminary analysis of comprehensibility and importance by a selected group of stakeholders (ie, health care experts, engineers, and potential patients), they did not undergo a validation process. Therefore, to address this issue, here we use the Delphi method [29,30] to analyze the validity of the MAG. By using this method, we also want to explore whether new criteria could be included to improve the guide. We also want to examine the importance of these criteria as perceived by the stakeholders.

Methods

Procedure

The Delphi method was created for people to reach consensus by answering questions in an iterative process [29]. Although the traditional Delphi process has an open initial phase [29], in this study we use a modified Delphi process, which provides a common starting point for discussion. This modified Delphi method is widely used, as it saves time and does not interfere with the original tenets of the method that participants can give suggestions and inputs at any stage [31]. It has been shown that results from Delphi-based studies offer more accurate answers to difficult questions than other prognostication techniques [32]. This modified Delphi method and the judgment of people are acknowledged as legitimate and valid inputs to generate forecasts, and have been used in many different areas to reach consensus on such strategic issues as the identification of health care quality indicators [33]; predictors of chronic pain and disability in children [34]; predictors of chronic pain in adults with cancer [35]; the needs of adolescents and young adult cancer survivors [36]; and, in the mHealth field, to develop an assessment framework for electronic mental health apps [37].

Participants

Our goal was to recruit 30 stakeholders, as this number has been shown to be sufficient for this kind of study [38,39], from any of the following groups: (1) health care professionals, (2) developers of health-related apps, and (3) users of health apps.

To identify potential participants and ensure an appropriate panel of stakeholders, we adopted five strategies: (1) we searched for national (Spain) and international organizations or associations of digital health professionals to make contact with health professionals knowledgeable about the topic; (2) we searched for medical health apps in the app stores of the main smartphone systems (Android and iOS), identified the most downloaded and best rated apps, and searched for their developers to ensure the participation of experienced individuals; (3) we searched for national (Spain) and international organizations to recruit patients with experience in the use of health-related apps or with an interest in this area; (4) we made a general call through the social networks of our research group to increase the likelihood of recruiting participants who satisfied the inclusion requirements; and (5) we asked researchers and clinicians who we personally knew were experts in the field to participate and help us identify other potential participants.

We identified 158 potential participants from Europe, Asia, Australia, and North and South America. They were multidisciplinary and included health care professionals, patients as potential end users, and developers.

Survey Content and Procedure

We developed a list of items on the basis of the criteria in the MAG [28]. Some of the criteria were broad and encompassed several issues and characteristics, so we broke them down into specific items to facilitate the comprehensibility and accuracy of responses. For example, the original criterion "The app can be consulted in more than one language. All languages adapt appropriately to the content interface" was divided into two items: "The app can be consulted in more than one language" and "All languages adapt appropriately to the content interface."

When the set of items was ready, it was moved to an online survey so that it could be distributed to participants more easily. Potential participants received an email with explanations about the study and a link to the survey. All the information was available in Spanish and English.

The survey included some sociodemographic questions and 56 items for rating, which were grouped in the same categories as the original guide, such as usability [28]. On a numerical rating scale from 0 (not important at all) to 10 (extremely important), participants had to report how important they considered each one to be for the quality of a health-related mobile app. Participants were also given instructions to include any other item they felt was important and missing from the original list. Like the original items, these new items were also rated. Participants were informed that only the criteria that received a score of 9 or higher from at least 75% of the participants would be included in the final list of criteria that a health-related app should meet. The rest were discarded.

Study data were collected and managed using LimeSurvey tools (LimeSurvey GmbH) [40]. We computed means and standard deviations of the demographics to describe the sample of participants. We used paired *t* tests (two-tailed) to study potential differences in the variance of the items between rounds and of the potential changes in the age or sex of participants. To study the consensus, mean, standard deviation, 95% confidence interval (with the lower and upper values for each item), and significance level ($P<.05$) were computed. All data analyses were performed using SPSS v.25 for Windows (IBM Corp).

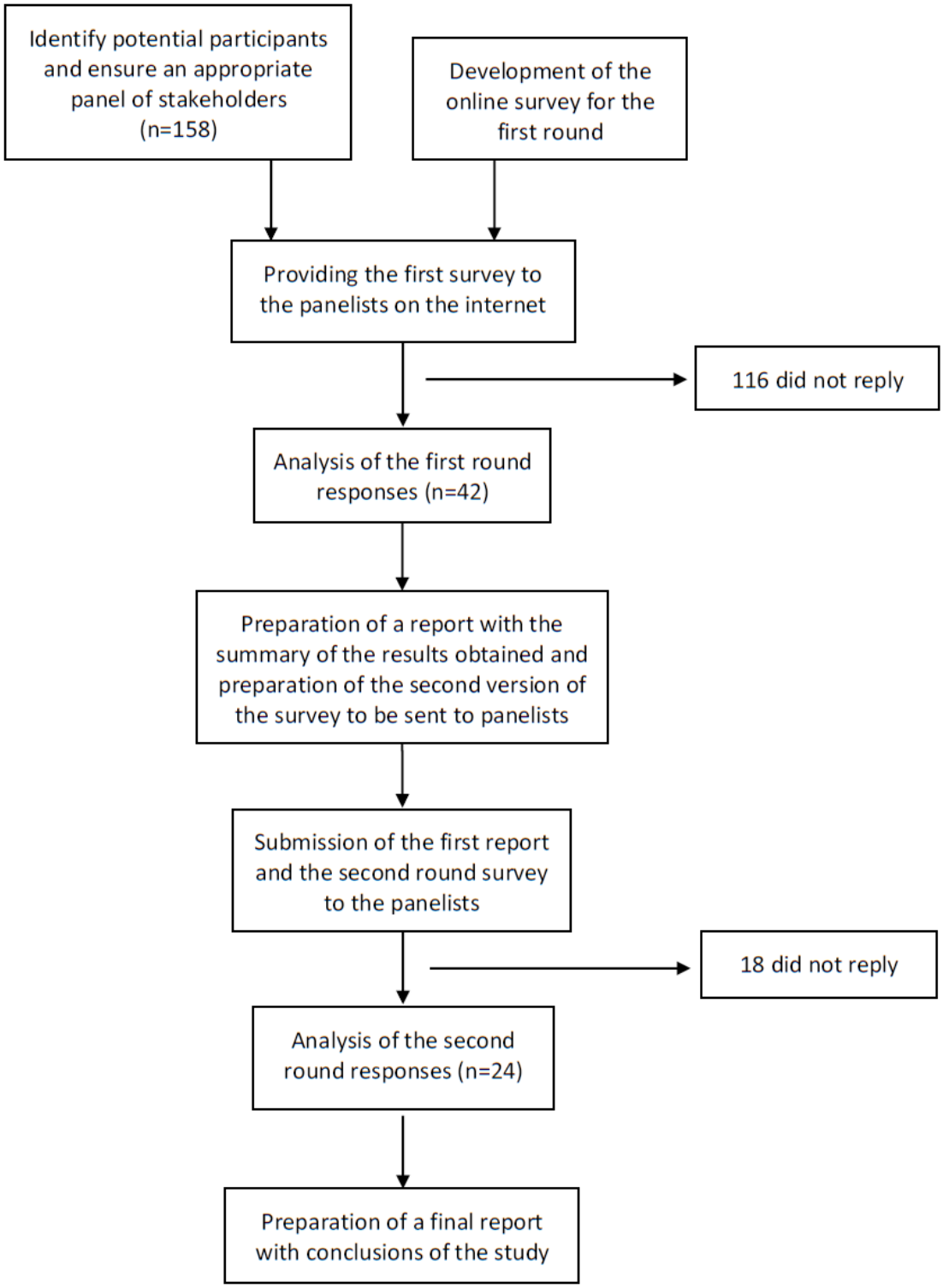
Delphi Rounds

In the first round, the survey was sent to 158 potential participants: 45 patients as potential end users, 41 health care professionals, and 72 developers. They were informed about the study and the requirements to participate. Participants were given 3 weeks to respond, during which time a reminder was sent each week to maximize the involvement of as many participants as possible. The survey took approximately 15 minutes. The answers were analyzed, and some new items were added in response to the suggestions of the participants.

In the second round, the results of the first iteration were sent by email to all the participants who had provided responses to the initial survey so that they could see their position and the position of the group as a whole on the items, as well as the level of agreement among the participants. This information was given so that participants in the group could re-examine their initial responses, in light of the group's opinion. Participants in this second round were asked to respond to the revised survey. Again, they were given 3 weeks to answer. The Delphi methodology requires that this procedure be repeated until participants' responses reach stability or when a point of diminishing returns is reached [39].

The stability of responses was the criterion used to identify that consensus had been reached on any given question [30,38]. In this study, stability was reached after two rounds (see the Results section), which is consistent with the findings of previous Delphi studies (eg, [35,41]). We considered that consensus was reached on a particular criterion when 75% of the participants rated it with at least a 9 [34]. If a criterion was rated with a 9 or more by at least 90% of the participants, we considered it to be of key importance for an mHealth-related app. The results only showed statistically significant differences for two items (see the Results section). Thus, given the stability of the responses, we decided to stop the iteration process after round 2. Figure 1 describes the steps of the study.

Figure 1. Steps in the Delphi poll.



Results

Round 1

Of the 158 stakeholders invited, 42 (27%) responded to the first round. The demographic characteristics of the participants in each round are shown in Table 1. There were no statistically significant differences in terms of sex or age between those who were invited and those who participated. Only a small increase in the mean age of participants and female participation was detected between rounds (see Table 1).

Multimedia Appendix 1 summarizes all the information about participants' responses to the initial 56 items.

To determine consensus on the items, we examined the percentage of participants who agreed on their importance. Items with an agreement ≥75% of participants were considered to have reached consensus. We also used confidence intervals,

instead of discrete estimation, because they have less error (see [34] for a similar procedure). Out of the total 56 items, participants reached consensus on the importance of 32 (57%) of the items (ie, at least 75% of the participating stakeholders rated their importance with a 9 or higher on the 0-10 scale).

In this first round, participants added 36 new items to the original list. As previously described, in response to participants' suggestions, changes were made to items 3, 51, 68, and 73. In addition, items 33 and 69 were divided in two, as several participants considered that they included two different clauses (see Multimedia Appendix 1).

Table 1. Sample characteristics in each round.

Characteristics	Invited participants (N=158)	Participants round 1 (n=42)	Participants round 2 (n=24)
Sex, n (%)[a]			
Male	17 (42.5)	15 (44.12)	9 (37.5)
Female	23 (57.5)	19 (55.88)	15 (62.5)
Stakeholder group, n (%)			
Health care professionals	45 (28.48)	16 (38.1)	9 (37.5)
Developers	41 (25.95)	14 (33.33)	8 (33.33)
Final users	72 (45.57)	12 (28.57)	7 (29.17)
Age (years), mean[a]	39.82	41.39	43.08
Citizenship, n (%)[a]			
Spain	78 (49.37)	28 (82.35)	20 (83.33)
United States	13 (8.23)	1 (2.94)	1 (4.17)
Argentina	1 (0.63)	1 (2.94)	1 (4.17)
Italy	1 (0.63)	1 (2.94)	1 (4.17)
Australia	6 (3.8)	1 (2.94)	1 (4.17)
United Kingdom	15 (9.49)	2 (5.88)	0 (0.00)
Canada	8 (5.06)	0 (0.00)	0 (0.00)
Brazil	3 (1.9)	0 (0.00)	0 (0.00)
India	6 (3.8)	0 (0.00)	0 (0.00)
Indonesia	1 (0.63)	0 (0.00)	0 (0.00)
Germany	5 (3.16)	0 (0.00)	0 (0.00)
Poland	2 (1.27)	0 (0.00)	0 (0.00)
Ireland	2 (1.27)	0 (0.00)	0 (0.00)
Turkey	1 (0.63)	0 (0.00)	0 (0.00)
Kazakhstan	1 (0.63)	0 (0.00)	0 (0.00)
Switzerland	3 (1.9)	0 (0.00)	0 (0.00)
Colombia	1 (0.63)	0 (0.00)	0 (0.00)
Romania	1 (0.63)	0 (0.00)	0 (0.00)
Hungary	1 (0.63)	0 (0.00)	0 (0.00)
Lithuania	1 (0.63)	0 (0.00)	0 (0.00)
Bulgaria	2 (1.27)	0 (0.00)	0 (0.00)
Greece	1 (0.63)	0 (0.00)	0 (0.00)
Sweden	2 (1.27)	0 (0.00)	0 (0.00)
Israel	1 (0.63)	0 (0.00)	0 (0.00)
Belgium	2 (1.27)	0 (0.00)	0 (0.00)

[a]Information was not available for all participants.

Round 2

Of the 42 that participated in the first round, 24 (57%) of the stakeholders participated in the second round. Out of the total 92 items, a total of 48 items (52%) were rated with a 9 or more by at least 75% of the participants (see Table 2 below).

The consensus on the importance of the 32 items in round 1 was maintained in round 2, except for item 69, which fell below the criteria of 75% agreement. On the other hand, items 8, 32, 46, and 47 did not reach consensus in round 1 but did in round 2. Consensus was also reached on the importance of 14 of the new items suggested by participants. Of all the items, 9 were particularly important (ie, at least 90% of participants rated their importance with a 9 or higher).

By categories, the number of items for which consensus was reached were *usability*: 8 of 18 items (44% of the total in the category); *privacy*: 14 of 19 items (74%); *security*: 9 of 13 items (69%); *appropriateness and suitability*: 2 of 5 items (40%); *transparency and content*: 2 of 11 items (18%); *safety*: 7 of 8 items (88%), *technical support and updates*: 2 of 9 items (22%); and *technology*: 4 of 9 items (44%).

Table 2. Items that reached consensus about their importance.

Category and item	Round 1 (n=42)			Round 2 (n=24)		
	Consensus, n (%)	Mean (SD)	95% CI	Consensus, n (%)	Mean (SD)	95% CI
Usability						
1. The app has been tested by potential users before being made available to the public.	33 (78.57)	9.14 (1.92)	8.56-9.72	20 (83.33)	9.21 (1.61)	8.56-9.85
2. It is easy to use (ie, navigation is intuitive).	39 (92.86)	9.67 (0.61)	9.48-9.85	21 (87.50)	9.50 (0.72)	9.21-9.79
3. Functionality is adapted to the purpose of the app.	40 (95.24)	9.74 (0.63)	9.55-9.93	21 (87.50)	9.42 (0.93)	9.05-9.79
4. Functionality is adjusted according to the profile and needs of the targeted user.	—[a]	—	—	19 (79.17)	9.08 (0.93)	8.71-9.45
5. Access is adapted for people with disabilities.	—	—	—	19 (79.17)	9.00 (1.41)	8.43-9.57
6. It complies with regulatory accessibility standards.	—	—	—	18 (75.00)	9.17 (1.13)	8.71-9.62
7. The language used makes it accessible to any user.	—	—	—	19 (79.17)	9.04 (1.55)	8.42-9.66
8. All users have access to all resources regardless of their capabilities.	28 (66.67)	8.60 (1.67)	8.09-9.10	18 (75.00)	9.08 (1.06)	8.66-9.51
Privacy						
9. The app gives information about the terms and conditions of purchases in the app.	35 (83.33)	9.29 (1.90)	8.71-9.86	21 (87.50)	9.50 (1.18)	9.03-9.97
10. It must only ask for user data that is essential for the app to operate.	34 (80.95)	9.26 (1.53)	8.80-9.72	18 (75.00)	8.92 (2.04)	8.10-9.73
11. It gives information about access policies and data processing, and ensures the right of access to recorded information.	34 (80.95)	9.02 (2.25)	8.34-9.70	18 (75.00)	9.38 (0.97)	8.99-9.76
12. It gives information about possible commercial agreements with third parties.	32 (76.19)	8.79 (2.54)	8.02-9.55	20 (83.33)	9.17 (2.12)	8.32-10.01
13. It clearly allows the user the option of non-transfer of data to third parties or for commercial purposes.	—	—	—	23 (95.83)	9.71 (0.55)	9.49-9.93
14. It guarantees the privacy of the information recorded.	39 (92.86)	9.71 (0.77)	9.48-9.95	20 (83.33)	9.46 (1.10)	9.02-9.90
15. It requires users to give their express consent.	36 (85.71)	9.12 (2.19)	8.46-9.78	19 (79.17)	9.38 (1.01)	8.97-9.78
16. It warns of the risks of using the app.	36 (85.71)	9.33 (1.86)	8.77-9.89	19 (79.17)	9.25 (1.19)	8.77-9.73
17. It tells users when it accesses other resources on the mobile device such as their accounts or their social network profiles.	36 (85.71)	9.33 (1.76)	8.80-9.87	22 (91.67)	9.71 (0.75)	9.41-10.01
18. It takes measures to protect minors in accordance with current legislation.	38 (90.48)	9.43 (1.74)	8.90-9.96	23 (95.83)	9.79 (0.51)	9.59-10.00
19. Confidential user data is protected and anonymized, and there is a privacy mechanism so that users can control their data.	38 (90.48)	9.60 (1.06)	9.27-9.92	21 (87.50)	9.46 (1.18)	8.99-9.93
20. It offers to erase the data when the service is finished.	—	—	—	19 (79.17)	9.04 (1.68)	8.37-9.71
21. It gives information about privacy policies in a simple and understandable way.	—	—	—	20 (83.33)	9.33 (1.09)	8.90-9.77
22. It complies with all current privacy laws.	—	—	—	22 (91.67)	9.54 (1.28)	9.03-10.06
Security						
23. The app has encryption mechanisms for storing, collecting, and exchanging information.	35 (83.33)	9.40 (1.33)	9.00-9.81	19 (79.17)	9.13 (1.57)	8.50-9.75
24. It has password management mechanisms.	33 (78.57)	9.05 (1.71)	8.53-9.56	19 (79.17)	9.04 (1.90)	8.28-9.80

Category and item	Round 1 (n=42)			Round 2 (n=24)		
	Consensus, n (%)	Mean (SD)	95% CI	Consensus, n (%)	Mean (SD)	95% CI
25. It states the terms and conditions of cloud services.	32 (76.19)	8.93 (2.23)	8.25-9.60	19 (79.17)	9.29 (1.08)	8.86-9.72
26. The cloud services used have the relevant security measures.	36 (85.71)	9.40 (1.47)	8.96-9.85	21 (87.50)	9.29 (1.60)	8.65-9.93
27. The authorization and authentication mechanisms protect users' credentials and allow access to their data.	37 (88.10)	9.57 (1.02)	9.26-9.88	21 (87.50)	9.42 (1.21)	8.93-9.90
28. It limits access to data that is only necessary for the user.	33 (78.57)	8.98 (2.25)	8.30-9.66	19 (79.17)	8.96 (2.10)	8.12-9.80
29. It detects and identifies cybersecurity vulnerabilities, possible threats, and the risk of being exploited.	36 (85.71)	9.33 (1.76)	8.80-9.87	18 (75.00)	8.96 (2.16)	8.10-9.82
30. It applies the appropriate security measures to cybersecurity vulnerabilities in the face of possible threats to reduce the risk of being exploited.	35 (83.33)	9.62 (0.82)	9.37-9.87	19 (79.17)	9.38 (0.92)	9.01-9.74
31. It informs users of the possible risks associated with the app's use of personal data.	—	—	—	20 (83.33)	9.25 (1.11)	8.80-9.70
Appropriateness and suitability						
32. The benefits and advantages of using the app are explained.	31 (73.81)	8.95 (1.58)	8.48-9.43	18 (75.00)	9.08 (1.53)	8.47-9.70
33. Experts have participated in the development of the app (for example, specialized professionals, health organizations, scientific societies, or specialized external organizations).	35 (83.33)	9.52 (1.02)	9.22-9.83	21 (87.50)	9.58 (0.72)	9.30-9.87
Transparency and content						
34. It uses scientific evidence to guarantee the quality of the content.	36 (85.71)	9.60 (0.86)	9.34-9.85	20 (83.33)	9.46 (0.78)	9.15-9.77
35. It is based on ethical principles and values.	39 (92.86)	9.71 (0.77)	9.48-9.95	22 (91.67)	9.75 (0.61)	9.51-9.99
Safety						
36. The possible risks to users are identified.	36 (85.71)	9.45 (1.15)	9.10-9.80	20 (83.33)	9.46 (0.88)	9.10-9.81
37. It ensures that there are no adverse effects.	—	—	—	18 (75.00)	8.92 (2.12)	8.07-9.77
38. It complies with regulatory standards as a medical device.	—	—	—	22 (91.67)	9.46 (1.67)	8.79-10.13
39. Users are warned when adverse events are identified so they can delete the app and avoid potential risks.	—	—	—	18 (75.00)	8.83 (1.93)	8.06-9.60
40. Users are warned that the app is not meant to replace the services provided by a professional.	40 (95.24)	9.74 (0.63)	9.55-9.93	22 (91.67)	9.67 (0.64)	9.41-9.92
41. It recommends always consulting a specialist in case of doubt.	—	—	—	22 (91.67)	9.33 (1.66)	8.67-10.00
42. Potential risks for users caused by incorrect use and possible adverse effects are explained.	34 (80.95)	9.48 (0.92)	9.20-9.75	20 (83.33)	9.38 (1.17)	8.91-9.84
Technical support and updates						
43. It gives a warning if updates can influence insensitive data (changes the use of the data or different data is collected).	32 (76.19)	8.90 (2.07)	8.28-9.53	19 (79.17)	9.17 (1.40)	8.60-9.73
44. Every time an update of a third-party component is published, the change is inspected and the risk evaluated.	33 (78.57)	8.98 (1.81)	8.43-9.52	20 (83.33)	8.96 (1.97)	8.17-9.75
Technology						

Category and item	Round 1 (n=42)			Round 2 (n=24)		
	Consensus, n (%)	Mean (SD)	95% CI	Consensus, n (%)	Mean (SD)	95% CI
45. It works correctly. It does not fail during use (blocks, etc).	36 (85.71)	9.36 (1.23)	8.99-9.73	23 (95.83)	9.75 (0.53)	9.54-9.96
46. Functions are correctly retrieved after context changes (switch to another app and return, etc), external interruptions (incoming calls or messages, etc), and switching off the terminal.	30 (71.43)	8.93 (1.50)	8.47-9.38	21 (87.50)	9.46 (0.83)	9.13-9.79
47. It does not waste resources excessively: battery, central processing unit, memory, data, network, etc.	29 (69.05)	8.88 (1.48)	8.43-9.33	19 (79.17)	9.25 (1.11)	8.80-9.70
48. It has a data recovery system in case of loss.	—	—	—	19 (79.17)	8.67 (2.28)	7.76-9.58

[a]These items were not available in round 1.

Discussion

Main Findings

The key finding from this study is that the MAG [28] is a valid tool to help guide the development of health-related mobile apps and assess their quality. The findings also indicate the items that are important to a health-related mobile app (the MAG is provided with this article; see Multimedia Appendix 2).

The data showed that 48 items on the MAG were considered to be of high importance (ie, they were rated with at least a 9 on a 0-10 numerical rating scale by at least 75% of the participants). Most of the items belonged to the categories *privacy* and *security*, thus showing that these are the issues that most concern stakeholders when assessing the quality of health mobile apps. In particular, the following items reached a consensus of 90%: *it clearly allows the user the option of nontransfer of data to third parties or for commercial purposes* (item 13); *it tells users when it accesses other resources on the mobile device, such as their accounts or their social network profiles* (item 17); *it takes measures to protect minors in accordance with current legislation* (item 18); *it complies with all current privacy laws* (item 22)*; it is based on ethical principles and values* (item 35); *it complies with regulatory standards as a medical device* (item 38); *users are warned that the app is not meant to replace the services provided by a professional* (item 40); *it recommends always consulting a specialist in case of doubt* (item 41); and *it works correctly, it does not fail during use (blocks, etc*; item 45).

Our work adds to previous proposals of quality guides or checklists by studying the validity of MAG, a comprehensive guide developed by Llorens-Vernet and Miró [28]. This guide was found to be a significant improvement on existing guides, as it had been developed with a comprehensive focus from a variety of sources (ie, research studies, recommendations from professional organizations, and standards governing the development of software for health or medical devices) and an international perspective (ie, resources used came from several regions worldwide). In addition, the guide was created to be of help to all stakeholders and not limited to a specific health problem.

Future Research

Additional research is needed to establish the applicability of the MAG as a guide for health-related mobile app development. Future studies will have to test the MAG with real apps and check their functionality and usability among the different stakeholders who are interested in using it. Furthermore, studies to determine the relative importance of the items and the reliability and suitability of the guide in assessing mobile apps are also warranted. In this regard, a user version of the MAG will be developed to study the association between the quality of the user experience and the score in MAG. In the future, it is highly likely that additional items or criteria will be required to be able to look into the new functions and actions included in mobile apps. Thus, revised and updated versions of the MAG are to be expected.

Limitations

The results of this study should be interpreted in the light of some limitations. The first of these is the representativeness of the participants. Although participants were an international sample of stakeholders, most of them were individuals living in Spain. We do not know if the results would have been the same with other experts. Nevertheless, for the most part, the group included individuals with extensive experience (in clinical work, research, and development), which suggests that their assessments are relevant and of good quality. Second, the number of participating experts changed from round 1 to round 2. However, this is quite normal and to be expected in all Delphi polls [23,28]. Although we cannot be certain that the results would have been the same had all participants in round 1 also responded to round 2, it is quite probable, as the differences between the rounds were minimal. Finally, the number of participants in each round was appropriate for the objectives (a minimum of 7 and maximum of 30 participants is recommended for studies like this; see [39,42]).

Conclusions

Despite the limitations, the results of this study will help advance the field by providing developers, health care professionals, and end users with a valid guide (the MAG) for developing and identifying quality mHealth-related apps. The data shows that the stakeholders reached a consensus on 48 items distributed in 8 categories to establish them as the important criteria for health apps.

The MAG provides stakeholders with a valid tool for systematically reviewing health-related mobile apps. The MAG can be readily used to develop new apps by pointing to the general requirements that a mobile app ought to have if it is to be of high quality. Furthermore, the guide can help to rate existing apps and identify those that are of most interest on the basis of quality criteria. The apps that meet most of the criteria will give users the confidence that their objectives will be fulfilled. It can be used to provide a checklist for the evaluation and development of quality health apps.

Acknowledgments

This work was partly supported by grants from the Spanish Ministry of Economy, Industry and Competitiveness (RTI2018-09870-B-I00; RED2018-102546-T); the European Regional Development Fund; and Obra Social de Caixabank. JM's work is supported by Fundación Grünenthal. PL-V benefited from a predoctoral fellowship (2019 FI_B2 00151) cofinanced by the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya, the European Union, and the European Social Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Importance of the items in each round.
[DOCX File , 53 KB - mhealth_v8i7e17760_app1.docx]

Multimedia Appendix 2
The Mobile App Development and Assessment Guide.
[DOCX File , 22 KB - mhealth_v8i7e17760_app2.docx]

References

1. Accenture. Meet today's healthcare team: patients + doctors + machines. Accenture. 2018. URL: https://www.accenture.com/_acnmedia/PDF-71/Accenture-Health-Meet-Todays-Healthcare-Team-Patients-Doctors-Machines.pdf [accessed 2020-07-14]
2. The growing value of digital health. IQVIA. 2017. URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2020-07-14]
3. Al Bawab AQ, AlQahtani F, McElnay J. Health care apps reported in newspapers: content analysis. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e10237 [FREE Full text] [doi: 10.2196/10237] [Medline: 30348632]
4. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
5. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
6. The socio-economic impact of mobile health. Group TBC. 2012. URL: https://www.telenor.com/wp-content/uploads/2012/05/BCG-Telenor-Mobile-Health-Report-May-20121.pdf [accessed 2020-07-14]
7. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
8. The Mobile Economy. GSMA. 2020. URL: https://www.gsma.com/r/mobileeconomy/ [accessed 2020-07-14]
9. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
10. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312. [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
11. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
12. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
13. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
14. AppSalut Marketplace. TIC Salut Social. 2018. URL: https://ticsalutsocial.cat/en/projectes/mhealth-en/appsalut-marketplace/ [accessed 2020-07-14]

15. López Seguí F, Pratdepàdua Bufill C, Rius Soler A, de San Pedro M, López Truño B, Aguiló Laine A, et al. Prescription and integration of accredited mobile apps in Catalan health and social care: protocol for the AppSalut site design. JMIR Res Protoc 2018 Dec 21;7(12):e11414 [FREE Full text] [doi: 10.2196/11414] [Medline: 30578234]
16. Distintivo AppSaludable. Agencia de Calidad Sanitaria de Andalucia. 2014. URL: http://www.calidadappsalud.com/listado-completo-recomendaciones-app-salud/ [accessed 2020-07-14]
17. Good practice guidelines on health apps and smart devices (mobile health or mhealth). Haute Autorité de Santé. 2016. URL: https://www.has-sante.fr/jcms/c_2681915/en/good-practice-guidelines-on-health-apps-and-smart-devices-mobile-health-or-mhealth [accessed 2020-07-14]
18. Health apps & co: safe digital care products with clearer regulations. IGES Institut GmbH. 2016. URL: https://www.iges.com/kunden/gesundheit/forschungsergebnisse/2016/digitalisierung-iii/index_ger.html [accessed 2020-07-14]
19. Medical devices: software applications (apps). Medicines and Healthcare products Regulatory Agency. 2014. URL: https://www.gov.uk/government/publications/medical-devices-software-applications-apps [accessed 2020-07-14]
20. Policy for device software functions and mobile medical applications. US Food and Drug Administration. 2019. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications [accessed 2020-07-14]
21. Regulation of software as a medical device. Therapeutic Goods Administration. 2019. URL: https://www.tga.gov.au/regulation-software-medical-device [accessed 2020-07-14]
22. Therapeutic G. Therapeutic goods legislation amendment (2019 measures no.1) regulations 2019. Therapeutic Goods Administration. 2019. URL: https://www.tga.gov.au/therapeutic-goods-legislation-amendment-2019-measures-no1-regulations-2019 [accessed 2020-07-14]
23. Green paper on mobile health ("mhealth"). European Commission. 2014. URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2020-07-14]
24. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
25. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
26. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
27. Gaziel-Yablowitz M, Schwartz DG. A review and assessment framework for mobile-based emergency intervention apps. ACM Comput Surv 2018 Apr 14;51(1):1-32. [doi: 10.1145/3145846]
28. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
29. Linstone H, Turoff M, Helmer O. The Delphi Method Techniques and Applications. London: Addison-Wesley; 1975.
30. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Inf Manage 2004 Dec;42(1):15-29. [doi: 10.1016/j.im.2003.11.002]
31. Landeta J. Current validity of the Delphi method in social sciences. Technological Forecasting Soc Change 2006 Jun;73(5):467-482. [doi: 10.1016/j.techfore.2005.09.002]
32. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 2003;56(12):1150-1156. [doi: 10.1016/s0895-4356(03)00211-7] [Medline: 14680664]
33. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011;6(6):e20476. [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
34. Miró J, Huguet A, Nieto R. Predictive factors of chronic pediatric pain and disability: a Delphi poll. J Pain 2007 Oct;8(10):774-792. [doi: 10.1016/j.jpain.2007.04.009] [Medline: 17627893]
35. Expósito-Vizcaíno S, Burjalés-Martí D, Miró J. Factores psicosociales en el dolor crónico oncológico: un estudio Delphi. Rev Soc Esp Dolor 2018;26(1):21-30. [doi: 10.20986/resed.2018.3658/2018]
36. Galán S, de la Vega R, Tomé Pires C, Racine M, Solé E, Jensen MP, et al. What are the needs of adolescents and young adults after a cancer treatment? A Delphi study. Eur J Cancer Care (Engl) 2017 Mar;26(2). [doi: 10.1111/ecc.12488] [Medline: 27007727]
37. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-Mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
38. Landeta J. El método Delphi: una técnica de previsión para la incertidumbre. Barcelona: Ariel; 1999:8434428369.
39. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]

40. LimeSurvey: an open source survey tool. LimeSurvey GmbH. Hamburg, Germany URL: https://www.limesurvey.org/ [accessed 2020-07-14]
41. Miró J, Nieto R, Huguet A. Predictive factors of chronic pain and disability in whiplash: a Delphi poll. Eur J Pain 2008 Jan;12(1):30-47. [doi: 10.1016/j.ejpain.2007.02.004] [Medline: 17400491]
42. Varela-Ruiz M, Díaz-Bravo L, García-Durán R. Descripción y usos del método Delphi en investigaciones del área de la salud. Investigación en Educación Médica 2012;1(2):90-95.

Abbreviations

MAG: Mobile App Development and Assessment Guide
mHealth: mobile health

Edited by G Eysenbach; submitted 10.01.20; peer-reviewed by A García-Palacios, R Warty, T van Mierlo, M Khalemsky; comments to author 03.03.20; revised version received 28.04.20; accepted 03.06.20; published 31.07.20.

Please cite as:
Llorens-Vernet P, Miró J
The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study
JMIR Mhealth Uhealth 2020;8(7):e17760
URL: http://mhealth.jmir.org/2020/7/e17760/
doi:10.2196/17760
PMID:32735226

Original Paper

A Wearable Technology Delivering a Web-Based Diabetes Prevention Program to People at High Risk of Type 2 Diabetes: Randomized Controlled Trial

Emily Staite[1], BSc, MSc; Adam Bayley[1], BSc, MSc; Ebaa Al-Ozairi[2], BMSc (Hons), MBChB, MRCP, CCT; Kurtis Stewart[1], BA, MSc; David Hopkins[3], BSc (Hons), MB CHB, MRCP, FRCP; Jennifer Rundle[4], BA (Hons), MRes, DClinPsy; Neel Basudev[5], MBBS, MRCGP, DCH, DFSRH, BSc, MRES; Zahra Mohamedali[1], BSc; Khalida Ismail[1], BA (Hons), BM BCh, MRCP, MRCPsych, MSc, PhD

[1]Institute of Psychiatry, Psychology and Neurosciences, King's College London, London, United Kingdom

[2]Faculty of Medicine, Department of Medicine & Dasman Diabetes Institute, Kuwait University, Al Kuwayt, Kuwait

[3]King's College Hospital NHS Foundation Trust, King's Health Partners, London, United Kingdom

[4]South Thames Cleft Service, St Thomas' Hospital, London, United Kingdom

[5]Health Innovation Network, London, United Kingdom

Corresponding Author:
Khalida Ismail, BA (Hons), BM BCh, MRCP, MRCPsych, MSc, PhD
Institute of Psychiatry, Psychology and Neurosciences
King's College London
Western Education Centre
10 Cutcombe Road
London, SE5 9RJ
United Kingdom
Phone: 44 207 848 5131
Email: khalida.2.ismail@kcl.ac.uk

Abstract

Background: Intensive lifestyle interventions are effective in reducing the risk of type 2 diabetes, but the implementation of learnings from landmark studies is expensive and time consuming. The availability of digital lifestyle interventions is increasing, but evidence of their effectiveness is limited.

Objective: This randomized controlled trial (RCT) aimed to test the feasibility of a web-based diabetes prevention program (DPP) with step-dependent feedback messages versus a standard web-based DPP in people with prediabetes.

Methods: We employed a two-arm, parallel, single-blind RCT for people at high risk of developing diabetes. Patients with a hemoglobin A_{1c} (HbA_{1c}) level of 39-47 mmol/mol were recruited from 21 general practices in London. The intervention integrated a smartphone app delivering a web-based DPP course with SMS texts incorporating motivational interviewing techniques and step-dependent feedback messages delivered via a wearable device over 12 months. The control group received the wearable technology and access to the web-based DDP but not the SMS texts. As this was a feasibility study, the primary aim was to estimate potential sample size at different stages of the study, including the size of the target study population and the proportion of participants who consented, were randomized, and completed follow-up. We also measured the main outcomes for a full-scale RCT, namely, change in weight and physical activity at 6- and 12-month follow-ups, and secondary outcomes, including changes in the HbA_{1c} level, blood pressure, waist circumference, waist-to-hip ratio, and lipid levels.

Results: We enrolled 200 participants: 98 were randomized to the intervention and 102 were randomized to the control group. The follow-up rate was higher in the control group (87/102, 85.3%) than in the intervention group (69/98, 70%) at 12 months. There was no treatment effect on weight at 6 months (mean difference 0.15; 95% CI −0.93 to 1.23) or 12 months (mean difference 0.07 kg; 95% CI −1.29 to 1.44) or for physical activity levels at 6 months (mean difference −382.90 steps; 95% CI −860.65 to 94.85) or 12 months (mean difference 92.64 steps; 95% CI −380.92 to 566.20). We did not observe a treatment effect on the secondary outcomes measured at the 6-month or 12-month follow-up. For the intervention group, the mean weight was 92.33 (SD 15.67) kg at baseline, 91.34 (SD 16.04) kg at 6 months, and 89.41 (SD 14.93) kg at 12 months. For the control group, the mean weight was 92.59 (SD 17.43) kg at baseline, 91.71 (SD 16.48) kg at 6 months, and 91.10 (SD 15.82) kg at 12 months. In

the intervention group, the mean physical activity was 7308.40 (SD 4911.93) steps at baseline, 5008.76 (SD 2733.22) steps at 6 months, and 4814.66 (SD 3419.65) steps at 12 months. In the control group, the mean physical activity was 7599.28 (SD 3881.04) steps at baseline, 6148.83 (SD 3433.77) steps at 6 months, and 5006.30 (SD 3681.1) steps at 12 months.

Conclusions: This study demonstrates that it is feasible to successfully recruit and retain patients in an RCT of a web-based DPP.

Trial Registration: ClinicalTrials.gov NCT02919397; http://clinicaltrials.gov/ct2/show/NCT02919397

(JMIR Mhealth Uhealth 2020;8(7):e15448) doi:10.2196/15448

KEYWORDS

motivational interviewing; lifestyle; diabetes prevention program; theory of planned behavior; type 2 diabetes mellitus; wearable technology; mobile phone

Introduction

Background

The prevalence of prediabetes is approximately 10% in the UK population [1], with higher rates in adults aged 40 years or older and those of South Asian or African Caribbean ethnicities [2]. Considering the rising prevalence of type 2 diabetes, which accounts for disproportionate and increasing costs to the individual, society, and health care systems globally [3,4], primary prevention is a current public health priority.

Landmark randomized controlled trials (RCTs) have repeatedly shown that intensive face-to-face diabetes prevention programs (DPP) are effective in reducing the risk of type 2 diabetes by approximately 50% [5-7]. The implementation of these landmark studies is expensive, time consuming for the patients and for health systems, and the uptake is often by those with the lowest risk for type 2 diabetes [8]. One solution is adapting the traditional DPP into a web-based DPP using wearable technologies and web-based programs. These are increasingly available from commercial providers at a low cost to the patient or health commissioners, but the evidence of their effectiveness in reducing the risk of type 2 diabetes is limited [9,10]. For example, in a recent RCT of 2062 people in India and the United Kingdom with impaired glucose tolerance, those who received 2 to 3 weekly SMS texts providing lifestyle advice did not have a significant reduction in diabetes conversion (defined by international criteria for fasting plasma glucose or hemoglobin [HbA_{1c}] levels) compared with controls who received standard lifestyle advice at baseline only [11]. Similarly, a pilot pre-post noncontrolled study of a 16-week web-based DPP was associated with reduced weight and lower HbA_{1c} levels, improvements that persisted at the 2-year follow-up [12]. An RCT of a fully automated and algorithm-driven email, web and mobile DPP found significant improvements in biomedical outcomes, weight, and diabetes risk at 6-months compared with the wait list control group [13].

Components of behavior change techniques considered to be most effective in improving diet and physical activity (PA) are based on self-regulatory behaviors, such as goal setting, self-monitoring, giving feedback, utilizing social support, and motivational interviewing (MI) [14,15]. Interventions based on a psychological theory, for example, the theory of planned behavior, are considered to be more effective and have better outcomes in high-risk populations [16], although the minimum threshold of intensity, such as the number of sessions or messages or duration of the intervention, is not known [14,17,18].

Aims

This RCT aimed to test the feasibility of a web-based DPP, consisting of a wearable technology that records PA, integrated with SMS texts based on MI techniques, and lifestyle education delivered via a smartphone app, over 12 months in participants with prediabetes. The primary aims were to assess (1) the potential size of the study population; (2) the proportion of those who consented to be screened for eligibility; (3) the proportion of those who were screened and who were eligible, consented, and randomized; (4) the proportion of those who were randomized and who completed the intervention; and (5) the proportion of those who completed the 6-month and 12-month follow-ups.

Our secondary aims were to measure the change in biomedical outcomes, including reducing weight and increasing PA, to inform the possible range of effect sizes and obtain outcome variance estimates required for sample size calculations in a full-scale trial.

Methods

Study Design

This was a two-arm, parallel, single-blind RCT conducted over 12 months. The trial has been reviewed and given favorable opinion by the London City and East Research Ethics Committee (16/LO/1505).

Setting

We recruited patients from 3 clinical commissioning groups in London (Lambeth, Southwark, and Lewisham), which comprise a population of 912,687 residents, with one of the highest prevalence rates of type 2 diabetes [19] in England and with broad socioeconomic and ethnic diversity [20]. Patients were recruited from participating primary care surgeries with list sizes greater than 6000 patients.

Participants

Patients with an HbA_{1c} level of 39 to 47 mmol/mol were defined as being in a prediabetes state according to the current American Diabetes Association criteria [21]. They were identified using a 2-stage process. First, patients at high risk of developing type

2 diabetes were identified by the general practitioner (GP) who conducted a search on the Egton Medical Information Systems (EMIS) web (clinical software where GPs can access patient health records) using HbA_{1c} results recorded in the previous 12 months. Additional data extracted from EMIS included anonymous ID, gender, date of birth, postcode, ethnic origin, QDiabetes score [22], and BMI. Potentially eligible patients (HbA_{1c} level of 39 to 47 mmol/mol) were invited to undergo screening for eligibility. All GPs used the same postal invitation. Second, the inclusion criteria at screening were as follows: adults aged between 18 and 65 years; BMI≥25 kg/m² (≥23 kg/m² if of Asian ethnicity) [19]; permanent residents in Lambeth, Southwark, or Lewisham; owning a smartphone (iPhone or Android models only)—defined as logging on at least once per day to the internet; being fluent in conversational English; and being ambulatory (eg, capable of walking without mobility support equipment).

The exclusion criteria included diabetes (not including past history of gestational diabetes); pregnancy; planning a pregnancy or lactating during the duration of the study; severe mental illness (severe depression with suicidal ideation, psychosis, bipolar affective disorder, dementia, learning difficulties, substance problem use, or dependence); severe physical disability (eg, that would prevent any increased uptake of physical exercise); advanced active disease, such as cancer or heart failure; any other condition that requires glucose-altering drugs; BMI≥50 kg/m²; and current participation in a weight loss program or DPP. When in doubt, we sought GP confirmation of eligibility.

Baseline Measures

We collected sociodemographic data, including age, gender, postcode of residence, employment status, educational level, and self-reported ethnicity. On the basis of the participant's postcode, we determined their indices of multiple deprivation (IMD) 2015 rank, which indicates the relative level of socioeconomic deprivation in their area [23].

Objective Physical Activity

Objective PA was measured using wearable technology (a wristband manufactured and provided by Buddi Ltd; *Wearable Technology* section below gives more details). Physical activity (number of steps per day) was recorded continuously by the wristband. Baseline PA was the mean step count of the first 7 days of wearing the wristband (starting from the baseline appointment). Days with no recorded steps were removed before calculating each participant's mean step count (Multimedia Appendix 1 shows the number of days used in calculating step counts).

Biomedical Data

We collected HbA_{1c} (mmol/mol) and lipid levels (total cholesterol, high- and low-density lipoproteins, and fasting triglycerides; all values in mmol/L). Weight was measured in light clothing, without shoes, to 0.01 kg, and height to 0.1 cm using a stadiometer (Class 3 Tanita SC240). Weight and height measurements were used to calculate the BMI (kg/m^2). Waist circumference (cm) was measured horizontally halfway between the lowest rib and the upper prominence of the pelvis using a nonextensible steel tape against the bare abdomen. Hip circumference was also measured to calculate the waist-to-hip ratio. Diastolic and systolic blood pressure (BP) and resting heart rate were measured with digital Omron BP monitors (Omron M7) using standardized procedures of the average of 2 readings taken 1 min apart while seated.

Self-Reported Data

Subjective PA was assessed using the International Physical Activity Questionnaire (IPAQ) [24], from which we derived 2 continuous summary scores (sitting minutes and total activity, given in metabolic equivalent of tasks min/week) and a categorical score (low, moderate, or high activity levels) of participants' PA levels in the past week. Depressive symptoms were collected using the patient health questionnaire-9 (PHQ-9) [25]. Readiness to change was measured using the University of Rhode Island change assessment scale (URICA) [26], adapted from 32 items to 12 items for this study, to ask specifically about participants' readiness to change dietary and activity behaviors with regard to their health; scores range from 2 to 14, with scores <8 indicating participants are in the *precontemplation* stage, 8 to 11 the *contemplation* stage, and >11 the *preparation or action* stage. Self-efficacy was measured using the self-efficacy for exercise scale, which has 9 items and a score range of 0 to 90, with higher scores indicating greater self-efficacy [27]. We collected data on smoking status and, if current, how many cigarettes per day. Alcohol intake was measured using the alcohol use disorders identification test (AUDIT) [28].

Intervention

The intervention was based on the theory of planned behavior, which states that to change behavior, people need to form an intention [16]. Intention formation is influenced by 3 constructs: expected value or positive attitude (people see the value in making the change), subjective norm (significant others and peers also value the change), and self-efficacy (people believe they are capable of making the change).

Wearable Technology

All participants were issued with a wristband (manufactured and provided by Buddi Ltd), its charger, and instructions for operating the wristband and downloading the associated study-specific smartphone app. In the baseline appointment, participants downloaded the app onto their smartphone and wirelessly connected it to the wristband via Bluetooth with the help of the researcher if needed. Participants were told that they must maintain the Bluetooth connection to facilitate the transfer of data captured by the Buddi wristband to the participants' smartphones. This allowed participants to track their activity in close to real time via the smartphone d as well as review past activity. If any technical issues arose, participants were able to contact a researcher (who was not blinded to the intervention allocation) for technical support, and any faulty devices were replaced.

Web-Based Education

We scheduled and delivered 22 web-based sessions over 12 months targeting diet, PA, and mental resilience. The curriculum

was based on the newly developed Centers for Disease Control and Prevention PreventT2 curriculum and handouts [28], which is an implementation version of the original landmark DPP studies [5,29,30]. The web-based sessions were available through the smartphone app, with PDF transcripts available to download for each session. SMS texts were sent to participants via the smartphone app to notify them when each module of the web-based DPP was available (1 text every 1-4 weeks approximately).

Motivational SMS Texts

SMS texts were generated and delivered via the smartphone app using principles and techniques from MI to support participants in forming healthy intentions, encourage self-monitoring of lifestyle behaviors, and promote social support [15]. MI is normally a face-to-face collaborative conversation style for strengthening a person's motivation, belief, and commitment to change. We adapted this principle for a virtual setting. The contents of the automated messages were temporally coordinated with the contents of the educational program. Participants typically received 3 to 4 MI-based messages/day (excluding Saturday and Sunday) for 12 months. There were 3 types of messages:

1. Messages targeting lifestyle behaviors encouraged *how* to make lifestyle changes (self-efficacy), for example, *Think about how many staircases you might be able to use today instead of the lift*, followed by *why* messages (expected value of change to the patient) as reinforcement, for example, *Exercise is best done little and often and will have a positive impact on your health*. The content was coordinated to mirror that of the educational program. One of each *how* and *why* message was sent daily.
2. One daily message giving feedback on the activity data received from the wearable technology; the content was based on the level of activity designed to reinforce or encourage an increase in activity levels.
3. Responsive messages were only sent if participants proactively selected the following on the app: *achieve* when they felt they had reached a goal, *crave* when they were thinking about pursuing an unhealthy behavior (eg, eating a high-calorie food or avoiding their exercise regime) but had not acted upon it, or *lapse* when they had acted upon their cravings and need support to re-engage with their good intentions. Participants also had the opportunity to record their achievements within the app.

Control Arm

The control group was provided with the Buddi wristband for the duration of the study and could access their activity data and web-based education material via the smartphone app. They received an automated message (via the smartphone app) informing them when the next educational session was available (ie, 22 messages in total), but they did not receive any other messages. This was weekly for modules 1 to 6, biweekly for modules 7 to 16, and monthly (4 weeks) for modules 17 to 22.

Outcome Measures

For feasibility parameters, the primary outcomes were proportion recruited and randomized and proportion followed up. The primary clinical outcomes were change in weight (kg) and PA (mean steps per day) from baseline to 12 months, with an interim measure at 6 months. Follow-up PA was the mean step count of the 7 days of wear leading up to and including the day of the follow-up appointment.

The secondary outcomes were a change in HbA_{1c} levels and BP at 6 and 12 months and waist circumference, waist-to-hip ratio, and lipid levels at 12 months. The HbA_{1c} level was analyzed as a continuous and categorical variable, with the following categories: normal (<42 mmol/mol), prediabetes (39-47 mmol/mol), and diabetes (>47 mmol/mol).

Sample Size

We aimed for 100 participants per arm, as this was a large enough sample to inform the practicalities of delivering the intervention, recruitment, uptake, and attrition to inform a full-scale trial rather than measures of intervention effects.

Randomization and Allocation Concealment

Before randomization, participants wore the Buddi wristband for 1 week to familiarize with the technology and to collect baseline activity data. All patients were offered a brief educational session on the use of the Buddi device, and an instruction manual was provided. Randomization of participants was conducted by the data manager from an independent clinical trials unit using computer-generated randomization blocks of random sizes and stratified by surgery in a 1:1 ratio. Allocation concealment of the randomization list was held in a password-locked computer. This was an open-label study, but outcome assessors, laboratory technicians, and researchers entering and scoring the data were blinded to patients' allocations.

Statistical Analysis Plan

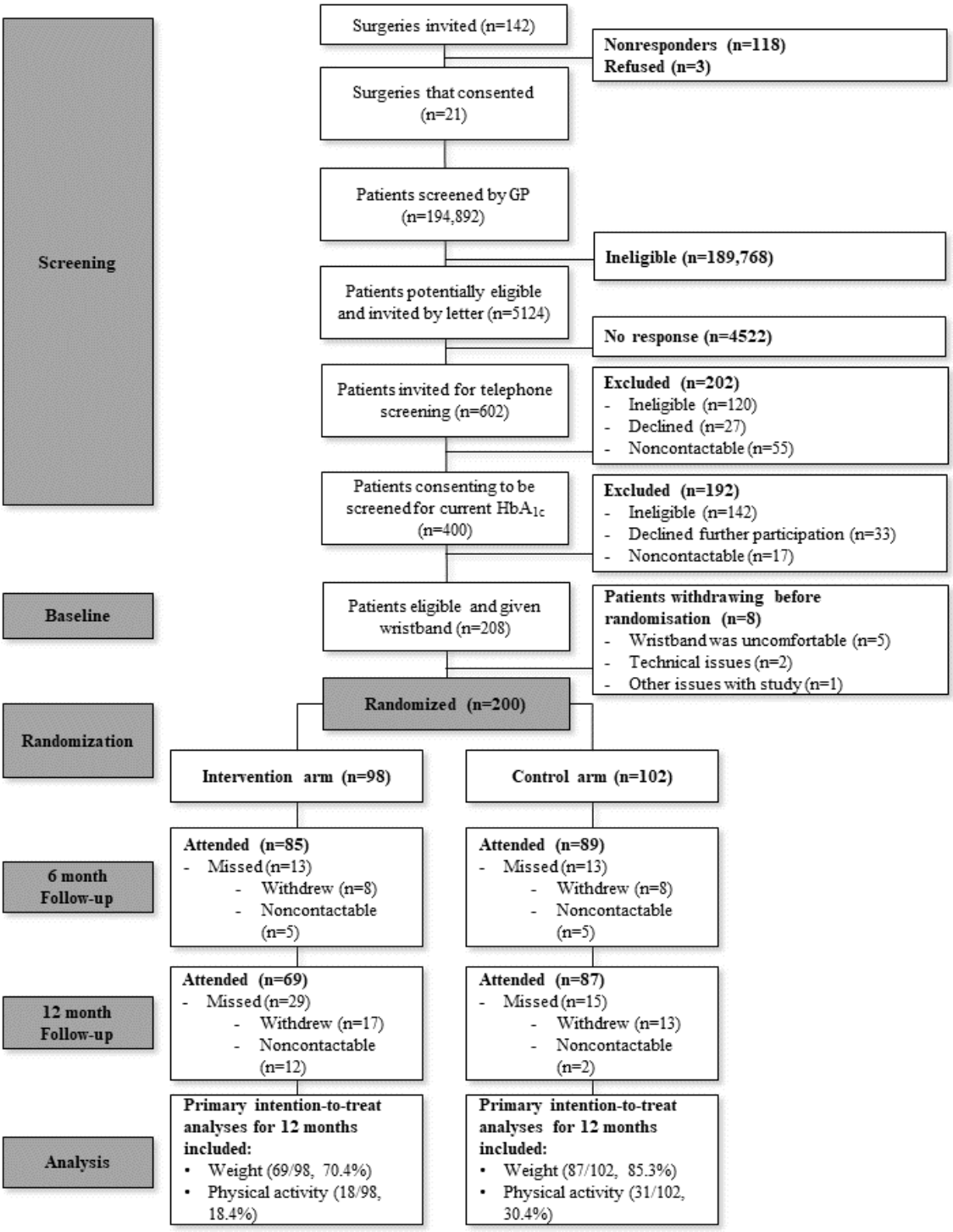
The full statistical analysis plan for this study is provided in Multimedia Appendix 2. Analysis and reporting were in line with the consolidated standards of reporting trials guidelines [29], including its extensions for pilot and feasibility trials; primary analyses were conducted on an intention-to-treat basis and using a two-sided significance level of .05. Statistical analyses were mainly descriptive, aiming to provide estimates of key feasibility parameters and to inform power calculations for a future definitive trial. Descriptive subanalyses were used to explore participation rates among participants based on ethnicity, education level, IMD 2015 score, BMI, depressive symptoms, readiness to change, and self-efficacy. The proportion of missing data for individual items and measures was examined to determine the suitability of instruments and the level of burden for a future full-scale trial. We compared baseline characteristics of (1) those who were eligible and who did and did not consent and (2) those who did and did not provide 12-month follow-up data for the primary clinical outcomes. The differences in treatment effect for the primary and secondary outcomes between the arms at 6- and 12-month follow-ups were analyzed using analysis of covariance–based, linear mixed effects models with prerandomization values as a covariate [30]; STATA's *mixed* command was used for estimation.

Results

Trial Flow

Figure 1 presents participant flow through the trial. Over a 6-month period, we sent postal invitations to 142 surgeries, whose patient list sizes were over 6000. A total 21 of 142 surgeries (14.8%) agreed to participate, from which 194,892 patients had the first-stage screening performed electronically by their GP. Out of these, 5124 were potentially eligible and invited for further screening by their GP by letter and telephone. The number of people who attended and consented for second-stage screening was approximately 11 patients per week, requiring 28 hours per week of a full-time equivalent research worker. Approximately half of those who attended the screening were eligible and consented to be randomized.

Figure 1. Participant flow diagram.



Baseline Characteristics

The baseline characteristics of the participants are presented in Table 1. The two arms were broadly similar (except for ethnicity) and representative of the local catchment in terms of high levels of deprivation, with over half being of nonwhite ethnicity.

Adherence to Intervention

Data were available for 192 participants (92 in the intervention arm). Of these 192, a total of 80 (80.0%) and 60 (65.2%) participants were adherent to the intervention in the intervention and control arms, respectively (X^2_1=4.6; *P*=.03). Participants used the smartphone app on average 41.5% (SD 27.5) and 33.2% (SD 27.5) of days in the control and intervention arms, respectively.

Table 1. Baseline characteristics.

Baseline characteristics	Intervention (n=98)[a]	Control (n=102)[b]	Total (N=200)
Age (years), mean (SD)	51.76 (7.68)	52.78 (8.20)	52.28 (7.94)
Gender, n (%)			
Female	56 (57)	51 (50.0)	106 (53.0)
Ethnicity, n (%)			
White	28 (28)	35 (34.3)	63 (31.5)
African or African Caribbean	50 (51)	58 (56.9)	108 (54.0)
Asian	16 (16)	4 (3.9)	20 (10.0)
Other	4 (4)	5 (4.9)	9 (4.5)
Highest qualification, n (%)			
No formal qualifications	3 (3)	5 (5.2)	8 (4.1)
GCSE[c] or equivalent	29 (29)	23 (23.7)	52 (26.8)
A level or higher	65 (67)	69 (71.1)	134 (69.1)
Currently employed, full- or part-time, n (%)	74 (77)	78 (78.8)	152 (77.9)
Relationship status, n (%)			
Married/cohabiting	61 (62)	59 (57.8)	120 (60.0)
Separated/divorced/widowed	12 (12)	13 (12.7)	25 (12.5)
Single	25 (25)	30 (29.4)	55 (27.5)
IMD[d] 2015 quintiles, n (%)			
1—most deprived	33 (33)	36 (35.3)	69 (34.5)
2	32 (32)	38 (37.3)	70 (35.0)
3	21 (21)	19 (18.6)	40 (20.0)
4	11 (11)	8 (7.8)	19 (9.5)
5—least deprived	1 (1)	1 (1.0)	2 (1.0)
Family history of diabetes, n (%)	52 (53)	51 (50.0)	103 (51.5)
Smoking status, n (%)			
Current smoker	13 (13)	10 (9.8)	23 (11.5)
Ex-smoker	32 (32)	41 (40.2)	73 (36.5)
Nonsmoker	53 (54)	51 (50.0)	104 (52.0)
Number of cigarettes per day for current smokers, mean (SD)	7.50 (5.76)	6.40 (5.21)	7.02 (5.43)
PHQ-9[e] depression score, mean (SD)	4.04 (4.61)	4.25 (4.01)	4.15 (4.31)
AUDIT[f] score category, n (%)			
Abstainer (0)	22 (22)	21 (20.6)	43 (21.5)
Low risk (1-7)	71 (72)	69 (67.6)	140 (70.0)
Possibly harmful (≥8)	5 (5)	12 (11.8)	17 (8.5)
IPAQ[g] total physical activity, MET[h] minutes/week, median (IQR)	2264.18 (2311.16)	2647.01 (2715.67)	2459.43 (2526.55)
URICA[i] readiness to change, n (%)			
Precontemplation (<8)	22 (22.)	12 (11.8)	34 (17.1)
Contemplation (8-11)	54 (55)	70 (68.6)	124 (62.3)
Preparation or action (>11)	21 (21)	20 (19.6)	41 (20.6)
SEE[j] self-efficacy score, mean (SD)	55.15 (22.11)	54.33 (21.95)	54.73 (21.97)

[a]Number of missing cases for the intervention arm are as follows: highest qualification (n=1), currently employed (n=3), PHQ-9 (n=1), URICA (n=1), and SEE (n=2).

[b]Number of missing cases for the control arm are as follows: highest qualification (n=5) and currently employed (n=2).

[c]GCSE: general certificate of secondary education.

[d]IMD: indices of multiple deprivation.

[e]PHQ-9: patient health questionnaire-9.

[f]AUDIT: alcohol use disorders identification test.

[g]IPAQ: international physical activity questionnaire.

[h]MET: metabolic equivalent of the task.

[i]URICA: University of Rhode Island change assessment scale.

[j]SEE: self-efficacy for exercise.

Summary of Outcome Data

Table 2 presents a descriptive summary of the clinical outcomes at each time point. We observed that the distributions for each primary outcome were positively skewed; however, log-transforming the data did not improve the distribution or alter the results of the analyses below. The pooled SD at 12 months for weight was 15.43 kg and for PA was 3588.76 steps.

When categorizing participants' metabolic status based on their HbA_{1c} values, 3 (3.5%) and 5 (7.3%) in intervention group and 2 (2.2%) and 6 (6.9%) in the control group met the cut-off for diabetes (>47 mmol/mol) at 6 and 12 months, respectively. Participants with HbA_{1c} >47 mmol/mol were referred to the GP. Participants were informed of their results via telephone and told to contact their GP. A total of 4 (4.7%) and 9 (10.1%) participants returned to the normal range (<39 mmol/mol) in the intervention and control arms, respectively, at 6 months. At 12 months, 0 and 4 (4.6%) participants were in the normal range in the intervention and control arms, respectively. There was no difference in the proportion of participants who developed type 2 diabetes or who returned to normal HbA_{1c} levels between the 2 groups at 6 months (X^2_2=2.0; *P*=.36) or 12 months (X^2_2=3.2; *P*=.20).

Table 2. Summary of primary and secondary outcomes and pairwise comparisons.

Outcomes and time	Intervention arm		Control arm		Mean difference (95% CI)
	n	Mean (SD)	n	Mean (SD)	
Weight (kg)					
Baseline	98	92.33 (15.67)	102	92.59 (17.43)	—[a]
6-month follow-up	85	91.34 (16.04)	89	91.71 (16.48)	0.15 (−0.93 to 1.23)[b]
12-month follow-up	69	89.41 (14.93)	87	91.10 (15.82)	0.07 (−1.29 to 1.44)
PA[c] (mean steps per day)[d]					
Baseline	87	7308.40 (4911.93)	93	7599.28 (3881.04)	—
6-month follow-up	36	5008.76 (2733.22)	51	6148.83 (3433.77)	−382.90 (−860.65 to 94.85)
12-month follow-up	18	4814.66 (3419.65)	31	5006.30 (3681.1)	92.64 (−380.92 to 566.20)
BP[e] diastolic (mm Hg)					
Baseline	98	82.92 (10.68)	102	81.47 (8.96)	—
6-month follow-up	82	81.41 (10.19)	89	83.22 (8.46)	−2.24 (−4.54 to 0.06)
12-month follow-up	68	83.03 (10.33)	87	83.87 (9.06)	−1.61 (−3.93 to 0.70)
BP systolic (mm Hg)					
Baseline	98	125.51 (17.39)	102	124.52 (12.19)	—
6-month follow-up	82	124.15 (16.33)	89	127.33 (13.30)	3.50 (−7.05 to 0.05)
12-month follow-up	68	125.37 (16.07)	87	127.54 (14.16)	−2.62 (−6.37 to 1.12)
Cholesterol:HDL[f] ratio					
Baseline	97	4.02 (0.99)	102	4.14 (1.02)	—
12-month follow-up	68	4.06 (1.20)	86	3.97 (1.04)	0.11 (−0.08 to 0.30)
HbA_{1c} (mmol/mol)					
Baseline	98	42.27 (2.32)	102	42.29 (1.98)	—
6-month follow-up	85	42.12 (2.44)	89	41.82 (3.05)	0.20 (−0.50 to 0.91)
12-month follow-up	68	44.06 (2.31)	87	43.54 (2.68)	0.53 (−0.19 to 1.25)
HDL (mmol/L)					
Baseline	97	1.39 (0.33)	102	1.34 (0.33)	—
12-month follow-up	68	1.39 (0.32)	86	1.38 (0.31)	0.00 (−0.07 to 0.06)
LDL[g] (mmol/L)					
Baseline	96	3.33 (0.84)	101	3.28 (0.83)	—
12-month follow-up	67	3.30 (0.85)	85	3.28 (0.85)	0.02 (−0.18 to 0.23)
Total cholesterol (mmol/L)					
Baseline	97	5.35 (0.96)	102	5.30 (0.93)	—
12-month follow-up	68	5.37 (1.03)	86	5.28 (0.99)	0.10 (−0.11 to 0.31)
Triglycerides (mmol/L)					
Baseline	97	1.39 (1.14)	102	1.39 (0.86)	—
12-month follow-up	68	1.72 (3.13)	86	1.35 (0.84)	0.39 (−0.02 to 0.81)
Waist circumference (cm)					
Baseline	97	103.31 (12.21)	100	103.92 (12.32)	—
12-month follow-up	69	96.73 (11.69)	86	97.38 (10.99)	−0.60 (−2.45 to 1.26)
Waist-to-hip ratio					
Baseline	97	0.92 (0.07)	100	0.93 (0.08)	—

Outcomes and time	Intervention arm		Control arm		Mean difference (95% CI)
	n	Mean (SD)	n	Mean (SD)	
12-month follow-up	69	0.91 (0.08)	86	0.90 (0.08)	0.00 (−0.02 to 0.02)

[a]Baseline comparisons are not done, as per the statistical analysis plan noted above.

[b]Pairwise comparison outputs were calculated by subtracting the control arm from the intervention arm, so a negative value indicates the control arm had a higher mean.

[c]PA: physical activity.

[d]Number of days included in the step count calculations is given in Multimedia Appendix 1.

[e]BP: blood pressure.

[f]HDL: high-density lipoprotein.

[g]LDL: low-density lipoprotein.

Primary Analyses

The fixed and random effects of the mixed effects models for weight and PA are presented in Multimedia Appendix 3. There was no treatment effect on weight or PA at 6 or 12 months (Table 2), or for HbA_{1c} levels, waist circumference, waist-to-hip ratio, lipid levels, or BP at 12 months.

Sensitivity Analyses

Our sensitivity analyses adjusting for baseline characteristics did not alter our conclusions of the primary outcome analysis above. Readiness to change (URICA stage) did not moderate the treatment-by-time effect for weight ($F_{1,2}$=0.10; *P*=.90) or PA ($F_{1,2}$=0.56; *P*=.57).

Our analysis of responders to the control or intervention arm indicated that improvements in weight and/or PA were associated with baseline smoking status (X^2_2=11.6; *P*=.003) and PHQ-9 categorical score (X^2_4=10.7; *P*=.03; Multimedia Appendix 4).

Testing the associations between IPAQ scores and mean step counts revealed a significant positive correlation between the intervention arm's IPAQ total activity score and step counts at baseline only (r_{85}=0.22; *P*=.04; Multimedia Appendix 5). The series of 1-way analysis of variances showed that the IPAQ activity category (low, moderate, or high) was only associated with step counts for the intervention arm at baseline ($F_{2,84}$=4.83; *P*=.01).

Discussion

We have demonstrated that conducting an RCT that tested web-based DPP delivered via a wearable technology and a smartphone app is feasible in terms of participation and retention of study patients over a period of 12 months.

Principal Findings

We successfully reached our target sample size in 36 weeks, with a recruitment rate of 5 to 6 patients per week per full-time research worker. For full-scale RCT testing the intervention with an expected effect size of 0.1 and 90% power, we estimate recruitment will take about 3.5 years per research worker or just over 1 year for 3 research workers (Multimedia Appendix 6).

We piloted the statistical plan that we would anticipate using for a full-scale RCT, which appeared to be valid, although hypothesis testing was only a secondary aim. The only significant difference observed was that those in the intervention group had lower PA levels at 6 months compared with those in the control group, but they were also less likely to complete the intervention and more likely to be noncontactable or have withdrawn at follow-up.

An important difference between our study and previous RCTs [11] is that we showed that it is feasible to use a wearable technology that continually records levels of PA (ie, step count). Furthermore, our study included male and female participants compared with another study that included only male participants [31]. We also had higher rates of retention than those in other studies. For instance, in an RCT of a fully automated and algorithm-driven web-based DPP [13], 72% of intervention participants were still interacting with the program at 6 months. We achieved 87% of participants attending their 6-month follow-up.

Strengths and Limitations

A key strength is that this is the first feasibility study in the United Kingdom, which tests an automated web-based DPP designed to mirror the landmark face-to-face DPPs. The sample was representative of the ethnic and social diversity common in prediabetes. We noted that there was variation by GP surgery in response, and this suggests that clustering should be considered in any full-scale RCT.

An important observation is the importance of sustaining the functionality of the technology. Despite a prior *road-test* of the wearable technology and its connectivity to the smartphone app by the commercial provider in a handful of volunteers, there were mechanical and technical failures that may have resulted in participants randomized to the intervention arm receiving less of the required *dose* and withdrawing because of these issues. For example, we observed a sharp decrease in the mean step counts for participants at 6 months and 12 months compared with baseline, suggesting that the wristband may not have been accurately recording PA over time or that participants were unaware of ongoing technical issues. Other explanations could be that participants reduced their step count over time despite being in the intervention group or that participants reduced the amount of time they wore the device.

Participant adherence to the intervention could not be comprehensively documented as the technology for this did not exist; therefore, we were not able to capture how the participants used the app, particularly if they accessed the DPP education materials (and for how long) or if they read the SMS messages sent to them.

The follow-up rates for the primary outcomes were, on average, 75%, and predictors of missing data included higher BMI, presence of depressive symptoms, and current smokers (ie, those more at risk of developing type 2 diabetes and increasing the risk of underestimation).

We assessed the degree to which the wristband-derived step counts (ie, objectively measured PA) were associated with a validated self-reported measure of PA levels. We observed that the 2 measures only corresponded to a limited extent in the control arm, with step counts being weakly and positively correlated with total self-reported activity at 2 time points. The self-reported scores were not associated with step counts in the intervention arm. There are several possible reasons for this discrepancy, one being that these participants disengaged from the intervention (ie, wearing the wristband because of the technical problems or the higher intensity of messages).

The responses of the participants to our implementation processes questionnaire were generally favorable of the intervention; however, they did note several key areas for improvement for a full-scale RCT. These included improving the wristband's clasp, offering the wristband in different styles, and improving the app's accessibility (eg, offering the educational material in other digital formats) and formatting (eg, improving the readability and precision of the activity graphs and personalizing the content and frequency of messages).

Research and Clinical Implications

This study found that there is a sufficiently large target population of patients for screening and a reasonably good participation rate (ie, patients are keen to receive support for diabetes prevention). Ensuring an optimum balance in the intensity of information sent and the functionality of the technology are potential key components to consider for a full-scale RCT.

Acknowledgments

Funding for this study was provided by Innovate UK (ID: 21370). The funder had no role in the design of the study or the collection, analysis, and interpretation of data, and the funder had no part in writing the manuscript. Buddi Ltd provided the wristband and worked with this King's College London (KCL) research team to develop the smartphone app. The paper does not necessarily represent the views of Innovate UK/Buddi. KI is funded in part by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London Maudsley Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. KI has received honorarium for educational lectures from Janssen, Sanofi, Eli Lilly, and Novo Nordisk. The authors thank Prof Daniel Stahl (Department of Biostatistics, Institute of Psychiatry, Psychology & Neuroscience, KCL) for feedback on the statistical plan in the protocol. The other authors have none to declare.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of days used in calculating step counts.
[DOCX File , 13 KB - mhealth_v8i7e15448_app1.docx]

Multimedia Appendix 2
Full statistical analysis plan.
[DOCX File , 14 KB - mhealth_v8i7e15448_app2.docx]

Multimedia Appendix 3
Fixed and random effects of the mixed effects models for weight and PA.
[DOCX File , 13 KB - mhealth_v8i7e15448_app3.docx]

Multimedia Appendix 4
Analysis of responders.
[DOCX File , 14 KB - mhealth_v8i7e15448_app4.docx]

Multimedia Appendix 5
Testing associations between the International Physical Activity Questionnaire scores and mean step count.
[DOCX File , 16 KB - mhealth_v8i7e15448_app5.docx]

Multimedia Appendix 6
Sample size calculations for a future full-scale randomized controlled trial.
[DOCX File , 16 KB - mhealth_v8i7e15448_app6.docx]

Multimedia Appendix 7
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 11844 KB - mhealth_v8i7e15448_app7.pdf]

References

1. Public Health England. Government of UK Developer Documentation. 2015. NHS Diabetes Prevention Programme (NHS DPP) Non-Diabetic Hyperglycaemia URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/456149/Non_diabetic_hyperglycaemia.pdf [accessed 2018-07-02]
2. Mainous A, Tanner R, Baker R, Zayas C, Harle CA. Prevalence of prediabetes in England from 2003 to 2011: population-based, cross-sectional study. BMJ Open 2014 Jun 9;4(6):e005002 [FREE Full text] [doi: 10.1136/bmjopen-2014-005002] [Medline: 24913327]
3. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. Diabet Med 2012 Jul;29(7):855-862. [doi: 10.1111/j.1464-5491.2012.03698.x] [Medline: 22537247]
4. Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. Pharmacoeconomics 2015 Aug;33(8):811-831 [FREE Full text] [doi: 10.1007/s40273-015-0268-9] [Medline: 25787932]
5. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
6. Lindström J, Peltonen M, Eriksson J, Ilanne-Parikka P, Aunola S, Keinänen-Kiukaanniemi S, Finnish Diabetes Prevention Study (DPS). Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish diabetes prevention study (DPS). Diabetologia 2013 Feb;56(2):284-293. [doi: 10.1007/s00125-012-2752-5] [Medline: 23093136]
7. Pan X, Li G, Hu Y, Wang J, Yang W, An Z, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and diabetes study. Diabetes Care 1997 Apr;20(4):537-544. [doi: 10.2337/diacare.20.4.537] [Medline: 9096977]
8. Aguiar E, Morgan P, Collins C, Plotnikoff R, Callister R. Efficacy of interventions that include diet, aerobic and resistance training components for type 2 diabetes prevention: a systematic review with meta-analysis. Int J Behav Nutr Phys Act 2014 Jan 15;11:2 [FREE Full text] [doi: 10.1186/1479-5868-11-2] [Medline: 24423095]
9. Ma J, Yank V, Xiao L, Lavori P, Wilson S, Rosas L, et al. Translating the diabetes prevention program lifestyle intervention for weight loss into primary care: a randomized trial. JAMA Intern Med 2013 Jan 28;173(2):113-121 [FREE Full text] [doi: 10.1001/2013.jamainternmed.987] [Medline: 23229846]
10. Johnson M, Jones R, Freeman C, Woods H, Gillett M, Goyder E, et al. Can diabetes prevention programmes be translated effectively into real-world settings and still deliver improved outcomes? A synthesis of evidence. Diabet Med 2013 Jan;30(1):3-15 [FREE Full text] [doi: 10.1111/dme.12018] [Medline: 22998334]
11. Nanditha A, Thomson H, Susairaj P, Srivanichakorn W, Oliver N, Godsland IF, et al. A pragmatic and scalable strategy using mobile technology to promote sustained lifestyle changes to prevent type 2 diabetes in India and the UK: a randomised controlled trial. Diabetologia 2020 Mar;63(3):486-496 [FREE Full text] [doi: 10.1007/s00125-019-05061-y] [Medline: 31919539]
12. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res 2015 Apr 10;17(4):e92 [FREE Full text] [doi: 10.2196/jmir.4052] [Medline: 25863515]
13. Block G, Azar K, Romanelli R, Block T, Hopkins D, Carpenter H, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
14. Baker M, Simpson K, Lloyd B, Bauman A, Singh MA. Behavioral strategies in diabetes prevention programs: a systematic review of randomized controlled trials. Diabetes Res Clin Pract 2011 Jan;91(1):1-12. [doi: 10.1016/j.diabres.2010.06.030] [Medline: 20655610]
15. Miller W, Rollnick S. Motivational Interviewing: Preparing People for Change. New York, USA: The Guilford Press; 2002.
16. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
17. Ebrahim S, Taylor F, Ward K, Beswick A, Burke M, Smith DG. Multiple risk factor interventions for primary prevention of coronary heart disease. Cochrane Database Syst Rev 2011 Jan 19(1):CD001561. [doi: 10.1002/14651858.CD001561.pub3] [Medline: 21249647]

18. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
19. NHS Digital. 2017. Quality and Outcomes Framework (QOF) - 2016-17 URL: https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data/quality-and-outcomes-framework-qof-2016-17 [accessed 2018-07-02]
20. Office of National Statistics. Census 2011 URL: https://www.ons.gov.uk/census/2011census/2011censusdata [accessed 2017-05-01]
21. American Diabetes Association. 2 Classification and diagnosis of diabetes. Diabetes Care 2018 Jan;41(Suppl 1):S13-S27. [doi: 10.2337/dc18-S002] [Medline: 29222373]
22. Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. Br Med J 2009 Mar 17;338:b880 [FREE Full text] [doi: 10.1136/bmj.b880] [Medline: 19297312]
23. Government of UK. 2015. English Indices of Deprivation 2015 URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 [accessed 2018-07-02]
24. 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]
25. Spitzer R, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. J Am Med Assoc 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
26. Dozois D, Westra H, Collins K, Fung T, Garry J. Stages of change in anxiety: psychometric properties of the University of Rhode Island change assessment (URICA) scale. Behav Res Ther 2004 Jun;42(6):711-729. [doi: 10.1016/S0005-7967(03)00193-1] [Medline: 15081886]
27. Resnick B, Jenkins LS. Testing the reliability and validity of the Self-Efficacy for Exercise scale. Nurs Res 2000;49(3):154-159. [doi: 10.1097/00006199-200005000-00007] [Medline: 10882320]
28. Saunders J, Aasland O, Babor T, 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. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
29. Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomised trials. Br Med J 2004 Mar 20;328(7441):702-708 [FREE Full text] [doi: 10.1136/bmj.328.7441.702] [Medline: 15031246]
30. Bauer DJ, Sterba SK, Hallfors DD. Evaluating group-based interventions when control participants are ungrouped. Multivariate Behav Res 2008 Apr 02;43(2):210-236 [FREE Full text] [doi: 10.1080/00273170802034810] [Medline: 20396621]
31. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013 Nov;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]

Abbreviations

AUDIT: alcohol use disorders identification test
BP: blood pressure
CDC: Centers for Disease Control and Prevention
DPP: diabetes prevention program
EMIS: Egton Medical Information Systems
GP: general practitioner
HbA_{1c}: hemoglobin A_{1c}
IMD: indices of multiple deprivation
IPAQ: International Physical Activity Questionnaire
KCL: King's College London
MI: motivational interviewing
PHQ-9: patient health questionnaire-9
RCT: randomized controlled trial
URICA: University of the Rhode Island change assessment scale

Edited by G Eysenbach; submitted 19.07.19; peer-reviewed by Y Pinillos, K Lukaschek, M Whelan; comments to author 21.10.19; revised version received 13.12.19; accepted 29.02.20; published 15.07.20.

Please cite as:
Staite E, Bayley A, Al-Ozairi E, Stewart K, Hopkins D, Rundle J, Basudev N, Mohamedali Z, Ismail K
A Wearable Technology Delivering a Web-Based Diabetes Prevention Program to People at High Risk of Type 2 Diabetes: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e15448
URL: https://mhealth.jmir.org/2020/7/e15448
doi:10.2196/15448
PMID:32459651

Original Paper

Mobile Personal Health Care System for Noninvasive, Pervasive, and Continuous Blood Pressure Monitoring: Development and Usability Study

Luis J Mena[1], PhD; Vanessa G Félix[1], PhD; Rodolfo Ostos[1], PhD; Armando J González[1], MSc; Rafael Martínez-Peláez[2], PhD; Jesus D Melgarejo[3], MD; Gladys E Maestre[4], MD, PhD

[1]Academic Unit of Computing, Master Program in Applied Sciences, Universidad Politecnica de Sinaloa, Mazatlan, Mexico

[2]Faculty of Information Technology, Universidad de la Salle Bajío, Leon, Mexico

[3]Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, University of Leuven, Leuven, Belgium

[4]Departments of Neurosciences and Human Genetics, and Rio Grande Valley Alzheimer´s Disease Resource Center for Minority Aging Research, University of Texas Rio Grande Valley, Brownsville, TX, United States

Corresponding Author:
Luis J Mena, PhD
Academic Unit of Computing, Master Program in Applied Sciences
Universidad Politecnica de Sinaloa
Carretera Municipal Libre Mazatlan-Higueras Km. 3
Mazatlan, 82199
Mexico
Phone: 52 6691800695 ext 140
Email: lmena@upsin.edu.mx

Abstract

Background: Smartphone-based blood pressure (BP) monitoring using photoplethysmography (PPG) technology has emerged as a promising approach to empower users with self-monitoring for effective diagnosis and control of hypertension.

Objective: This study aimed to develop a mobile personal health care system for noninvasive, pervasive, and continuous estimation of BP level and variability, which is user friendly for elderly people.

Methods: The proposed approach was integrated by a self-designed cuffless, calibration-free, wireless, and wearable PPG-only sensor and a native purposely designed smartphone app using multilayer perceptron machine learning techniques from raw signals. We performed a development and usability study with three older adults (mean age 61.3 years, SD 1.5 years; 66% women) to test the usability and accuracy of the smartphone-based BP monitor.

Results: The employed artificial neural network model had good average accuracy (>90%) and very strong correlation (>0.90) (P<.001) for predicting the reference BP values of our validation sample (n=150). Bland-Altman plots showed that most of the errors for BP prediction were less than 10 mmHg. However, according to the Association for the Advancement of Medical Instrumentation and British Hypertension Society standards, only diastolic blood pressure prediction met the clinically accepted accuracy thresholds.

Conclusions: With further development and validation, the proposed system could provide a cost-effective strategy to improve the quality and coverage of health care, particularly in rural zones, areas lacking physicians, and areas with solitary elderly populations.

(JMIR Mhealth Uhealth 2020;8(7):e18012) doi:10.2196/18012

KEYWORDS

mHealth; photoplethysmography; blood pressure monitoring; hypertension

Introduction

Background

Although hypertension is the most important modifiable risk factor for global disability-adjusted life-years among diseases, injuries, and risk factors [1], its prevalence continues to increase globally. Currently, an estimated 1.13 billion people worldwide have hypertension (blood pressure [BP] ≥140/90 mmHg) [2]. It is known as a "silent killer," as its signs and symptoms can be very subtle and slow until after key organs are severely affected. To identify BP dysregulation and ascertain if treatment or lifestyle modifications are appropriate for control, BP needs to be measured frequently and in real-life conditions. Hypertension prevalence rates are usually based on clinic BP measures, but these can exclude a substantial number of people who have "masked" hypertension (normotensive by clinical measurement, but hypertensive by ambulatory or home BP readings) [3], which confers cardiovascular risk similar to that of sustained hypertension [4]. Therefore, the development and implementation of new strategies to stratify population-level risk and identify people who are treated but still have uncontrolled hypertension or have "masked" hypertension are desperately needed.

Ambulatory BP monitoring is the noninvasive gold standard for the clinical diagnosis of BP dysregulation, as it records BP over a period of 24 hours or more using automated upper-arm cuff oscillometric devices during normal daily activities, thereby providing a much more complete and representative picture of BP patterns than traditional single clinic-based measurement [5]. Continuous BP monitoring records can provide an estimate of the true BP level, the circadian BP profile, and information on inherent BP variability, which is considered a potential independent predictor of cardiovascular events or complications [6-8]. Nevertheless, the periodic inflation of the cuff to register BP readings is associated with moderate discomfort or pain and severe restriction in everyday activities during the monitoring period [9].

The use of noninvasive and cuffless BP monitors has emerged as a more comfortable and user-friendly alternative to estimate BP measurement [10]. The adoption of photoplethysmography (PPG) technology has led to the development of small, inexpensive, and wearable optoelectronic sensors for long-term BP monitoring [11-13]. The underlying principle of the PPG method is based on illumination of the skin by a light-emitting diode (LED) and measurement of the amount of light that is absorbed or reflected from living tissue using a luminosity sensor. Since the tissue changes its vascular tone according to the volume of blood pumped by the heart during each cardiac cycle, the PPG signal can detect changes in volumetric blood flow by the photoelectric technique [11,13]. Thus, variations in volume and distension of the arteries can relate to the heart's systolic and diastolic periods to estimate BP.

Most PPG-based BP monitors have been developed through synchronicity-dependent approaches that measure the PPG signal simultaneously with another biosignal or two PPG signals from different body sites. This type of approach requires two sensors to calculate the time that a single heartbeat pulse takes to travel from the heart to a peripheral location or from one arterial site to another. Generally, an initial or periodic calibration step using an additional BP monitor as a reference is required to obtain a more reliable BP measurement [14-17]. Another simple and less obtrusive approach uses a single PPG sensor. BP is estimated by extracting time and frequency features derived from the PPG waveform to examine the similarity between PPG and arterial BP morphologies or directly using the raw PPG signal [18]. However, PPG-only approaches need more robust methods to accurately describe the inherent but not well-established relationship between PPG signal variations and BP changes [11,19].

Parallel advances in sensing technologies combined with the growth of mobile computing and wireless communication are shifting health care from traditional clinic monitoring to real-time personal monitoring. The development and application of mobile personal health care systems could improve quality and coverage, while reducing costs through early detection [20,21]. Thus, the use of a smartphone equipped with built-in or external sensors is considered a promising approach to empower users with self-monitoring for the effective diagnosis and control of hypertension and to facilitate timely patient and health care provider communication for medical feedback and clinical support [22-27]. Underserved populations (mainly elderly people living in developing countries) will benefit from mobile technology [28,29].

On the other hand, with the increasing advancement and popularity of machine learning methods, novel ways to improve BP estimation from a single PPG signal without any additional sensor have emerged [13,19]. Given that light emitted by LEDs can penetrate an area that involves skin, arteries, veins, blood, bone, and other tissues, optical absorption changes detected by a PPG sensor represent a complex mixture of pulsatile and nonpulsatile blood flow components [13,30]. Therefore, heuristic modeling based on advanced artificial neural networks (ANNs) using nonlinear regression could aid in dealing with confounding factors and improve the result of BP estimation from feature extraction or raw PPG signals [19,31-34].

Recent Work

We provide an overview of recent work on smartphone-based BP monitors using single or combined PPG approaches, whose estimates were compared with measures recorded via a reference BP device (Table 1) [35-42].

Table 1. Overview of smartphone-based blood pressure monitors using single or combined photoplethysmography approaches.

Smartphone PPG[a] approach	Correlation coefficient		Error bias (mmHg), mean (SD)		BP[b] measures	Reference BP device
	SBP[c]	DBP[d]	SBP	DBP		
Chandrasekaran et al [35]	NR[e]	NR	2.45[f]	1.71[f]	500	Mercury sphygmomanometer
Plante et al [36]	0.44[g]	0.41[g]	12.40 (10.50)[h]	10.10 (8.10)[h]	101	Omron HEM-907
Wang et al [37]	NR	0.81[i]	NR	6.70[j]	196	Microlife BP3NA1-1x
Chandrasekhar et al [38]	0.76[i]	0.79[i]	3.30 (8.80)[f]	−5.60 (7.70)[f]	32	Omron BP7650N
Chandrasekhar et al [39]	0.79[i]	0.78[i]	−4.00 (11.40)[f]	−9.40 (9.70)[f]	18	Omron BP786
Raichle et al [40]	0.40[g]	NR	5.00 (14.50)[h]	NR	96	Omron HBP-1300
Dey et al [41]	NR	NR	6.90 (9.00)[h]	5.00 (6.10)[h]	205	Mercury sphygmomanometer
Luo et al [42]	0.67[i]	0.47[i]	0.39 (7.30)[f]	−0.20 (6.00)[f]	1328	CNAP Monitor 500

[a]PPG: photoplethysmogram.
[b]BP: blood pressure.
[c]SBP: systolic blood pressure.
[d]DBP: diastolic blood pressure.
[e]NR: not reported.
[f]Mean error.
[g]Spearman correlation.
[h]Mean absolute error.
[i]Pearson correlation.
[j]Root mean square error.

PPG-Combined Approaches

Chandrasekaran et al [35] made the earliest attempts to monitor BP changes using an external microphone and smartphone camera for recording heart sounds and corresponding finger pulse. From preprocessed phonocardiography and PPG signals, they calculated the vascular transit time [43] (transmission delay of blood from the heart to a body peripheral point for one heartbeat) to estimate BP values. Approach accuracy was assessed by separately measuring BP using a commercial sphygmomanometer. However, the data collection procedure and correlation analysis with BP reference measures were not reported.

A validation study of this method was performed by Plante et al [36] using a commercial iOS app to estimate BP from PPG and phonocardiography signals collected with a smartphone camera and microphone. Simultaneous BP measures were taken using a validated oscillometric device and smartphone to calculate the mean and SD of absolute differences, correlation coefficient, and British Hypertension Society (BHS) accuracy grade [44] between reference and test values. The findings indicated a highly inaccurate BP estimation by showing weak correlation, the lowest possible BHS score, and 78% misclassification of hypertensive measures.

In 2018, Wang et al [37] developed a smartphone BP monitoring app using a built-in accelerometer and camera to capture vibration caused by heartbeats and fingertip pulse. Only diastolic BP (DBP) measures were estimated by calculating pulse transit time [17] (time taken by a heart pulse to travel between two arterial sites), using seismocardiography and PPG signals. DBP estimated with initial calibration showed very strong correlation and adequate average error against reference values measured simultaneously with a commercial oscillometric device.

Additionally, in 2018, Chandrasekhar et al [38] developed a calibration-free BP measurement smartphone device based on an extension of the oscillometric principle. The approach involved external PPG and force sensors to detect blood flow and applied pressure from a finger, and an Android app to calculate and display BP estimates. Comparison with reference values obtained by a standard oscillometric device showed a strong correlation and error range close to the limits established by the Association for the Advancement of Medical Instrumentation (AAMI) [45].

Thereafter, the same authors [45] developed an iOS app to estimate BP via a similar method, but using the smartphone camera to detect blood volume oscillations and a strain gauge array under the screen to measure finger pressure [39]. The assessment of accuracy also demonstrated high correlation and good agreement in terms of error bias and precision with respect to standard reference measures.

PPG-Only Approaches

In 2018, Raichle et al [40] conducted a study to assess the performance of an iOS app estimating systolic BP (SBP) via PPG signals recorded by a smartphone camera from the finger. The estimation algorithm determined BP based on morphology and frequency analysis of PPG signals. Validation results reported a moderate correlation, an adequate mean absolute

error but with high SD, and the worst BHS accuracy grade in comparison with reference SBP measured using an oscillometric device.

In the same year, Dey et al [41] developed an ensemble of BP prediction models based on PPG feature extraction and values of demographic and physiological profiles. PPG signals were acquired using a heart rate sensor embedded in a smartphone, and an Android app was used to display BP estimates. A machine learning approach using the Lasso regression model [46] was adopted for initial DBP estimation, and subsequently, this output was used as feedback to derive SBP. The combined model of PPG features and demographic and physiological partitioning showed adequate absolute error and precision with respect to reference BP measured using a mercury-based cuff device.

More recently, Luo et al [42] proposed a BP monitoring approach using a variant of the remote PPG technique to detect facial blood flow changes from videos captured with a smartphone camera, and a multilayer perceptron machine learning algorithm was used for BP estimation. The prediction model used orthogonal eigenvectors as input by applying principal component analysis on features extracted from facial blood flow signals, metafeatures for imaging condition normalization, environmental temperature variations, and individual physical characteristics. Assessment against reference measures from a continuous finger BP monitor indicated moderate and strong correlations for DBP and SBP, and high BP prediction accuracy and precision according to the AAMI standard [45].

Goal of This Study

In this study, we address these issues by developing a mobile personal health care system for noninvasive, pervasive, and continuous BP monitoring, with the goal of improving the early diagnosis and control of hypertension and identifying a potential cardiovascular predictor (abnormal BP variability). The proposed approach is integrated by a self-designed cuffless, calibration-free, wireless, and wearable PPG-only sensor and a native smartphone app designed to be user friendly for elderly people, which is based on machine learning techniques for BP level and variability estimation from raw PPG signals.

Methods

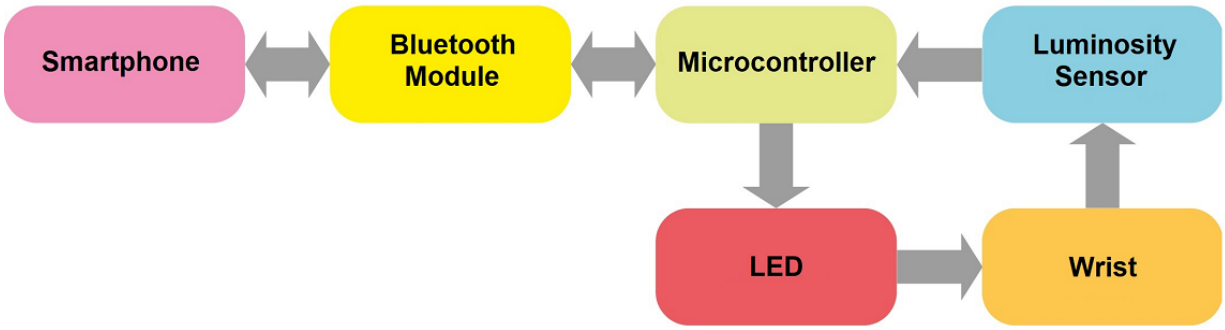
Brief Description

This section provides an overview of our approach, which is based on the development of a PPG-only sensor to acquire and transfer raw signals to a smartphone and a mobile app using multilayer perceptron machine learning techniques to estimate BP level and variability.

Approach

The mobile personal health care system consists of a wearable sensor based on a portable microcontroller provided with Bluetooth transmission and PPG technology to detect light absorption changes from the wrist and a smartphone using ANN algorithms to estimate BP parameters. Communication between the PPG sensor and smartphone occurs via a master-slave scheme controlled by the mobile device. Therefore, all BP estimations are started by the master device (smartphone), which receives responses from the slave device (PPG sensor).The basic block diagram of the proposed approach is shown in Figure 1.

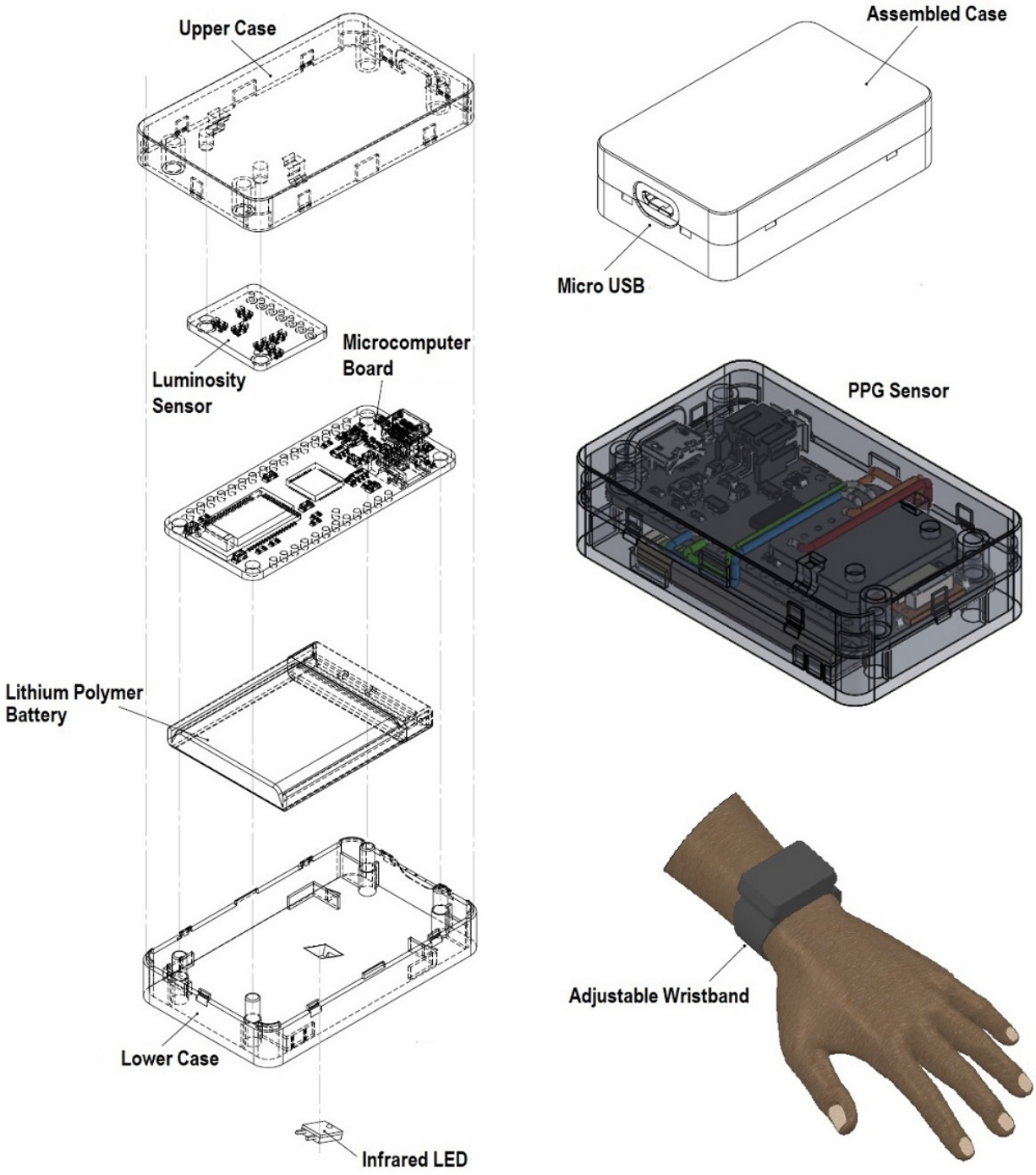
Figure 1. Block diagram of the blood pressure estimation approach.



PPG Sensor

The modular development of the sensor was performed on an Arduino-compatible microcomputer board from Adafruit [47], which includes an ATmega32u4 portable microcontroller clocked at 8 MHz, a Bluetooth Low Energy (BLE) [48] module, a connector for a lithium polymer battery, and a built-in micro-USB jack for power charging. A TSL2561 luminosity sensor (Adafruit Industries) [49], an infrared LED LTE-302 (Lite-On) with a wavelength of 940 nm, and a rechargeable lithium polymer battery (3.7 V, 500 mAh) were added to the development board. The electronic circuit was embedded in a 3D-printed case and attached to an adjustable wristband. Figure 2 shows the modular design of our PPG sensor.

Figure 2. Modular design of the photoplethysmography sensor. LED: light-emitting diode; PPG: photoplethysmography.



Estimation of BP Parameters

BP was estimated using the principle of the oscillometric method that determines SBP and DBP from the value of the mean arterial pressure (MAP), which is assumed to be nearly equal to the maximum oscillation of the cuff pressure applied to an arterial segment [50]. On the other hand, MAP is also estimated as the maximum vascular unloading detected by the PPG signal [51,52]. Therefore, we estimated BP through a combinatorial ANN model in the following two stages: (1) direct estimation of MAP from raw PPG data and (2) feedback of MAP output to estimate SBP and DBP approximations.

In addition, our approach calculated the variability of continuous BP estimations using the average real variability index [53], which provides superior predictive value over conventional BP variability estimators [54]. This method focuses on changes occurring over short time intervals, and thus, it corrects some of the limitations of SD, which only reflects the dispersion of BP measures around the mean. The average real variability index calculates the average of absolute changes between consecutive BP readings as follows:

where *N* denotes the number of valid BP measures and *k* is the order of the measures.

Data Collection Procedure

Fit data for MAP→BP estimation were obtained from the Maracaibo Aging Study [55], which is an ongoing population-based longitudinal study that includes participants aged 55 years or above, who underwent 24-hour ambulatory BP monitoring for the follow-up of cardiovascular outcomes. BP measures were recorded with a validated fully automatic SpaceLabs 90207 device [56], which was programmed to register ambulatory BP readings at 15-minute intervals during the daytime (6:00 AM-10:59 PM) and 30-minute intervals at night (11:00 PM-05:59 AM). We selected subjects with 24-hour BP recordings of good technical quality (>70% valid readings) and obtained a data set of 43,552 BP records belonging to 662 participants (mean age 67.3 years, SD 7.9 years; 69% women). The mean 24-hour BP values were as follows: SBP, 131.1 (SD 21.4) mmHg; DBP, 74.9 (SD 14.4) mmHg; and MAP, 93.6 (SD 15.3) mmHg. The ethical review board of the Institute of Cardiovascular Diseases of the University of Zulia approved the protocol. Written informed consent was obtained from each subject or a close family member.

To collect data for the PPG→MAP problem and test PPG sensor usability, we performed a development and usability study with three older adults (mean age 61.3 years, SD 1.5 years; 66% women) without a history of cardiovascular disease or hypertension. The protocol consisted of individual sessions involving three rounds (8:00 AM-12:00 PM, 2:00 PM-6:00 PM, and 8:00 PM-10:00 PM) to record simultaneous measures with our PPG sensor and a validated semiautomated upper-arm cuff oscillometric device (Omron HEM-4030; BP accuracy ±3 mmHg) [57]. Each round included 10 minutes of rest before starting to acquire measures from participants sitting in the upright position with the back supported, legs uncrossed, and nondominant arm on a desk. BP and PPG readings were taken every 5 minutes by a trained observer until obtaining a specific number of valid records for each round (20, 20, and 10 for the three rounds). The nondominant arm of each participant was equipped with the oscillometric cuff device and PPG sensor placed on top of the wrist with moderate pressure. The BP reference device required a maximum of 30 seconds to detect SBP and DBP. During each BP acquisition, the PPG sensor was programed to collect light absorption readings every 0.6 seconds; values greater than zero were saved and averaged to obtain one PPG measurement for each reference BP measure. The corresponding MAP value was calculated as follows:

$$MAP = (2 \times DBP + SBP) / 3 \quad (2)$$

The fit data set contained a total of 150 values (three participants × one session × 50 measures/session) for MAP (mean 88.9 mmHg, SD 16.1 mmHg) and PPG (mean 4100.7 a.u., SD 107.7 a.u.) measurements. The data acquisition procedure was performed in a private indoor room maintained at a controlled temperature of 25°C and supervised by medical staff. Written informed consent was obtained from all participants. They were compensated US $50 per session, and the protocol was approved by the ethical review board of the Polytechnic University of Sinaloa.

Machine Learning Model

The ANN model was built using the Neural Network Fitting App from Matlab [58]. Data fitting problems (PPG→MAP and MAP→BP) were addressed by separate two-layer feed-forward networks. In both cases, we used sigmoid and linear transfer functions for the hidden and output layers, and the Bayesian regularization backpropagation algorithm with random weights and bias initialization as the training function. Each fit data set was divided into three sets randomly (70%, 15%, and 15%) for training, validation, and testing. The performance of network outputs with respect to targets was assessed using linear regression analysis. In order to avoid overfitting, training stopped when we reached 1000 epochs or the validation error failed to decrease for six iterations.

Software Development

We used Matlab Compiler SDK [59] to deploy the trained combinatorial ANN model in a smartphone. The mobile app to display BP estimates from network output was developed in the integrated development environment Android Studio [60]. Communication with the wearable sensor to acquire input PPG data was enabled through the Android Bluetooth serial port profile library [61]. This connection was exclusive, because a BLE peripheral can be paired to only one central device at a time. Android multithreading [62] allowed the smartphone to maintain normal operations while receiving real-time PPG signals. The smartphone app included time setting parameters for the start of monitoring and the measurement frequency of continuous BP readings, and a battery power indicator for the PPG sensor. To provide a user-friendly app for elderly subjects with reduced vision and manual dexterity, we used a simplified graphical user interface with a bright screen, large text, and numbers, and simple input buttons with touchscreen technology, all of which have been proven to be suitable for older users [63].

Performance Assessment of BP Estimation

To assess performance, we compared outputs of the PPG→MAP→BP combinatorial ANN model (n=150) against BP reference measures recorded with the oscillometric device. The accuracy and precision of comparisons were evaluated through mean error (ME) (SD), Pearson and Spearman correlations, and percentage accuracy. The ME was calculated as the average of differences between reference and predicted BP values. For percentage accuracy, we quantified the error proportion as the absolute difference between reference and predicted values, with division by the reference BP value. We then subtracted this result from 1 and multiplied it by 100. Thereafter, we calculated the mean across all results to obtain the percentage accuracy for BP estimation [42]. Additionally, we assessed BP prediction according to the following: (1) cumulative percentage of absolute differences between reference and predicted values within 5, 10, and 15 mmHg [36]; (2) maximum error of 5 mmHg and 8 mmHg for ME and SD values, respectively [45]; and (3) agreement between reference and predicted values by a Bland-Altman plot [64].

Results

PPG Sensor Device

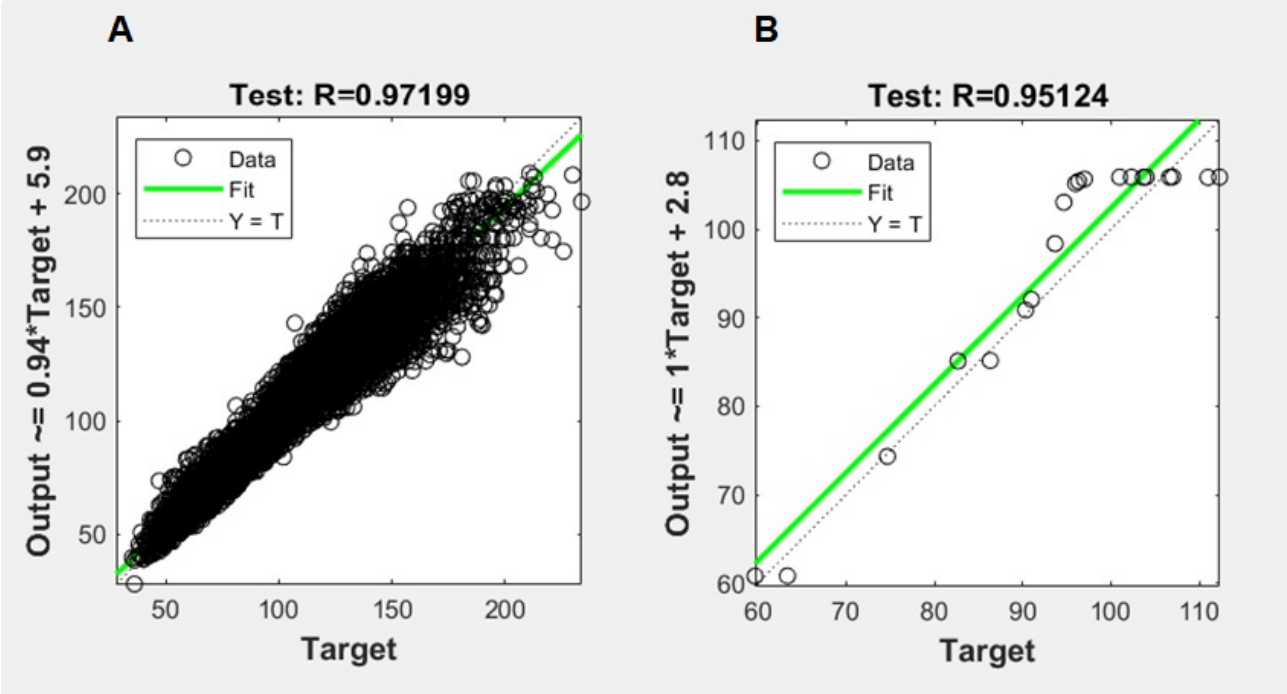
The prototype of the PPG sensor is shown in Figure 3. The 3D-printed case (black PLA printer filament) containing the electronic circuit measured 55 mm × 40 mm × 15 mm and weighed 35 g. The distance between the light source and the luminosity sensor was 3 mm. Power consumption ranged from 0.066 W to 0.231 W for the passive and active modes. The circumference range for the adjustable wristband attached to the case was 160 to 240 mm. The cost of all electronic components was US $32.

Figure 3. Prototype of the self-designed photoplethysmography sensor device.

ANN Performance for Fitting Problems

Fit data sets corresponding to MAP→BP (n=43,552) and PPG→MAP (n=150) problems were divided respectively into 30,486, 6553, and 6553 and 104, 23, and 23 samples for training, validation, and testing. The best performance for network outputs according to regression analysis was reached using 30 and 10 hidden neurons to separately map the BP and MAP targets. Figure 4 shows regression plots displaying the correlation of network outputs with respect to targets for each test set of MAP→BP (R=0.97) and PPG→MAP (R=0.95) fitting ANN models.

Figure 4. Regression plots for the test set of (A) mean arterial pressure→blood pressure and (B) photoplethysmography→mean arterial pressure fitting artificial neural network models.



Accuracy and Precision of BP Prediction

Table 2 shows the performance of our combinatorial ANN model in terms of predicting all BP measures recorded with the oscillometric device. The mean error bias (SD) was −7.77 (8.58) mmHg for SBP and −1.02 (4.21) mmHg for DBP. Network outputs predicted SBP and DBP with an average accuracy of 91.72% and 96.67%, respectively. Pearson and Spearman coefficients indicated a very strong correlation (>0.90, *P*<.001) between prediction and BP measurements. The percentages of all predicted values that absolutely differed from reference BP measures by 5, 10, and 15 mmHg or less were 25.33% (38/150), 54.67% (82/150), and 81.33% (122/150) for SBP and 76% (114/150), 98.67% (148/150), and 99.33% (149/150) for DBP. The average of DBP predictions fell within the tolerable error of 5 (SD 8) mmHg, and SBP predictions were close to this accuracy threshold. Overestimation and underestimation of all predicted values with respect to reference BP measures according to a Bland-Altman plot are shown in Figure 5. Table 3 shows the prediction assessment for each participant.

Table 2. Accuracy and precision of predicted values with respect to reference blood pressure measures.

BP[a] estimate	Error bias (mmHg), mean (SD)	Accuracy (%)	Correlation coefficient	
			Pearson	Spearman
SBP[b]	−7.77 (8.58)	91.72	0.91[c]	0.94[c]
DBP[d]	−1.02 (4.21)	96.67	0.97[c]	0.98[c]

[a]BP: blood pressure.

[b]SBP: systolic blood pressure.

[c]*P*<.001.

[d]DBP: diastolic blood pressure.

Figure 5. Bland-Altman plots with multiple (A) systolic blood pressure and (B) diastolic blood pressure measures per subject.

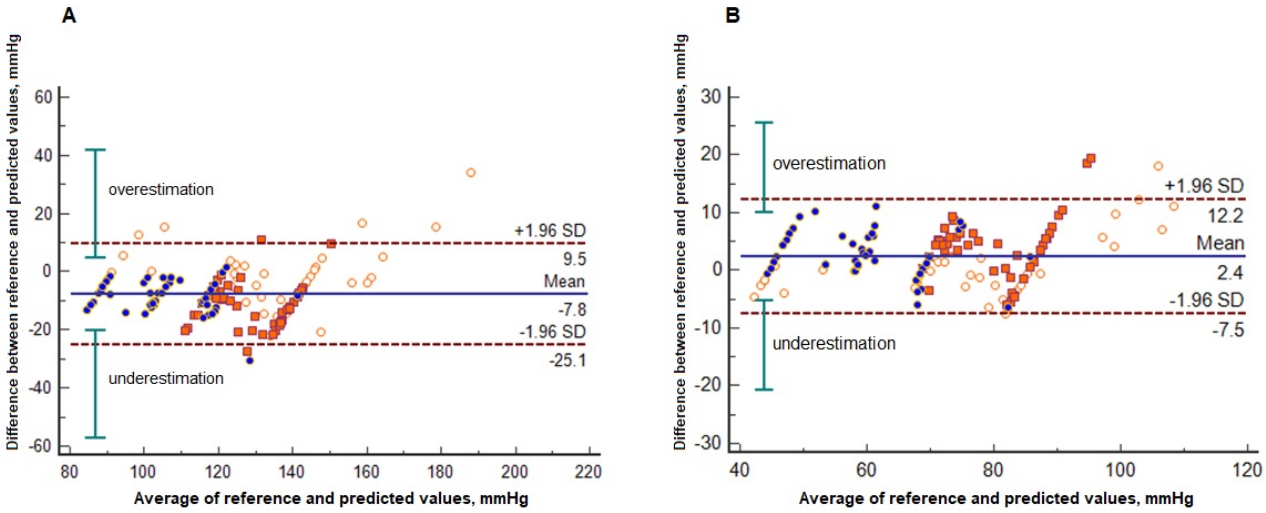


Table 3. Assessment of blood pressure prediction for each participant.

Participant	Error bias (mmHg), mean (SD)		Accuracy (%)	
	SBP[a]	DBP[b]	SBP	DBP
A	−3.75 (10.34)	−0.93 (5.29)	93.58	96.03
B	−11.24 (7.59)	−1.21 (3.98)	90.07	97.01
C	−8.32 (5.51)	−0.91 (3.16)	91.51	96.97

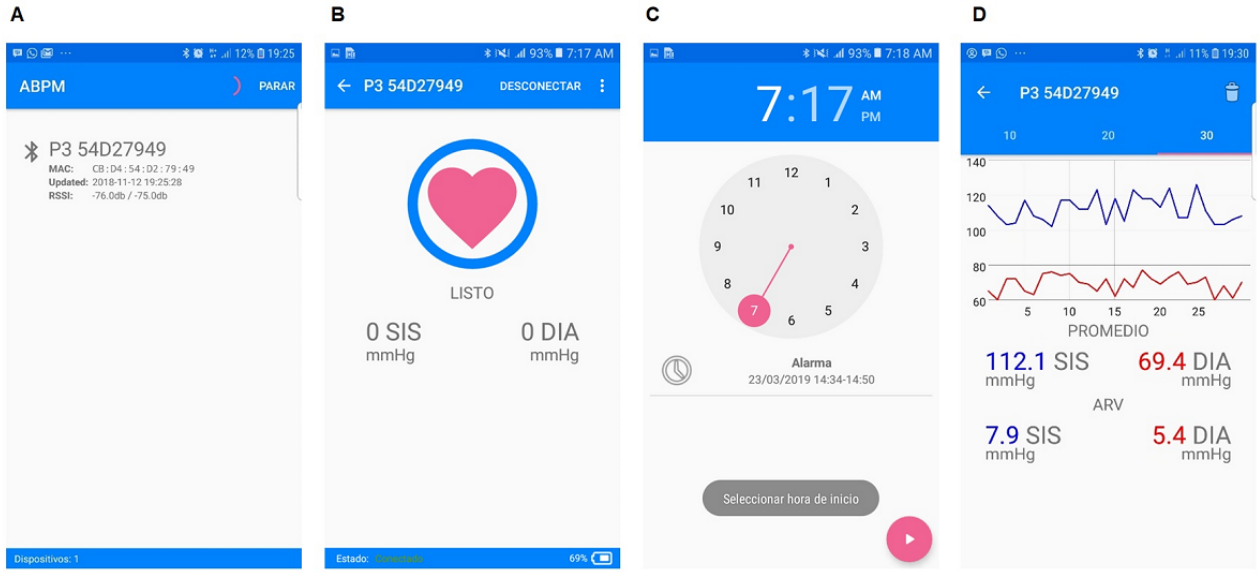
[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

Mobile App Interface

Screenshots of the interface of the developed mobile app running on a smartphone using Android operating system are depicted in Figure 6. We present the following four basic operations: pairing the smartphone with the PPG sensor through BLE, the smartphone ready to acquire PPG signals, setting the start time for BP monitoring, and deployment of the BP estimates.

Figure 6. Screenshots of the mobile app operations for (A) Bluetooth pairing between the smartphone and photoplethysmography (PPG) sensor, (B) established connection showing the battery power indicator of the PPG sensor, (C) setting of the start time for blood pressure (BP) monitoring, and (D) deployment of continuous BP measures, average BP level, and BP variability estimation.



Discussion

Principal Findings

Most PPG approaches to estimate BP use pulse transit time calculation, a feature extraction stage, and PPG signals reflected from finger vascular tone [12,14,17]. However, the nonlinearity of the blood volume-pressure relation indicates that the robust performance of these PPG-based BP monitors depends on specific sensing and biological parameters, such as signal recording of high quality, and individual physical property valuation. Therefore, obtaining acceptable BP estimation requires a filtering process to mitigate the effects of contaminated signals by voluntary or involuntary body movements and a calibration step to identify personal cardiovascular and respiratory factors that vary the propagated PPG signal from living tissue.

On the other hand, the portability and location of PPG sensors are key factors to continuously, comfortably, and accurately monitor the BP of users during normal daily activities. In this sense, it is necessary to reduce the number of biosensors for minimizing invasiveness and operational complexity, to avoid PPG measurement from obtrusive body locations that may restrict user movement, and to diminish ambient light influence or displacement of a sensor that can affect PPG signal quality.

To overcome these issues, we proposed a smartphone PPG-only approach for continuous BP level and variability estimation using raw PPG signals, without a calibration procedure. In order to address nonlinear, complex, and dynamic relationships between blood volumetric pulsations and BP changes of the artery, we initially analyzed the correlation between PPG and MAP readings [51,65] for identifying mutual information that can be used to derive BP approximation, instead of examining correspondence on morphological features of PPG and BP waveforms. Furthermore, we developed a small, light, convenient, and wearable PPG sensor wrapped with moderate pressure around the wrist, with the goal of optimizing user comfort and reducing the effects of motion artifacts or inaccurate measurement of PPG signals caused by poorly estimated finger pressing against the PPG sensor [13,39,66].

The proposed combinatorial ANN model showed good accuracy in terms of predicting the reference BP values of our validation sample. On average, our approach predicted SBP and DBP measures with accuracy over 90% and correlations over 0.90 (P<.001). Bland-Altman plots showed that most of the errors for BP prediction were less than 10 mmHg. According to the AAMI standard, predicted values successfully fell within the reference measurements for DBP estimation and were very close to the established limit for SBP estimation. In reference to the BHS protocol, the prediction of DBP consistently reached the best accuracy grade, whereas that of SBP reached approximately the boundary accuracy grade. In agreement with previous studies, the error bias range for DBP was narrower than that for SBP in reference to the AAMI and BHS criteria, likely because the variability of DBP is regularly less than that of SBP [17]. On the other hand, another smartphone-based BP monitor that also estimates the variability of BP has not been reported at the moment.

Limitations

The main limitation of our study is the relatively small size of the validation cohort (n=3). However, we increased the dimension of the final data set by collecting 50 PPG and MAP measures per subject. Moreover, the key stage for approach accuracy was BP mapping from MAP values, which was sustained by an extensive database with more than 43,500 records (n=662).

Second, our selection strategy of only normotensive participants avoided achieving an even distribution of different BP values. Nevertheless, the heterogeneous BP measures recorded by participants (Figure 5) helped cover a wider range of BP values to mitigate part of this limitation.

Third, the development and usability study focused only on estimating BP measures when individuals were in a calm state and under controlled conditions. Therefore, for effective assessment of continuous BP monitoring, it is necessary to validate our findings by upgrading the record of BP estimates during everyday activities in different environments.

Fourth, we included only older subjects for training, validation, and testing of our machine learning approach, and consequently, PPG signals were very susceptible to vascular aging. Since it is one of the factors that can lead to arterial stiffness [67,68], this could represent an overfitting problem to accurately estimate BP in the general population. Nevertheless, one of the major advantages of ANN models is their generalization ability for fitting outputs from new inputs. On the other hand, we mainly addressed continuous monitoring of BP in older adults because they represent a high-risk population for hypertension development.

Finally, despite the potential correspondence between MAP and PPG signals, this relationship may be more complex than previously anticipated [69]. Therefore, future approaches, including additional input user parameters, such as age, gender, and body mass index, may be required to develop a more robust ANN model for BP estimation.

Conclusion

In this paper, a novel smartphone-based BP monitor was proposed for noninvasive, pervasive, and continuous BP level and variability estimation, using a cuffless, calibration-free, wireless, and wearable PPG-only sensor. We addressed the nonlinear relationship between BP and PPG signals through a multilayer perceptron machine learning approach for estimating BP from raw PPG signals and without a conventional feature extraction stage. The findings of this development and usability study indicated that the employed ANN model performed with good average accuracy and very strong correlation. However, according to the AAMI and BHS standards, only DBP prediction met the clinically accepted accuracy thresholds. Therefore, with further development and validation in real-life conditions, the developed mobile personal health care system could provide a cost-effective strategy for the early diagnosis and control of hypertension and an independent cardiovascular predictor (abnormal BP variability), particularly in rural zones, areas lacking physicians, and areas with solitary elderly populations.

Acknowledgments

This study was financially supported by Grants FNE-170531-C2-1-000008979 and PFCE-2019-0857/19 from Secretaria de Economia and Secretaria de Educacion Publica, Mexico and Grants 1R01AG036469-01A1 and R03AG054186 from the National Institute on Aging and Fogarty International Center.

Authors' Contributions

LJM led the concept, development, and analysis of this study and wrote the initial draft of the manuscript. RO, RMP, and AJG developed the PPG sensor. VGF and AJG developed the mobile app. VGF, AJG, and JDM performed data acquisition and analysis. GEM was the supervisor of this study, as well as joint senior author. All authors read, reviewed, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1659-1724 [FREE Full text] [doi: 10.1016/S0140-6736(16)31679-8] [Medline: 27733284]
2. Hypertension. World Health Organization. 2009. URL: https://www.who.int/news-room/fact-sheets/detail/hypertension [accessed 2020-05-28]
3. Pickering TG, Davidson K, Gerin W, Schwartz JE. Masked hypertension. Hypertension 2002 Dec;40(6):795-796. [doi: 10.1161/01.hyp.0000038733.08436.98] [Medline: 12468559]
4. Hansen TW, Kikuya M, Thijs L, Bj??rklund-Bodeg??rd K, Kuznetsova T, Ohkubo T, et al. Prognostic superiority of daytime ambulatory over conventional blood pressure in four populations: a meta-analysis of 7030 individuals. Journal of Hypertension 2007;25(8):1554-1564. [doi: 10.1097/hjh.0b013e3281c49da5]
5. Pickering TG, Shimbo D, Haas D. Ambulatory blood-pressure monitoring. N Engl J Med 2006 Jun 01;354(22):2368-2374. [doi: 10.1056/NEJMra060433] [Medline: 16738273]
6. Kikuya M, Hozawa A, Ohokubo T, Tsuji I, Michimata M, Matsubara M, et al. Prognostic significance of blood pressure and heart rate variabilities: the Ohasama study. Hypertension 2000 Nov;36(5):901-906. [doi: 10.1161/01.hyp.36.5.901] [Medline: 11082164]
7. Sander D, Kukla C, Klingelhöfer J, Winbeck K, Conrad B. Relationship between circadian blood pressure patterns and progression of early carotid atherosclerosis: A 3-year follow-up study. Circulation 2000 Sep 26;102(13):1536-1541. [doi: 10.1161/01.cir.102.13.1536] [Medline: 11004145]
8. Pringle E, Phillips C, Thijs L, Davidson C, Staessen JA, de Leeuw PW, Syst-Eur investigators. Systolic blood pressure variability as a risk factor for stroke and cardiovascular mortality in the elderly hypertensive population. J Hypertens 2003 Dec;21(12):2251-2257. [doi: 10.1097/00004872-200312000-00012] [Medline: 14654744]
9. Nasothimiou EG, Karpettas N, Dafni MG, Stergiou GS. Patients' preference for ambulatory versus home blood pressure monitoring. J Hum Hypertens 2014 Apr;28(4):224-229. [doi: 10.1038/jhh.2013.104] [Medline: 24152822]
10. Bard DM, Joseph JI, van Helmond N. Cuff-Less Methods for Blood Pressure Telemonitoring. Front Cardiovasc Med 2019;6:40 [FREE Full text] [doi: 10.3389/fcvm.2019.00040] [Medline: 31157236]
11. Slapni?ar G, Lu?trek M, Marinko M. Continuous blood pressure estimation from PPG signal. Informatica 2018;42(1):33-42 [FREE Full text]
12. Martínez G, Howard N, Abbott D, Lim K, Ward R, Elgendi M. Can Photoplethysmography Replace Arterial Blood Pressure in the Assessment of Blood Pressure? J Clin Med 2018 Sep 30;7(10) [FREE Full text] [doi: 10.3390/jcm7100316] [Medline: 30274376]
13. Elgendi M, Fletcher R, Liang Y, Howard N, Lovell NH, Abbott D, et al. The use of photoplethysmography for assessing hypertension. NPJ Digit Med 2019;2:60 [FREE Full text] [doi: 10.1038/s41746-019-0136-7] [Medline: 31388564]
14. Stojanova A, Koceski S, Koceska N. Continuous Blood Pressure Monitoring as a Basis for Ambient Assisted Living (AAL) - Review of Methodologies and Devices. J Med Syst 2019 Jan 02;43(2):24. [doi: 10.1007/s10916-018-1138-8] [Medline: 30603777]
15. Gaurav A, Maheedhar M, Tiwari V, Narayanan R. Cuffless PPG based continuous blood pressure monitoring: a smartphone based approach. 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 17-20, 2016; Orlando, USA p. 607-610 URL: https://ieeexplore.ieee.org/document/7590775 [doi: 10.1109/embc.2016.7590775]

16. Dias Junior A, Murali S, Rincon F, Atienza D. Methods for reliable estimation of pulse transit time and blood pressure variations using smartphone sensors. Microprocessors and Microsystems 2016 Oct;46:84-95. [doi: 10.1016/j.micpro.2016.06.001]
17. Mukkamala R, Hahn J, Inan OT, Mestha LK, Kim C, Toreyin H, et al. Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. IEEE Trans. Biomed. Eng 2015 Aug;62(8):1879-1901. [doi: 10.1109/tbme.2015.2441951]
18. Liang Y, Chen Z, Ward R, Elgendi M. Hypertension Assessment Using Photoplethysmography: A Risk Stratification Approach. J Clin Med 2018 Dec 21;8(1) [FREE Full text] [doi: 10.3390/jcm8010012] [Medline: 30577637]
19. Slapničar G, Mlakar N, Luštrek M. Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network. Sensors (Basel) 2019 Aug 04;19(15) [FREE Full text] [doi: 10.3390/s19153420] [Medline: 31382703]
20. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2020-05-27]
21. Garcia-Perez C, Diaz-Zayas A, Rios A, Merino P, Katsalis K, Chang C, et al. Improving the efficiency and reliability of wearable based mobile eHealth applications. Pervasive and Mobile Computing 2017 Sep;40:674-691. [doi: 10.1016/j.pmcj.2017.06.021]
22. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a Culturally Tailored mHealth Medication Regimen Self-Management Program upon Blood Pressure among Hypertensive Hispanic Adults. Int J Environ Res Public Health 2019 Apr 06;16(7) [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
23. Parati G, Dolan E, McManus RJ, Omboni S. Home blood pressure telemonitoring in the 21st century. J Clin Hypertens (Greenwich) 2018 Jul;20(7):1128-1132 [FREE Full text] [doi: 10.1111/jch.13305] [Medline: 30003701]
24. Goldberg EM, Levy PD. New Approaches to Evaluating and Monitoring Blood Pressure. Curr Hypertens Rep 2016 Jun;18(6):49. [doi: 10.1007/s11906-016-0650-9] [Medline: 27137524]
25. 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]
26. Ciemins EL, Arora A, Coombs NC, Holloway B, Mullette EJ, Garland R, et al. Improving Blood Pressure Control Using Smart Technology. Telemed J E Health 2018 Mar;24(3):222-228. [doi: 10.1089/tmj.2017.0028] [Medline: 28930497]
27. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone Apps to Support Self-Management of Hypertension: Review and Content Analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
28. Searcy RP, Summapund J, Estrin D, Pollak JP, Schoenthaler A, Troxel AB, et al. Mobile Health Technologies for Older Adults with Cardiovascular Disease: Current Evidence and Future Directions. Curr Geri Rep 2019 Jan 28;8(1):31-42. [doi: 10.1007/s13670-019-0270-8]
29. Kitt J, Fox R, Tucker KL, McManus RJ. New Approaches in Hypertension Management: a Review of Current and Developing Technologies and Their Potential Impact on Hypertension Care. Curr Hypertens Rep 2019 Apr 25;21(6):44 [FREE Full text] [doi: 10.1007/s11906-019-0949-4] [Medline: 31025117]
30. Asada H, Shaltis P, Reisner A, Rhee S, Hutchinson R. Mobile monitoring with wearable photoplethysmographic biosensors. IEEE Eng Med Biol Mag 2003;22(3):28-40. [doi: 10.1109/memb.2003.1213624] [Medline: 12845817]
31. Hassani A, Foruzan AH. Improved PPG-based estimation of the blood pressure using latent space features. SIViP 2019 Mar 19;13(6):1141-1147. [doi: 10.1007/s11760-019-01460-1]
32. Rundo F, Ortis A, Battiato S, Conoci S. Advanced Bio-Inspired System for Noninvasive Cuff-Less Blood Pressure Estimation from Physiological Signal Analysis. Computation 2018 Aug 28;6(3):46. [doi: 10.3390/computation6030046]
33. Xing X, Sun M. Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. Biomed Opt Express 2016 Aug 01;7(8):3007-3020 [FREE Full text] [doi: 10.1364/BOE.7.003007] [Medline: 27570693]
34. Li P, Liu M, Zhang X, Hu X, Pang B, Yao Z, et al. Novel wavelet neural network algorithm for continuous and noninvasive dynamic estimation of blood pressure from photoplethysmography. Sci. China Inf. Sci 2015 Sep 25;59(4). [doi: 10.1007/s11432-015-5400-0]
35. Chandrasekaran V, Dantu R, Jonnada S, Thiyagaraja S, Subbu KP. Cuffless Differential Blood Pressure Estimation Using Smart Phones. IEEE Trans. Biomed. Eng 2013 Apr;60(4):1080-1089. [doi: 10.1109/tbme.2012.2211078]
36. Plante TB, Urrea B, MacFarlane ZT, Blumenthal RS, Miller ER, Appel LJ, et al. Validation of the Instant Blood Pressure Smartphone App. JAMA Intern Med 2016 May 01;176(5):700-702 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0157] [Medline: 26938174]
37. Wang E, Zhu J, Jain M, Lee T, Saba E, Nachman L. Seismo: bood pressure monitoring using built-in smartphone accelerometer and camera. 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Canada p. 1-9. [doi: 10.1145/3173574.3173999]
38. Chandrasekhar A, Kim C, Naji M, Natarajan K, Hahn J, Mukkamala R. Smartphone-based blood pressure monitoring via the oscillometric finger-pressing method. Sci Transl Med 2018 Mar 07;10(431) [FREE Full text] [doi: 10.1126/scitranslmed.aap8674] [Medline: 29515001]

39. Chandrasekhar A, Natarajan K, Yavarimanesh M, Mukkamala R. An iPhone Application for Blood Pressure Monitoring via the Oscillometric Finger Pressing Method. Sci Rep 2018 Sep 03;8(1):13136 [FREE Full text] [doi: 10.1038/s41598-018-31632-x] [Medline: 30177793]
40. Raichle CJ, Eckstein J, Lapaire O, Leonardi L, Brasier N, Vischer AS, et al. Performance of a Blood Pressure Smartphone App in Pregnant Women. Hypertension 2018 Jun;71(6):1164-1169. [doi: 10.1161/hypertensionaha.117.10647]
41. Dey J, Gaurav A, Tiwari V. InstaBP: Cuff-less Blood Pressure Monitoring on Smartphone using Single PPG Sensor. Conf Proc IEEE Eng Med Biol Soc 2018 Jul;2018:5002-5005. [doi: 10.1109/EMBC.2018.8513189] [Medline: 30441464]
42. Luo H, Yang D, Barszczyk A, Vempala N, Wei J, Wu SJ, et al. Smartphone-Based Blood Pressure Measurement Using Transdermal Optical Imaging Technology. Circ: Cardiovascular Imaging 2019 Aug;12(8):e008857. [doi: 10.1161/circimaging.119.008857]
43. Foo JY, Lim CS, Wang P. Evaluation of blood pressure changes using vascular transit time. Physiol Meas 2006 Aug;27(8):685-694. [doi: 10.1088/0967-3334/27/8/003] [Medline: 16772667]
44. O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, O'Malley K, et al. The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. J Hypertens 1990 Jul;8(7):607-619. [doi: 10.1097/00004872-199007000-00004] [Medline: 2168451]
45. Non-invasive sphygmomanometers–Part 2: clinical investigation of automated measurement type. Association for the Advancement of Medical Instrumentation. URL: https://my.aami.org/aamiresources/previewfiles/8106002_1306_preview.pdf [accessed 2020-05-28]
46. Muthukrishnan R, Rohini R. LASSO: a feature selection technique in predictive modeling for machine learning. 2016 Presented at: 2016 IEEE International Conference on Advances in Computer Application; October 24, 2016; Coimbatore, India p. 18-20 URL: https://ieeexplore.ieee.org/document/7887916 [doi: 10.1109/icaca.2016.7887916]
47. Adafruit Feather 32u4 Bluefruit LE. Adafruit Industries. 2015. URL: https://learn.adafruit.com/adafruit-feather-32u4-bluefruit-le [accessed 2020-05-28]
48. Gomez C, Oller J, Paradells J. Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. Sensors 2012 Aug 29;12(9):11734-11753. [doi: 10.3390/s120911734]
49. Adafruit TSL2561 Luminosity Sensor. Adafruit Industries. 2012. URL: https://learn.adafruit.com/tsl2561 [accessed 2020-05-27]
50. Drzewiecki G, Hood R, Apple H. Theory of the oscillometric maximum and the systolic and diastolic detection ratios. Ann Biomed Eng 1994 Jan;22(1):88-96. [doi: 10.1007/bf02368225]
51. Yamakoshi K, Shimazu H, Shibata M, Kamiya A. New oscillometric method for indirect measurement of systolic and mean arterial pressure in the human finger. Part 1: Model experiment. Med. Biol. Eng. Comput 1982 May;20(3):307-313. [doi: 10.1007/bf02442797]
52. Laurent C, Jönsson B, Vegfors M, Lindberg L. Non-invasive measurement of systolic blood pressure on the arm utilising photoplethysmography: development of the methodology. Med. Biol. Eng. Comput 2005 Feb;43(1):131-135. [doi: 10.1007/bf02345134]
53. Mena L, Pintos S, Queipo NV, Aizpúrua JA, Maestre G, Sulbarán T. A reliable index for the prognostic significance of blood pressure variability. J Hypertens 2005 Mar;23(3):505-511. [doi: 10.1097/01.hjh.0000160205.81652.5a] [Medline: 15716690]
54. Mena LJ, Felix VG, Melgarejo JD, Maestre GE. 24‐Hour Blood Pressure Variability Assessed by Average Real Variability: A Systematic Review and Meta‐Analysis. JAHA 2017 Oct;6(10). [doi: 10.1161/jaha.117.006895]
55. Maestre GE, Pino-Ramírez G, Molero AE, Silva ER, Zambrano R, Falque L, et al. The Maracaibo Aging Study: population and methodological issues. Neuroepidemiology 2002;21(4):194-201. [doi: 10.1159/000059524] [Medline: 12065882]
56. Iqbal P, Fotherby MD, Potter JF. Validation of the SpaceLabs 90207 automatic non-invasive blood pressure monitor in elderly subjects. Blood Press Monit 1996 Aug;1(4):367-373. [Medline: 10226261]
57. Belghazi J, El Feghali RN, Moussalem T, Rejdych M, Asmar RG. Validation of four automatic devices for self-measurement of blood pressure according to the International Protocol of the European Society of Hypertension. Vasc Health Risk Manag 2007;3(4):389-400 [FREE Full text] [Medline: 17969368]
58. Neural Net Fitting. The MathWorks, Inc. URL: https://www.mathworks.com/help/deeplearning/ref/neuralnetfitting-app.html [accessed 2020-05-27]
59. Matlab Compiler SDK. The MathWorks, Inc. URL: https://www.mathworks.com/products/matlab-compiler-sdk.html [accessed 2020-05-28]
60. Android Studio. Android Developers. URL: https://developer.android.com/studio [accessed 2020-05-28]
61. BluetoothSocket. Android Developers. URL: https://developer.android.com/reference/android/bluetooth/BluetoothSocket [accessed 2020-05-28]
62. Sending operations to multiple threads. Android Developers. URL: https://developer.android.com/training/multiple-threads [accessed 2020-05-28]
63. Nurgalieva L, Jara Laconich JJ, Baez M, Casati F, Marchese M. A Systematic Literature Review of Research-Derived Touchscreen Design Guidelines for Older Adults. IEEE Access 2019;7:22035-22058. [doi: 10.1109/access.2019.2898467]

64. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/S0140-6736(86)90837-8]
65. Penáz J. Criteria for set point estimation in the volume clamp method of blood pressure measurement. Physiol Res 1992;41(1):5-10. [Medline: 1610779]
66. Matsumura K, Rolfe P, Toda S, Yamakoshi T. Cuffless blood pressure estimation using only a smartphone. Sci Rep 2018 May 08;8(1):7298 [FREE Full text] [doi: 10.1038/s41598-018-25681-5] [Medline: 29740088]
67. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202 [FREE Full text] [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
68. Liang Y, Chen Z, Ward R, Elgendi M. Photoplethysmography and Deep Learning: Enhancing Hypertension Risk Stratification. Biosensors (Basel) 2018 Oct 26;8(4) [FREE Full text] [doi: 10.3390/bios8040101] [Medline: 30373211]
69. Chin KY, Panerai RB. Relating external compressing pressure to mean arterial pressure in non-invasive blood pressure measurements. J Med Eng Technol 2015 Jan;39(1):79-85. [doi: 10.3109/03091902.2014.979953] [Medline: 25429784]

Abbreviations

AAMI: Association for the Advancement of Medical Instrumentation
ANN: artificial neural network
BHS: British Hypertension Society
BLE: Bluetooth Low Energy
BP: blood pressure
DBP: diastolic blood pressure
LED: light-emitting diode
MAP: mean arterial pressure
PPG: photoplethysmography
SBP: systolic blood pressure

Edited by G Eysenbach; submitted 29.01.20; peer-reviewed by S Omboni, E Bellei; comments to author 09.04.20; revised version received 11.04.20; accepted 26.04.20; published 20.07.20.

Please cite as:
Mena LJ, Félix VG, Ostos R, González AJ, Martínez-Peláez R, Melgarejo JD, Maestre GE
Mobile Personal Health Care System for Noninvasive, Pervasive, and Continuous Blood Pressure Monitoring: Development and Usability Study
JMIR Mhealth Uhealth 2020;8(7):e18012
URL: https://mhealth.jmir.org/2020/7/e18012
doi:10.2196/18012
PMID:32459642

Original Paper

Effects of a 12-Week Multifaceted Wearable-Based Program for People With Knee Osteoarthritis: Randomized Controlled Trial

Linda C Li[1,2], PT, PhD; Lynne M Feehan[1], PT, PhD; Hui Xie[2,3], PhD; Na Lu[2], MPH; Christopher D Shaw[4], PhD; Diane Gromala[4], PhD; Siyi Zhu[2,5,6], MD, PhD; J Antonio Aviña-Zubieta[2,7], MD, PhD; Alison M Hoens[1,2], PT, MSc; Cheryl Koehn[8], BA; Johnathan Tam[2], BSc; Stephanie Therrien[2], BA; Anne F Townsend[9], PhD; Gregory Noonan[10], BSc, PT; Catherine L Backman[2,11], OT, PhD

[1]Department of Physical Therapy, University of British Columbia, Vancouver, BC, Canada
[2]Arthritis Research Canada, Richmond, BC, Canada
[3]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada
[4]School of Interactive Art & Technology, Simon Fraser University, Burnaby, BC, Canada
[5]Department of Rehabilitation Medicine, West China Hospital, Sichuan University, Chengdu, China
[6]Rehabilitation Medicine Key Laboratory of Sichuan Province, West China Hospital, Sichuan University, Chengdu, China
[7]Department of Medicine, University of British Columbia, Vancouver, BC, Canada
[8]Arthritis Consumer Experts, Vancouver, BC, Canada
[9]Division of Health Research, Faculty of Health & Medicine, Lancaster University, Lancashire, United Kingdom
[10]Mary Pack Arthritis Program, Vancouver General Hospital, Vancouver, BC, Canada
[11]Department of Occupational Science & Occupational Therapy, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Linda C Li, PT, PhD
Department of Physical Therapy
University of British Columbia
2177 Wesbrook Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 6042074020
Email: lli@arthritisresearch.ca

Abstract

Background: Current guidelines emphasize an active lifestyle in the management of knee osteoarthritis (OA), but up to 90% of patients with OA are inactive. In a previous study, we demonstrated that an 8-week physiotherapist (PT)-led counseling intervention, with the use of a Fitbit, improved step count and quality of life in patients with knee OA, compared with a control.

Objective: This study aimed to examine the effect of a 12-week, multifaceted wearable-based program on physical activity and patient outcomes in patients with knee OA.

Methods: This was a randomized controlled trial with a delay-control design. The immediate group (IG) received group education, a Fitbit, access to FitViz (a Fitbit-compatible app), and 4 biweekly phone calls from a PT over 8 weeks. Participants then continued using Fitbit and FitViz independently up to week 12. The delay group (DG) received a monthly electronic newsletter in weeks 1 to 12 and started the same intervention in week 14. Participants were assessed in weeks 13, 26, and 39. The primary outcome was time spent in daily moderate-to-vigorous physical activity (MVPA; in bouts ≥10 min) measured with a SenseWear Mini. Secondary outcomes included daily steps, time spent in purposeful activity and sedentary behavior, Knee Injury and OA Outcome Score, Patient Health Questionnaire-9, Partners in Health Scale, Theory of Planned Behavior Questionnaire, and Self-Reported Habit Index.

Results: We enrolled 51 participants (IG: n=26 and DG: n=25). Compared with the IG, the DG accumulated significantly more MVPA time at baseline. The adjusted mean difference in MVPA was 13.1 min per day (95% CI 1.6 to 24.5). A significant effect was also found in the adjusted mean difference in perceived sitting habit at work (0.7; 95% CI 0.2 to 1.2) and during leisure activities (0.7; 95% CI 0.2 to 1.2). No significant effect was found in the remaining secondary outcomes.

Conclusions: A 12-week multifaceted program with the use of a wearable device, an app, and PT counseling improved physical activity in people with knee OA.

Trial Registration: ClinicalTrials.gov NCT02585323; https://clinicaltrials.gov/ct2/show/NCT02585323

(JMIR Mhealth Uhealth 2020;8(7):e19116) doi:10.2196/19116

KEYWORDS

physical activity; counseling; knee osteoarthritis; physiotherapy; wearables

Introduction

Background

Arthritis is the most common cause of severe chronic pain and disability worldwide [1,2]. Analysis by the Arthritis Alliance of Canada estimates a new diagnosis of osteoarthritis (OA) every 60 seconds [3]. Current evidence supports the use of physical activity to manage OA because of its benefits on pain, mobility, and quality of life [4-6]. Guidelines by the OA Research Society International recommend the use of physical activity and therapeutic exercise as first-line treatment for knee OA [7]. Canadian physical activity guidelines recommend ≥150 min a week of moderate-to-vigorous physical activity (MVPA), performed in bouts of ≥10 min [8]. A study using accelerometers, however, found that over 90% of people with knee OA did not meet the physical activity guidelines [9]. This concurs with a systematic review that only 13% of people with OA accumulated ≥150 min per week of MVPA in bouts of ≥10 min [10].

Several modifiable factors are associated with low physical activity participation in patients with arthritis. These include lack of motivation [11], doubts about the effectiveness of exercise [12], and lack of health professional advice [13]. To promote an active lifestyle, the use of activity tracking devices has been explored based on the assumption that providing direct feedback on the amount of physical activity encourages people to meet specific targets. A systematic review of 14 randomized controlled trials (RCTs) of accelerometer-based interventions reported a small effect on physical activity participation (standardized mean difference 0.26; 95% CI 0.04 to 0.49). Notably, consumer-grade accelerometers promote physical activity behaviors through the use of behavioral change techniques [14], such as goal setting, self-monitoring, feedback, and rewards [15]. However, effective techniques such as action planning and problem-solving are absent from the use of these devices alone [15]. These techniques require contact with a health professional with counseling experience.

Our previous study [16] and Lyons et al [17] have demonstrated the feasibility of a physical activity counseling program with the use of a Fitbit (Fitbit Inc), a consumer-grade wearable device. Our subsequent study showed that an 8-week physiotherapist (PT)-led counseling program improved step count and quality of life in people with knee OA, compared with a control [18].

Study Aim

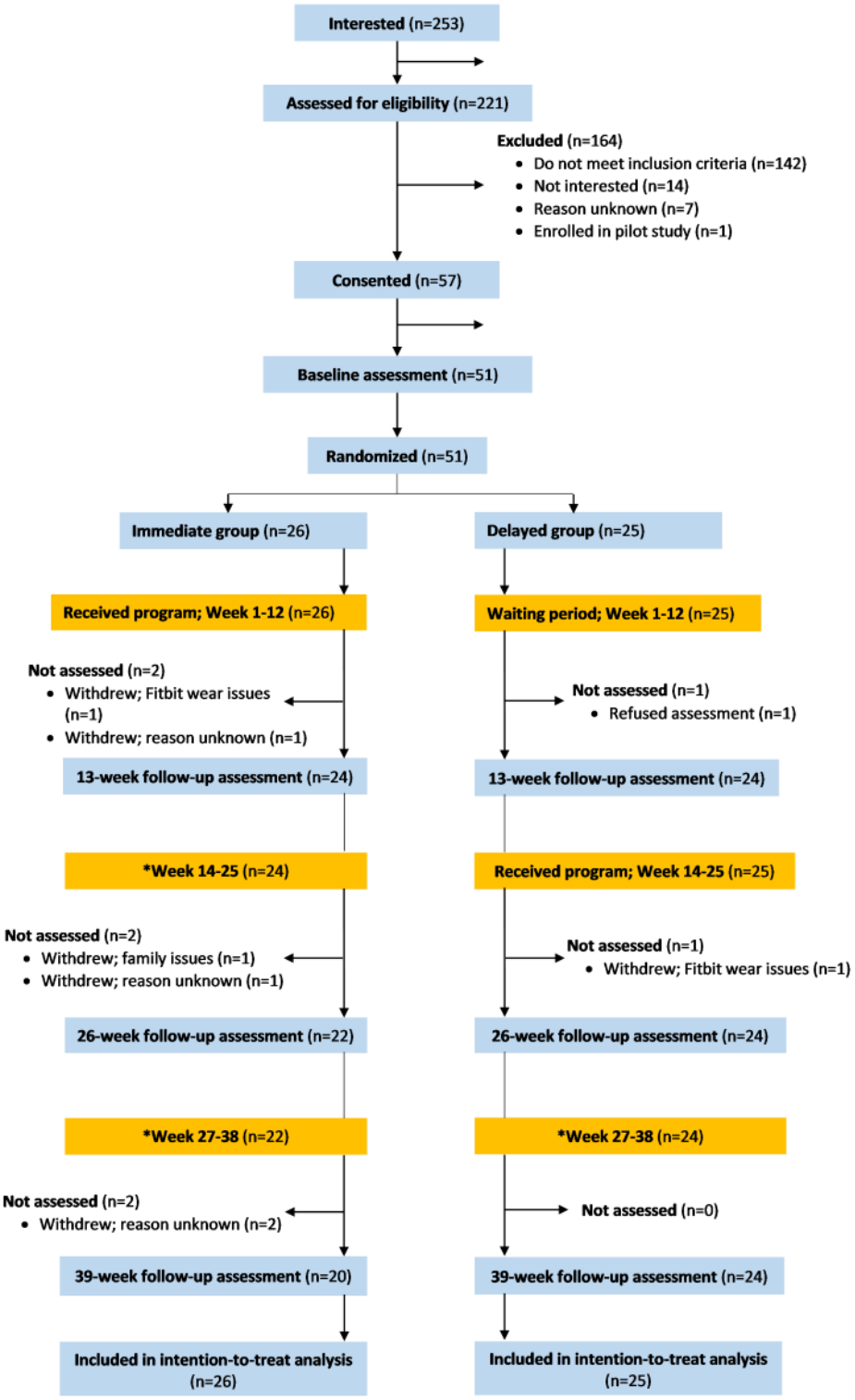
In this study, we aimed to assess the effect of a 12-week multifaceted intervention on improving activity participation in people with knee OA. Our primary hypothesis was that, compared with controls, those who received the program would increase mean daily MVPA time as determined by an objective measure. In addition, we explored the effect of the program on OA disease status, depressive symptoms, perceived habitual behaviors, and psychological constructs of being active.

Methods

Study Design

The Supporting Physical activity & Reducing sedentary behaviour in Arthritis (SuPRA) project was a proof-of-concept RCT with a delay-control design. Participants were randomly assigned to start the intervention either immediately (*immediate group* [IG]) or 14 weeks (*delay group* [DG]) after completing the baseline assessment. All participants were reassessed in weeks 13 (primary end point), 26, and 39 (Figure 1).

Figure 1. Consolidated Standards of Reporting Trials flowchart.



Participants

Participants were recruited from the Mary Pack Arthritis Program (Vancouver Coastal Health Authority) and the Fraser Health Authority in British Columbia, Canada. We also posted study information on Facebook, Twitter, Kajiji, and Craigslist. Individuals were eligible if they had a diagnosis of knee OA or met 2 criteria for early OA [19]. People with other chronic musculoskeletal conditions or contraindications to be physically active without medical supervision were excluded (Textboxes 1 and 2).

Textbox 1. Inclusion criteria.

Inclusion criteria

- Patients who had a physician-confirmed diagnosis of knee osteoarthritis or were aged ≥50 years and had felt pain or discomfort in or around the knee during the previous year lasting >28 separate or consecutive days [19]
- Patients who had no previous diagnosis of rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, polymyalgia rheumatica, connective tissue diseases, fibromyalgia, or gout
- Patients who had no history of using disease-modifying antirheumatic drugs or gout medications
- Patients who had no prior knee arthroplasty and not on a waiting list for total knee or hip replacement surgery
- Patients who did not have surgery in the back, hip, knee, foot, or ankle joints in the past 12 months
- Patients who had no history of acute injury to the knee in the past 6 months
- Patients who had an email address and access to the internet daily
- Patients who were able to attend a 1.5-hour group education session

Textbox 2. Exclusion criteria.

Exclusion criteria

- Patients who had previously used a physical activity wearable tracker
- Patients who received a steroid injection in a knee in the last 6 months
- Patients who received a hyaluronate injection in a knee in the last 6 months
- Patients who used medication that may impair activity tolerance (eg, beta blockers)
- Patients who faced a level of risk by exercising as identified by the Physical Activity Readiness Questionnaire [20]. If a participant did not pass the Physical Activity Readiness Questionnaire, a physician's note was requested to determine the eligibility

Randomization and Blinding

After completing the baseline measures, participants were randomly assigned to the IG or the DG in a 1:1 allocation ratio. Randomization was performed by a research staff not involved in the study using numbers generated by SAS version 9.4 (SAS Institute Inc) in variable block sizes to ensure allocation concealment. Participants were not blinded as they knew whether they received the program during the intervention period. Daily MVPA time (the primary outcome) was measured with a wearable multisensor device couriered to the participant to wear for 7 days. All research personnel processing the objectively measured physical activity data were blinded to the participants' group assignment. Self-reported measures were completed by the participant via a web-based questionnaire.

Intervention

The intervention has 3 components: (1) an in-person session with 20 min of group education and 30 min of individual counseling with a PT, (2) the use of a Fitbit Flex-2 wristband, and (3) PT counseling by phone to review physical activity goals (20-30 min). The in-person session was held in a meeting room at either the Mary Park Arthritis Centre, Fraser Health Authority, or Arthritis Research Canada. Fitbit is a consumer-grade wearable device that tracks and displays steps walked, gross level of physical exertion, and time spent being active. It is easy for users to put on and remove for charging. Fitbit Flex-2 is splash proof but cannot be used during water-based activities (eg, shower, swimming). Participants may record activities not captured by the device on the Fitbit website. Participants could view their physical activity goal attainment on FitViz, a new Fitbit-compatible web-based app developed for this study [21].

All 8 study PTs completed a 2-day basic training in motivational interviewing at the University of British Columbia [22]. In addition, they attended an orientation session, received a counseling guide, and shadowed at least one education and counseling session before they were paired with a participant.

Participants in the IG were scheduled to attend a PT-led group education session that focused on physical activity in OA management and strategies to manage joint symptoms. Next, each participant was paired with the next available study PT. During the individual counseling, PTs used the Brief Action Planning approach [23] to guide participants to set *specific, measurable, attainable, relevant, and time-bound* (SMART) physical activity goals. An example of a SMART goal is "attending a pool exercise class in a community center every Tuesday and Saturday for the next three months."

Study PTs then set the parameters on their assigned participants' FitViz accounts based on the participant's goals. These parameters included the following: (1) the upper and lower limits of intensity and duration of MVPA (ie, to promote physical activity based on the participant's goal), (2) the duration when a sedentary behavior should be interrupted (ie, to promote *less sitting*), and (3) the rest time in between sessions of MVPA (ie, to promote pacing).

During weeks 1 to 8, the study PT remotely reviewed participants' progress on FitViz and counseled them to modify their physical activity goals via 4 biweekly phone calls. A

counseling guide was provided, and the discussion was documented by the PT. During weeks 9 to 12, participants continued using their Fitbit and FitViz but had no counseling calls with their PTs. However, participants could email their PTs if there were questions related to being physically active. At the end of the program, participants could keep their Fitbit and FitViz account.

The DG received the same intervention in week 14. During the waiting period, they received monthly emails of arthritis news that were unrelated to physical activity.

Outcome Measures

The primary outcome measure was mean daily MVPA time measured with SenseWear Mini (BodyMedia Inc). SenseWear integrates triaxial accelerometer data, physiological sensor data, and personal demographic information to provide estimates of steps and energy expenditure. A strong relationship has been found between SenseWear and indirect calorimetry measures of energy expenditure for activities of daily living (Pearson r=0.85) [24]. The device can be worn 24 hours a day and can capture a full picture of physical activity and off-body time throughout the day [25]. An important feature of SenseWear is its ability to differentiate between sedentary and light physical activities [26]; hence, it is an ideal instrument to assess both active and sedentary behaviors.

Participants wore a SenseWear Mini over the triceps for 7 days at each assessment. Almeida et al [27] determined that a minimum of 4 days of wear was required to reliably assess energy expenditure from different levels of physical activity in people with rheumatoid arthritis (intraclass correlation coefficient >0.80). We calculated the average MVPA accumulated in bouts (min per day). A bout was defined as ≥10 min at the level of ≥3 metabolic equivalent of the task (MET), with an allowance for interruption of up to 2 min below the threshold [28].

The secondary outcomes included the following:

1. Average daily time in purposeful activity performed in ≥4 MET in bouts of ≥10 min, with allowance for interruption of up to 2 min below the threshold (eg, brisk walking) [29].
2. Average daily step count.
3. Average daily time in sedentary behavior performed in ≤1.5 MET in bouts of ≥20 min during waking hours [30-33].
4. Knee Injury and OA Outcome Score (KOOS) [34,35].
5. Partners in Health Scale (PIHS) [36].
6. Theory of Planned Behavior Questionnaire [37,38].
7. Patient Health Questionnaire-9 (PHQ-9) [39].
8. Self-Reported Habit Index [40,41].

We used SenseWear to measure purposeful activity time, steps, and sedentary behavior time. The KOOS consists of 5 subscales: *pain*, *symptoms*, *activities of daily living*, *sports/recreation*, and *knee-related quality of life*. It was originally developed for people recovering from injuries such as the anterior cruciate ligament and meniscus injury and was validated in patients with OA [34,35]. The PIHS is a 12-item measure designed to assess self-efficacy, knowledge of health conditions and treatment, and self-management behavior, such as adopting a healthy lifestyle (Cronbach α=.82) [36]. Motivation for engaging in physical activity was measured using the Rhodes 7-point Likert-type Theory of Planned Behavior questionnaire [37,38]. The questionnaire consists of 16 items measuring all components of the theory. Previous studies using this measure showed good predictive validity and internal consistency in adult populations [37,38].

The PHQ-9 consists of 9 questions that correspond to the diagnostic criteria for major depressive disorder. A score greater than 11 indicates a major depressive disorder [39]. The Self-Reported Habit Index is a 12-item scale, rated on a 7-point Likert scale, which measures characteristics of habitual behavior (reliability minimum α=.81) [40,41]. We asked participants to rate their strength of habit for 3 specific activity-related behaviors: sitting during leisure time at home, sitting during usual occupational activities, and walking outside for 10 min. A higher score indicates a stronger habit or behavior that is done frequently and automatically. Demographic variables and comorbid conditions were collected at baseline.

Adverse Event and Intervention Fidelity Monitoring

We tracked adverse events (falls as well as cardiovascular and musculoskeletal events) related to their physical activity [42] in the follow-up questionnaire at weeks 13, 26, and 39. Participants were deemed adhering to the 12-week intervention protocol if they (1) attended the education session, (2) used their Fitbit ≥5 days per week in ≥11 weeks, and (3) participated in ≥3 of 4 counseling calls. We monitored participants' Fitbit wear using FitViz, which wirelessly synchronized physical activity data recorded by a Fitbit 150 times per hour [43]. We calculated the percentage of participants meeting each criterion and all 3 criteria.

Data Analysis

Descriptive analysis was used to summarize participant characteristics and comorbid conditions. We performed an intention-to-treat analysis using SAS version 9.4. The analysis of covariance (ANCOVA) was used for the main analysis to estimate an adjusted mean difference comparing time in MVPA (primary outcome) at 13 weeks between groups, adjusting for baseline MVPA and blocking. In a secondary analysis, we used a longitudinal mixed effects model that allowed us to additionally examine the intervention effects at 26 and 39 weeks after intervention initiation. The mixed effects models included the following variables as fixed effects: (1) the randomization group indicator for baseline difference, (2) a set of indicator variables for follow-up assessment time points (weeks 13, 26, and 39) to account for secular trend, and (3) a set of indicator variables for the lengths of time since intervention initiation (weeks 12, 25, or 38) to estimate intervention effects after these amounts of time postintervention initiation. The models additionally included participants as random effects to account for the repeated measures nature of the data. We used the sandwich estimators for linear mixed models [44] to compute empirical standard errors that were robust to model specifications. We examined the secondary outcomes at 13 weeks using ANCOVA and over weeks 26 and 39 using the mixed effects models mentioned earlier.

Sample Size

We estimated that our collaboration with health authorities and patient groups allowed the study to recruit 60 eligible participants within 12 months. Our previous study of a similar physical activity counseling program resulted in an estimated MVPA time of 75.5 min per day (SD 54.3) in the intervention group and 50.0 min per day (SD 46.8) in the controls [18]. Assuming approximately 15% attrition, we anticipated that 50 of the 60 participants would complete the study. With a sample size of 50, we would have 74% power at an α level of 0.1 (via a one-sided test).

Ethics

The research protocol was approved by the University of British Columbia Behavioural Research Ethics Board (application number: H15-02038) and published in ClinicalTials.gov (NCT02585323).

Results

Baseline Characteristics

In the years 2017 to 2019, 253 people indicated an interest to participate, and 221 met the eligibility criteria (Figure 1). Of these, we recruited 51 participants (IG: 23/26, 88% were women; DG: 19/25, 76% were women). Both groups were similar in age (IG: mean 65.0, SD 8.3 years; DG: mean 64.8, SD 9.0 years) and BMI (IG: mean 29.8, SD 9.0 kg/m^2; DG: mean 28.9, SD 6.2 kg/m^2). Approximately 55% (28/51) of the participants did not meet the Canadian physical activity guidelines at baseline (Table 1).

Table 1. Baseline characteristics of participants.

Variables	All (N=51)	Immediate group (n=26)	Delay group (n=25)
Women, n (%)	42 (82)	23 (89)	19 (76)
Age (years), mean (SD)	64.9 (8.5)	65.0 (8)	64.8 (9)
Marital status, n (%)			
Married/common law	30 (59)	18 (69)	12 (48)
Separated/divorced	10 (20)	5 (19)	5 (20)
Widowed/never married/other	11 (22)	3 (12)	8 (32)
University degree or trades certificate, n (%)	25 (49)	14 (54)	11 (44)
Gross annual household income (US $), n (%)			
≤24,000	2 (4)	0 (0)	2 (8)
24,001-40,000	6 (12)	2 (8)	4 (16)
40,001-60,000	9 (18)	4 (15)	5 (20)
60,001-80,000	8 (16)	4 (15)	4 (16)
80,001-100,000	7 (14)	5 (19)	2 (8)
>100,000	7 (14)	3 (12)	4 (16)
No answer	12 (24)	8 (31)	4 (16)
Diagnosed with OA[a], n (%)			
Yes	37 (73)	20 (77)	17 (68)
No, but met the likely OA criteria	14 (28)	6 (23)	8 (32)
In general, would you say your health is, n (%)			
Excellent	6 (12)	1 (4)	5 (20)
Very good	14 (27)	10 (39)	4 (16)
Good	22 (43)	11 (42)	11 (44)
Fair	8 (16)	3 (12)	5 (20)
Poor	1 (2)	1 (4)	0 (0)
Compared with 1 year ago, how would you rate your health in general now? n (%)			
Much better	3 (6)	1 (4)	2 (8)
Somewhat better	8 (16)	4 (15)	4 (16)
About the same	24 (47)	11 (42)	13 (52)
Somewhat worse	16 (31)	10 (39)	6 (24)
Much worse	0 (0)	0 (0)	0 (0)
Number of comorbid conditions, median (25th, 75th percentile)	3.0 (2.0, 5.0)	4.0 (3.0, 5.0)	3.0 (2.0, 4.0)
BMI (kg/m^2), mean (SD)	29.4 (7.7)	29.8 (9.0)	28.9 (6.2)
Participants did not meet the Canadian physical activity guideline (≥150 min of MVPA[b] in bouts of ≥10 min per week), n (%)	28 (55)	15 (58)	13 (52)

[a]OA: osteoarthritis.

[b]MVPA: moderate-to-vigorous physical activity.

Comparison of the Immediate Group With Delay Group

Table 2 and Figure 2 present the results of the primary outcome from 4 time points. At baseline, the mean MVPA time was 31.0 min per day (SD 37.3) for the IG and 71.3 min per day (SD 99.8) for the DG. The DG accumulated significantly more MVPA time—2 outliners accumulated a mean of >300 min per day (Figure 3). At 13 weeks (the primary end point), the IG accumulated a mean MVPA of 37.7 min per day (SD 30.5), whereas the DG had 49.4 min per day (SD 63.6). The adjusted mean difference in time spent in MVPA between groups following the intervention at 13 weeks was 13.1 min per day (95% CI 1.6 to 24.5), favoring the IG. Analyses adjusted for

blocking yielded nearly identical results; thus, it was removed from subsequent analyses. A secondary analysis using a mixed effects model revealed smaller intervention effects at 12 weeks (9.4 min per day; 95% CI −3.0 to 21.7), 25 weeks (−3.0 min per day; 95% CI −34.9 to 29.0), and 38 weeks (0.2 min per day; 95% CI −44.0 to 44.4) postprogram initiation (Table 3).

Table 2. Participant outcomes and results of the primary analysis.

Measures	Immediate group, mean (SD)				Delay group, mean (SD)				Adjusted difference in mean change at T0-T1	
	Baseline (T0; n=26)	13 weeks (T1; n=24)	26 weeks (T2; n=22)	39 weeks (T3; n=20)	Baseline (T0; n=25)	13 weeks (T1; n=24)	26 weeks (T2; n=22)	39 weeks (T3; n=23)	Mean difference (95% CI)	*P* value
Time in MVPA[a,b] (min)	31.0 (37.3)	37.7 (30.5)	37.0 (32.3)	34.0 (25.2)	71.3 (99.8)	49.4 (63.6)	74.6 (102.1)	54.8 (66.2)	13.1 (1.6 to 24.5)	.03
Time in purposeful activity[c] (min)	11.1 (19.5)	13.3 (20.0)	13.7 (18.8)	12.8 (16.8)	42.1 (80.2)	23.1 (37.1)	36.5 (62.6)	22.0 (42.3)	1.6 (−3.0 to 6.1)	.50
Daily steps	6294.0 (3418.0)	7133.3 (3603.3)	6381.6 (3492.0)	5845.3 (2575.5)	7030.1 (3921.6)	6232.7 (3086.1)	8162.3 (6642.3)	7445.1 (4713.2)	1106.5 (−19.9 to 2232.9)	.05
Sedentary time[d] (min)	567.5 (183.1)	531.4 (173.5)	492.5 (156.6)	502.8 (135.3)	551.1 (234.9)	558.3 (224.9)	499.5 (248.9)	483.1 (225.4)	−29.5 (−75.8 to 16.7)	.21
Knee Injury and Osteoarthritis Outcome Score (0-100; higher=better)										
Symptoms	68.5 (10.7)	69.3 (12.7)	65.4 (45.8)	68.4 (16.3)[e]	65.7 (12.3)	66.9 (14.9)[f]	72.2 (16.8)[g]	72.5 (13.3)[g]	−1.4 (−7.8 to 4.9)	.66
Pain	72.6 (13.5)	73.1 (15.3)	72.5 (18.3)	72.1 (19.8)[e]	65.1 (13.7)	65.9 (15.6)[f]	74.8 (15.4)[g]	72.8 (13.2)[g]	2.5 (−4.2 to 9.5)	.49
Activity of daily living	75.5 (14.7)	75.0 (13.1)	77.7 (18.9)	74.8 (20.7)[e]	72.2 (15.8)	70.3 (16.9)[f]	80.3 (13.1)[g]	80.0 (14.0)[g]	2.8 (−3.3 to 8.8)	.37
Sports and recreation	47.9 (23.7)	47.1 (22.4)	50.5 (30.1)	51.6 (30.4)[e]	46.8 (25.4)	52.5 (22.7)[f]	62.5 (22.5)[g]	55.8 (26.3)[g]	−3.8 (−14.9 to 7.2)	.50
Quality of life	44.0 (16.0)	48.7 (17.5)	49.4 (15.7)	49.3 (19.1)[e]	47.5 (16.0)	46.9 (13.6)[f]	54.7 (14.7)[g]	54.4 (14.6)[g]	1.4 (−5.0 to 7.9)	.66
Partners in Health (0-96; higher=better)	76.8 (11.0)	76.7 (11.5)	78.9 (9.0)	81.0 (7.0)[e]	78.0 (12.0)	81.6 (9.6)[f]	82.6 (9.3)[g]	82.6 (10.0)[g]	−2.3 (−6.6 to 1.9)	.28
Patient Health Questionnaire-9 (0-27; lower=better)	5.2 (4.6)	4.0 (3.0)	3.6 (3.3)	4.2 (4.1)[e]	5.4 (5.3)	4.7 (4.9)[f]	4.5 (5.2)[g]	4.5 (5.3)[g]	−0.4 (−1.7 to 0.8)	.47
Self-Reported Habit Index (1-7; higher=stronger habit)										
Sitting at work subscale	5.0 (1.4)	5.2 (1.6)	4.7 (2.1)	5.0 (1.9)[e]	4.5 (2.1)	4.4 (2.0)[f]	4.3 (2.1)[g]	4.1 (2.1)[g]	0.7 (0.2 to 1.2)	.004
Sitting at leisure subscale	4.8 (1.1)	5.1 (1.1)	5.0 (1.1)	5.5 (1.1)[e]	5.1 (1.4)	4.7 (1.6)[f]	4.4 (1.9)[g]	4.7 (1.7)[g]	0.7 (0.2 to 1.2)	.006
Walking subscale	4.3 (1.8)	4.4 (1.6)	4.6 (1.8)	4.7 (1.8)[e]	4.8 (2.0)	4.6 (2.0)[f]	4.8 (1.6)[g]	4.6 (1.8)[g]	0.3 (−0.3 to 0.9)	.27
Theory of Planned Behavior Questionnaire (1-7; higher=more positive)										
Attitude toward physical activity	6.0 (0.6)	5.9 (0.7)	6.0 (0.5)	6.0 (0.5)[e]	6.1 (0.6)	6.1 (0.7)[f]	6.2 (0.6)[g]	6.1 (0.7)[g]	−0.1 (−0.4 to 0.2)	.63
Subjective norm	6.2 (0.6)	6.2 (0.8)	6.1 (0.8)	6.2 (1.1)[e]	6.3 (0.8)	6.2 (0.7)[f]	6.3 (0.8)[g]	6.3 (0.7)[g]	0.2 (−0.1 to 0.5)	.13
Perceived control	5.8 (1.0)	5.6 (1.4)	5.6(1.5)	5.7 (1.4)[e]	6.1 (0.7)	5.8 (1.2)[f]	6.2 (1.0)[g]	6.2 (0.9)[g]	0.1 (−0.6 to 0.8)	.80
Intention	6.2 (0.8)	5.9 (1.0)	5.7 (1.1)	5.8 (0.8)[e]	6.3 (0.8)	6.4 (0.6)[f]	6.4 (0.6)[g]	6.3 (0.9)[g]	0.1 (−0.1 to 0.3)	.22

[a]MVPA: moderate-to-vigorous physical activity.

[b]MVPA was performed at ≥3 metabolic equivalent of tasks and in bouts ≥10 min.

[c]Purposeful activity was performed at ≥4 METs and in bouts ≥10 min.

[d]Sedentary behavior was performed at ≤1.5 METs in bouts ≥20 min.

[e]n=19.

[f]n=22.

[g]n=24.

Figure 2. Time in moderate-to-vigorous physical activity.

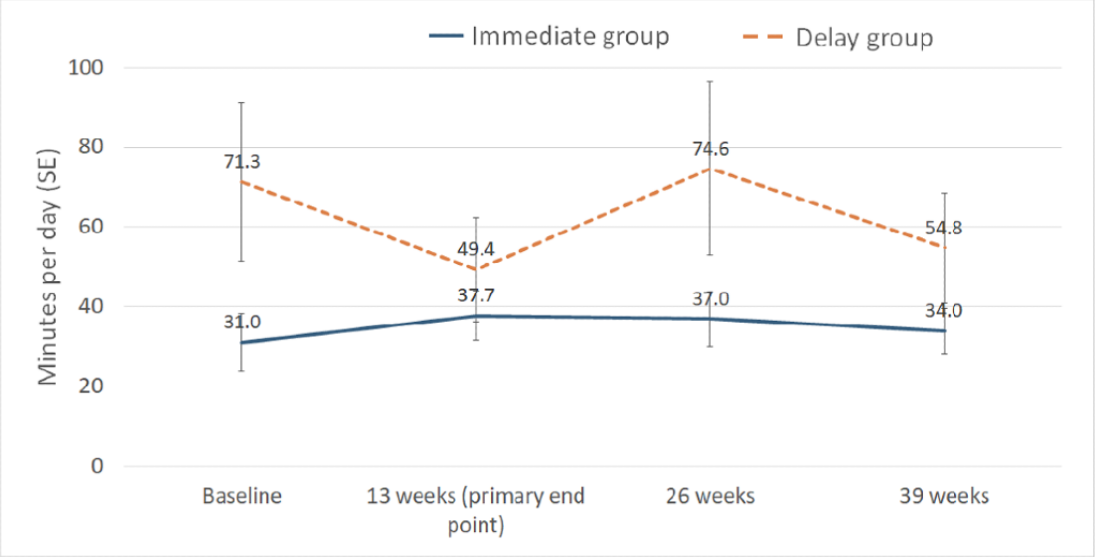


Figure 3. Boxplot of time in moderate-to-vigorous physical activity at baseline and 13 weeks.

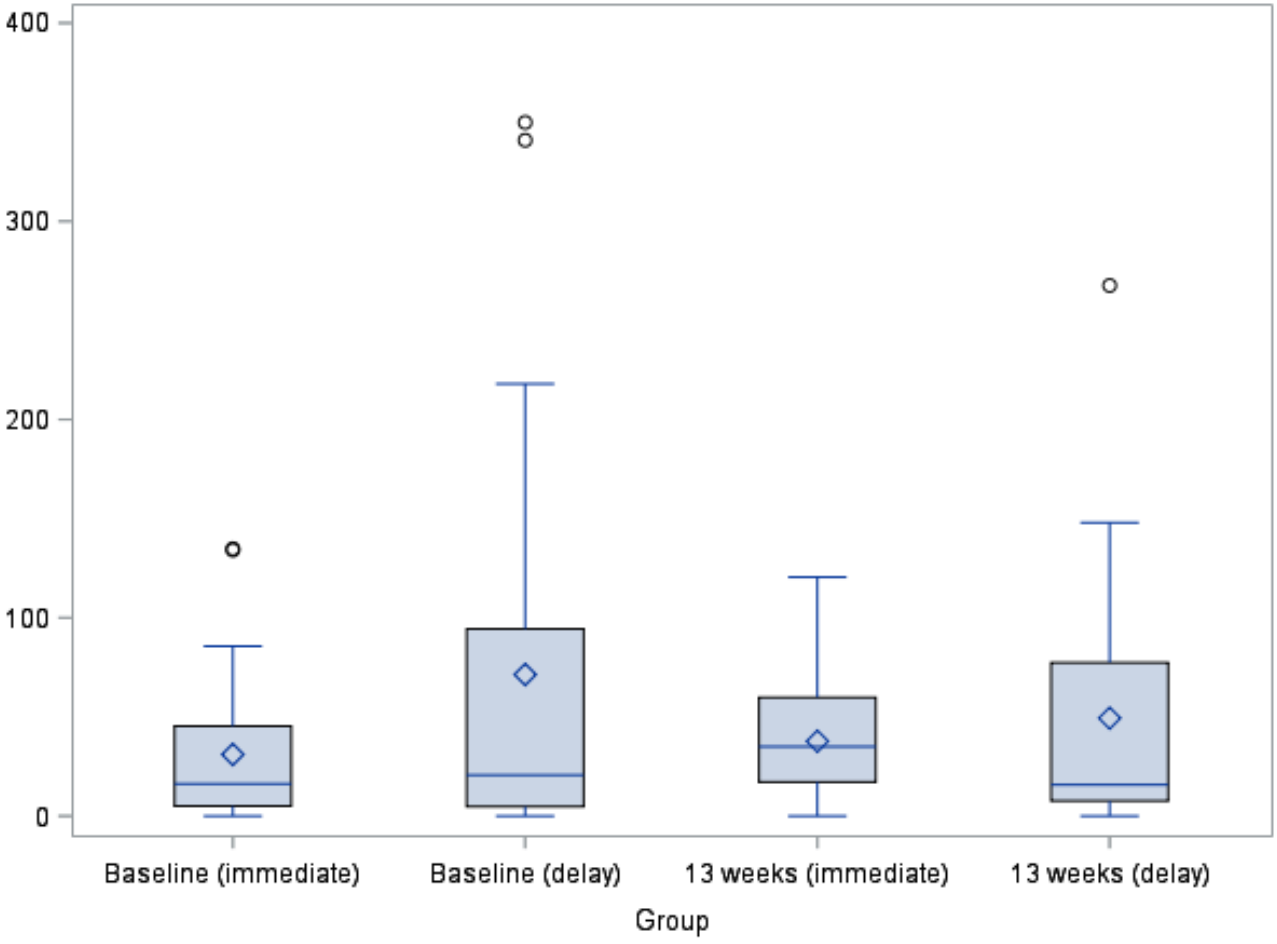


Table 3. Results of mixed effects models.

Measures	Intervention effect postprogram initiation (95% CI)			Secular trend (95% CI)	
	12 weeks	25 weeks	38 weeks	T2[a] effect vs T1[b]	T3[c] effect vs T1
Time in MVPA[d,e] (min)	9.4 (−3.0 to 21.7)	−3.0 (−34.9 to 29.0)	0.2 (−44.0 to 44.4)	11.9 (−12.4 to 36.1)	6.4 (−29.7 to 42.5)
Time in purposeful activity[f] (min)	1.5 (−8.4 to 11.5)	−9.0 (−33.5 to 15.5)	−6.8 (−42.3 to 28.6)	10.0 (−8.0 to 28.0)	6.9 (−21.3 to 35.1)
Daily steps	1461.2 (433.8 to 2488.6)	715.0 (−1995.9 to 3425.9)	71.4 (−4063.6 to 4206.3)	101.7 (−2142.6 to 2346.0)	336.9 (−3261.7 to 3935.4)
Sedentary time[g] (min)	−32.2 (−78.1 to 13.7)	−62.6 (−162.8 to 37.6)	−63.4 (−217.6 to 90.8)	−19.8 (−83.7 to 44.1)	−7.6 (−117.0 to 101.9)
Symptoms	−2.2 (−8.2 to 3.4)	−12.4 (−22.6 to −2.2)	−21.1 (−36.8 to −5.4)	6.1 (0.2 to 11.9)	16.5 (6.2 to 26.9)
Pain	0.3 (−6.2 to 6.8)	−7.8 (−19.7 to 4.1)	−15.1 (−33.1 to 3.0)	7.5 (1.3 to 13.8)	13.7 (2.0 to 25.4)
Activity of daily living	1.1 (−5.0 to 7.1)	−4.6 (−16.2 to 7.0)	−13.2 (−32.0 to 5.6)	8.1 (2.7 to 13.6)	13.5 (2.1 to 24.9)
Sports and recreation	−7.4 (−18.7 to 3.9)	−21.2 (−42.4 to 0.1)	−27.2 (−59.2 to 4.9)	16.5 (5.1 to 27.9)	23.6 (2.5 to 44.7)
Quality of life	2.4 (−4.6 to 9.3)	−2.1 (−13.6 to 9.4)	−6.9 (−24.8 to 11.0)	4.8 (−3.2 to 12.7)	9.0 (−3.2 to 21.2)
Partners in Health (0-96; higher=better)	−1.7 (−5.7 to 2.4)	−3.8 (−10.9 to 3.3)	−4.2 (−15.2 to 6.8)	3.4 (−1.1 to 7.9)	5.6 (−2.7 to 13.9)
Patient Health Questionnaire-9 (0-27; lower=better)	−0.9 (−2.4 to 0.6)	−1.3 (−4.0 to 1.3)	−1.3 (−5.5 to 2.9)	0.1 (−1.2 to 1.5)	0.6 (−2.2 to 3.3)
Sitting at work subscale	0.4 (−0.3 to 1.1)	0.3 (−1.1 to 1.6)	0.6 (−1.5 to 2.7)	−0.5 (−1.2 to 0.3)	−0.5 (−1.9 to 0.9)
Sitting at leisure subscale	0.6 (0.1 to 1.0)	1.4 (0.5 to 2.4)	2.4 (0.9 to 3.9)	−1.0 (−1.6 to −0.4)	−1.5 (−2.5 to −0.4)
Walking subscale	0.3 (−0.3 to 0.9)	0.4 (−0.6 to 1.4)	0.7 (−1.1 to 2.4)	0 (−0.6 to 0.5)	−0.3 (−1.4 to 0.8)
Beliefs toward physical activity	−0.1 (−0.4 to 0.2)	−0.2 (−0.7 to 0.3)	−0.3 (−1.1 to 0.4)	0.2 (−0.1 to 0.4)	0.2 (−0.4 to 0.7)
Subjective norm	0.1 (−0.3 to 0.5)	0.0 (−0.6 to 0.6)	0.1 (−0.9 to 1.1)	0.0 (−0.4 to 0.4)	0.0 (−0.6 to 0.6)
Perceived control	0.0 (−0.6 to 0.7)	−0.4 (−1.6 to 0.7)	−0.7 (−2.5 to 1.1)	0.3 (−0.3 to 1.0)	0.8 (−0.3 to 2.0)
Intention	−0.4 (−0.7 to −0.1)	−1.1 (−1.7 to −0.4)	−1.5 (−2.6 to −0.5)	0.4 (0.1 to 0.7)	1.0 (0.3 to 1.7)

[a]T2: 26-week assessment.

[b]T1: 13-week assessment.

[c]T3: 39-week assessment.

[d]MVPA: moderate-to-vigorous physical activity.

[e]MVPA was performed at ≥3 metabolic equivalent of tasks and in bouts of ≥10 min.

[f]Purposeful activity was performed at ≥4 METS and in bouts of ≥10 min.

[g]Sedentary behavior was performed at ≤1.5 METS in bouts of ≥20 min.

For the secondary outcome, a trend favoring the IG, but not statistically significant, was found from baseline to 13 weeks in purposeful activity time (1.6 min per day; 95% CI −3.0 to 6.1), step count (1106.5; 95% CI −19.9 to 2232.9), and sedentary time (−29.5 min per day; 95% CI −75.8 to 16.7). The results from the KOOS, PIHS, and PHQ-9 were also not statistically significant (Table 2). We found a small effect in perceived sitting habit while at work (0.7; 95% CI 0.2 to 1.2) or during leisure activities (0.7; 95% CI 0.2 to 1.2).

Secondary analysis demonstrated an effect attributable to being in the program for daily steps at 12 weeks (1461.2; 95% CI 433.8 to 2488.6; Table 3). Knee symptoms (measured by the KOOS symptoms subscale), perceived sitting habit during leisure activities, and intention to be physically active also showed statistically significant effects attributable to being in the program for different durations, although statistically significant secular trends were also observed over the assessment period.

Intervention Adherence and Adverse Events

Intervention adherence in the IG was 100% (26/26) for education session attendance, 96% (25/26) for PT counseling phone calls, and 81% (21/26) for Fitbit use (Table 4). In all, 81% (21/26) of participants met all 3 fidelity criteria. Adherence rates were similar in the DG when participants received the program in week 13. During the 4 weeks when the PT counseling ended, 2 participants each from the IG and the DG contacted their study PT via email with further questions regarding their physical activity.

After starting the program, 10 of the 51 participants reported adverse events because of physical activity; of those, 7 reported muscle pain (IG: n=5 and DG: n=2). Falls were reported by 3 in the IG; of those, 2 fell while being physically active (1 had

an ankle sprain). Three participants from the DG also reported a fall; 1 occurred while being physically active. Of the remaining 2 participants who had a fall, 1 sustained a vertebral compression fracture.

Table 4. Summary of intervention adherence.

Adherence criterion	Immediate group (n=26), n (%)	Delay group (n=25), n (%)	All (N=51), n (%)
Attended the initial session with group education and met with a physiotherapist to set physical activity goals	26 (100)	25 (100)	51 (100)
Completed ≥3 of 4 counseling phone calls with a physiotherapist	25 (96)	22 (88)	47 (92)
Met Fitbit use criteria[a] ≥11 weeks out of the 12-week intervention period	21 (81)	20 (80)	41 (80)
Met 2 of 3 criteria	25 (96)	24 (96)	49 (96)
Met all 3 criteria	21 (81)	18 (72)	39 (77)

[a]Participants had steps recorded in their Fitbit ≥5 days per week.

Discussion

Principal Findings

More than 1 in 6 people in the United States are using wearable devices to monitor their health [45], but the integration of these tools in chronic disease management is at an early stage. This study demonstrated the potential of a multifaceted wearable-based program for promoting MVPA in people with knee OA. We found an effect in the adjusted mean difference in time spent in MVPA between groups after the 12-week program. Furthermore, the mixed effects model analysis suggests a significant effect in daily steps attributable to the program.

These results, however, should be viewed in the context that the DG accumulated significantly more daily MVPA time than the IG at baseline, and the observed effect was primarily driven by a decline in daily MVPA time in the DG at week 13. More than 80% of participants rated their health as “good,” “very good,” or “excellent” at baseline, suggesting that they might have few health constraints to be physically active. Nonetheless, the results extend those of our previous RCT on a similar program, whereby a significant improvement in participation in physical activity was found in people with knee OA at the end of an 8-week intervention [18]. We also found a small effect on the awareness of sitting habits at work and during leisure activities at 13 weeks. The reason for this observation is unclear, but it is likely too small to be clinically important.

Our results contribute to the literature on physical activity promotion in arthritis management. Recommendations by the European League Against Rheumatism endorse the use of behavior change techniques to promote physical activity among people with arthritis [46]. The 2019 American College of Rheumatology/Arthritis Foundation guidelines further highlight the involvement of health professionals, including PTs, to deliver nonpharmacological treatment [47]. The optimal approach for supporting an active lifestyle in people with arthritis remains to be unclear, but research in healthy adults has shown that low-dose health coaching had little effect on physical activity behavior [48]. Although PTs are skilled in exercise prescription, a recent survey in Canada revealed up to 71% of the respondents wished to acquire further training in physical activity counseling.[49] This suggests an opportunity for professional development for PTs to master skills in counseling techniques that match their patients’ readiness to acquire a health-related behavior. In our study, training on motivational interviewing and the opportunity to shadow a PT with experience in counseling participants are essential to the program. The PTs’ written record for each interaction with participants allowed us to ensure that counseling followed the Brief Action Planning approach. Participants continuing to use their Fitbits suggests that the behavior of self-monitoring was sustained even after the PT counseling ended. Future research can refine and compare different implementation strategies of physical activity counseling for this population.

Strengths and Limitations

A strength of this study was intervention fidelity, with an overall intervention adherence of 81%. There is no consensus on what constitutes good intervention adherence, but 80% to 100% has been deemed as high fidelity in delivery [50]. Furthermore, more than 80% of participants adhered to Fitbit use over the 12-week period, indicating that it is feasible to transition individuals from a multifaceted program to a wearable-only intervention after the initial 8-week counseling from a PT.

This study has some limitations. With the use of a delay-control design in which the participants in the control arm received the program after a 13-week delay, the efficacy of the counseling program could only be unequivocally assessed at 13 weeks. At 26 and 39 weeks, both groups had already received the intervention, and there was an absence of a control group at these 2 time points, which can cause intervention effect estimates at 26 and 39 weeks to be less robust and more susceptible to small sample bias. Hence, the long-term effects of the program remain unclear. Furthermore, the results may not be generalizable to men because 82% of the participants were women.

Conclusions

Supporting a physically active lifestyle is a core component of physiotherapy practice. The Exercise is Medicine initiative advocates for the creation and worldwide implementation of effective physical activity promotion strategies in treatment plans for patients [51]. With the ubiquitous use of wearables, health professionals can leverage the use of these tools to motivate, monitor, and counsel people with arthritis to reach

physical activity goals. To this end, we have shown that a 12-week multifaceted counseling program, with the use of a wearable device, can improve physical activity participation in people with knee OA.

Acknowledgments

The authors would like to thank Navi Grewal, Morgan Barber, Ankit Gupta, Juliane Chien, Eric C Sayre, Hussein Mamdani, and Amrit Sandhu for their important contributions to this project. The authors would also like to thank Jacob McIvor for his support with the SenseWear Mini data processing. LCL was supported by the Harold Robinson/Arthritis Society Chair in Arthritic Diseases award, the Canada Research Chair Program, and the Michael Smith Foundation for Health Research Scholar Award. HX was supported by the Maureen and Milan Ilich/Merck Chair in Statistics for Arthritis and Musculoskeletal Diseases award. JAAZ was supported by the BC Lupus Society Research Scholar and the Walter & Marilyn Booth Research Scholar. DG was supported by the Canada Research Chair Program during the study. SuPRA project was funded by the Canadian Institutes of Health Research (CIHR; KAL-135772) and Preventing Complications from Inflammatory Skin, Joint and Bowel Conditions, a CIHR team grant (THC-135235). The latter provided funding for the development of FitViz and statistical analysis for the RCT. The funders had no role in the study design, collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Authors' Contributions

LCL, LMF, CDS, DG, AMH, CK, JAAZ, AFT, GN, and CLB obtained funding. LCL, LMF, HX, CDS, DG, AMH, JAAZ, AFT, GN, and CLB contributed to the study design and proposal. LCL contributed to study oversight. LCL, SZ, JT, and ST contributed to data collection. LCL, LMF, HX, and NL analyzed the data. LCL, LMF, HX, NL, and AMH interpreted the data. LCL, LMF, HX, NL, CDS, DG, SZ, JAAZ, AMH, CK, JT, ST, AFT, GN, and CLB reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 735 KB - mhealth_v8i7e19116_app1.pdf]

References

1. Woolf AD, Akesson K. Understanding the burden of musculoskeletal conditions. The burden is huge and not reflected in national health priorities. Br Med J 2001 May 5;322(7294):1079-1080 [FREE Full text] [doi: 10.1136/bmj.322.7294.1079] [Medline: 11337425]
2. Health Canada. Arthritis in Canada: An Ongoing Challenge. Ottawa, Canada: Health Canada; 2003.
3. Bombardier C, Hawker G, Mosher D. The Impact of Arthritis in Canada: Today and Over the Next 30 Years. Arthritis Alliance of Canada. 2011. URL: http://www.arthritisalliance.ca/images/PDF/eng/Initiatives/20111022_2200_impact_of_arthritis.pdf [accessed 2020-06-19]
4. Brosseau L, MacLeay L, Robinson V, Wells G, Tugwell P. Intensity of exercise for the treatment of osteoarthritis. Cochrane Database Syst Rev 2003(2):CD004259. [doi: 10.1002/14651858.CD004259] [Medline: 12804510]
5. Ottawa Panel. Ottawa panel evidence-based clinical practice guidelines for therapeutic exercises and manual therapy in the management of osteoarthritis. Phys Ther 2005 Sep;85(9):907-971. [Medline: 16117601]
6. Zhang W, Moskowitz R, Nuki G, Abramson S, Altman R, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007 Sep;15(9):981-1000 [FREE Full text] [doi: 10.1016/j.joca.2007.06.014] [Medline: 17719803]
7. Zhang W, Moskowitz R, Nuki G, Abramson S, Altman R, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008 Feb;16(2):137-162 [FREE Full text] [doi: 10.1016/j.joca.2007.12.013] [Medline: 18279766]
8. Tremblay MS, Warburton DE, Janssen I, Paterson DH, Latimer AE, Rhodes RE, et al. New Canadian physical activity guidelines. Appl Physiol Nutr Metab 2011 Feb;36(1):36-46; 47. [doi: 10.1139/H11-009] [Medline: 21326376]
9. Dunlop DD, Song J, Semanik PA, Chang RW, Sharma L, Bathon JM, et al. Objective physical activity measurement in the osteoarthritis initiative: are guidelines being met? Arthritis Rheum 2011 Nov;63(11):3372-3382 [FREE Full text] [doi: 10.1002/art.30562] [Medline: 21792835]
10. Wallis J, Webster K, Levinger P, Taylor N. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. Osteoarthritis Cartilage 2013 Nov;21(11):1648-1659 [FREE Full text] [doi: 10.1016/j.joca.2013.08.003] [Medline: 23948979]

11. Gyurcsik NC, Brawley LR, Spink KS, Brittain DR, Fuller DL, Chad K. Physical activity in women with arthritis: examining perceived barriers and self-regulatory efficacy to cope. Arthritis Rheum 2009 Aug 15;61(8):1087-1094 [FREE Full text] [doi: 10.1002/art.24697] [Medline: 19644901]
12. der Ananian C, Wilcox S, Saunders R, Watkins K, Evans A. Factors that influence exercise among adults with arthritis in three activity levels. Prev Chronic Dis 2006 Jul;3(3):A81 [FREE Full text] [Medline: 16776882]
13. Henchoz Y, Zufferey P, So A. Stages of change, barriers, benefits, and preferences for exercise in RA patients: a cross-sectional study. Scand J Rheumatol 2013;42(2):136-145. [doi: 10.3109/03009742.2012.724707] [Medline: 23244196]
14. 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]
15. 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 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
16. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
17. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Mar 6;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
18. 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]
19. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and therapeutic criteria committee of the american rheumatism association. Arthritis Rheum 1986 Aug;29(8):1039-1049. [doi: 10.1002/art.1780290816] [Medline: 3741515]
20. Thomas S, Reading J, Shephard RJ. Revision of the physical activity readiness questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
21. Gupta A, Tong X, Shaw CD, Li LC, Feehan LM. FitViz: A Personal Informatics Tool for Self-Management of Rheumatoid Arthritis. In: Proceedings of the International Conference on Human-Computer Interaction.: Springer International Publishing; 2017 Presented at: HCI'17; July 9-14, 2017; Vancouver Canada URL: https://link.springer.com/chapter/10.1007%2F978-3-319-58753-0_35 [doi: 10.1007/978-3-319-58753-0_35]
22. Rollnick S, Miller WR, Butler CC. Motivational Interviewing in Health Care: Helping Patients Change Behavior. New York, USA: Guilford Press; 2008.
23. Gutnick D, Reims K, Davis C, Gainforth H, Jay M, Cole S. Brief action planning to facilitate behavior change and support patient self-management. J Clin Outcomes Manag 2014;21:17-29 [FREE Full text]
24. Tierney M, Fraser A, Purtill H, Kennedy N. Study to determine the criterion validity of the sensewear armband as a measure of physical activity in people with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2013 Jun;65(6):888-895 [FREE Full text] [doi: 10.1002/acr.21914] [Medline: 23213019]
25. Holsgaard-Larsen A, Roos EM. Objectively measured physical activity in patients with end stage knee or hip osteoarthritis. Eur J Phys Rehabil Med 2012 Dec;48(4):577-585 [FREE Full text] [Medline: 22641252]
26. Feehan LM, Goldsmith CH, Leung AY, Li LC. Sensewearmini and actigraph GT3X accelerometer classification of observed sedentary and light-intensity physical activities in a laboratory setting. Physiother Can 2016;68(2):116-123 [FREE Full text] [doi: 10.3138/ptc.2015-12] [Medline: 27909358]
27. Almeida GJ, Wasko MC, Jeong K, Moore CG, Piva SR. Physical activity measured by the SenseWear Armband in women with rheumatoid arthritis. Phys Ther 2011 Sep;91(9):1367-1376 [FREE Full text] [doi: 10.2522/ptj.20100291] [Medline: 21719635]
28. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
29. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett Jr DR, Tudor-Locke C, et al. Compendium of Physical Activities Tracking Guide. Google Sites. 2011. URL: https://sites.google.com/site/compendiumofphysicalactivities [accessed 2020-06-15]
30. Owen N. Sedentary behavior: understanding and influencing adults' prolonged sitting time. Prev Med 2012 Dec;55(6):535-539. [doi: 10.1016/j.ypmed.2012.08.024] [Medline: 22968124]
31. Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983 [FREE Full text] [doi: 10.2337/dc11-1931] [Medline: 22374636]

32. Latouche C, Jowett JB, Carey AL, Bertovic DA, Owen N, Dunstan DW, et al. Effects of breaking up prolonged sitting on skeletal muscle gene expression. J Appl Physiol (1985) 2013 Feb 15;114(4):453-460 [FREE Full text] [doi: 10.1152/japplphysiol.00978.2012] [Medline: 23271697]
33. Howard BJ, Fraser SF, Sethi P, Cerin E, Hamilton MT, Owen N, et al. Impact on hemostatic parameters of interrupting sitting with intermittent activity. Med Sci Sports Exerc 2013 Jul;45(7):1285-1291. [doi: 10.1249/MSS.0b013e318285f57e] [Medline: 23439415]
34. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
35. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and osteoarthritis outcome score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports 1998 Dec;8(6):439-448. [doi: 10.1111/j.1600-0838.1998.tb00465.x] [Medline: 9863983]
36. Petkov J, Harvey P, Battersby M. The internal consistency and construct validity of the partners in health scale: validation of a patient rated chronic condition self-management measure. Qual Life Res 2010 Sep;19(7):1079-1085. [doi: 10.1007/s11136-010-9661-1] [Medline: 20437206]
37. Rhodes RE, Courneya KS, Blanchard CM, Plotnikoff RC. Prediction of leisure-time walking: an integration of social cognitive, perceived environmental, and personality factors. Int J Behav Nutr Phys Act 2007 Oct 31;4:51 [FREE Full text] [doi: 10.1186/1479-5868-4-51] [Medline: 17974022]
38. Rhodes RE, Blanchard CM, Courneya KS, Plotnikoff RC. Identifying belief-based targets for the promotion of leisure-time walking. Health Educ Behav 2009 Apr;36(2):381-393. [doi: 10.1177/1090198107308376] [Medline: 18077658]
39. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the patient health questionnaire (PHQ-9): a meta-analysis. Can Med Assoc J 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
40. Gardner B, de Bruijn G, Lally P. A systematic review and meta-analysis of applications of the self-report habit index to nutrition and physical activity behaviours. Ann Behav Med 2011 Oct;42(2):174-187. [doi: 10.1007/s12160-011-9282-0] [Medline: 21626256]
41. Gardner B, Abraham C, Lally P, de Bruijn G. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the self-report habit index. Int J Behav Nutr Phys Act 2012 Aug 30;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
42. Ory M, Resnick B, Jordan PJ, Coday M, Riebe D, Garber CE, et al. Screening, safety, and adverse events in physical activity interventions: collaborative experiences from the behavior change consortium. Ann Behav Med 2005 Apr;29(Suppl):20-28. [doi: 10.1207/s15324796abm2902s_5] [Medline: 15921486]
43. Rate Limits. Fitbit SDK. 2019. URL: https://dev.fitbit.com/build/reference/web-api/basics/#rate-limits [accessed 2019-01-01]
44. Diggle PG, Heagerty P, Liang KY, Zeger S. Analysis of Longitudinal Data (Oxford Statistical Science Series - 25). Oxford, UK: Oxford University Press; 2013.
45. iHealth: How Consumers Are Using Tech to Stay Healthy. The Nielsen Company. 2014 Apr 10. URL: http://www.nielsen.com/us/en/insights/news/2014/ihealth-how-consumers-are-using-tech-to-stay-healthy.html [accessed 2020-06-15]
46. Osthoff AR, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018 Sep;77(9):1251-1260. [doi: 10.1136/annrheumdis-2018-213585] [Medline: 29997112]
47. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020 Feb;72(2):220-233. [doi: 10.1002/art.41142] [Medline: 31908163]
48. Ellingson LD, Lansing JE, DeShaw KJ, Peyer KL, Bai Y, Perez M, et al. Evaluating motivational interviewing and habit formation to enhance the effect of activity trackers on healthy adults' activity levels: randomized intervention. JMIR Mhealth Uhealth 2019 Feb 14;7(2):e10988 [FREE Full text] [doi: 10.2196/10988] [Medline: 30762582]
49. O'Brien MW, Shields CA, Campbell KL, Crowell SJ, Fowles JR. Perceptions and practices of providing physical activity counselling and exercise prescriptions among physiotherapists in Nova Scotia. Phys Can 2019 Aug 29;-:1-9. [doi: 10.3138/ptc-2018-0098]
50. Borrelli B, Sepinwall D, Ernst D, Bellg AJ, Czajkowski S, Breger R, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. J Consult Clin Psychol 2005 Oct;73(5):852-860. [doi: 10.1037/0022-006X.73.5.852] [Medline: 16287385]
51. Lobelo F, Stoutenberg M, Hutber A. The exercise is medicine global health initiative: a 2014 update. Br J Sports Med 2014 Dec;48(22):1627-1633. [doi: 10.1136/bjsports-2013-093080] [Medline: 24759911]

Abbreviations

ANCOVA: analysis of covariance
CIHR: Canadian Institutes of Health Research

DG: delay group
IG: immediate group
KOOS: Knee Injury and Osteoarthritis Outcome Score
MET: metabolic equivalent of task
MVPA: moderate-to-vigorous physical activity
OA: osteoarthritis
PHQ-9: Patient Health Questionnaire-9
PIHS: Partners in Health Scale
PT: physiotherapist
RCT: randomized controlled trial
SMART goal: specific, measurable, attainable, relevant, and time-bound goal
SuPRA: Supporting Physical activity & Reducing sedentary behaviour in Arthritis

Edited by G Eysenbach; submitted 04.04.20; peer-reviewed by R Warty, AS Aslam, G Signorelli; comments to author 27.04.20; revised version received 01.05.20; accepted 03.06.20; published 03.07.20.

Please cite as:
Li LC, Feehan LM, Xie H, Lu N, Shaw CD, Gromala D, Zhu S, Aviña-Zubieta JA, Hoens AM, Koehn C, Tam J, Therrien S, Townsend AF, Noonan G, Backman CL
Effects of a 12-Week Multifaceted Wearable-Based Program for People With Knee Osteoarthritis: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(7):e19116
URL: https://mhealth.jmir.org/2020/7/e19116
doi:10.2196/19116
PMID:32618578

Original Paper

Effectiveness of Wearable Trackers on Physical Activity in Healthy Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Matilda Swee Sun Tang[1], BA, MD; Katherine Moore[1], BSc, PhD, MD; Andrew McGavigan[1,2], MBChB, MD; Robyn A Clark[3,4], RN, RM, ICUcert, DipAppliSci, BN, MEd, PhD, ACCCN, FCNA, FAHA, FCSANZ; Anand N Ganesan[1,2,4], MBBS, PhD, FRACP, FCSANZ

[1]College of Medicine and Public Health, Flinders University, Adelaide, Australia

[2]Department of Cardiovascular Medicine, Flinders Medical Centre, Adelaide, Australia

[3]College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

[4]South Australian Health and Medical Research Institute, Adelaide, Australia

Corresponding Author:
Anand N Ganesan, MBBS, PhD, FRACP, FCSANZ
College of Medicine and Public Health
Flinders University
Level 5, Room 5E209 Flinders Medical Centre
Adelaide
Australia
Phone: 61 (08) 7221 8200
Email: anand.ganesan@flinders.edu.au

Abstract

Background: Wearable trackers are an increasingly popular tool among healthy adults and are used to facilitate self-monitoring of physical activity.

Objective: We aimed to systematically review the effectiveness of wearable trackers for improving physical activity and weight reduction among healthy adults.

Methods: This review used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology and reporting criteria. English-language randomized controlled trials with more than 20 participants from MEDLINE, CINAHL, Cochrane Library, Web of Science, PubMed, and Scopus (2000-2017) were identified. Studies were eligible for inclusion if they reported an intervention group using wearable trackers, reporting steps per day, total moderate-to-vigorous physical activity, activity, physical activity, energy expenditure, and weight reduction.

Results: Twelve eligible studies with a total of 1693 participants met the inclusion criteria. The weighted average age was 40.7 years (95% CI 31.1-50.3), with 64.4% women. The mean intervention duration was 21.4 weeks (95% CI 6.1-36.7). The usage of wearable trackers was associated with increased physical activity (standardized mean difference 0.449, 95% CI 0.10-0.80; *P*=.01). In the subgroup analyses, however, wearable trackers demonstrated no clear benefit for physical activity or weight reduction.

Conclusions: These data suggest that the use of wearable trackers in healthy adults may be associated with modest short-term increases in physical activity. Further data are required to determine if a sustained benefit is associated with wearable tracker usage.

(JMIR Mhealth Uhealth 2020;8(7):e15576) doi:10.2196/15576

KEYWORDS

wearable activity tracker; physical activity; healthy adults; randomized controlled trials

Introduction

Wearable activity trackers have rapidly emerged in the past decade as consumer devices to support self-monitoring of physical activity [1,2]. The use of these devices has increased exponentially, and the global sales of wearables in health care are expected to reach US $4.4 billion in 2019 and US $4.5 billion by 2020 [3].

In the past, structured lifestyle interventions have utilized education with behavior change techniques, provision of written information materials, and telephone counseling in a series of combination and permutation [4,5]. These interventions are successful in the short term but not in the long term, and they tend to be labor-intensive and costly [2,6,7]. Today, the availability and accessibility of wearable trackers equip consumers with the ability to monitor their physical activity along with online applications with motivational and tracking tools. Several systematic reviews have shown that wearable trackers are effective [4,6,8].

Contemporary wearable trackers differ from conventional pedometers as they are sophisticated devices providing real-time multidimensional feedback on physiological and health parameters including steps, calories burned, distance covered, active time, sleep assessment, and heart rate, and may include mobile connectivity or an internet application to provide personalized feedback reports [9].

To date, most of the literature on wearable trackers has focused on their feasibility, validity, and reliability [10] with limited data on the impact using these devices has on improving physical activity [6,11]. The primary purpose of this study is to evaluate the effectiveness of wearable trackers and their impact on physical activity levels in healthy adult populations with secondary outcomes of weight change in overweight populations.

Methods

Protocol and Registration

The protocol for this study is registered under PROSPERO with registration number CRD42019131868.

Eligibility Criteria, Information Sources, Search and Study Selection

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [12] methods and reporting were used to perform a systematic literature search with a professional medical librarian (Multimedia Appendix 1) on English-language randomized controlled trials published between January 1, 2000, and August 1, 2017.

We considered English-language studies eligible for inclusion if they reported an intervention with at least one of the groups using wearable trackers to provide objective feedback on physical activity to the wearer, alone or in combination with other interventions to enhance physical activity. Only randomized controlled trials with more than 20 participants in the adult outpatient and community setting that reported a change in physical activity behavior (total steps, total activity, the proportion of participants at activity goal) were included. We are looking at the healthy population and therefore excluded studies that required participants to be hospitalized or confined to a research center, studies in disease populations, and obese populations. We excluded studies that were predominantly pedometer-based interventions since we were only considering the effect of wearable trackers.

Data Collection Process and Data Items

Two authors independently abstracted three categories of variables from each of the included studies, with differences resolved by consensus: intervention variables (intervention duration); participant variables; quality variables (method of blinding control participants to step counts, the use of validity- and reliability-tested wearable trackers, the extent of affordability of wearable trackers and the extent to which co-interventions may have affected physical activity). If the study reported results from a different period, we used the final immediate post-intervention data in our primary analysis. For studies that reported different intensities of physical activity instead of steps per day, we chose the walking intensity results for the primary analysis.

Risk of Bias and Quality Assessment

The risk of bias was assessed using the Cochrane risk of bias tool [13] across seven domains. Each domain was scored with low (L), unclear (U), or high (H) risk of bias. The domains assessed were as follows:

- Random sequence generation: Was there selective bias (biased allocation to interventions) due to inadequate generation of a randomized sequence?
- Allocation concealment: Was there selective bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment?
- Selective outcome reporting: Was there reporting bias due to selective outcome reporting?
- Blinding outcome assessment: Was there detection bias due to knowledge of the allocated interventions by outcome assessors?
- Blinding participants and personnel: Was there performance bias due to knowledge of the allocated interventions by participants and personnel during the study?
- Incomplete outcome data: Was there attribution bias due to knowledge of the allocated interventions by outcome assessors?
- Other sources of bias: Was there bias due to problems not covered elsewhere?

The GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) system was used to rank the quality of evidence for each study [14]. The GRADE approach uses five considerations (study limitations, consistency of effect, imprecision, indirectness, and publication bias). The criteria for the grade of evidence are as follows:

- High: We are very confident that the true effect lies close to that of the estimated effect.
- Moderate: We are moderately confident in the effect estimate; the true effect is likely close to the estimated effect, but there is a possibility that it is substantially different
- Low: Our confidence in the effect estimate is limited; the true effect may be substantially different from the estimated effect.
- Very low quality: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of the effect.

Synthesis of Results

Statistical analysis was performed with Comprehensive Meta-Analysis Version 3 [15]. For each of the included studies, we calculated the effect sizes of physical activity for our primary interest outcome using steps/day, total moderate-vigorous physical activity (MVPA), activity units or physical activity energy expenditure, depending on the available results from the studies with steps/day taking priority. We calculated the summary outcomes or the standardized mean difference (SMD) (95% CI) using random-effects calculations. SMD is used since the included studies all assess the same outcome, physical activity, but they measure it in a variety of ways (steps/day, MVPA, energy expenditure). Hence, differences in means that are the same proportion of the standard deviation will have the same SMD, regardless of the actual scales used to make the measurements [16].

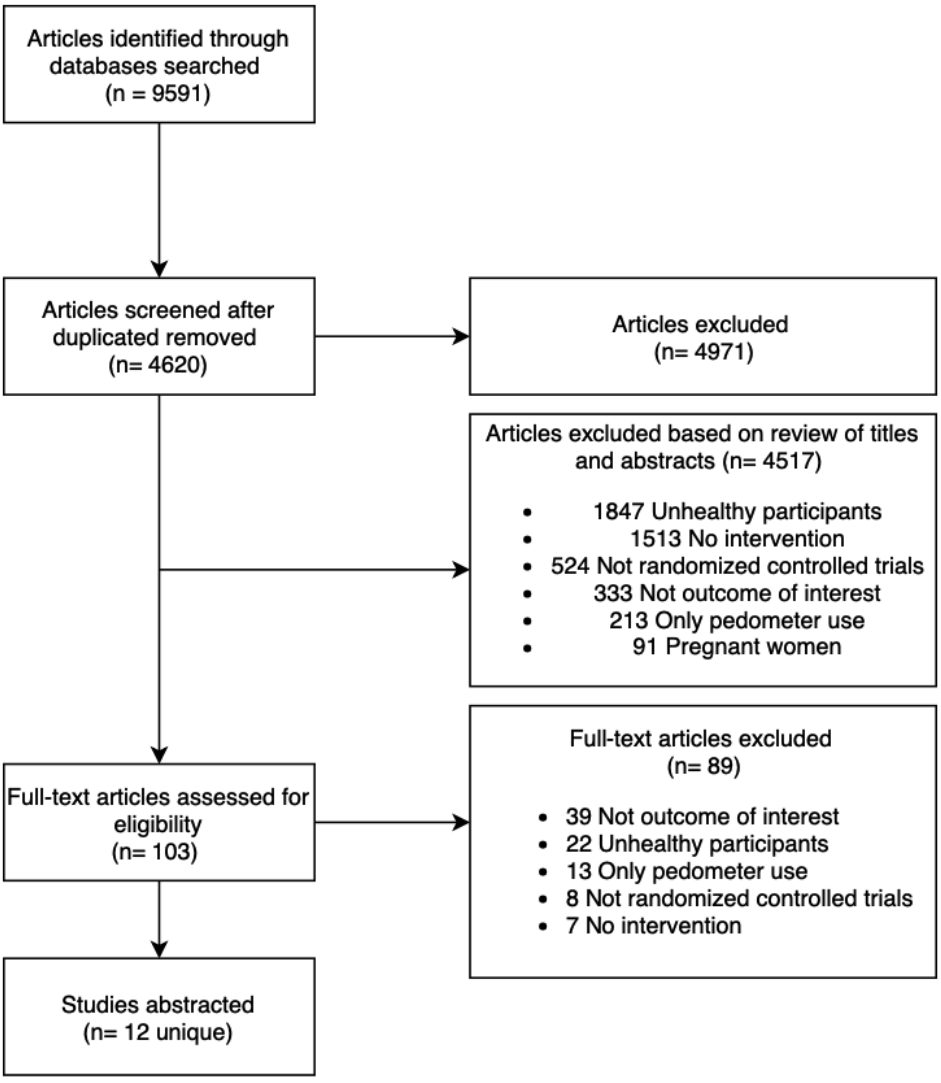
The I^2 statistic was used as a measure of variability in observed effects estimates attributable to between-study heterogeneity [17]. For variables exhibiting mild heterogeneity ($I^2 \leq 25\%$), pooled estimates were derived with fixed effects models. For variables exhibiting more than moderate heterogeneity ($I^2 > 25\%$), pooled estimates were derived with random-effects models. Sub-group analysis was used to assess secondary outcomes.

Results

Study Selection

The primary search identified a total of 9591 studies from MEDLINE (n=1910), CINAHL (n=62), Cochrane Library (n=2871), Web of Science (n=1633), PubMed (n=1280), and Scopus (n=1835), as detailed in Multimedia Appendix 1. There were 4620 unique citations in which 103 full texts were retrieved for further review after applying the Population, Intervention, Comparison, Outcome, and Study type [18] criteria of inclusion and exclusion studies (Multimedia Appendix 2) at the title-and-abstract screening level. A total of 12 unique papers were retained for data abstraction (Figure 1) [19].

Figure 1. PRISMA Flow Chart.



Study Characteristics

Characteristics of the included studies are presented in Table 1, and the intervention and comparators are presented in Table 2. A total of 1693 participants were included from 12 randomized trials. Included studies were published from 2007 to 2017.

Study sample sizes varied from 21 to 471 participants. The participants' weighted average age was 40.7 years (95% CI 31.1-50.3), and 64.4% of the participants were women. The duration of interventions ranged from 6 to 104 weeks, with a mean intervention duration of 21.4 weeks (95% CI 6.1-36.7).

Table 1. Characteristics of participants in the included studies.

Study name, publication year	Sample size	Participant characteristics			Study duration (weeks)
		Health status	Mean age (years)	Proportion	
Ashe, 2016 [20]	25	Healthy participants	40.1	Women: 100%	26
Buis, 2017 [21]	40	Healthy overweight adults	61.5	Women: 66%; white ethnicity: 66%	12
Cadmus-Bertram, 2015 [22]	51	Healthy and overweight participants	60.0	White ethnicity: 92%	16
Godino, 2013 [23]	466	Healthy participants	47.7	Women: 46%; white ethnicity: 96%	8
Hurling, 2007 [24]	77	Healthy participants	40.3	Women: 66%; white ethnicity: 99%	9
Jakicic, 2016 [25]	471	Healthy overweight adults	30.9	Women: 71%; white ethnicity: 77%	104
Martin, 2015 [26]	49	Healthy overweight adults	58.0	Women: 46%; white ethnicity: 79%	5
Melton, 2016 [27]	69	Healthy participants	19.7	Women: 100%; black ethnicity: 100%	6
Poirier, 2007 [28]	264	Healthy participants	39.9	Women: 66%; white ethnicity: 77%	7
Shrestha, 2013 [29]	28	Healthy overweight adults	32.1	Women: 54%	26
Thompson, 2014 [30]	49	Healthy participants	79.1	Women: 91%; white ethnicity: 66%	26
Thorndike, 2014 [31]	104	Healthy participants	29.0	Women: 54%	12

Table 2. Characteristics of intervention and comparators of included studies.

Study Name, Year	Intervention Device	Intervention	Comparator
Ashe, 2016 [20]	Fitbit	26 weeks of group-based education, social support, individualized physical activity prescription, given Fitbit	26 weeks: only received health-related information
Buis, 2017 [21]	Jawbone Up24	Received a Jawbone Up24 monitor, a tablet with Jawbone Up app installed, and brief weekly telephone counseling	Waitlist control (did not receive any intervention until after their final assessment where they were provided the intervention in full)
Cadmus-Bertram, 2015 [22]	Fitbit	16 weeks of Web-Based Tracking Group: Fitbit, instructional session, follow-up call at the fourth week	16 weeks of standard pedometer
Godino, 2013 [23]	Combined HR monitor and accelerometer (Actiheart)	8 weeks of wearing of Actiheart with one of three different types of feedback (simple, visual, contextualized)	8 weeks of wearing of Actiheart but with no feedback until the end of the trial
Hurling, 2007 [24]	Wrist-worn accelerometer	9 weeks of wristworn accelerometer, weekly exercise schedule, email reminders, real-time feedback via the internet	9 weeks of wrist-worn accelerometer with no feedback
Jakicic, 2016 [25]	FIT Core; BodyMedia	24 weeks of enhanced intervention: wearable technology, accompanying web interface to monitor diet and physical activity	24 weeks of standard intervention: website for self-monitoring of diet and physical activity
Martin, 2015 [26]	Fitbug Orb	3-arm study • Phase 1 (1 week): blinded run-in • Phase 2 (2 weeks): unblinded versus blinded tracking • Phase 3 (2 weeks): smart texts versus no texts	Blinded participants with no feedback
Melton, 2016 [27]	Jawbone UP	6 weeks of wearing Jawbone UP band and engaging with the application daily with weekly reminders	6 weeks of using MyFitnessPal application
Poirier, 2007 [28]	Variety of activity trackers	2-arm study • 6 weeks of walking program, Walkadoo, and wireless activity tracker • 1 week of follow-up with wearing of activity tracker for at least 10 hours a day	2-arm study • 6 weeks of not wearing activity trackers and maintaining daily activity routine • 1 week of follow-up wearing of activity tracker for 10 hours a day
Shrestha, 2013 [29]	Polar FA20 accelerometer	1 time 1.5-hour lifestyle instruction and 26 weeks of continuous accelerometer use and feedback	26 weeks of self-directed exercise and/or US Army mandated physical training
Thompson, 2014 [30]	Fitbit	26 weeks of accelerometer use and feedback, weekly brief telephone counseling sessions focused on accelerometer feedback, 6 in-person brief counseling sessions	26 weeks of accelerometer without feedback
Thorndike, 2014 [31]	Fitbit e3	2-arm study • Phase 1: 6 weeks RCT[a] comparing daily steps displaying feedback about steps and energy consumed • Phase 2: 6 weeks non-RCT team steps competition where all participants wore monitors with feedback	2-arm study • Phase 1: 6 weeks blinded monitor • Phase 2: 6 weeks non-RCT team steps competition where all participants wore monitor with feedback

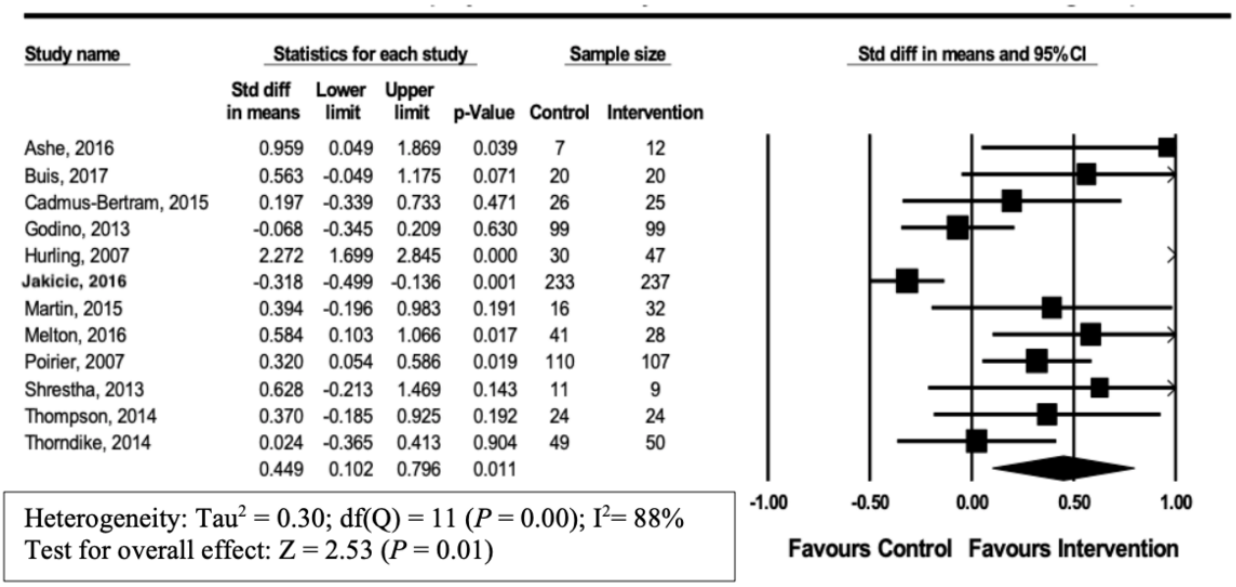
[a]RCT: randomized controlled trial.

Impact of Wearable Tracker Use and Physical Activity

The primary outcome for this review was the impact of Wearable Tracker Use and Physical Activity. The overall summary estimate from 12 studies [20-31] showed a modest increase in physical activity with the usage of wearable trackers (SMD 0.449, 95% CI 0.10-0.80; *P*=.01). There was significant heterogeneity (I^2=88%) (Figure 2). Subgroup analyses were performed for studies using steps/day or weight as reported endpoints, and in healthy versus overweight populations to explore mechanisms of heterogeneity.

We performed subgroup analyses and assessed heterogeneity to evaluate the robustness of our results [32].

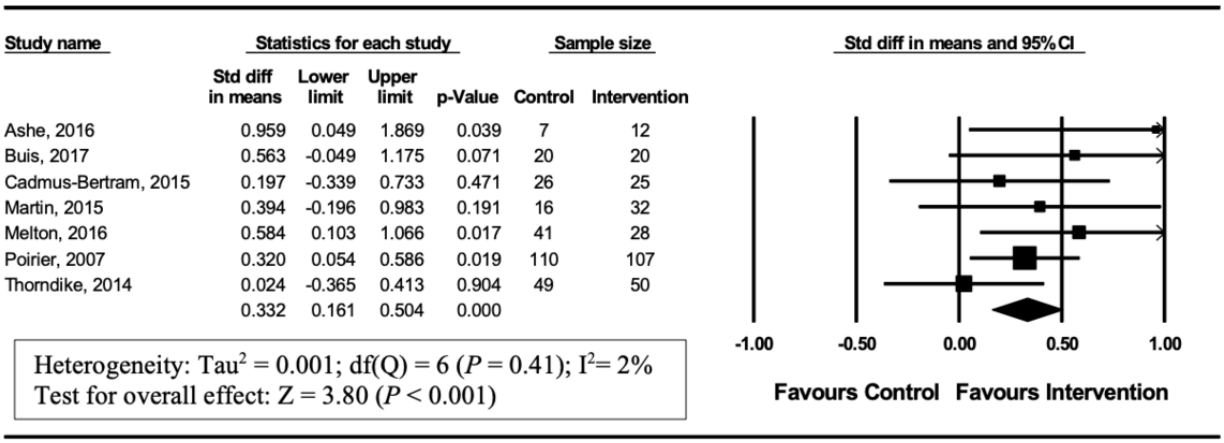
Figure 2. Forest plot of standardized mean difference (95% CI) in the effect of wearable trackers on physical activity.



Impact of Wearable Tracker Use in Studies With Steps/Day as Primary Outcome Variable

A total of 7 [20-22,26-28,31] of 12 studies utilized steps/day as the primary endpoint. In these studies, steps/day were significantly increased by the end intervention (SMD 0.332, 95% CI 0.16-0.50; $P<.001$). Heterogeneity in this analysis was low ($I^2=2\%$) (Figure 3).

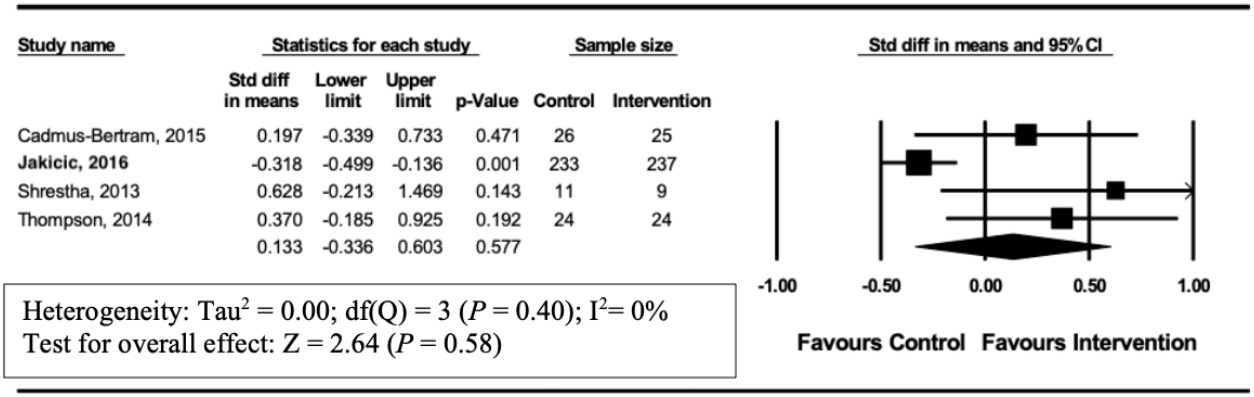
Figure 3. Forest plot of standardized mean difference (95% Cl) in the effect of wearable trackers on steps/day.



Impact of Wearable Tracker Use in Studies With Weight Loss as an Outcome

A total of 4 [22,25,29,30] of 12 studies reported weight change. However, no significant effect on weight change was observed (SMD 0.133, 95% CI –0.34 to 0.60, P=.58). Heterogeneity in this analysis was low ($I^2=0\%$) (Figure 4).

Figure 4. Forest plot of standardized mean difference (95% Cl) in the effect of wearable trackers on weight loss for intervention and control group.



Impact of Wearable Tracker Use in Studies on Overweight Reduction

A total of 5 [21,22,25,26,29] of 12 studies reported physical activity outcomes in overweight adults. In these studies, no significant increase in physical activity occurred (SMD 0.225, 95% CI –0.23 to 0.68, *P*=.33). Heterogeneity in this analysis was high (I^2=76%) (Figure 5). Seven [20,23,24,27,28,30,31] out of 12 studies reported physical activity outcomes in healthy adults unselected by weight. In these studies, a significant increase in physical activity was observed (SMD 0.594, 95% CI 0.10-1.09; *P*=.02) Heterogeneity in this analysis was high (I^2=90%; Figure 6).

Figure 5. Forest plot of standardized mean difference (95% CI) in the effect of wearable trackers on physical activity in overweight adults.

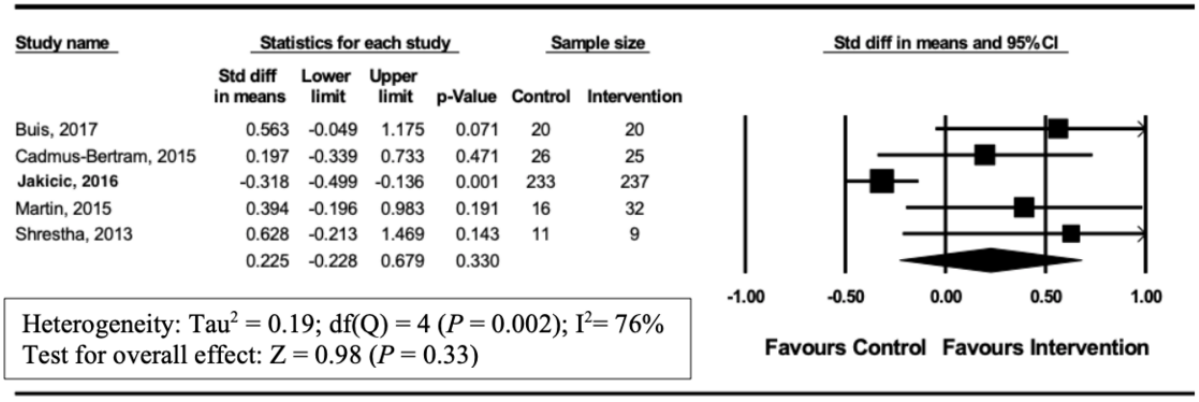
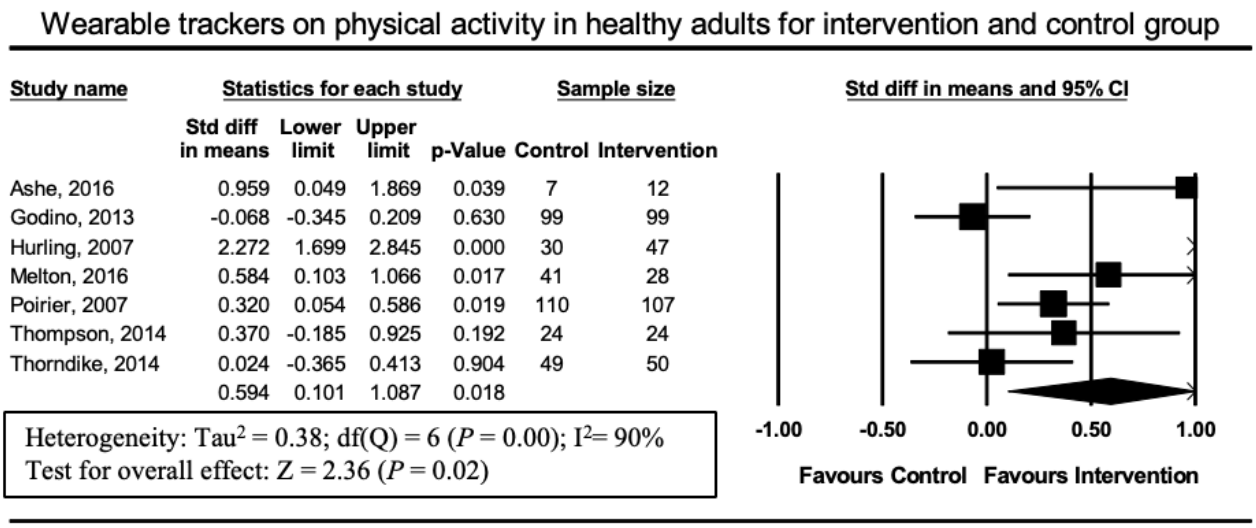


Figure 6. Forest plot of standardized mean difference (95% CI) in the effect of wearable trackers on healthy adults.



Synthesis of Results

A GRADE summary of results is presented in Figure 7. We assessed outcomes using the Cochrane GRADE approach. Results were downgraded when there was serious risk of bias, inconsistency, indirectness, imprecision, upgrading of a large effect size, or a dose-response gradient, all of which are possible confounding effects. Such confounding effects may create the appearance of an effect when there is none or reduce the appearance of an effect [33]. A GRADE Summary of Evidence table is provided in Multimedia Appendix 4.

Figure 7. GRADE Working Group grades of evidence summary.

Outcomes	Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Anticipated absolute effects* (95% CI) Comments
Effect on Physical Activity	-	1693 (12 RCTs)	⊕⊕⊕◯ MODERATE [a]	Five studies revealed significant differences on physical activity between intervention group and control group and seven studies showed no statistical significance.
Effect on Steps/Day	-	543 (7 RCTs)	⊕⊕⊕◯ MODERATE	Three studies revealed significant differences on steps/day between intervention group and control group and four studies showed no statistical significance.
Effect on Weight Loss	-	589 (4 RCTs)	⊕⊕◯◯ LOW [a,b]	One study revealed statistically significant (*P*= 0.001) benefit of physical activity on overweight population. Three studies showed no statistical significance.
Effect of Physical Activity on Overweight Population	-	629 (5 RCTs)	⊕⊕◯◯ LOW [a,b]	One study revealed statistically significant (*P*= 0.001) benefit of physical activity on overweight population. Four studies showed no statistical significance.
Effect of Physical Activity on Healthy Adult	-	727 (7 RCTs)	⊕⊕◯◯ LOW [a,b]	Four studies revealed significant differences in healthy adults between intervention group and control group and three studies showed no statistical significance.

Explanations
a. The blinding of physical activity interventions was highly variable and hence, not possible due to the nature of the intervention.
b. Effect size was used due to the various quality variable used.

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

Risk of Bias Across Results

Risk of bias (ROB) is measured using the Cochrane risk of bias tool [13] for randomized controlled trials, and the summary of the evaluation is shown in Table 3. Selection bias was low as randomization was considered high in all of the studies. In 2/12 studies [20,28], methods for allocation concealment were described in insufficient details resulting in high ROB, and one [21] of the studies showed an unclear ROB. For the outcome of physical activity, the blinding of participants and personnel was highly variable as there are challenges to blinding physical activity interventions.

Most of the trials (9/12, 58%) [20-25,27,28,30] did not provide sufficient methodological detail to judge bias not covered within other domains mentioned and were given judged to have an unclear ROB. The remaining studies (25%) provided sufficient details and judged to be low ROB [26,29,31].

Reporting bias was judged to be at low ROB because most of the trials (9/12, 75%) [21-26,29-31] reported details of the measured outcomes that were sufficient.

Detection bias was judged to be low since most of the trials (10/12, 83%) [20-30] provided sufficient information regarding outcome blinding assessment, and the remaining trials (16%) provided insufficient information.

Attrition bias was assessed to be low since all the trials reported the numbers reported to each group. The majority of trials (11/12, 92%) [20,21,23-31] included information on attrition and excisions from the analysis. One trial [22] did not disclose the reason for attrition/exclusion in sufficient detail, resulting in a judgment of high ROB.

Of the 12 studies measuring physical activity as an outcome, only 2 were judged to be of high ROB in terms of performance bias due to lack of blinding, and 1 study showed inadequate information regarding blinding.

Overall ROB was assessed, and the majority of studies (9/12, 75%) [21,23-30] were judged to be at low ROB and the remaining studies [20,22,31] were judged at unclear ROB.

Publication bias was determined by visual inspection of funnel plots comparing physical activity against effect size. There was visual evidence of publication bias with at least two studies falling outside the range of expected precision for their effect size (Multimedia Appendix 3).

Table 3. Risk of bias (Cochrane Critical Appraisal Skills Program Tool[a]).

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total score/7	Quality Score, %
Ashe, 2016 [20]	L[b]	L	H[c]	U[d]	H	L	L	4.5	64
Buis, 2017 [21]	L	L	U	U	U	L	L	5.5	79
Cadmus-Bertram, 2015 [22]	L	L	L	U	U	U	H	4.5	64
Godino, 2013 [23]	L	L	L	U	L	L	L	6.5	93
Hurling, 2007 [24]	L	L	L	U	H	L	L	5.5	79
Jakicic, 2016 [25]	L	L	L	U	L	L	L	6.5	93
Martin, 2015 [26]	L	L	L	U	H	L	L	5.5	79
Melton, 2016 [27]	L	L	L	L	U	L	L	6.5	93
Poirier, 2007 [28]	L	L	H	U	L	L	L	5.5	79
Shrestha, 2013 [29]	L	H	L	L	L	L	L	6	86
Thorndike, 2014 [31]	L	U	L	L	H	U	L	5	71
Thompson, 2014 [30]	L	H	L	U	L	L	L	6	86
Category Score (%)	100	88	79	45	54	88	100		

[a]Cochrane risk of bias tool. Q1: Were there selection bias (biased allocation to interventions) due to inadequate generation of a randomized sequence? Q2: Were there selective bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment? Q3: Were there reporting bias due to selective outcome report? Q4: Were there bias due to problems not covered elsewhere in the table? Q5: Were there performance bias due to knowledge of the allocated interventions by participants and personnel during the study? Q6: Were there detection bias due to knowledge of the allocated interventions by outcome assessors? Q7: Were there attribution bias due to amount, nature, or handling or incomplete outcome data?

[b]L: low risk.

[c]H: high risk.

[d]U: unclear risk.

Discussion

This meta-analysis examined the effects of wearable trackers on physical activity and is based on 12 randomized controlled trials involving 1693 participants. Overall, wearable tracker usage was associated with improvements in physical activity (SMD 0.594, 95% CI 0.10-1.09; *P*=.018). Interventions that included a consumer-based wearable tracker demonstrated an improvement in physical activity as compared to control groups, especially with daily steps. No clear evidence of benefit was seen from wearable tracker use in the endpoints of weight reduction or physical activity of overweight populations. Indeed, a recent review showed a potential increase in physical activity but no evidence for its effectiveness in weight loss [34].

Public Health Implications

This data may have both individual and public health implications. Wearable trackers can support continuous health monitoring at both the individual and the population level. Wearable trackers are activity monitoring tools that help to engage patients as advocates in their personalized care and have been proposed to encourage healthy behavior. Benefits are thought to include prevention or reduction of health problems, support of chronic disease self-management, enhanced provider knowledge, reduced number of healthcare visits, and personalized, localized, and on-demand interventions in ways not previously possible [35]. The low cost of delivery and the feasibility of wearable trackers makes them an attractive potential tool to facilitate self-monitoring of physical activity. The data presented in this meta-analysis demonstrates that wearable tracker usage was associated with short-term gains in physical activity [36].

A wide range of wearable trackers was used. Four [20,22,30,31] of 12 studies used Fitbit, two [21,27] used Jawbone Up, and the remaining studies used a variety of other wearable trackers. The relative proportion of commercial wearable trackers in the included studies is similar to global market shares, with Fitbit having the largest market share (20%) and hence applicable to the real world [37]. The Fitbit and Jawbone Up used in this meta-analysis have the same selected measures of “steps, distance, calories, and sleep” and are worn on the wrist [21,22,27,30,31]. A systematic review assessing the validity and reliability of Fitbit and Jawbone found that the validity and inter-device reliability of steps counts was generally high [10].

Study Limitations

This study has several limitations. First, most studies included in this meta-analysis included small study sizes with short

intervention durations and limited follow-up, highlighting the necessity for longer-term studies. The variety of study designs may have been reflected by the statistical heterogeneity of the outcomes. Hence, it might be challenging to generalize the results due to heterogeneity. Future research should explore the long-term effectiveness of wearable trackers in increasing physical activity. Second, statistical estimates of publication bias identified individual, small studies with relatively large effect sizes, which may be a reflection of a file drawer effect. Third, it is difficult to establish the independent contribution of adjunctive interventions (eg, behavioral counseling, interactive health coach, weekly reminders, and text messages) that were often offered alongside wearable tracker usage. The last study was obtained in 2017, since which time there have been changes in market share due to the volatile nature of the industry. The entrance of new brands like Apple, Xiaomi, and Samsung means that an updated review is needed. Next, the skewed demographics to white females would mean that the translation to other demographics might be limited. Lastly, we restricted our search to only full-text published articles in the English language and, thus, may have excluded relevant studies outside our current scope.

Recommendations for Further Research

The current study examined the utility of wearable activity trackers in healthy adults and found a modest but measurable benefit in the population studied. Important questions for future research could include the identification of appropriate deployment strategies for these novel technologies in cardiac rehabilitation, aged care, and youth. A related question that we could not address was the role of social engagement in modulating the response of participants to wearable tracking devices.

Recommendations for Clinical Practice

The current study provides qualified support for the use of wearable activity trackers in healthy populations, showing evidence of short-term gains in physical activity, but not weight. As technology advances and these devices improve over time, future studies will be necessary to delineate the optimal use case for these devices.

Conclusion

In conclusion, this meta-analysis demonstrates the efficacy of wearable trackers in facilitating short-term increases in consumer physical activity. Future studies will be required to determine the durability of the influence of wearable tracker use on consumer physical activity behavior.

Acknowledgments

AG is supported by a Future Leader Fellowship of the National Heart Foundation of Australia (111086) and Prof Robyn A Clark Heart Foundation Future Leader Fellowship (APP ID. 100847).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 17 KB - mhealth_v8i7e15576_app1.docx]

Multimedia Appendix 2
PICOS criteria for inclusion and exclusion of studies.
[DOCX File , 13 KB - mhealth_v8i7e15576_app2.docx]

Multimedia Appendix 3
Funnel plot showing publication bias.
[DOCX File , 35 KB - mhealth_v8i7e15576_app3.docx]

Multimedia Appendix 4
GRADE Summary of Evidence Table.
[DOCX File , 16 KB - mhealth_v8i7e15576_app4.docx]

References

1. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Health Promot 2018 Dec 26;33(5):778-791. [doi: 10.1177/0890117118816278]
2. van der Bij AK, Laurant MGH, Wensing M. Effectiveness of physical activity interventions for older adults: a review. Am J Prev Med 2002 Feb;22(2):120-133. [doi: 10.1016/s0749-3797(01)00413-5] [Medline: 11818183]

3. Dias D, Paulo Silva Cunha J. Wearable Health Devices-Vital Sign Monitoring, Systems and Technologies. Sensors (Basel) 2018 Jul 25;18(8) [FREE Full text] [doi: 10.3390/s18082414] [Medline: 30044415]
4. Howlett NT, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 01;9(1):147-157 [FREE Full text] [doi: 10.1093/tbm/iby010] [Medline: 29506209]
5. Cradock KA, ÓLaighin G, Finucane FM, Gainforth HL, Quinlan LR, Ginis KAM. Behaviour change techniques targeting both diet and physical activity in type 2 diabetes: A systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Feb 08;14(1):18 [FREE Full text] [doi: 10.1186/s12966-016-0436-0] [Medline: 28178985]
6. Coughlin SS, Stewart J. Use of Consumer Wearable Devices to Promote Physical Activity: A Review of Health Intervention Studies. J Environ Health Sci 2016 Nov;2(6) [FREE Full text] [doi: 10.15436/2378-6841.16.1123] [Medline: 28428979]
7. Hillsdon MF, Foster C, Thorogood M. Interventions for promoting physical activity. Cochrane Database Syst Rev 2005 Jan 25(1):CD003180 [FREE Full text] [doi: 10.1002/14651858.CD003180.pub2] [Medline: 15674903]
8. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
9. Hickey AM, Freedson PS. Utility of Consumer Physical Activity Trackers as an Intervention Tool in Cardiovascular Disease Prevention and Treatment. Prog Cardiovasc Dis 2016;58(6):613-619. [doi: 10.1016/j.pcad.2016.02.006] [Medline: 26943981]
10. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
11. 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]
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(4):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
13. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928-d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
14. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
15. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive Meta-Analysis Version 3. Comprehensive Meta-Analysis Version 3. URL: https://www.meta-analysis.com/downloads/Meta-Analysis%20Manual%20V3.pdf [accessed 2017-07-13] [WebCite Cache ID 6ruZZzoFb]
16. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook. In: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 The Cochrane Collaboration. United States: John Wiley & Sons Ltd; 2011.
17. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
18. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14(1). [doi: 10.1186/s12913-014-0579-0]
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 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]
20. Ashe MC, Winters M, Hoppmann CA, Dawes MG, Gardiner PA, Giangregorio LM, et al. "Not just another walking program": Everyday Activity Supports You (EASY) model-a randomized pilot study for a parallel randomized controlled trial. Pilot Feasibility Stud 2015 Jan 12;1(1):4 [FREE Full text] [doi: 10.1186/2055-5784-1-4] [Medline: 27175291]
21. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
22. 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]
23. Godino JG, Watkinson C, Corder K, Marteau TM, Sutton S, Sharp SJ, et al. Impact of personalised feedback about physical activity on change in objectively measured physical activity (the FAB study): a randomised controlled trial. PLoS One 2013 Sep 16;8(9):e75398 [FREE Full text] [doi: 10.1371/journal.pone.0075398] [Medline: 24066178]

24. 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 Apr 27;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
25. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial. JAMA 2016 Sep 20;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
26. 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. JAHA 2015 Oct 29;4(11). [doi: 10.1161/jaha.115.002239]
27. Melton BF, Buman MP, Vogel RL, Harris BS, Bigham LE. Wearable Devices to Improve Physical Activity and Sleep. Journal of Black Studies 2016 Jul 27;47(6):610-625. [doi: 10.1177/0021934716653349]
28. 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 Feb 09;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]
29. Shrestha M, Combest T, Fonda SJ, Alfonso A, Guerrero A. Effect of an accelerometer on body weight and fitness in overweight and obese active duty soldiers. Mil Med 2013 Jan;178(1):82-87. [doi: 10.7205/milmed-d-12-00275] [Medline: 23356124]
30. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. Arch Gerontol Geriatr 2014 May;58(3):314-319. [doi: 10.1016/j.archger.2014.01.004] [Medline: 24485546]
31. Thorndike AN, Mills S, Sonnenberg L, Palakshappa D, Gao T, Pau CT, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. PLoS One 2014 Jun 20;9(6):e100251 [FREE Full text] [doi: 10.1371/journal.pone.0100251] [Medline: 24950218]
32. Handbook C. Undertaking subgroup analyses. United States: John Wiley & Sons Ltd; 2019.
33. Ryan R, Hill S. How to GRADE. How to GRADE the quality of the evidence Cochrane Consumers and Communication 2018 Jul 25 [FREE Full text] [doi: 10.26181/5b57d95632a2c]
34. de Vries HJ, Kooiman TJ, 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 26;24(10):2078-2091 [FREE Full text] [doi: 10.1002/oby.21619] [Medline: 27670401]
35. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13(1):77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
36. 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 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
37. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]

Abbreviations

GRADE: Grades of Recommendation, Assessment, Development, and Evaluation
MVPA: moderate-vigorous physical activity
RCT: randomized controlled trial
ROB: risk of bias
SMD: standardized mean difference

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by A Henriksen, A Rickman, L Toth, S Haghayegh; comments to author 30.09.19; revised version received 18.11.19; accepted 14.05.20; published 22.07.20.

Please cite as:
Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN
Effectiveness of Wearable Trackers on Physical Activity in Healthy Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2020;8(7):e15576
URL: http://mhealth.jmir.org/2020/7/e15576/
doi:10.2196/15576
PMID:32706685

Original Paper

Deep Learning Approach for Imputation of Missing Values in Actigraphy Data: Algorithm Development Study

Jong-Hwan Jang[1*], BA; Junggu Choi[1*], BE; Hyun Woong Roh[2], MD; Sang Joon Son[3], MD, PhD; Chang Hyung Hong[3], MD, PhD; Eun Young Kim[1,4], PhD; Tae Young Kim[1], BE; Dukyong Yoon[1,4], MD, PhD

[1]Department of Biomedical Informatics, School of Medicine, Ajou University, Suwon, Gyeonggi-do, Republic of Korea

[2]Department of Brain Science, School of Medicine, Ajou University, Suwon, Gyeonggi-do, Republic of Korea

[3]Department of Psychiatry, School of Medicine, Ajou University, Suwon, Gyeonggi-do, Republic of Korea

[4]Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, Gyeonggi-do, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Dukyong Yoon, MD, PhD
Department of Biomedical Informatics
School of Medicine
Ajou University
World cup-ro 206, Yeongtong-gu
Suwon, Gyeonggi-do, 16499
Republic of Korea
Phone: 82 031 219 4476
Email: yoon8302@gmail.com

Abstract

Background: Data collected by an actigraphy device worn on the wrist or waist can provide objective measurements for studies related to physical activity; however, some data may contain intervals where values are missing. In previous studies, statistical methods have been applied to impute missing values on the basis of statistical assumptions. Deep learning algorithms, however, can learn features from the data without any such assumptions and may outperform previous approaches in imputation tasks.

Objective: The aim of this study was to impute missing values in data using a deep learning approach.

Methods: To develop an imputation model for missing values in accelerometer-based actigraphy data, a denoising convolutional autoencoder was adopted. We trained and tested our deep learning–based imputation model with the National Health and Nutrition Examination Survey data set and validated it with the external Korea National Health and Nutrition Examination Survey and the Korean Chronic Cerebrovascular Disease Oriented Biobank data sets which consist of daily records measuring activity counts. The partial root mean square error and partial mean absolute error of the imputed intervals (partial RMSE and partial MAE, respectively) were calculated using our deep learning–based imputation model (zero-inflated denoising convolutional autoencoder) as well as using other approaches (mean imputation, zero-inflated Poisson regression, and Bayesian regression).

Results: The zero-inflated denoising convolutional autoencoder exhibited a partial RMSE of 839.3 counts and partial MAE of 431.1 counts, whereas mean imputation achieved a partial RMSE of 1053.2 counts and partial MAE of 545.4 counts, the zero-inflated Poisson regression model achieved a partial RMSE of 1255.6 counts and partial MAE of 508.6 counts, and Bayesian regression achieved a partial RMSE of 924.5 counts and partial MAE of 605.8 counts.

Conclusions: Our deep learning–based imputation model performed better than the other methods when imputing missing values in actigraphy data.

(JMIR Mhealth Uhealth 2020;8(7):e16113) doi:10.2196/16113

KEYWORDS

accelerometer; actigraphy; imputation; autoencoder; deep learning

Introduction

An accelerometer-based actigraphy device can measure the movement of the person wearing the device by capturing acceleration in a single axis or in multiple axes of motion. By mounting an actigraphy device on the wrist, ankle, or waist, researchers or physicians can measure the amount of movement or movement patterns. One study [1] has used this tool to detect changes in activity resulting from neurological diseases, including Parkinson disease, depression, dementia, and affective disorder. In addition to its utility for certain diseases, this device also has been used in the analysis of sleeping patterns or activity recognition (such as running, walking, or sitting) in healthy participants [2,3]. Furthermore, because accelerometer-based data are objective whereas self-reported data are subjective and can be biased by participant recall, accelerometers and actigraphy devices have increasingly been used in clinical trials to collect real-world data [4,5].

Because data recorded over a few days to a few months are usually required to be able to analyze activity patterns, studies have frequently faced the issue of participant adherence to use [6]. Many indices of activity such as interdaily stability, intradaily variability, or relative amplitude require full records over a day, so even a small number of missing values hinders the use of the records captured on that day [7]. In one previous study, the researchers collected 17,542 day-long activity records, but only 2003 (11.41%) records could be used in their study because of missing data [6].

To overcome this issue, statistical models have been suggested to impute missing data [8,9]; however, these models require assumptions about the distribution of the missing values (such as missing at random, missing completely at random, or not missing at random assumptions) and using a specific assumption may not be appropriate for a given study.

Recent studies [10-12] have confirmed that deep learning models which are able to learn features from raw input without manual feature engineering or statistical assumptions perform well in many tasks, including data denoising, detecting and classifying objects in images, and predicting values. An autoencoder is a deep-learning method that can extract key features from input data and restore them; therefore, an autoencoder can also be used for reconstructing original data from corrupted data [13]. This approach has been used to impute missing values in traffic data, which are similar to actigraphy data—both data sets have time-series characteristics [14]. Furthermore, another study [15] showed that the autoencoder-based approach can extract core features from activity data. These findings suggested the possibility of improving the performance of imputation in activity data using a deep-learning model. In contrast to statistical approaches, the deep-learning method can discover hidden patterns from the data themselves without any specific assumptions about the missing values. Assuming that missing values could be treated as a type of noise in the data, we developed an autoencoder-based model to restore missing values in actigraphy data. We used three actigraphy data sets and compared imputed data from our model with those of other methods.

Methods

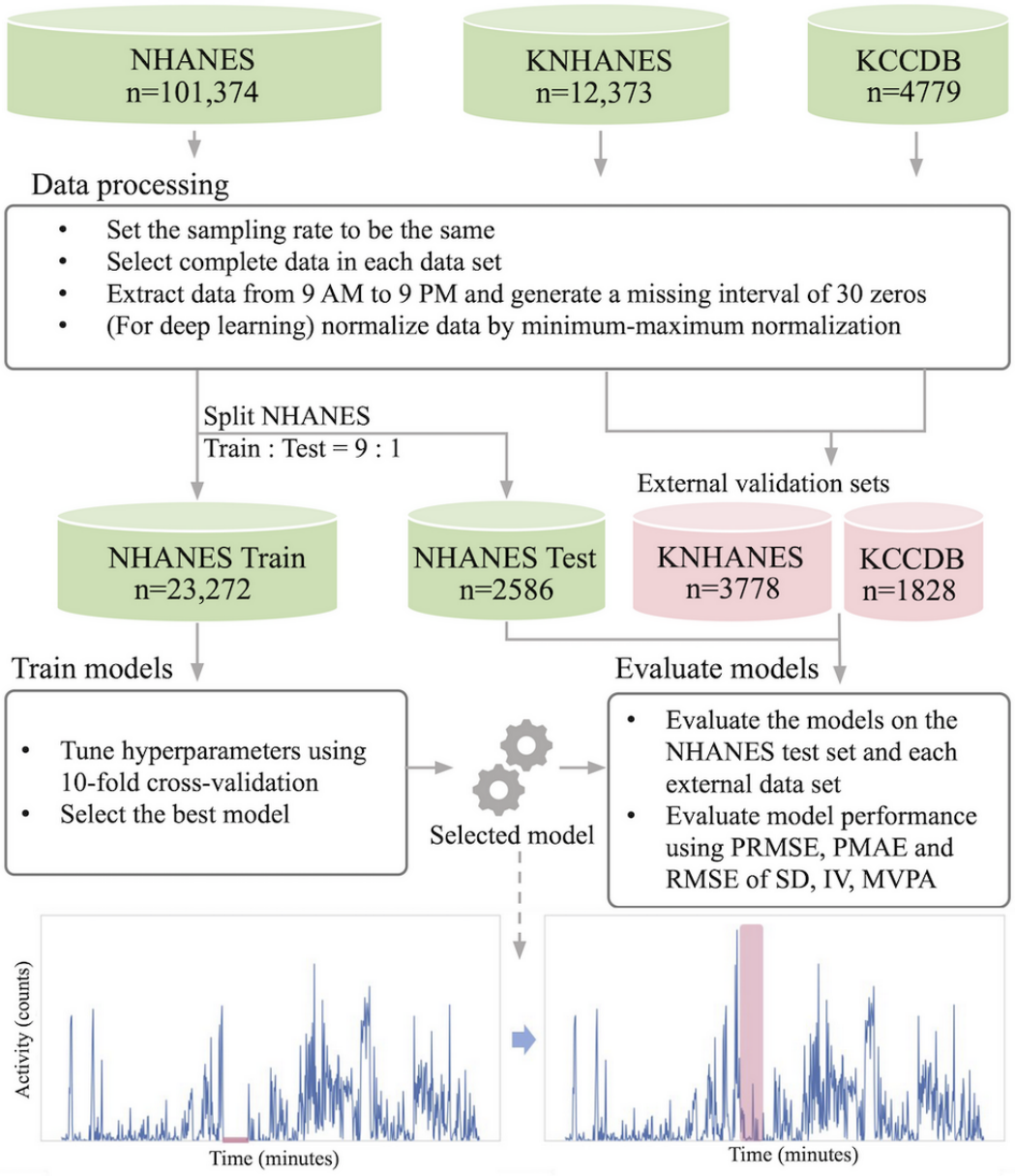
Ethics

This study was approved by the Ajou University Hospital institutional review board (AJIRB-MED-EXP-17-470). Before collecting actigraphy data from patients at Ajou University Hospital, we received their informed consent (Korean Chronic Cerebrovascular Disease Oriented Biobank, KCCDB). The other two data sets used in this study, the National Health and Nutrition Examination Survey (NHANES) [16,17] and Korea National Health and Nutrition Examination Survey (KNHANES) [18-21], are publicly available. All data were deidentified and used only for retrospective study.

Overview

This study consisted of four phases. First, we collected activity data from three different actigraphy data sets. Second, complete data for each day were selected from the data sets and preprocessed to generate artificially corrupted data. Third, models were constructed from the data, including our deep learning–based imputation model. Finally, models were evaluated with performance indices. Details of the procedures are shown in Figure 1.

Figure 1. Overview of the study and data where n indicates number of records (days). IV: intradaily variability; KCCDB: Korean Chronic Cerebrovascular Disease Oriented Biobank; KNHANES: Korea National Health and Nutrition Examination Survey; MVPA: moderate-to-vigorous physical activity; NHANES: National Health and Nutrition Examination Survey; PRMSE: partial root mean square error; PMAE: partial mean absolute error; RMSE: root mean square error.



Data Sources

Three actigraphy data sets were used for this study—NHANES, accelerometer-based actigraphy data collected over four years (2003-2006) from 14,482 individuals living in the United States; KNHANES, accelerometer-based actigraphy data collected over two years (2014-2015) from 1768 people living in South Korea; and KCCDB, accelerometer-based actigraphy data were collected over two years (2014-2015) from 177 patients who had visited Ajou University Hospital for evaluation or treatment of cerebrovascular disease.

The NHANES data set was collected using a uniaxial accelerometer-based actigraphy device (ActiGraph AM-7164, ActiGraph LLC) which gathered only z-axis data. The KNHANES data set, though collected using a triaxial accelerometer-based actigraphy device (ActiGraph GTX3+, ActiGraph LLC), also consisted of only z-axis data (only these data were made available to the public). In contrast, the KCCDB data set was collected using a triaxial accelerometer-based actigraphy device (Fitmeter, Fit.Life Inc). The triaxial data were aggregated into a single magnitude vector using the formula (Multimedia Appendix 1). In this study, data from each device are used in their original format and values from one device are not comparable to those from others since activity units, device type, and device range (±g) vary [22]. We developed the autoencoder-based imputation model with 90% of the NHANES data set (the training data set); the performance of the model was evaluated using the other 10% of the NHANES data set (test data set) and was externally validated using the KNHANES data set. The KCCDB data set was used to test whether the

model, trained with uniaxial data, could also be used with triaxial data.

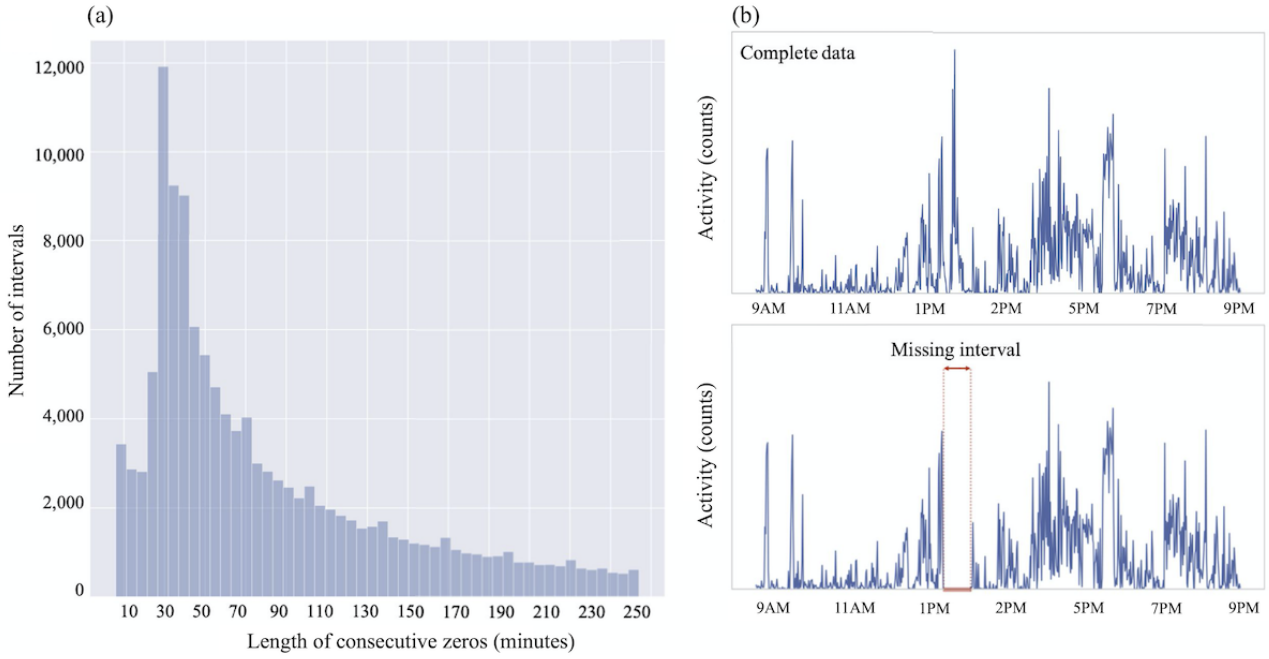
Data Processing

First, because the KCCDB data set has a higher sampling rate (0.1 Hz) than those of the other data sets (0.016 Hz), the KCCDB data were downsampled by averaging the values for each minute. As a result, each record consisted of 1440 values per day.

In previous studies, a missing interval has been defined as a period (of 20, 30, or 60 minutes, depending on the study) over which zero values are continuously repeated [23-25]. For this study, we defined a missing interval to be 30 minutes or more of consecutive zeros. We found that intervals of such lengths occurred most frequently in the NHANES data set (Figure 2). We found the frequency of missing intervals less than 30 minutes was the highest in both datasets (Multimedia Appendix 2). Only complete records from the data sets were used so that we could calculate the reconstruction error for the missing intervals by comparing the imputed values with the original values.

Finally, we extracted the data from 9 AM to 9 PM, which is the most active period for humans [23-25], and overwrote a random 30-minute interval with zeros in each record. Consequently, each record used in the study had 720 values (over the 12-hour period from 9 AM to 9 PM), in which 30 consecutive zero values were included as a missing interval. Moreover, we conducted additional experiments with 90 and 180 consecutive zero values to confirm the imputation robustness for longer missing intervals. Preprocessing algorithms were written in Python (version 3.6; scikit-learn).

Figure 2. (a) Frequency of missing data intervals found in the NHANES data set. The interval of approximately 30 minutes occurred most frequently. (b) Example of a complete data record and of a record with missing data interval.



Models for Imputation

Mean Imputation

Mean imputation is a method of replacing a missing value with the mean value from the other instances of valid data at that time [26]. For example, if there is a value missing at 12:30 PM in a certain record, the missing value is replaced with the mean of the values from 12:30 PM in the other records. Mean imputation, algorithms were written in Python.

Zero-Inflated Poisson Regression Model

To compensate for the disadvantage of a single imputation method—where missing values are replaced with a single value—the multiple imputation method generates several data sets and the results are combined into a single result to replace the missing values. In this study, the multiple imputation by chained equation approach (also called fully conditional specification) was used for multiple imputation. (1) All missing values of each variable were filled using the mean imputation method. (2) The regression equation, in which the dependent variable is the variable to be imputed and the independent variables are the other variables surrounding the dependent variables, was developed. Then, missing values were replaced by a value estimated by the regression equation [27]. After changing the dependent and independent variables, the process was repeated. The zero-inflated Poisson model has previously been used as a regression model for activity data in which a large portion of the data had zero values [28]; similar to the previous study, the lead and lag options were added to reflect the values of the activity before and after the missing variable. In this study, we applied the zero-inflated Poisson regression model using R (3.3.1) and a package (accelmissing) developed in a previous study [28,29].

Bayesian Linear Regression Model

The process was the same as the zero-inflated Poisson model, but Bayesian multiple imputation utilized Bayesian linear regression. Because the Bayesian model aims to find the parameter for the posterior distribution and take a sample from the estimated distribution, the imputed values of this approach can be negative, which cannot exist for the units defined for these devices. Hence, the negative values were replaced with zero values. Bayesian regression algorithms were written in Python.

Zero-Inflated Denoising Convolutional Autoencoder (New Method)

An autoencoder is an unsupervised deep-learning method. Its aim is to make its output X' approximate the input X. An autoencoder consists of an encoder that compresses the information of the input into a compressed-information (or hidden state) Z and a decoder that restores Z back to X as closely as possible:

Because Z has fewer dimensions than the input, and because the decoder must restore a value close to the input value by utilizing the information in Z, the autoencoder must learn the key information of the input in the hidden state in order to reduce the reconstruction error.

A denoising autoencoder is a type of autoencoder that restores noisy input X_{noised}, which contains masking noise, to the denoised output X' [30]. Here, noise refers to data that distort the original data. The zero values were treated as noise distorting the data in this study.

A convolutional neural network is a deep-learning method commonly adopted for analyzing images [31]. A convolutional neural network can extract the information from a specific area of the data by applying filters to the input data. This approach compensates for the disadvantage—that of being unable to utilize positional information—found in simple neural networks. One-dimensional sequential data can be analyzed using a one-dimensional convolutional neural network; studies on this approach have been actively conducted in various fields such as voice synthesis or biosignal analysis [32,33].

Our model, the zero-inflated denoising convolutional autoencoder, consisted of an autoencoder that encoded and decoded the data using a convolutional neural network and a unique activation function designed for the zero distribution at the last layer. The zero-inflated denoising convolutional autoencoder received corrupted data as input, then compressed and recovered these data using a convolutional autoencoder, as shown in Figure 3.

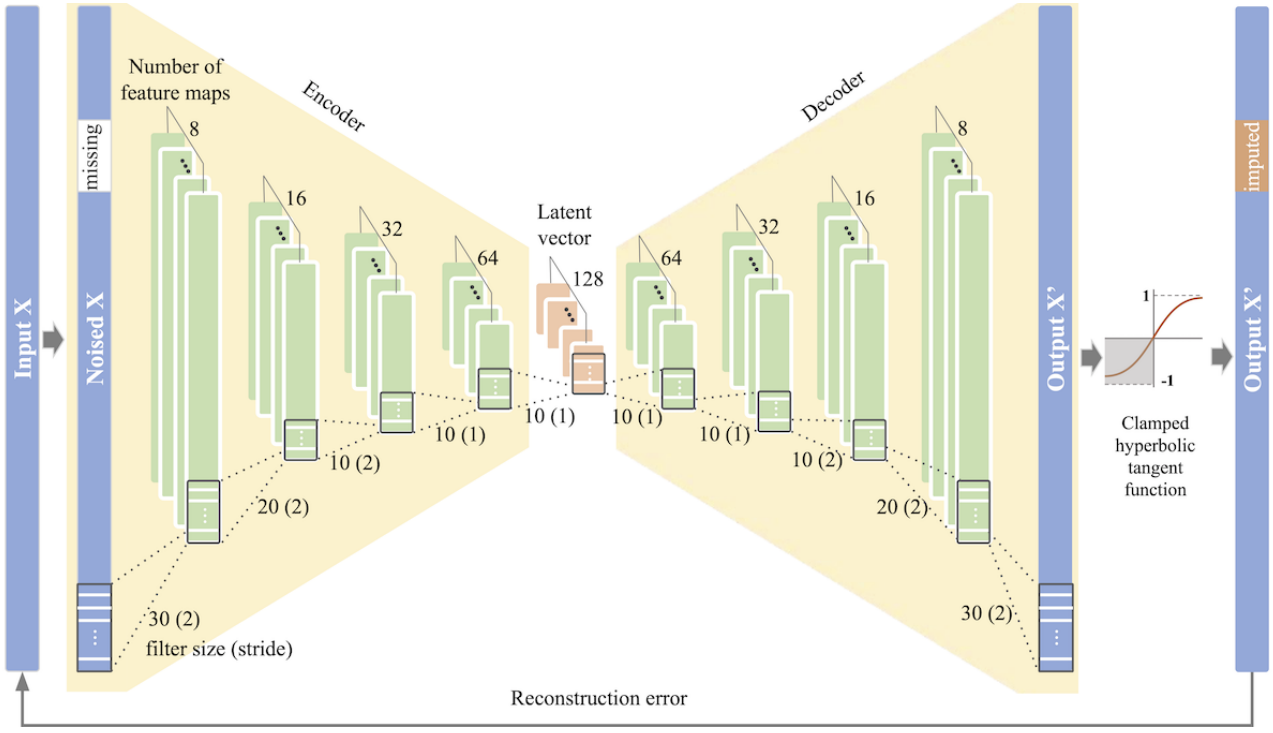
We used the PyTorch (version 1.4.1) framework to construct the deep learning model [34]. Input data for zero-inflated denoising convolutional autoencoder were normalized (minimum-to-maximum normalization); then, each zero in the missing interval was replaced with a value of 0.5. As the input data were encoded, the filters of each convolutional neural network layer in the encoder compressed the regional information into a feature map. The size of the feature map decreased and the number of feature maps increased every time data passed through the convolutional neural network layer so that latent vectors with various perspectives could be extracted. After the corrupted data had been encoded in latent vectors, transconvolution layers were applied symmetrically to decode the latent vectors using the same hyperparameters used in the encoder. The size of the output was set to be equal to the size of the input data at the end of the decoder layer. Batch normalization and a hyperbolic tangent function were applied at the end of each layer. The clamped hyperbolic tangent was applied to the output of the decoder, which limited the output to values in the range of –1 to 1 and converted negative values to zeros. This caused the model to treat the distribution of negative values as a distribution of zeros. Finally, the recovered data, which were as close to the uncorrupted data as possible, were output.

Hyperparameters, filter size (m=20 and 30) and size of the latent vector (k=40, 60, and 80), were determined by a grid search. The other hyperparameters (the number of layers, q, and number of filters, n) were gradually increased and the best-performing values were selected. Stride was set adaptively according to the latent vector size.

For training and selecting the final hyperparameter settings, the training data set was used. The hyperparameters of the zero-inflated denoising convolutional autoencoder were tuned using 10-fold cross-validation. Hyperparameters exhibiting the lowest root mean square error (RMSE) were selected as the final hyperparameter settings.

When training the zero-inflated denoising convolutional autoencoder model, we used RMSE as a loss function to provide feedback to modify the weights. The loss between the original data and restored data was calculated for the entire record (imputed and original parts). Applying the loss function to all values allowed the zero-inflated denoising convolutional autoencoder to learn to impute the missing intervals and to reconstruct the other observed parts.

Figure 3. Model architecture for zero-inflated denoising convolutional autoencoder consisting of encoder with 5 convolutional layers in which the filter size and stride decrease and the number of feature maps increases and decoder with 5 transconvolutional layers in which hyperparameters are symmetrically the same as those used in the encoder.



Performance Evaluation

Metrics

Performance of the imputation methods were evaluated using two metrics on the imputed portion of the outputs. Partial RMSE

and partial MAE

where

were calculated. In the equations, i and j represent the ith data record in a data set and the jth value in a data record, respectively. I_{ij} was a binary variable that represented whether a variable was observed or missing, M was the number of records in the data set, N was the length of the data record, X_{ij} indicated an original value, and X'_{ij} indicated its imputed value. Lower values of partial RMSE and partial MAE indicated better imputation.

We calculated the standard deviation

and the intradaily variability index

where N was the total number of values in the data record, X_j was the jth value in the data record, and  was the mean of the values in that record.

Intradaily variability index is a nonparametric index of the circadian rhythm and represents the fragmentation of activity. Index values range from 0 to 2, and higher values indicate higher variability.

To evaluate reconstruction of the original variability in the data, we calculated RMSE of the standard deviation,

and RMSE of the intradaily variability,

where M was the total number of records in a data set, SD_i^{origin} was the standard deviation of the ith original record, and $SD_i^{imputed}$ was the standard deviation of the ith imputed record.

We also evaluated the RMSE in the missing intervals of a moderate-to-vigorous physical activity measure that was derived from the data. Moderate-to-vigorous physical was used to represent activity intensity. We applied a cutoff=1267 counts [35] to uniaxial data in NHANES and KNHANES and applied a cutoff=2691 counts [36], which was designed for a single magnitude vector, to the KCCDB data set. We measured the duration of moderate-to-vigorous physical activity using the 10-minute bout method, which identifies 10 minutes of moderate-to-vigorous physical activity if over 80% of the values are greater than the cutoff point [25]. We evaluated the

moderate-to-vigorous physical activity measure from original and imputed data using the RMSE,

where, unlike the RMSE of intradaily variability or standard deviation, the duration of moderate-to-vigorous physical activity was calculated for only missing intervals. *M* is the total number of records in a data set, denotes the moderate-to-vigorous physical activity for the original values in the missing interval of *i*th data record, and $MVPA_i^{imputed}$ denotes the moderate-to-vigorous physical activity for the imputed values in the missing interval of the *i*th data record.

Mean Imputation

The mean value at each minute was determined from the NHANES training set. For external validation, a model was constructed and evaluated for each data set without dividing them into training and testing sets.

Zero-Inflated Poisson Regression and Bayesian Regression Imputation

The multiple imputation-based models were constructed without dividing the data set into training, validation, and testing data sets. These models utilize the expectation–maximization algorithm, which requires the entire data set and fills in the missing values with values inferred from existing values in other data records. After imputation was performed, the performance was estimated using the NHANES test data set only. External validation was performed using the same process that was used to evaluate mean imputation.

Zero-Inflated Denoising Convolutional Autoencoder Imputation

The performance of zero-inflated denoising convolutional autoencoder was evaluated using the model generated by the training and validation sets of NHANES to impute the test set. For external validation, the performance was evaluated by applying the model trained on the NHANES data set to the external validation data sets without retraining.

Results

To visualize the data and results, both Python (matplotlib, version 3.0.0; seaborn, version 0.9.0) and R (ggplot2) were used [37-39]. After preprocessing, the NHANES data set comprised 25,858 days of data from 9236 persons, the KNHANES data set comprised 3778 days of data from 1301 persons, and the KCCDB data set comprised 1829 days of data from 169 persons (Table 1).

After 10-fold cross-validation for hyperparameter tuning, we selected the hyperparameter conditions with the lowest mean RMSE (Table 2). Detailed results of 10-fold cross-validation are described in Multimedia Appendix 3.

We used the zero-inflated denoising convolutional autoencoder model with a filter length of 30 in the first convolutional layer and a latent vector size of 60 in the subsequent experiments in this study (q=10, n=8, m=30, k=60, Table 3).

Table 1. Baseline characteristics of the data sets.

Characteristics	NHANES[a] (n=12,475)	KNHANES[b] (n=1768)	KCCDB[c] (n=177)
Age (years), mean (SD)	39.04 (22.27)	42.88 (13.04)	74.07 (7.05)
Gender, n (%)			
Male	6077 (48.71)	662 (37.44)	56 (31.63)
Female	6398 (51.28)	1106 (62.55)	121 (68.36)
Weight (kg), mean (SD)	75.26 (21.73)	63.35 (11.97)	59.03 (10.04)
Height (cm), mean (SD)	166.01 (11.72)	163.45 (8.55)	156.96 (8.33)
BMI (kg/m^2), mean (SD)	27.03 (6.56)	23.62 (3.48)	22.66 (7.19)
Activity (count), mean (SD)	344 (694.23)	433 (586.78)	637 (1121.27)

[a]NHANES: National Health and Nutrition Examination Survey data set (device: ActiGraph AM-7164; type: uniaxial; sample rate: 0.016 Hz).

[b]KNHANES: Korea National Health and Nutrition Examination Survey data set (device: ActiGraph GTX3+; type: triaxial; sample rate: 0.016 Hz).

[c]KCCDB: Korean Chronic Cerebrovascular Disease Oriented Biobank data set (device: Fitmeter; type: triaxial; sample rate: 0.01 Hz).

Table 2. Result of 10-fold cross-validation.

Experiment	Hyperparameters		Result
	Latent vector size, k	Filter size, m	RMSE[a] (count), mean
1	40	20	830.5
2	40	30	838.0
3	60	20	858.7
4	60	30	788.4
5	80	20	825.1
6	80	30	831.0

[a]RMSE: root mean square error.

Table 3. Hyperparameter settings for the zero-inflated denoising convolutional autoencoder.

Component layer (order of layer)	Filter, n[a] ×size (stride)	Feature map output size, n × size
Input		1×720
Encoder		
Convolution (1)	8×30[b] (2)	8×346
Convolution (2)	16×20 (2)	16×164
Convolution (3)	32×10 (2)	32×78
Convolution (4)	64×10 (1)	64×69
Convolution (5)	128×10 (1)	128×60[c]
Decoder		
Transconvolution (6)	64×10 (1)	64×69
Transconvolution (7)	32×10 (1)	32×78
Transconvolution (8)	16×10 (2)	16×164
Transconvolution (9)	8×20 (2)	8 ×346
Transconvolution (10[d])	1 ×30 (2)	1×720

[a]number of filters, n.

[b]filter size, m.

[c]latent vector size, k, extracted by the encoder.

[d]number of layers, q.

Examples for the zero-inflated denoising convolutional autoencoder, mean imputation, zero-inflated Poisson regression and Bayesian regression methods on the NHANES test data set are shown in Figure 4, and the results are given in Table 4. These results indicate that zero-inflated denoising convolutional autoencoder performed better and more accurately reflected the natural variation in human activity. In addition, the zero-inflated denoising convolutional autoencoder trained with the NHANES data set was tested on the KNHANES and KCCDB data set, and the results are also shown in Table 4. As with the results on the KNHANES and KCCDB data set, the values imputed by zero-inflated denoising convolutional autoencoder showed the lowest RMSE of standard deviation (24.4 counts and 27.1 counts, respectively).

In addition, we evaluated the RMSE of intradaily variability and moderate-to-vigorous physical activity for each data set to evaluate imputation. Zero-inflated denoising convolutional autoencoder yielded the lowest RMSE of intradaily variability for both the NHANES and KNHANES data sets with values of 0.047 and 0.037, respectively. In contrast, for KCCDB, zero-inflated denoising convolutional autoencoder had the second-lowest RMSE of intradaily variability with a value of 0.02. Moreover, the RMSE of moderate-to-vigorous physical activity for zero-inflated denoising convolutional autoencoder was the lowest (13.4 minutes) on KCCDB and the second lowest on the NHANES and KNHANES data sets (12.3 minutes and 12.9 minutes, respectively).

Additional analyses were performed including confidence interval calculation (Multimedia Appendix 4), additional experiments with 90- and 180-minute missing intervals (Multimedia Appendix 5 and Multimedia Appendix 6), other models (Multimedia Appendix 7), and other analyses (Multimedia Appendix 8 and Multimedia Appendix 9).

Figure 4. Examples of (a) NHANES, (b) KNHANES, and (c) KCCDB data sets for zero-inflated denoising convolutional autoencoder (left), zero-inflated Poisson regression (center), and Bayesian regression (right) with imputed (red) and original (green) intervals within the record (blue). KCCDB: Korean Chronic Cerebrovascular Disease Oriented Biobank; KNHANES: Korea National Health and Nutrition Examination Survey; ZI-DCAE: zero-inflated denoising convolutional autoencoder; ZIP: zero-inflated Poisson regression.

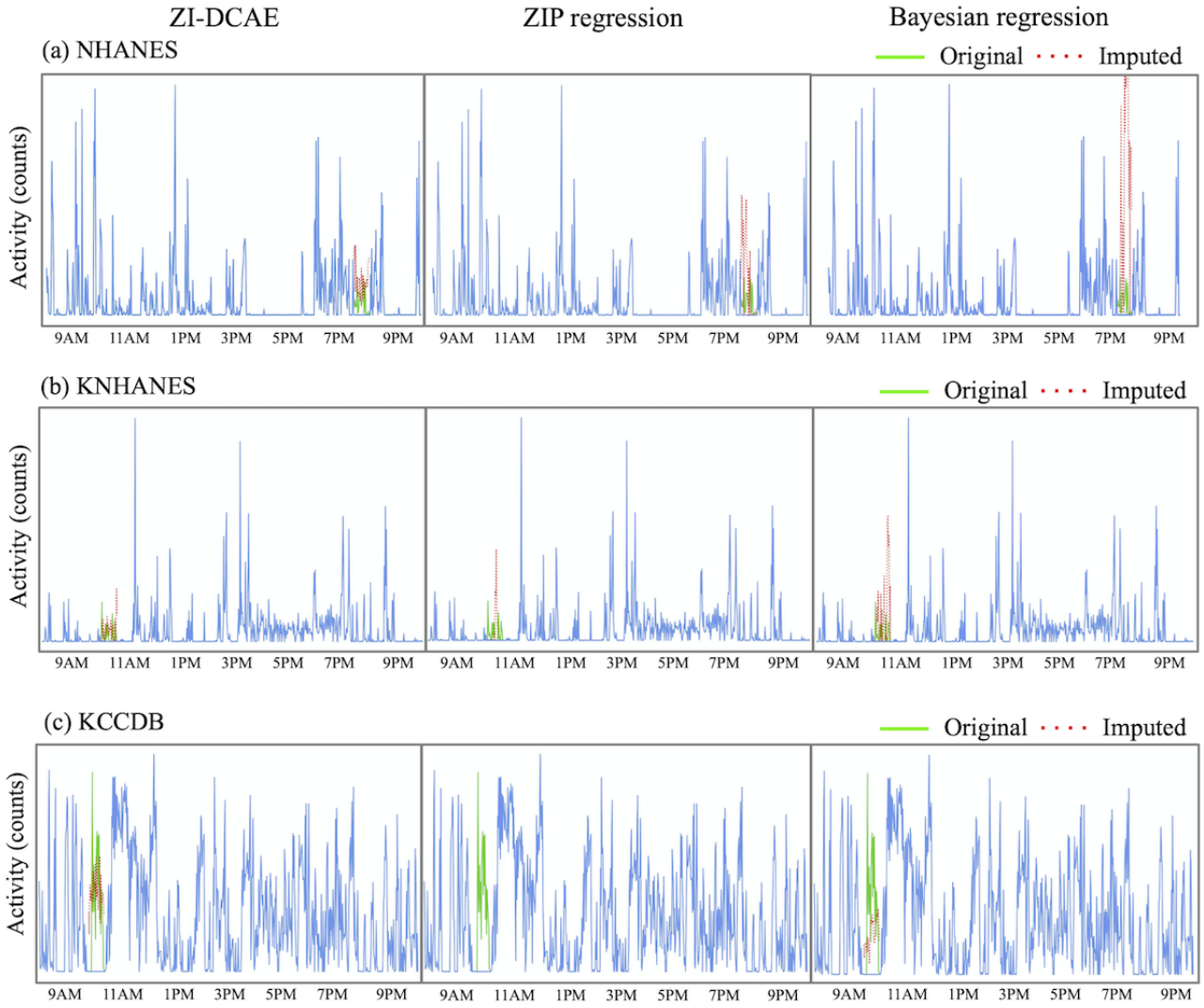


Table 4. Imputation performance results for the comparison methods.

Dataset measurement	ZI-DCAE[a]	Mean imputation	Zero-inflated Poisson regression	Bayesian regression
NHANES[b]				
partial RMSE[c] (counts)	839.3	1053.2	1255.6	924.5
partial MAE[d] (counts)	431.1	545.4	508.5	605.8
RMSE of SD (counts)	35.1	65.2	69.2	34.2
RMSE of intradaily variability index	0.047	0.067	0.060	0.071
RMSE of moderate-to-vigorous physical activity (minutes)	12.3	16.2	16.2	11.0
KNHANES[e]				
partial RMSE (counts)	672.1	660.0	778.8	824.1
partial MAE (counts)	396.3	419.7	395.5	555.5
RMSE of SD (counts)	24.4	26.5	26.0	24.7
RMSE of intradaily variability index	0.037	0.039	0.040	0.050
RMSE of moderate-to-vigorous physical activity (minutes)	12.9	14.7	14.6	12.2
KCCDB[f]				
partial RMSE (counts)	1217.2	1313.2	1638.4	1139.4
partial MAE (counts)	819.6	1045.8	1161.6	810.7
RMSE of SD (counts)	27.1	30.8	29.6	27.7
RMSE of intradaily variability index	0.02	0.036	0.041	0.018
RMSE of moderate-to-vigorous physical activity (minutes)	13.4	14.9	14.9	13.6

[a]ZI-DCAE: zero-inflated denoising convolutional autoencoder.

[b]NHANES: National Health and Nutrition Examination Survey data set.

[c]RMSE: root mean square error.

[d]MAE: mean absolute error.

[e]KNHANES: Korea National Health and Nutrition Examination Survey data set.

[f]KCCDB: Korean Chronic Cerebrovascular Disease Oriented Biobank data set.

Discussion

Principal Findings

Our study was the first to attempt to impute activity count data using a deep-learning approach. For activity data with intervals of zero values, in statistical models such as a zero-inflated Poisson distribution, the probability distribution of zeros is assumed to have a specific distribution. In contrast, the zero-inflated denoising convolutional autoencoder model created the distribution of zeros by using a clamped hyperbolic tangent activation function which caused the model to transform negative outputs to zero values. The zero-inflated denoising convolutional autoencoder model can learn the distribution of zeros from the data themselves. The results confirm that this approach performs better than previous approaches.

By testing the performance with an external data set that was not related to the training data set, we found that the model could be generalized. On the KNHANES and KCCDB data sets, our model exhibited better performance than those of the other imputation algorithms. Although our model was trained with the NHANES data set, which was collected with a uniaxial device, we confirmed that the zero-inflated denoising convolutional autoencoder model also worked well on triaxial data (the KCCDB data set). This result indicated that our model did not overfit to the NHANES data set, but instead was able to learn important features in the activity data.

In addition to predicting the missing values, our model was also able to reproduce the variability of activity data. On the NHANES data set, the zero-inflated denoising convolutional autoencoder had the second-lowest RMSE for standard deviation and the lowest RMSE for standard deviation on the external validation data sets. This characteristic is desirable for subsequent analysis using the imputed data set because some activity indices such as intradaily variability use variability in the activity data to evaluate the rhythms of human activity. We compared the RMSE of intradaily variability for the original data and imputed data. The zero-inflated denoising convolutional autoencoder showed a lower RMSE of intradaily variability

than those of the other imputation methods. Generally, it seemed that the zero-inflated denoising convolutional autoencoder was able to restore the variability of the original data more accurately than other methods. These results suggest that the zero-inflated denoising convolutional autoencoder model can not only impute the missing data while reflecting the meaningful variability of the activity data in the general population (NHANES and KNHANES data sets), but can also reflect the variability of activity data from patients with cerebrovascular disease (KCCDB data set).

Moderate-to-vigorous physical activity is an index for evaluating the intensity of activity. We evaluated how well the model restored the duration of the original moderate-to-vigorous physical activity. With the lowest RMSE of moderate-to-vigorous physical activity in KCCDB and the second lowest in other data sets, the zero-inflated denoising convolutional autoencoder also demonstrated the ability to restore measures of activity intensity.

Ad Hoc Fine-Tuning

Fine-tuning the approach could be considered to better reflect the unique characteristics of new data if there are enough complete cases available after training. To test whether fine-tuning could improve imputation performance, we conducted the following ad hoc analysis. The KNHANES data set was split into training, validation, and testing data sets in a 9:1:1 ratio. The fine-tuning process was conducted with the training data of the KNHANES data set using the fully developed model used in this study. Training was stopped when the performance of the model was best in the KNHANES validation set. When the performance was evaluated on the KNHANES testing set, the fine-tuned model performed better, with a partial RMSE of 663.6 counts, partial MAE of 391.2 counts, whereas those of the baseline model were 672.1 counts, 396.3 counts, respectively.

Limitations

There are some limitations to be discussed. First, activity patterns can depend on demographic factors such as age, BMI, and other metrics. Although demographic information may help improve the performance of the imputation model, it was not used in this study because of a lack of data. If we can obtain sufficient data sets with demographic information in future, we will be able to improve the performance of our imputation model by including it in the training; however, because the demographic information of the user is often unknown in actual studies, it could be more practical for the model to restore missing data using only activity data. Second, some activities that cause participants to remove the device cannot be collected in data set, and these data are difficult to impute correctly. This limitation is always carefully considered before imputation methods are applied. Although all imputation methods have this limitation, our model performed better than other methods did. Third, although many imputation methods exist, we only compared our model with three methods. We conducted Gaussian process regression, but it predicted only zero values for the imputed data (Multimedia Appendix 9). Further studies should include a wider range of imputation methods. Finally, the use of a count-based algorithm to define missing intervals could be too simple for representing nonwear time; however, the aim of our work to impute missing intervals, and we only utilized a count-based nonwear detection algorithm to construct complete data sets. Regardless of which nonwear detection algorithms is used to generate corruption, the model should be able learn the pattern and impute the corrupted data.

Conclusions

To our knowledge, this is the first study to develop a deep-learning model for imputing missing values in actigraphy data. The results of this study suggest that the deep learning approach is useful for imputing missing values in activity data. We expect that our model will contribute to studies of human activity by decreasing the amount of discarded data due to missing values.

Acknowledgments

This study was conducted with biospecimens and data from the biobank of the Chronic Cerebrovascular Disease Consortium. The consortium was supported and funded by the Korea Centers for Disease Control and Prevention (#4845-303). This work was also supported by the faculty research fund of Ajou University School of Medicine and grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health & Welfare, Republic of Korea (Government-wide R&D Fund project for infectious disease research HG18C0067).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed information about accelerometer device.
[DOCX File , 52 KB - mhealth_v8i7e16113_app1.docx]

Multimedia Appendix 2
The distribution of length of consecutive zeros in each dataset.
[DOCX File , 232 KB - mhealth_v8i7e16113_app2.docx]

Multimedia Appendix 3
Detailed result of 10-fold cross validation.
[DOCX File , 16 KB - mhealth_v8i7e16113_app3.docx]

Multimedia Appendix 4
Confidence intervals of evaluation measurement by the bootstrapping method.
[DOCX File , 973 KB - mhealth_v8i7e16113_app4.docx]

Multimedia Appendix 5
Experimental results for 90- and 180-min missing intervals for each imputation method.
[DOCX File , 23 KB - mhealth_v8i7e16113_app5.docx]

Multimedia Appendix 6
Example imputed result with 90- and 180-min missing intervals.
[DOCX File , 2470 KB - mhealth_v8i7e16113_app6.docx]

Multimedia Appendix 7
Comparison of model performance between ZI-DCAE model and the naïve convolutional autoencoder with sigmoid function.
[DOCX File , 16 KB - mhealth_v8i7e16113_app7.docx]

Multimedia Appendix 8
Accuracy of restoring Moderate-to-Vigorous Physical Activity (MVPA).
[DOCX File , 16 KB - mhealth_v8i7e16113_app8.docx]

Multimedia Appendix 9
The result of Gaussian process imputation.
[DOCX File , 1345 KB - mhealth_v8i7e16113_app9.docx]

References

1. Tranah GJ, Blackwell T, Stone KL, Ancoli-Israel S, Paudel ML, Ensrud KE, et al. Circadian activity rhythms and risk of incident dementia and mild cognitive impairment in older women. Ann Neurol 2011 Nov;70(5):722-732 [FREE Full text] [doi: 10.1002/ana.22468] [Medline: 22162057]
2. Kinder JR, Lee KA, Thompson H, Hicks K, Topp K, Madsen KA. Validation of a hip-worn accelerometer in measuring sleep time in children. J Pediatr Nurs 2012 Apr;27(2):127-133 [FREE Full text] [doi: 10.1016/j.pedn.2010.11.004] [Medline: 22341191]
3. Ravi N, Dandekar N, Mysore P, Littman M. Activity recognition from accelerometer data. URL: https://dl.acm.org/doi/10.5555/1620092.1620107 [accessed 2020-06-24]
4. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun;37(3):197-206; discussion 206 [FREE Full text] [Medline: 12782543]
5. Zhu W, Wadley VG, Howard VJ, Hutto B, Blair SN, Hooker SP. Objectively measured physical activity and cognitive function in older adults. Med Sci Sports Exerc 2017 Jan;49(1):47-53 [FREE Full text] [doi: 10.1249/MSS.0000000000001079] [Medline: 27580146]
6. Cho C, Lee T, Kim M, In HP, Kim L, Lee H. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J Med Internet Res 2019 Apr 17;21(4):e11029 [FREE Full text] [doi: 10.2196/11029] [Medline: 30994461]
7. Witting W, Kwa IH, Eikelenboom P, Mirmiran M, Swaab DF. Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. Biol Psychiatry 1990 Mar 15;27(6):563-572. [doi: 10.1016/0006-3223(90)90523-5] [Medline: 2322616]
8. Catellier DJ, Hannan PJ, Murray DM, Addy CL, Conway TL, Yang S, et al. Imputation of missing data when measuring physical activity by accelerometry. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S555-S562 [FREE Full text] [doi: 10.1249/01.mss.0000185651.59486.4e] [Medline: 16294118]
9. Yue Xu S, Nelson S, Kerr J, Godbole S, Patterson R, Merchant G, et al. Statistical approaches to account for missing values in accelerometer data: applications to modeling physical activity. Stat Methods Med Res 2018 Apr;27(4):1168-1186. [doi: 10.1177/0962280216657119] [Medline: 27405327]
10. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: a deep learning approach. 2016 Presented at: Proceedings of the 2016 SIAM International Conference on Data Mining; May 5-7, 2016; Miami, Florida, Unite States of America p. A URL: https://epubs.siam.org/doi/10.1137/1.9781611974348.49 [doi: 10.1137/1.9781611974348.49]

11. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
12. Yoon D, Lim HS, Jung K, Kim TY, Lee S. Deep learning-based electrocardiogram signal noise detection and screening model. Healthc Inform Res 2019 Jul;25(3):201-211 [FREE Full text] [doi: 10.4258/hir.2019.25.3.201] [Medline: 31406612]
13. Baldiditor. Autoencoders, unsupervised learning, and deep architectures. Proceedings of ICML workshop on unsupervised and transfer learning. 2012. URL: http://proceedings.mlr.press/v27/baldi12a.html [accessed 2020-06-24]
14. Duan Y, Lv Y, Liu Y, Wang F. An efficient realization of deep learning for traffic data imputation. Transportation Research Part C: Emerging Technologies 2016 Nov;72:168-181. [doi: 10.1016/j.trc.2016.09.015]
15. Wang L. Recognition of human activities using continuous autoencoders with wearable sensors. Sensors (Basel) 2016 Feb 04;16(2):189 [FREE Full text] [doi: 10.3390/s16020189] [Medline: 26861319]
16. National Center for Health Statistics. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey Data. 2003. URL: https://wwwn.cdc.gov/nchs/nhanes/2003-2004/PAXRAW_C.htm [accessed 2020-04-17]
17. National Center for Health Statistics. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey Data. 2005. URL: https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/PAXRAW_D.htm [accessed 2020-04-17]
18. Kweon S, Kim Y, Jang MJ, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). Int J Epidemiol 2014 Feb;43(1):69-77 [FREE Full text] [doi: 10.1093/ije/dyt228] [Medline: 24585853]
19. The Sixth Korea National Health and Nutrition Examination Survey (KNHANES VI-2). Korea Centers for Disease Control and Prevention. 2014. URL: https://knhanes.cdc.go.kr/knhanes/eng/index.do [accessed 2020-06-24]
20. The Sixth Korea National Health and Nutrition Examination Survey (KNHANES VI-3). Korea Centers for Disease Control and Prevention. 2015. URL: https://knhanes.cdc.go.kr/knhanes/eng/index.do [accessed 2020-06-24]
21. Survey Overview of Korean National Health and Nutrition Examination (KNHANES). Korea Centers for Disease Control and Prevention. URL: https://knhanes.cdc.go.kr/knhanes/eng/index.do [accessed 2020-06-24]
22. Kozey SL, Staudenmayer JW, Troiano RP, Freedson PS. Comparison of the ActiGraph 7164 and the ActiGraph GT1M during self-paced locomotion. Med Sci Sports Exerc 2010 May;42(5):971-976 [FREE Full text] [doi: 10.1249/MSS.0b013e3181c29e90] [Medline: 19997000]
23. Cradock AL, Wiecha JL, Peterson KE, Sobol AM, Colditz GA, Gortmaker SL. Youth recall and TriTrac accelerometer estimates of physical activity levels. Med Sci Sports Exerc 2004 Mar;36(3):525-532. [doi: 10.1249/01.mss.0000117112.76067.d3] [Medline: 15076797]
24. Evenson KR. Towards an understanding of change in physical activity from pregnancy through postpartum. Psychol Sport Exerc 2011 Jan;12(1):36-45 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.010] [Medline: 21278835]
25. Metzger JS, Catellier DJ, Evenson KR, Treuth MS, Rosamond WD, Siega-Riz AM. Patterns of objectively measured physical activity in the United States. Med Sci Sports Exerc 2008 Apr;40(4):630-638. [doi: 10.1249/MSS.0b013e3181620ebc] [Medline: 18317384]
26. Pratama I, Permanasari A, Ardiyanto I, Indrayani R. A review of missing values handling methods on time-series data. 2016 Presented at: 2016 International Conference on Information Technology Systems and Innovation (ICITSI); Oct 24-27, 2016; Bandung - Bali, Indonesia. [doi: 10.1109/icitsi.2016.7858189]
27. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 2011 Mar;20(1):40-49 [FREE Full text] [doi: 10.1002/mpr.329] [Medline: 21499542]
28. Lee JA, Gill J. Missing value imputation for physical activity data measured by accelerometer. Stat Methods Med Res 2018 Feb;27(2):490-506. [doi: 10.1177/0962280216633248] [Medline: 26994215]
29. Jung A. accelmissing: Missing Value Imputation for Accelerometer Data. R package version 1. URL: https://CRAN.R-project.org/package=accelmissing [accessed 2020-06-24]
30. Vincent P, Larochelle H, Bengio Y, Manzagol P, editors. Extractingcomposing robust features with denoising autoencoders. 2008 Presented at: Proceedings of the 25th international conference on Machine learning;ICML; July 5-9, 2008; Helsinki, Finland p. A. [doi: 10.1145/1390156.1390294]
31. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017 May 24;60(6):84-90. [doi: 10.1145/3065386]
32. Abdel-Hamid O, Mohamed A, Jiang H, Deng L, Penn G, Yu D. Convolutional neural networks for speech recognition. IEEE/ACM Trans. Audio Speech Lang. Process 2014 Oct;22(10):1533-1545. [doi: 10.1109/taslp.2014.2339736]
33. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 2016 Mar;63(3):664-675. [doi: 10.1109/TBME.2015.2468589] [Medline: 26285054]
34. PyTorch (version 1.4.1). URL: https://pytorch.org/ [accessed 2020-04-17]
35. Leenders NYJM, Nelson TE, Sherman WM. Ability of different physical activity monitors to detect movement during treadmill walking. Int J Sports Med 2003 Jan;24(1):43-50. [doi: 10.1055/s-2003-37196] [Medline: 12582951]
36. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
37. Visualization with Python. matplotlib. URL: https://matplotlib.org/ [accessed 2020-04-17]

38. Statistical data visualization. seaborn. URL: https://seaborn.pydata.org/ [accessed 2020-04-17]
39. ggplot2. URL: https://ggplot2.tidyverse.org [accessed 2020-04-17]

Abbreviations

BMI: body mass index
KCCDB: Korean Chronic Cerebrovascular Disease Oriented Biobank
KNHANES: Korea National Health and Nutrition Examination Survey
MAE: mean absolute error
NHANES: National Health and Nutrition Examination Survey
RMSE: root mean square error

Edited by G Eysenbach; submitted 06.09.19; peer-reviewed by BK Yi, P Lee, J Migueles; comments to author 04.11.19; revised version received 30.12.19; accepted 18.05.20; published 23.07.20.

Please cite as:
Jang JH, Choi J, Roh HW, Son SJ, Hong CH, Kim EY, Kim TY, Yoon D
Deep Learning Approach for Imputation of Missing Values in Actigraphy Data: Algorithm Development Study
JMIR Mhealth Uhealth 2020;8(7):e16113
URL: http://mhealth.jmir.org/2020/7/e16113/
doi:10.2196/16113
PMID:32445459

Original Paper

Telemonitoring Versus Usual Care for Elderly Patients With Heart Failure Discharged From the Hospital in the United States: Cost-Effectiveness Analysis

Xinchan Jiang[1], BSc, MPhil; Jiaqi Yao[1], BSc, MPhil; Joyce HS You[1], PharmD

School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, NT, China (Hong Kong)

Corresponding Author:
Joyce HS You, PharmD
School of Pharmacy
Faculty of Medicine
The Chinese University of Hong Kong
8/F
Lo Kwee-Seong Integrated Biomedical Sciences Building, CUHK
Shatin, NT,
China (Hong Kong)
Phone: 852 39436830
Fax: 852 26035295
Email: joyceyou@cuhk.edu.hk

Abstract

Background: Telemonitoring-guided interventional management reduces the need for hospitalization and mortality of patients with chronic heart failure (CHF).

Objective: This study aimed to analyze the cost-effectiveness of usual care with and without telemonitoring-guided management in patients with CHF discharged from the hospital, from the perspective of US health care providers.

Methods: A lifelong Markov model was designed to estimate outcomes of (1) usual care alone for all postdischarge patients with CHF (New York Heart Association [NYHA] class I-IV), (2) usual care and telemonitoring for all postdischarge patients with CHF, (3) usual care for all postdischarge patients with CHF and telemonitoring for patients with NYHA class III to IV, and (4) usual care for all postdischarge patients with CHF plus telemonitoring for patients with NYHA class II to IV. Model inputs were derived from the literature and public data. Sensitivity analyses were conducted to assess the robustness of model. The primary outcomes were total direct medical cost, quality-adjusted life years (QALYs), and incremental cost-effectiveness ratio (ICER).

Results: In the base case analysis, universal telemonitoring group gained the highest QALYs (6.2967 QALYs), followed by the telemonitoring for NYHA class II to IV group (6.2960 QALYs), the telemonitoring for NYHA class III to IV group (6.2450 QALYs), and the universal usual care group (6.1530 QALYs). ICERs of the telemonitoring for NYHA class III to IV group (US $35,393 per QALY) and the telemonitoring for NYHA class II to IV group (US $38,261 per QALY) were lower than the ICER of the universal telemonitoring group (US $100,458 per QALY). One-way sensitivity analysis identified five critical parameters: odds ratio of hospitalization for telemonitoring versus usual care, hazard ratio of all-cause mortality for telemonitoring versus usual care, CHF hospitalization cost and monthly outpatient costs for NYHA class I, and CHF hospitalization cost for NYHA class II. In probabilistic sensitivity analysis, probabilities of the universal telemonitoring, telemonitoring for NYHA class II to IV, telemonitoring for NYHA class III to IV, and universal usual care groups to be accepted as cost-effective at US $50,000 per QALY were 2.76%, 76.31%, 18.6%, and 2.33%, respectively.

Conclusions: Usual care for all discharged patients with CHF plus telemonitoring-guided management for NYHA class II to IV patients appears to be the preferred cost-effective strategy.

(JMIR Mhealth Uhealth 2020;8(7):e17846) doi:10.2196/17846

KEYWORDS

telemedicine; heart failure; hospitalization; cost; quality-adjusted life year; cost-effectiveness analysis

Introduction

Background

Chronic heart failure (CHF) is a condition that imposes a major burden on health care systems and the society. Globally, it is estimated that 37.3 million individuals suffer from CHF, and the prevalence continues to rise [1]. In the United States, the number of patients with CHF was nearly 6.2 million in 2016 and is projected to be over 8 million in the year 2030 [2]. CHF is the most common reason for admission, and hospitalization for CHF is highly associated with readmission within 30 days (20%-30%) and 6 months (50%) [3]. The total cost of care for CHF is expected to rise from US $30.7 billion in 2012 to US $69.8 billion in 2030 [4].

The use of digital health interventions in the provision of remote health care services is a promising strategy to improve the clinical outcomes of CHF. Telemonitoring allows remote daily monitoring of patients' vital signs and, therefore, enables detection of clinical deterioration and early clinical interventions. The Efficacy of Telemedical Interventional Management in Patients with Heart Failure II (TIM-HF2) study compared the efficacy of the telemedical interventional management for CHF patients (New York Heart Association [NYHA] class II or III) along with usual care versus usual care only [5]. It was reported that the structured telemonitoring-based management reduced the percentage of days lost to unplanned cardiovascular admission and all-cause death (4.88% vs 6.64%; *P*=.05) and lowered the all-cause mortality rate (7.86% vs 11.34%; *P*=.03).

Telemonitoring also engages patients in CHF self-care and improves the quality of patients' self-management at home. In the Trans-European Network–Home-Care Management System study, the telemonitoring intervention consisted of twice-daily patient self-management and monitoring of weight, blood pressure, heart rate, and rhythm with automated devices linked to a cardiology center [6]. The study found that the mean duration of admissions was reduced by 6 days (95% CI 1-11 days) in the home telemonitoring group versus the usual care group. One-year mortality was higher in the usual care group (45% vs 29%; *P*=.03).

Objectives

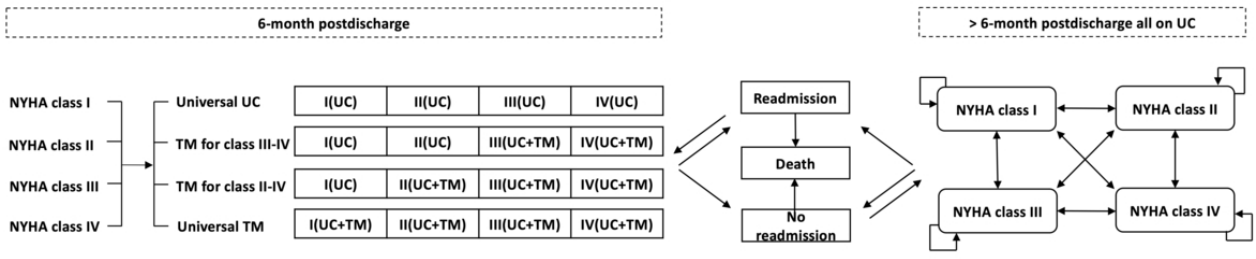
Clinical findings have indicated telemonitoring to be beneficial to patients with CHF, yet few studies examined the cost-effectiveness of telemonitoring on the management of CHF [7]. The aim of this study was to analyze the cost-effectiveness of usual outpatient care with and without telemonitoring for patients with CHF who were discharged from the hospital, from the perspective of the US health care providers.

Methods

Study Design

A lifelong Markov model was constructed using TreeAge Pro 2019 (TreeAge Software Inc) to project the long-term clinical and economic outcomes of a hypothetical cohort of 65-year-old patients with CHF who were recently discharged for CHF-related hospitalization (Figure 1) in the US health care setting. The Markov model is an analytical framework in which the hypothetical cohort of patients transits between the mutually exclusive health states, with costs and health outcomes aggregated over successive cycles. The following strategies were examined in the present model: (1) universal usual care only for all patients (NYHA class I-IV), (2) universal usual care for all patients and telemonitoring for NYHA class III to IV patients (telemonitoring for class III-IV group), (3) universal usual care for all patients and telemonitoring for NYHA class II to IV patients (telemonitoring for class II-IV group), and (4) universal usual care plus telemonitoring for all patients (universal telemonitoring group). The model time horizon was 35 years with 6-month cycles for lifelong simulation of 100 years of age maximum. The primary outcome measures were direct medical costs, quality-adjusted life years (QALYs), and incremental cost-effectiveness ratio (ICER) as additional cost per QALY gained.

Figure 1. Simplified model structure. NYHA: New York Heart Association; TM: telemonitoring; UC: usual care; universal UC: UC only for all patients; TM for class III to IV: UC for all patients and TM for NYHA class III to IV patients; TM for class II to IV: UC for all patients and TM for NYHA class II to IV patients; universal TM: UC plus TM for all patients.



As reported by the ancillary Digitalis Investigation Group (DIG) trial, a retrospective cohort study of 988 patients with NYHA class I to IV heart failure, the NYHA classification was associated with the risk of hospitalization and mortality [8]. In this model, we stratified the hypothetical cohort of recently discharged patients with CHF by the NYHA classification. Subjects of all study arms entered the present model at the Markov status *NYHA class I*, *NYHA class II*, *NYHA class III*, or *NYHA class IV*. In each 6-month cycle, all patients might be readmitted for CHF-related hospitalization and might experience all-cause death. For those who survived at the end of cycle, they might remain in the same NYHA classification or transit (improve or progress) to another NYHA classification.

Patients in all study arms received usual care: physician's office visits (for evaluation of CHF) and medication prescriptions,

without structured follow-up or service of specialized nurses [9]. Patients in the telemonitoring groups received telemonitoring-guided CHF intervention in addition to usual care for 6 months post discharge. The 6-month duration is most commonly used for examination of the telemonitoring-guided intervention in clinical trials [10]. The telemonitoring-guided intervention included a telemonitoring system (installed at patients' home) usually with a digital tablet as a central structural element to transmit vital elements from patients' home to the hospital. Algorithms were programmed and implemented in this system to guide patient management (concomitant medication change, initiation of an ambulatory assessment by a home physician, or to hospitalize the patient). The patients in the telemonitoring groups might or might not adhere to the telemonitoring-guided management. If patients were not readmitted to the hospital during the 6-month telemonitoring-guided management period, they would be followed up by usual care only in the following cycles until rehospitalization occurred. Those who survived the next hospitalization would receive usual care alone or usual care plus 6-month telemonitoring-guided management according to patients' NYHA classification and the corresponding study group.

Clinical Inputs

The clinical parameters are listed in Multimedia Appendix 1. Literature search on MEDLINE over the period 2000 to 2019 was conducted using the keywords *heart failure*, *telemonitoring*, *remote patient monitoring*, *telehealth*, *telemedicine*, *telemedical center*, *hospitalization*, *admission*, and *mortality*. The inclusion criteria for clinical trials of CHF management were (1) reports written in English, (2) patients aged 18 years or older, and (3) the incidence of hospitalization or mortality was reported. A study was included if the data relevant to the model inputs were available. Case studies were excluded. Preferred studies were meta-analyses or randomized controlled trials. If multiple sources were found for a model input, the weighted average was used as the base case value and the high or low values formed as the range for sensitivity analysis.

The distribution of the hypothetical cohort of recently discharged patients with CHF among NYHA classes (upon entry to the present model) was estimated from the findings of the TIM-HF2 trial that the percentages of NYHA classes I, II, III, and IV in 1538 patients with CHF were 0.7%, 51.8%, 47.2%, and 0.3%, respectively [5]. The CHF-related hospitalization rate of patients with NYHA class I and the hazard ratios of hospitalization for patients with NYHA classes II, III, and IV (vs NYHA class I) were reported by the DIG trial, which demonstrated higher NYHA classification to be associated with increased risk of hospitalization. The hazard ratios of hospitalization for patients with NYHA classes II, III, and IV were 1.16 (95% CI 0.76-1.77; *P*=.50), 2.27 (95% CI 1.45-3.56; *P*<.001), and 3.71 (95% CI 1.25-11.02; *P*=.02), respectively [8]. The 3-year hospitalization rate of NYHA class I patients was 14.3% [8], and it was converted into a 6-month probability of 2.36% by the equation [11] $p=1-e^{-rt}$ (p=probability, r=event rate, and t=cycle length). In the present model, the probabilities of hospitalization for patients with NYHA classes II, III, and IV were estimated by multiplying the probability of hospitalization for NYHA class I with the corresponding hazard ratios of hospitalization. A retrospective outcome study, including 2176 patients with heart failure, reported the hazard ratio of readmission for patients with prior admission to be 1.25 (95% CI 1.05-1.48; *P*=.01) [12]. We, therefore, estimated the readmission rate of patients in each NYHA class using the hazard ratio of readmission for patients with prior hospitalization and corresponding hospitalization rate of each NYHA class.

A network meta-analysis on telemonitoring interventions for heart failure patients included 13 randomized clinical trials with total 10,913 patients [10]. The telemonitoring with transmission of physiologic measurements (weight, blood pressure, and heart rate) was found to reduce CHF-related hospitalization (odds ratio 0.64, 95% CI 0.39-0.95) when compared with usual care alone, and these findings were adopted as the model input for the odds of hospitalization during telemonitoring-guided management versus usual care alone.

The 6-month probabilities of mortality in patients with CHF receiving usual care (0.65%, 3.56%, 11.68%, and 11.68% for NYHA classes I, II, III, and IV, respectively) were estimated by the yearly all-cause mortality rate of each NYHA class, retrieved from the outcomes of the control group in a clinical trial of 2737 patients with heart failure with 4783 patient-years of follow-up [13,14]. The probabilities of all-cause mortality in the telemonitoring groups were approximated by the probabilities of mortality in the usual care group and hazard ratio of mortality with telemonitoring versus usual care (0.70; 95% CI 0.50-0.96; *P*=.03, reported by the TIM-HF2 trial) [5]. Adherence rate (81%) for the telemonitoring-guided CHF management was estimated from the findings of the Telemedical Interventional Monitoring in Heart Failure trial [15]. In this trial, 81% of the study patients (n=354) in the telemonitoring arm achieved at least 70% of daily data transfer. In this model, we therefore adopted 70% as the lower threshold of daily data transfer for telemonitoring, and adherence rate of 81% (achieving lower threshold of daily data transfer) was used as the base case value. The hospitalization and mortality rates in the usual care group were applied to patients in the telemonitoring groups who did not achieve the lower threshold of data transfer.

The annual transition probabilities between NYHA classes were also retrieved from a prior clinical trial on patients with heart failure [13,14] and were converted to 6-month probabilities by the eigen decomposition approach using MATLAB (MathWorks) [16]. A previous cost-effectiveness analysis of patients with systolic heart failure assumed the same transition probabilities from NYHA class IV to classes I to III as of NYHA class III to three other classes, and the present model had adopted similar assumption [14]. Transition between NYHA classes was assumed to be the same for the telemonitoring and usual care groups in this study because of paucity of evidence to indicate influence of telemonitoring-guided management (if any) on the changes of NYHA classes.

Utility Inputs

The QALY gain expected by each subject in the model was estimated from subject-time spent in various health statuses and

the corresponding utility values. The utility inputs are shown in Multimedia Appendix 1. The base case utility and the decrement utility of CHF-related admission were retrieved from the findings of the standard care group in a randomized controlled trial, including 6505 patients with prior hospitalization for heart failure within 12 months [17]. The duration of hospitalization was approximated from the average total days per discharge in the diagnosis-related groups (DRGs) of heart failure reported by the Centers for Medicare and Medicaid Services [18]. The expected lifelong QALYs were discounted to the year 2019 with an annual rate of 3%.

Cost Inputs

The health economic analysis was performed from the perspective of US health care providers, and the cost inputs were retrieved from the literature review and public data. The CHF inpatient cost was retrieved from the 2016 DRGs data [18]. The costs of hospitalization for NYHA classes I, II, and III to IV (including death occurred during hospitalization) were approximated by the charges per discharge for heart failure without complication, with complication, and with major complication, respectively. The monthly outpatient costs in different NYHA classes were estimated from findings of a resource utilization study of 117,870 patients with CHF in 2010 and the relative difference in inpatient costs between NYHA classes [18,19]. The monthly telemonitoring cost per patient was estimated by the total annual expense on home telehealth and the number of patients served reported by the Department of Veterans Affairs in 2018 [20,21]. The utilization of telemonitoring-guided management in patients who were nonadherence or died during telemonitoring treatment were both assumed to be 3 months (and examined over a range of 1-6 months in sensitivity analysis) for cost estimation. All cost inputs and the expected lifelong cost were discounted to the year 2019 with an annual rate of 3%.

Cost-Effectiveness Analysis

Expected direct medical cost and QALYs gain were calculated for each strategy. The cost per QALY saved by each strategy versus universal usual care (the common baseline) was reported. A strategy was dominated by the comparator when it gained fewer QALYs at higher cost, and the dominated strategy was eliminated from further cost-effectiveness analysis. If a strategy gained additional QALYs at higher cost than the comparator, the incremental cost per QALY saved (ICER) of the more effective strategy was calculated using the following equation: △cost/△QALYs. The willingness-to-pay (WTP) threshold of US $50,000 per QALY was adopted in this analysis [22]. The study group gained the highest QALYs with ICER less than US$ 50,000 per QALY. This was considered as the preferred cost-effective option.

Sensitivity Analysis

Sensitivity analysis was conducted to examine the robustness of the model results. One-way sensitivity analysis was conducted over the range specified in Multimedia Appendix 1. To evaluate the impact of all variables simultaneously, probabilistic sensitivity analysis was performed with 10,000 Monte Carlo simulations by randomly drawing each of the model inputs from the specific probability distribution indicated in Multimedia Appendix 1. The probability of each study arm to be accepted as the preferred option was determined over a wide range of WTP from US $0 to US $100,000 per QALY in the acceptability curve.

Results

Base Case Analysis

Base case results are shown in Table 1. All telemonitoring groups incurred higher QALYs at higher costs when compared with the universal usual care group. The universal telemonitoring group gained the highest QALYs (6.2967 QALYs), followed by the telemonitoring for class II to IV group (6.2960 QALYs), the telemonitoring for class III to IV group (6.2450 QALYs), and the universal usual care group (6.1530 QALYs).

When compared with universal usual care (the common baseline), the cost per QALY saved by each telemonitoring group ranged between US $35,393 and US $36,720 per QALY and was lower than the WTP threshold (US $50,000 per QALY). Comparing with the next less costly option (in an ascending order), the ICERs of the telemonitoring for class III to IV group (US $35,393 per QALY) and the telemonitoring for class II to IV group (US $38,261 per QALY) were lower than the WTP threshold, and the ICER of the universal telemonitoring group (US $100,458 per QALY) exceeded the WTP threshold. The telemonitoring for class II to IV group gained the highest QALYs with ICER less than the WTP threshold and was therefore the preferred cost-effective option.

Table 1. Base case results.

Treatment option	Direct medical cost (US $)	Incremental cost (US $)[a]	Incremental cost (US $)[b]	QALYs[c]	Incremental QALYs[a]	Incremental QALYs[b]	ICER[d] (US $ per QALY)	Cost per QALY saved (US $ per QALY)[b]
Universal usual care	238,146	N/A[e]	N/A	6.1530	N/A	N/A	N/A	N/A
TM[f] for class III to IV	241,401	3256	3256	6.2450	0.0920	0.0920	35,393	35,393
TM for class II to IV	243,354	1953	5209	6.2960	0.0510	0.1430	38,261	36,416
Universal TM	243,423	68	5277	6.2967	0.0007	0.1437	100,458	36,720

[a]ICER=(Total cost$_{TM}$−Total cost$_{\text{next less costly option}}$)/(QALY saved$_{TM}$−QALY saved$_{\text{next less costly option}}$).

[b]Cost per QALY saved=(Total cost$_{TM}$−Total cost$_{\text{universal UC}}$)/(QALY saved$_{TM}$−QALY saved $_{\text{universal UC}}$).

[c]QALYs: quality-adjusted life years.

[d]ICER: incremental cost-effectiveness ratio.

[e]N/A: not applicable.

[f]TM: telemonitoring.

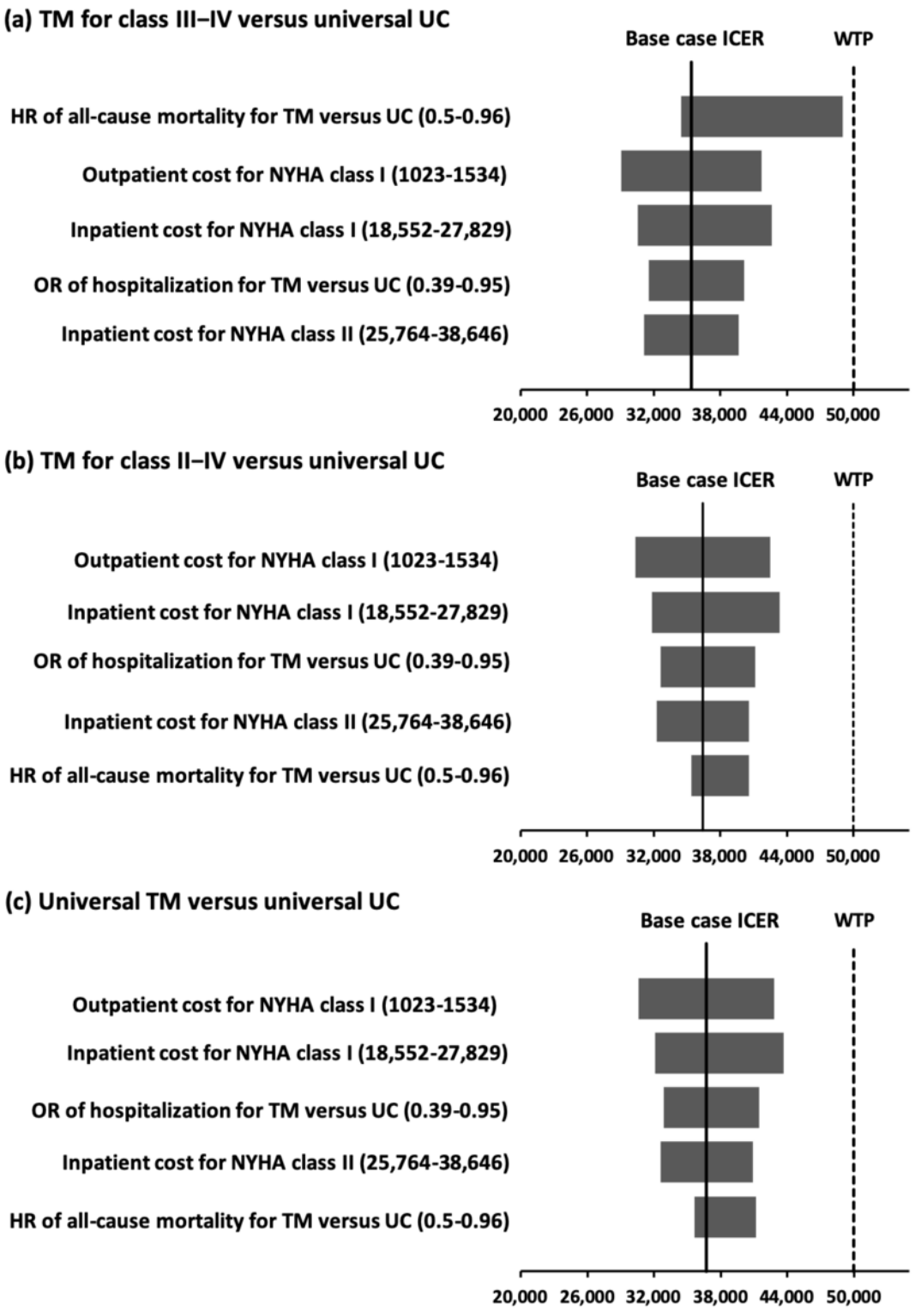
Sensitivity Analysis

One-Way Sensitivity Analysis

The ICER of each telemonitoring strategy, compared with universal usual care, was examined by one-way sensitivity analysis. No threshold value was identified in the one-way sensitivity analysis. Five critical parameters with impact on the ICER of each telemonitoring strategy by 10% or greater were identified (Figure 2). Two of the critical parameters were clinical inputs: odds ratio of hospitalization for telemonitoring versus usual care and hazard ratio of all-cause mortality for telemonitoring versus usual care. The remaining three critical parameters were cost inputs: CHF hospitalization cost and monthly outpatient costs for NYHA class I and CHF hospitalization cost for NYHA class II.

To further explore the impact of adherence of telemonitoring-guided management and monthly telemonitoring cost, extended one-way sensitivity analyses were conducted over the lower limits of these variables (Multimedia Appendix 2). When the adherence to achieve lower threshold of data transfer was 20.9% to 30.8%, the telemonitoring for class III to IV group became the preferred option. Universal usual care was the preferred option if the patient adherence declined to less than 20.9%. When monthly telemonitoring cost reduced from US $193 (base case value) to less than US $87.5, universal telemonitoring would become the preferred strategy. The findings of the extended one-way sensitivity analysis are shown in Multimedia Appendix 2.

Figure 2. Tornado diagrams: (A) TM for class III to IV versus universal UC, (B) TM for class II to IV versus universal UC, (C) universal TM versus universal UC. HR: hazard ratio; ICER: incremental cost-effectiveness ratio; NYHA: New York Heart Association; OR: odds ratio; TM: telemonitoring; UC: usual care; WTP: willingness-to-pay.



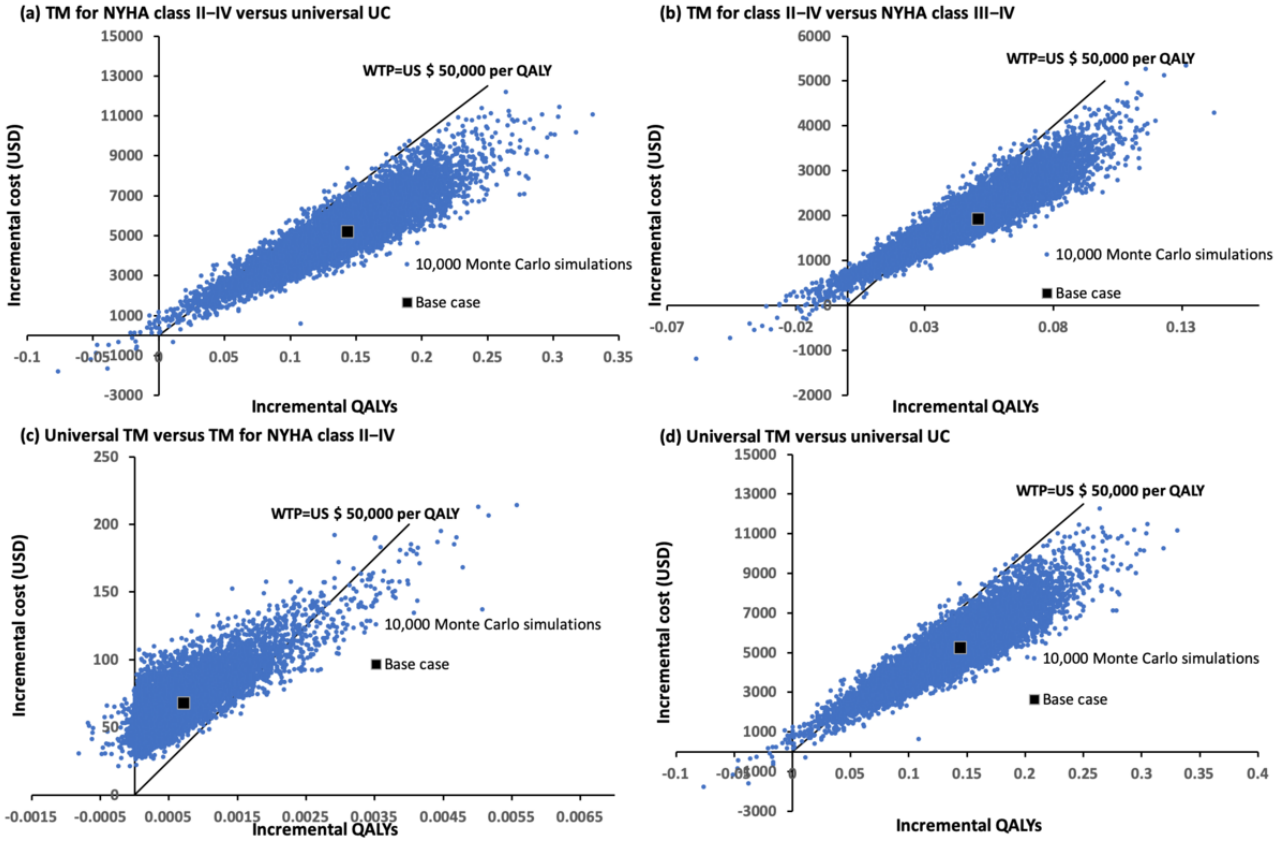
Probabilistic Sensitivity Analysis

The probabilistic sensitivity analysis was performed by the 10,000 Monte Carlo simulations (Figure 3). Compared with the universal usual care group, the telemonitoring for class II to IV group gained a mean QALY of 0.1343 (95% CI 0.1334-0.1352; *P*<.001) with an additional mean cost of US $5062 (95% CI US $5031-US $5092; *P*<.001). Of the 10,000 simulations, the ICERs of the telemonitoring for class II to IV group were below the WTP threshold in 95.91% of time. Compared with the telemonitoring for class III to IV group, the telemonitoring for class II to IV group was more costly by US $2045 (95% CI US $2032-US $2058; *P*<.001) and gained 0.0490 QALYs (95% CI 0.0486-0.0494; *P*<.001). The telemonitoring for class II to IV group had ICERs (for additional QALYs gain) below WTP in

79.86% of the simulations. Compared with the telemonitoring for NYHA class II to IV group, the universal telemonitoring group was costlier by US $68 (95% CI US $68-US $69; *P*<.001) and gained 0.00064 QALYs (95% CI 0.00063-0.00065; *P*<.001).

The ICERs of the universal telemonitoring group (vs telemonitoring for class II-IV group) were below the WTP in 3.30% of time.

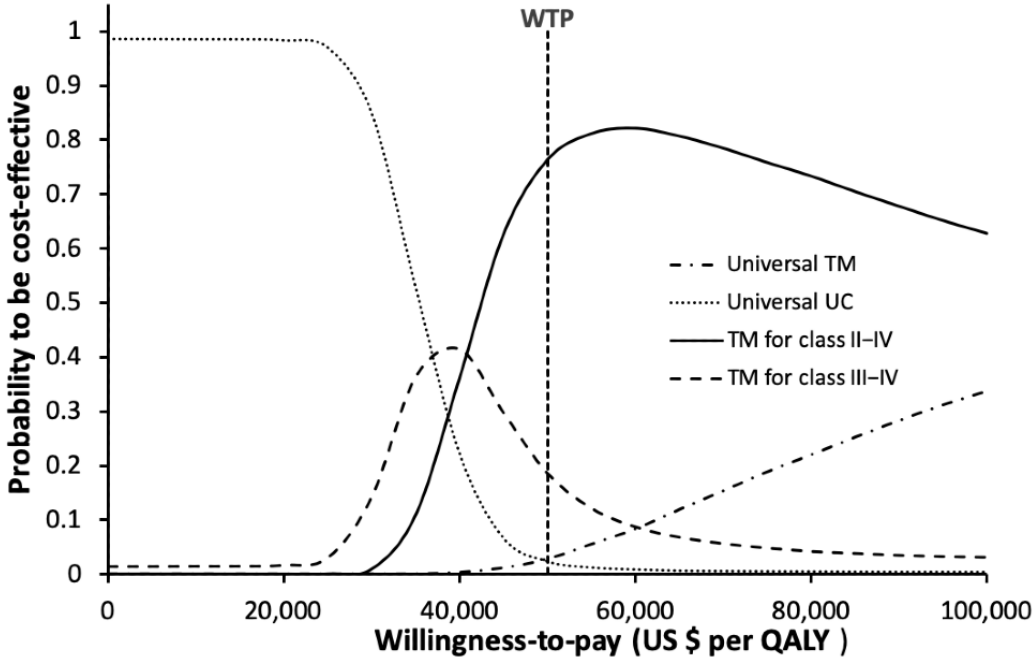
Figure 3. Scatter plots: (A) TM for class II to IV versus universal UC, (B) TM for class II to IV versus TM for class III to IV, (C) universal TM versus TM for class II to IV, (D) universal TM versus universal UC. QALY: quality-adjusted life year; TM: telemonitoring; UC: usual care; WTP: willingness-to-pay.



As the ICER of universal telemonitoring versus the universal usual care group (US $36,720 per QALY) was below the WTP threshold in the base case analysis, a probabilistic sensitivity analysis was further conducted. The universal telemonitoring group incurred higher cost of US $5130 (95% CI 5099-5161; *P*<.001) and gained 0.1349 QALYs (95% CI 0.1340-0.1358; *P*<.001). The ICERs (for additional QALYs gained) of the universal telemonitoring group were below the WTP threshold in 95.52% of 10,000 simulations.

The probability of each study group to be accepted as cost-effective was examined in the acceptability curves over a wide range of WTP (US $0-US $100,000 per QALY; Figure 4). At WTP of US $50,000 per QALY, the probabilities of the universal telemonitoring, telemonitoring for class II to IV, telemonitoring for class III to IV, and universal usual care groups to be accepted as the preferred option were 2.76%, 76.31%, 18.6%, and 2.33%, respectively.

Figure 4. Cost-effectiveness acceptability curve for each strategy to be preferred option against the willingness-to-pay threshold. QALY: quality-adjusted life year; TM: telemonitoring; UC: usual care; WTP: willingness-to-pay.



Discussion

Principal Findings

The present model examined the potential clinical and economic outcomes of providing telemonitoring-guided CHF management with usual outpatient care to postdischarge patients with different NYHA classifications. Our findings showed that all telemonitoring plus usual care strategies versus usual care alone were cost-effective using US $50,000 per QALY as the WTP threshold from the perspective of US health care providers. Further comparison between telemonitoring groups showed that universal telemonitoring was more effective than telemonitoring for NYHA class II to IV in QALY gain, yet the ICER (US $100,458 per QALY) exceeded the WTP threshold. Telemonitoring for NYHA class II to IV was the most effective strategy with a WTP-acceptable ICER in the base case analysis. The robustness of base case findings was supported by the one-way analysis that no influential factor with threshold value was identified. The probabilistic sensitivity analysis further demonstrated that the strategy of providing telemonitoring for NYHA class II to IV had the highest probabilities to be accepted as cost-effective at the WTP threshold of US $40,830 to US $100,000 per QALY, as indicated by the acceptability curves.

Limitations

This study was limited by the uncertainties on data availability and model assumptions. Rigorous sensitivity analyses were therefore performed to examine the impact of model input uncertainties on the robustness of the model results. The US health care costs were used as model inputs, potentially limiting the findings' generalizability in health care systems of other countries. To enhance the transferability of the decision model to other countries, acquisition of country- and region-specific health care cost as model inputs is necessary. Indirect cost (eg, cost of caregivers and loss of productivity) was not included and might, therefore, underestimate the health economic impact of telemonitoring. A cost-effectiveness analysis from the perspective of society to include both direct and indirect costs on telemonitoring-guided management in patients discharged for CHF is highly warranted.

Comparison With Prior Work

This study was the first cost-effectiveness analysis that examined the cost and QALY gained by telemonitoring-guided management for patients with CHF from the perspective of US health care providers. Previously, a cost-consequence analysis conducted from the perspective of US public payers over a 20-year time horizon found the telehealth program for CHF management to be likely to save cost (from US $3422 to US $4456) and gain 0.46 to 0.50 life-years for high-risk (including prior hospitalization) patients [23]. The findings of this prior analysis supported our results that telemonitoring-guided management was effective in saving life-years and QALYs. The incremental cost incurred to telemonitoring groups in this analysis was attributed to the higher current (year 2019) costs of CHF care in both inpatient and outpatient settings, whereas the cost saving reported in the prior US analysis was generated by lower CHF inpatient and outpatient costs (originated in the year 2010 and earlier). A cost-effectiveness analysis on telemonitoring, usual care, and nurse telephone support for CHF patients was performed from the perspective of a third-party payer of the Netherlands [24]. A Markov model was used to examine the disease progression over four classes of NYHA classifications (I-IV) and death in the time horizon of 20 years. The analysis reported that the telemonitoring group gained higher QALYs at an additional cost than the usual care group. Our findings were consistent with the Netherlands study, and we further examined the impact of two factors, which were less transferable from region to region: patient adherence to

telemonitoring and monthly telemonitoring cost on the cost-effective acceptance of telemonitoring-guided management in various levels of CHF severity.

In the Telemonitoring to Improve Heart Failure Outcomes and Better Effectiveness After Transition–Heart Failure studies, the adherence to telemonitoring decreased by nearly half within the first 30 days [25,26]. A threshold of 20% to 30% of patients to adhere to 70% of data transfer was identified in the present model. For a health care system in which patients' adherence to 70% or greater daily data transfer for telemonitoring is less than 20%, the telemonitoring-guided service might not be acceptable as a cost-effective option for all patients with CHF. If the adherence ranged between 20.9% and 30.8%, telemonitoring-guided management for the patients with more severe CHF (NYHA class III-IV) was likely to be cost-effective. If the adherence was higher than 30.8%, telemonitoring-guided management would likely be the preferred cost-effective option for patients in NYHA class II and above. The adherence of patients varies among different health care settings, and the collection of local adherence data is therefore critical to inform the decision-making process of the health care providers on the implementation of telemonitoring-guided service.

With the advancement of digital technology, the costs of devices and the computational techniques applied in the telemonitoring-guided management are anticipated to decrease over time. If the cost of telemonitoring-guided interventions is decreased to US $87.5 per patient per month or less (as indicated by the extended sensitivity analysis), the universal telemonitoring-guided management would be acceptable as the preferred option for all patients with CHF from the perspective of health service providers.

Conclusions

In conclusion, universal usual outpatient care for all discharged patients with CHF plus telemonitoring-guided management for those with NYHA class II to IV appears to be the preferred cost-effective strategy from the perspective of US health care providers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Model inputs.
[DOCX File , 40 KB - mhealth_v8i7e17846_app1.docx]

Multimedia Appendix 2
Results of extended one-way sensitivity analysis.
[DOCX File , 174 KB - mhealth_v8i7e17846_app2.docx]

References

1. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nat Rev Cardiol 2016 Jun;13(6):368-378 [FREE Full text] [doi: 10.1038/nrcardio.2016.25] [Medline: 26935038]
2. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2019 update: a report from The American Heart Association. Circulation 2019 Mar 5;139(10):e56-528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
3. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? Circulation 2012 Jul 24;126(4):501-506. [doi: 10.1161/CIRCULATIONAHA.112.125435] [Medline: 22825412]
4. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, American Heart Association Advocacy Coordinating Committee, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Radiology and Intervention, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
5. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
6. Cleland JG, Louis AA, Rigby AS, Janssens U, Balk AH, TEN-HMS Investigators. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. J Am Coll Cardiol 2005 May 17;45(10):1654-1664 [FREE Full text] [doi: 10.1016/j.jacc.2005.01.050] [Medline: 15893183]
7. Jiang X, Ming W, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]

8. Ahmed A, Aronow WS, Fleg JL. Higher New York Heart Association classes and increased mortality and hospitalization in patients with heart failure and preserved left ventricular function. Am Heart J 2006 Feb;151(2):444-450 [FREE Full text] [doi: 10.1016/j.ahj.2005.03.066] [Medline: 16442912]
9. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. J Card Fail 2017 Aug;23(8):628-651. [doi: 10.1016/j.cardfail.2017.04.014] [Medline: 28461259]
10. Kotb A, Cameron C, Hsieh S, Wells G. Comparative effectiveness of different forms of telemedicine for individuals with heart failure (HF): a systematic review and network meta-analysis. PLoS One 2015;10(2):e0118681 [FREE Full text] [doi: 10.1371/journal.pone.0118681] [Medline: 25714962]
11. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-338. [doi: 10.1177/0272989X9301300409] [Medline: 8246705]
12. Krumholz HM, Chen YT, Wang Y, Vaccarino V, Radford MJ, Horwitz RI. Predictors of readmission among elderly survivors of admission with heart failure. Am Heart J 2000 Jan;139(1 Pt 1):72-77. [doi: 10.1016/s0002-8703(00)90311-9] [Medline: 10618565]
13. Zannad F, McMurray JJ, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, EMPHASIS-HF Study Group. Eplerenone in patients with systolic heart failure and mild symptoms. N Engl J Med 2011 Jan 6;364(1):11-21. [doi: 10.1056/NEJMoa1009492] [Medline: 21073363]
14. Ademi Z, Pasupathi K, Liew D. Cost-effectiveness of eplerenone compared to usual care in patients with chronic heart failure and NYHA class II symptoms, an Australian perspective. Medicine (Baltimore) 2016 May;95(18):e3531 [FREE Full text] [doi: 10.1097/MD.0000000000003531] [Medline: 27149456]
15. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 3;123(17):1873-1880. [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
16. Chhatwal J, Jayasuriya S, Elbasha EH. Changing cycle lengths in state-transition models: challenges and solutions. Med Decis Making 2016 Nov;36(8):952-964 [FREE Full text] [doi: 10.1177/0272989X16656165] [Medline: 27369084]
17. Griffiths A, Paracha N, Davies A, Branscombe N, Cowie MR, Sculpher M. Analyzing health-related quality of life data to estimate parameters for cost-effectiveness models: an example using longitudinal EQ-5D data from the shift randomized controlled trial. Adv Ther 2017 Mar;34(3):753-764 [FREE Full text] [doi: 10.1007/s12325-016-0471-x] [Medline: 28205056]
18. Center for Medicare and Medicaid services. 100% MEDPAR Inpatient Hospital Data for Fiscal Year 2016 URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/Downloads/DRGState16.pdf [accessed 2019-09-15]
19. Yoon J, Fonarow GC, Groeneveld PW, Teerlink JR, Whooley MA, Sahay A, et al. Patient and facility variation in costs of VA heart failure patients. JACC Heart Fail 2016 Jul;4(7):551-558 [FREE Full text] [doi: 10.1016/j.jchf.2016.01.003] [Medline: 26970829]
20. Department of Veterans Affairs. 2020 Feb. Volume II Medical Programs and Information Technology Programs URL: https://www.va.gov/budget/docs/summary/fy2021VAbudgetVolumeIImedicalProgramsAndInformationTechnology.pdf [accessed 2019-09-28]
21. Congressional Research Service. Federation Of American Scientists. 2019 Jul 26. Department of Veterans Affairs (VA): A Primer on Telehealth URL: https://fas.org/sgp/crs/misc/R45834.pdf [accessed 2019-09-29]
22. Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev Pharmacoecon Outcomes Res 2008 Apr;8(2):165-178. [doi: 10.1586/14737167.8.2.165] [Medline: 20528406]
23. Liu SX, Xiang R, Lagor C, Liu N, Sullivan K. Economic modeling of heart failure telehealth programs: when do they become cost saving? Int J Telemed Appl 2016;2016:3289628 [FREE Full text] [doi: 10.1155/2016/3289628] [Medline: 27528868]
24. Grustam AS, Severens JL, de Massari D, Buyukkaramikli N, Koymans R, Vrijhoef HJ. Cost-effectiveness analysis in telehealth: a comparison between home telemonitoring, nurse telephone support, and usual care in chronic heart failure management. Value Health 2018 Jul;21(7):772-782 [FREE Full text] [doi: 10.1016/j.jval.2017.11.011] [Medline: 30005749]
25. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
26. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]

Abbreviations

CHF: chronic heart failure

DIG: Digitalis Investigation Group
DRG: diagnosis-related group
ICER: incremental cost-effectiveness ratio
NYHA: New York Heart Association
QALY: quality-adjusted life year
TIM-HF2: Efficacy of Telemedical Interventional Management in Patients with Heart Failure II
WTP: willingness-to-pay

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by T Muto, H Chih; comments to author 20.02.20; revised version received 26.02.20; accepted 29.02.20; published 06.07.20.

Please cite as:
Jiang X, Yao J, You JHS
Telemonitoring Versus Usual Care for Elderly Patients With Heart Failure Discharged From the Hospital in the United States: Cost-Effectiveness Analysis
JMIR Mhealth Uhealth 2020;8(7):e17846
URL: https://mhealth.jmir.org/2020/7/e17846
doi:10.2196/17846
PMID:32407288

Original Paper

Cognition in Context: Understanding the Everyday Predictors of Cognitive Performance in a New Era of Measurement

Emma Weizenbaum[1], BA, MA; John Torous[2], BSc, MD, MBI; Daniel Fulford[1,3], BA, MA, PhD

[1]Department of Psychological and Brain Sciences, Boston University, Boston, MA, United States

[2]Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[3]Department of Occupational Therapy and Rehabilitation Sciences, Boston University, Boston, MA, United States

Corresponding Author:
Emma Weizenbaum, BA, MA
Department of Psychological and Brain Sciences
Boston University
2nd Floor
900 Commonwealth Avenue
Boston, MA, 02215
United States
Phone: 1 512 217 2825
Email: eweizen@bu.edu

Abstract

Background: Research suggests that variability in attention and working memory scores, as seen across time points, may be a sensitive indicator of impairment compared with a singular score at one point in time. Given that fluctuation in cognitive performance is a meaningful metric of real-world function and trajectory, it is valuable to understand the internal state-based and environmental factors that could be driving these fluctuations in performance.

Objective: In this viewpoint, we argue for the use of repeated mobile assessment as a way to better understand how context shapes moment-to-moment cognitive performance. To elucidate potential factors that give rise to intraindividual variability, we highlight existing literature that has linked both internal and external modifying variables to a number of cognitive domains. We identify ways in which these variables could be measured using mobile assessment to capture them in ecologically meaningful settings (ie, in daily life). Finally, we describe a number of studies that have already begun to use mobile assessment to measure changes in real time cognitive performance in people's daily environments and the ways in which this burgeoning methodology may continue to advance the field.

Methods: This paper describes selected literature on contextual factors that examined how experimentally induced or self-reported contextual variables (ie, affect, motivation, time of day, environmental noise, physical activity, and social activity) related to tests of cognitive performance. We also selected papers that used mobile assessment of cognition; these papers were chosen for their use of high-frequency time-series measurement of cognition using a mobile device.

Results: Upon review of the relevant literature, it is evident that contextual factors have the potential to meaningfully impact cognitive performance when measured in laboratory and daily life environments. Although this research has shed light on the question of what gives rise to real-life variability in cognitive function (eg, affect and activity), many of the studies were limited by traditional methods of data collection (eg, involving retrospective recall). Furthermore, cognition has often been measured in one domain or in one age group, which does not allow us to extrapolate results to other cognitive domains and across the life span. On the basis of the literature reviewed, mobile assessment of cognition shows high levels of feasibility and validity and could be a useful method for capturing individual cognitive variability in real-world contexts via passive and active measures.

Conclusions: We propose that, through the use of mobile assessment, there is an opportunity to combine multiple sources of contextual and cognitive data. These data have the potential to provide individualized digital signatures that could improve diagnostic precision and lead to meaningful clinical outcomes in a wide range of psychiatric and neurological disorders.

(JMIR Mhealth Uhealth 2020;8(7):e14328) doi:10.2196/14328

KEYWORDS

smartphone; mobile phone; neuropsychology; individualized medicine

Introduction

Background

Thinking occurs dynamically. From day to day and moment to moment, our ability to hold information in mind or regulate attention is constantly in a state of flux. The sources of these fluctuations are varied—whether coming in the wake of mid-afternoon sleepiness, an anxiety-provoking presentation, or the confusion of navigating a crowded store, cognitive performance is influenced by the contexts in which a person operates. Despite this inherent variability, researchers and clinicians often assume that cognitive performance reflects internal processes that are constant and stable. This assumption leads to the characterizations of cognitive performance as static, treating the *mean* of an individual's or group's cognitive performance as the output of internal processes and dismissing variance in performance as *noise*. Research has found that these within-person changes in performance, commonly referred to as *intraindividual cognitive variability* [1,2], can serve as highly sensitive markers of cognitive dysfunction [3], such as in attention deficit hyperactivity disorder [4] and cognitive decline [5] or dementia [6,7]. Intraindividual variability has been used to describe change or dispersion across domains of function (eg, memory vs attention) [8], variability or inconsistency in response within a single measurement time point [9], or across a number of time points of varying scales (eg, moment-to-moment fluctuations) [10]. In this paper, we seek to identify the factors that give rise to changes in attention and working memory (given these domains' susceptibility to fluctuation) [11] *within individuals*, *across time* (eg, 1-2 weeks) and *contexts* (eg, physical environment and emotional state).

To effectively measure how contexts shape attention and working memory, dense sampling of cognitive performance across time and in various settings is essential. Measuring performance across a series of time points in traditional laboratory or clinic settings is arduous and expensive and represents a limited and unnatural context for cognitive function. Instead, by using an experience sampling method (ESM) or ecological momentary assessment (EMA) of frequent testing over the course of days or weeks [12], high-resolution measurements can be obtained, which give rise to a rich picture of the dynamics and patterns of an individual's cognitive functioning within a period of time [13]. Until recently, methods of studying frequency and patterns of behavior in daily life relied on pen-and-paper responses or frequent calls to participants that were burdensome and intrusive. However, the increasing ubiquity of mobile devices (eg, smartphones and tablets) allows for the feasible acquisition of high-frequency data [14,15], which can be collected in vivo, without the limitations of retrospective reporting. Mobile assessment data, in turn, produce a *digital phenotype* [16] (ie, a reflection of a person's moment-to-moment cognitive and behavioral function in the context of everyday life). The goal of digital phenotyping aligns with the advancement of precision medicine, where personalized models predict function based on relevant internal and environmental variables. The use of this method has become increasingly common in the fields of clinical psychology and psychiatry to understand the temporal patterns of symptoms [17,18] and how contextual factors, such as a person's physical and social environment, influence important clinical features, such as mood dysregulation [19]. However, digital phenotyping is just beginning to be validated within the field of neuropsychology as digital cognitive assessments are being developed and implemented [15,20,21].

Objective

On the basis of the research available to date, we propose that significant strides could be made in predicting meaningful health outcomes by assessing the everyday factors that influence cognition via smartphones or other mobile technologies. In this viewpoint, we begin by describing the importance of measuring intraindividual cognitive variability, focusing on the domains of attention and working memory. To better understand the factors that give rise to this variability, we explore how internal state-based variables (eg, affect and motivation) and external contextual variables (eg, time of day, surrounding noise, and activity) affect cognitive performance. Finally, we propose using mobile assessment as a means of integrating ambulatory cognitive assessment tasks with contextual and environmental data to produce digital phenotypes that could aid in diagnostic precision.

Intraindividual Variability in Cognition

Intraindividual variability can be observed at multiple time scales [1], and yet, the majority of literature has focused on trial-to-trial response variability within a single time point [11,22-25]. However, there is some evidence to suggest that intraindividual variability within and across days is positively correlated, signaling potentially similar processes, although the extent of this generalization is unclear [26]. Broadly defined, variability in attention and inhibition of behavioral response is thought to be supported by the prefrontal cortex, which serves to differentially activate the strength of specific networks depending on factors of context, state, and task demands [23]. One study found the regulation of task performance and adaptive modulation of attention to be localized to the right dorsolateral frontal regions that exert top-down attentional control [22]. Cognitive variability (ie, performance across course of a day) in attention and memory is known to increase in late adulthood [9,27]. Although the function and mechanisms of this increase are not fully understood [11], it seems to occur in both normal aging and varying disease processes [28], resulting in decreased frontal-executive function [29] and an observed decrease in dopamine response within extrastriatal regions [30]. Given that a non-negligible amount of response variability can occur from one time point to another [26], this may limit the generalizability of scores created from aggregate measures that are interpreted to reflect global functioning across many time points [31]. Instead of dismissing the error or noise in these estimates, it is valuable to understand contributors that give rise to the variability seen within one's performance, specifically how internal state-based and environmental factors shape one's cognitive performance from moment to moment.

Modifying Factors of Moment-to-Moment Cognition

Traditional approaches to measuring cognition seek to remove *noise*, which might impact an individual's performance, such

as time of day and distractors in the physical environment. However, in the context of daily life, these factors likely exert meaningful influence on one's overall function. With the emergence of mobile technology, we now have the opportunity to measure these factors in real time and quantify their influence on cognitive function. As there are only few prior studies that measure the influence of contextual factors on repeat cognitive performance [32-36], whether these factors produce reliable and meaningful patterns of variability remains equivocal. If consistent patterns of within-person variability at either the nomothetic or idiographic level are evident, this could signal opportunities for uniquely tailored real-time intervention points. The potential modifiers of cognitive performance that we will focus on in the following section include internal state-based factors (eg, affect and motivation) and external contextually based factors (eg, time of day, surrounding noise, and recent activity level; Table 1). In addition, we will discuss how each of these modifiers could be more optimally measured in relation to cognitive performance using mobile assessment.

Table 1. Internal and external modifiers of cognitive performance.

Reference	Modifier	Sample	Test location and task domain	Length	Result
Jefferies et al [37]	Mood	100 younger adults	In laboratory; visual attention	1 day	Low arousal with negative affect associated with best performance
Ellis et al [38]	Mood	160 younger adults	In laboratory; semantic recall	1 day	Depressed mood reduced semantic processing, interaction between depressed mood and task difficulty
Brose et al [39]	Mood and motivation	101 younger adults	In laboratory; working memory	100 days	Negative affect and low task motivation reduced working memory performance
Brose et al [40]	Mood and motivation	101 younger adults	In laboratory; working memory	100 days	Positive affect and high task motivation improved working memory performance
Sliwinski et al [41]	Mood	184 younger and older adults	In laboratory; working memory	1-2 weeks	Higher stress ratings associated with slower response time on working memory tasks, greater effect in older adults
Salthouse and Berish [36]	Mood	420 adults	Palm-pilot; reaction time	7 days	No relation between mood and reaction time scores
Krawczyk [42]	Motivation	16 younger adults	In scanner; working memory	1 day	Modulation of reward potential correlates with response time and functional magnetic resonance imaging blood oxygen level–dependent response
Yeo and Neal [43]	Motivation	99 younger adults	In laboratory; executive function	1 day	Motivation had the strongest influence on multistep task performance once task was learned and familiar
van der Heijden et al [44]	Time of day	2167 children	In laboratory; sustained attention	1 day	Performance on sustained attention slower but more accurate in the morning
Manly et al [45]	Time of day	10 younger adults	At home; sustained attention	4 days	Positive correlation between errors and sleepiness rating
West et al [46]	Time of day	40 younger and older adults	In laboratory; computer task	4 days	Younger adults performed best on working memory tasks in the evening, older adults performed best in the morning
Lange [47]	Noise	34 younger adults	In laboratory; working memory	1 day	Noise disrupted verbal but not visuospatial working memory performance
Bell and Buchner [48]	Noise	182 younger and 193 older adults	In laboratory; working memory	1 day	Same level of impairment on visual working memory from noise versus silence in younger and older adults
Ljungberg and Neely [49]	Noise	24 adults	In laboratory; reasoning and working memory	1 day	No significant effect of noise on performance, but higher levels of subjective task difficulty and stress ratings
Sibley and Beilock [48]	Activity	48 younger adults	In laboratory; working memory	2 days	Cardiovascular exercise significantly improves working memory
Whitbourne et al [50]	Activity	59 younger adults	Daily diary; subjective complaints	8 days	Older adults report fewer memory failures on days with exercise
Phillips et al [51]	Activity	51 older adults	In laboratory; reasoning and processing speed tasks	5 days	Physical activity accounted for significant within-person variance in cognition, especially processing speed
Allard et al [32]	Activity	60 older adults	Personal digital assistant; semantic memory task	7 days	Intellectually stimulating activities improved semantic memory performance measured later on the same-day
Bielak et al [35]	Activity	146 older adults	Web-based; processing speed, memory, and reasoning	7 days	Faster memory and processing speed on days with individual or small group social activities

Internal State-Driven Modifiers of Cognitive Performance

In this section, we will cover how a handful of previous studies have measured individuals' state affect, either through self-report or experimental manipulation, and how it relates to cognitive performance on specific tasks of attention, working memory, and recall. We then examine how momentary motivation, measured either from self-report or via reward manipulations, influences how well people perform on tasks of decision making and working memory. The studies described in this section are listed in Table 1.

Affect

The valence (ie, positive or negative) of one's momentary affect has been shown to impact cognitive performance in the domains of attention, working memory, and recall. Seminal work by Ellis et al [38] demonstrated that when negative affect was induced in healthy undergraduates, using self-referential negative statements, word recall was reduced compared with those in a neutral affect condition, particularly in challenging trials. The authors proposed that in states where strong affect is present, attention is diverted and fewer cognitive resources are allocated for the task at hand; the effects of this become most apparent on difficult tasks that require greater cognitive resources [38]. A separate study of cognition and affect found that visual attention (target detection) varied in relation to the combination of level of arousal and affect valence in a sample of healthy adults. The primary finding was that low arousal-negative affect (eg, sadness) was associated with greater accuracy on second but not the first target detection, indicating that sadness enhanced attention prioritization but not overall improvement of attention [37]; this could be because affect serves to shape the strength of attentional control or allocation [22].

In a study in which self-reported affect and working memory were assessed daily over a period of 6 months, Brose et al [39] found poorer working memory performance on days with greater negative affect. Conversely, on days when positive affect was higher, working memory performance was improved, which also related to higher task-related motivation [40]. In line with these findings, Sliwinski et al [41] identified a within-person association between higher daily stress ratings and slowed response time on a working memory task; this effect was particularly pronounced in older adults. However, one of the first studies to use repeat cognitive assessment of attention with a reaction time task using palm pilots (prompted 100 times over 7 days) found no association between momentary affect ratings and reaction times [36].

Taken together, there is evidence from the existing literature that suggests negative affect may reduce working memory and semantic recall abilities, at least in controlled settings. However, considerable heterogeneity in the main effects and interaction effects suggests that there is still much to be learned about the role of affect in cognitive performance, particularly when examined outside of the laboratory context and in clinical populations where affect dysregulation may be a primary symptom. In this new digital era, the consistency with which people carry smartphones could be useful in allowing for quick, momentary probes about an individual's current affect as it changes in daily life. Such assessments of affect could then be paired with smartphone cognitive assessments of different types (eg, attention, working memory, and recall) to provide a more detailed picture of how people's affect varies with task performance.

Motivation

Motivation is thought to drive the allocation of effort, which in turn affects cognitive performance [52]. Measuring the influence of motivation has long been a core feature of clinical neuropsychological assessments with the inclusion of measures designed to assess effort and engagement as checks for validity in testing [53]. However, although these standard measures of effort capture motivation in a given moment in a laboratory or clinic setting, in daily life, motivation varies and depends on factors such as affect, the given rewards or costs of a task, and practice. Ellis et al [38] theorized that strong affect decreased performance by diverting attentional resources to the affective experience versus the task at hand. Adding nuance to this explanation, Pessoa [54] proposed a dual-competition framework of affect, motivation, and executive control. In this framework, both affect and motivation are hypothesized to enhance or impair executive control depending on whether the emotion or locus of reward is aligned or divergent from the task objective. This conceptual model was exemplified through a study by another group and showed how the modulation of reward potential (ie, money) on each trial was positively correlated with both behavioral response time and blood oxygen level–dependent response (ie, functional magnetic resonance imaging–measured neural activation), with the hypothesized mechanism being increased motivation to perform at one's best when there is a relatively high payout [42]. Another factor underlying motivation's influence on cognitive performance is practice or task familiarity [43]. When accuracy was measured across 30 trials of a complex multistep judgment task, a person's self-reported motivation had a greater positive effect on performance in later trials. In other words, when a task is well learned, differences in motivation level have more of an effect on accuracy, compared with when a task is unfamiliar [43]. One study examining self-reported motivation and cognitive performance serially over 100 days found that higher daily motivation ratings were positively correlated with higher scores on verbal and spatial working memory tasks [40].

Although daily retrospective self-reports of motivation have been assessed in relation to working memory [39,40], what remains unknown is the extent to which naturally occurring fluctuations in state-based motivation throughout a given day might influence performance on tasks of working memory or other cognitive domains (eg, processing speed and recall). Using smartphone assessment, one could probe for real-time self-reported motivation before or after a test of cognitive function. Alternatively, tests could include passive or built-in measures of effort that determine the level of engagement in a given task. Simultaneously, sensing metrics, such as time of day and GPS location, could be used to determine the contextual factors that correlate with task engagement. Finally, motivation and cognition could be better studied in tandem via smartphone assessment through the gamification of cognitive tasks that are

modeled after smartphone games that participants may have prior familiarity with. In contrast to traditional tests of cognition that provide no immediate reward for the patient or participant, smartphone tasks can easily incorporate point systems, potentially tied to monetary or other rewards, which could allow for observational or experimentally induced modulation of task engagement.

External Contextual Modifiers of Cognitive Performance

In this section, we have selected studies that measure the influence that factors, outside of the person, have on cognitive performance in various domains, including attention, working memory, executive function, and memory. We will describe studies that seek to understand the impact of a person's environment, specifically the time of day (morning, afternoon, and evening), the quality and amplitude of environmental noise, and the impact of recent social and physical activity on individuals' task performance. The studies included in this section are also listed in Table 1.

Time of Day

Internal circadian rhythms, driven by the time of day, have been shown to impact cognitive performance. One study of healthy participants aged 10-12 years found that time of day impacted performance on challenging trials of visual working memory and processing speed tasks [44]. Another study measured working memory in the morning and at night across multiple days, in both younger and older adults. Rather than time of day, working memory performance varied with self-reported alertness, which was higher for younger adults in the evening and older adults in the morning [46]. Similarly, a study of young adults found that errors on a task of sustained attention increased with self-reported sleepiness [45]. A review of circadian rhythms and cognitive performance suggested that time of day had a significant effect on a wide range of cognitive tasks, including those of attention, executive functioning, and memory. Notably, performance fluctuation was linked to individual differences in peak circadian arousal [55]. In summary, this literature suggests there is a relation between cognitive function and time of day; however, this relationship seems to rely mostly on time of day's connection to intraindividual alertness. Mobile assessment can be particularly useful in relating the time of day to cognition using time stamps captured with assessments that can be scheduled at specific times or in random intervals throughout the course of a day. Furthermore, subjective alertness could be captured through momentary reports of wakefulness at the time of the cognitive assessment. Eventually, passive measurement of alertness may be collected using touch screen latencies when typing or through smartwatch sensors that measure heart rate and other biometrics.

Surrounding Noise

Distractions are inherent in a noisy or chaotic environment and intuitively impact cognitive performance. In the literature on auditory distractions, the primary area of study has been in the domain of working memory, where a series of findings suggests that the irregularity of sound is most impairing to cognitive performance, rather than the absolute volume of the noise in the environment [56]. Of all types of cognitive tasks, auditory working memory tasks involving maintenance of a series or order of information appear to be most affected by auditory distraction [56]. Similar to the proposed mechanism for affect and motivation, noise is thought to impair performance when attention is pulled away from the task at hand and toward task-irrelevant stimuli, particularly an irregular nonhabituated stimulus [57]. One study demonstrated a particularly large impact of sound on attentional abilities when the distractor and task were similar (eg, auditory-verbal), as this requires more resources for differentiation and suppression than when the task is more unique from the distractor (eg, visuospatial) [47]. Another study found that although distracting ambient noise did not affect objective performance on working memory and reasoning tasks, the addition of environmental noise was related to significantly higher subjective stress ratings and greater perceived difficulty of a given task [49]. Although most studies of noise and cognitive performance have used younger adults in experiments, there has been some investigation into age as a moderator of the influence of irrelevant noise on working memory [48]. When recorded ambient office-noise was played during a visual working memory task, there was a similar level of impairment for both younger and older adults [48]. All the studies above experimentally constructed a *noise condition* to study its effect on cognitive performance, and yet, noise is a product of unpredictable real-world environments that can be challenging to replicate in laboratory settings. Here, mobile assessment could allow for a real-time capture of ambient noise via the use of a phone's microphone or real-time self-reports of the characteristics of a given environment. Moving forward, sensing technology using advanced data analytics (eg, machine learning algorithms for sound detection) could be helpful in determining the type of audio in a given environment, for example, whether the sounds are of a noisy subway train or a conversation. These categories of sound could then be classified, measured in duration, and linked to cognitive function outcomes, providing individualized assessment results in the context of a person's unique set of daily environments.

Recent Physical and Social Activity

The frequency and recency of physical and social activities appear to impact intraindividual variability in cognitive performance. For example, in a study of healthy adults, those with lower baseline working memory capacity performed significantly better on working memory tasks when they engaged in 30 min of cardiovascular exercise immediately before the tasks [58]. Furthermore, in an 8-day daily diary study in young, middle-aged, and older adults, older adults reported fewer memory failures on the day of and day after physical activity; there was no effect of physical activity in younger adults [50]. In another study examining relationships between physical activity and cognitive performance in older adults, no significant correlation was found between same-day or day-to-day average physical activity and cognition. However, physical activity did explain significant variance in within-person performance in certain cognitive domains, most notably processing speed [51]. Furthermore, in mental illnesses that largely affect cognition, such as schizophrenia, a recent meta-analysis showed that aerobic exercise in these patients

was associated with improved attention and working memory [59]. Using smartphone assessment, physical activity can be measured objectively and unobtrusively via an accelerometer. This passive measurement could be enhanced with heart rate detection in associated devices such as smartwatches. Recording physical activity using mobile devices has the potential to add precision to our understanding of how duration, intensity, and recency of physical activity may influence subsequent cognitive performance.

Few studies have looked at the impact of one's recent recreational or social activities on moment-to-moment cognitive performance, and existing studies are generally specific to samples of healthy older adults. In one study, 60 older adults were prompted via personal digital assistant (PDA) 5 times per day for 1 week to answer questions about their location and to choose a category of their most recent activity. In addition, 2 out of these 5 daily assessments included a measure of semantic reasoning, where an overall category was selected in relation to a list of words. Results indicated that when participants reported having recently engaged in *intellectual activities* (eg, reading and crossword puzzles), scores on the semantic reasoning task were significantly higher [32]. A 7-day study of healthy older adults examined short-term associations between a number of daily activities and performance on web-based tests of processing speed, memory, and reasoning. Intellectual activities were not significantly correlated with cognitive scores, as seen in the study described above; instead, same-day individual or small group social interaction was associated with higher memory scores and response times. Of note, recent physical activity in this study was not associated with better cognitive performance [35]. These findings suggest a need to further explore the mechanisms by which recent social interaction improves cognitive performance and whether or not this finding is limited to older adults or to other specific individual differences.

Social activity could be measured through subjective reports of recent or current activity from smartphone assessments administered throughout the day. Furthermore, the identification of activities could be enhanced and less burdensome through GPS tagging of specific centers of activity such as school, work, friends' houses, or other locations where socialization takes place (eg, church, gym, and restaurants). Finally, smartphone assessments of cognition can be paired not only with identification of recent activity but also with important subjective ratings such as how important or enjoyable a person found their most recent activity, as this may be even more essential to cognitive performance than the activity type itself.

Limitations of Research on Modifiers of Cognitive Performance

In summary, research to date suggests that both internal (eg, affect and motivation) and external (eg, time of day, surrounding noise, and recent activity) variables can impact cognitive performance at any given moment. Although this research has shed light on the question of what gives rise to real-life variability in cognitive function, many of the abovementioned studies were limited by traditional methods of data collection. For example, testing cognition in a laboratory or clinical setting, across several time points, ignores the influence of one's real-world environment, activity levels, and social engagement. Furthermore, performance was often measured in one cognitive domain or in one age group, which does not allow us to extrapolate results to other cognitive constructs and to people across the life span. Finally, some studies relied on retrospective reports of contextual factors or subjective cognition, which are vulnerable to biases. To better understand the unique impact that potential internal and external modifying factors can have on intraindividual variability in cognition over time, we will need to improve the quality and ecological validity of our assessments, including gathering data at a higher temporal frequency and in real time.

Studying Intraindividual Variability in Cognition via Mobile Technology: Current Evidence

From the existing body of research, it is clear that internal and external factors of one's environment contribute to cognitive performance. Furthermore, a number of the studies described above suggest that individual differences play a role in determining which, and to what extent, contextual factors affect cognitive performance. Nonetheless, relatively few studies have used repeated measurements across days or weeks to establish temporal relationships between cognitive performance and internal and external modifiers of cognition. Limited research in this area is, at least in part, due to constraints accompanying traditional laboratory or clinic-based research designs. By employing repeated measurements via mobile assessment, there is an opportunity to better understand how cognition and its modifiers temporally covary in real-world environments.

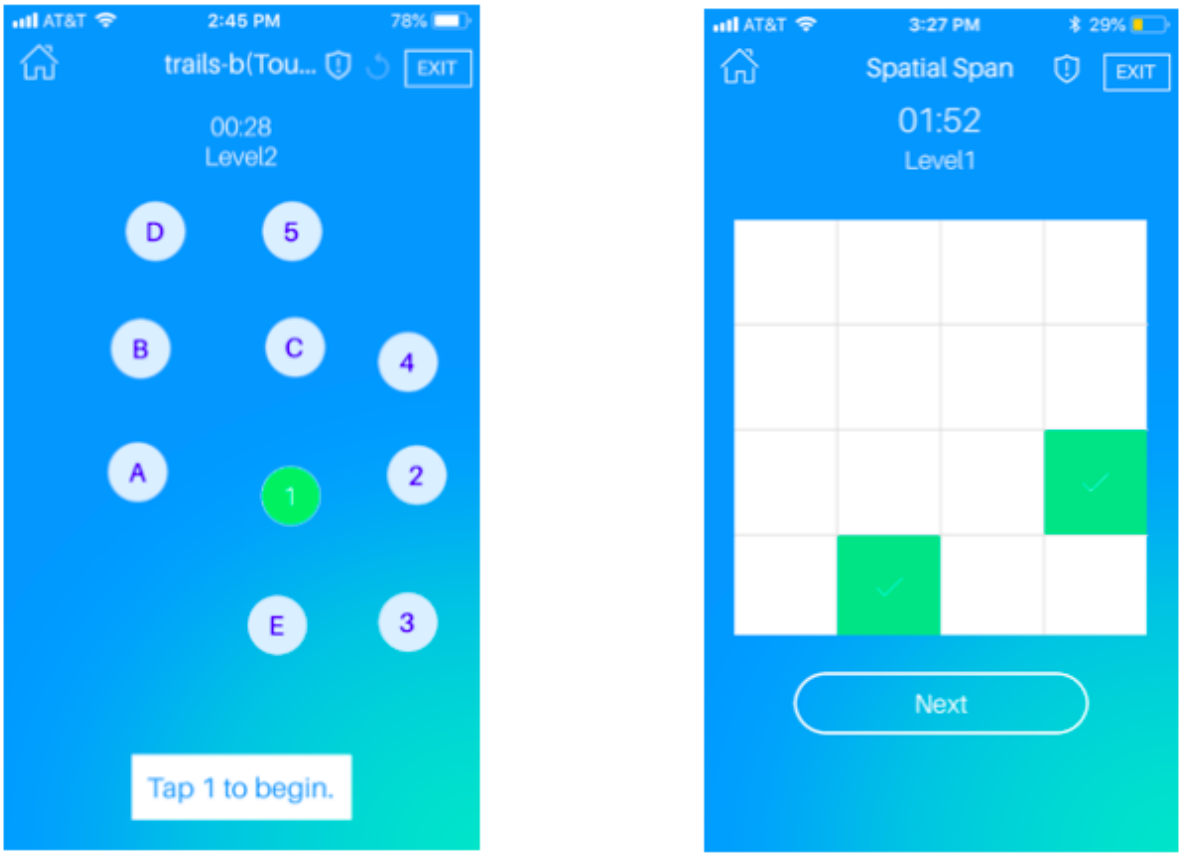
Building off research using ESM and EMA [12,60], mobile devices have been increasingly used in mental health assessment. Smartphone and tablet apps have the capacity to capture a person's unique fingerprint of response in different domains, ranging from establishing new behavioral habits to symptoms of severe mental illness or progressive neurological disorders. Mobile app–based assessments allow researchers and clinicians to reach people who face barriers to participating in laboratory-based studies or attending regular mental health care appointments [61]. One of the greatest benefits to using mobile devices, specifically smartphones, for ambulatory assessment is capitalizing on their ubiquity: owners carry them on-person most of the time and check them up to hundreds of times in a given day [62].

The incorporation of smartphone-based sensors, such as accelerometers, GPS, and microphones, is a new application of digital assessment that specifically allows for passive measurement of a person's real-world context and environment. For example, accelerometers capture a proxy for physical activity, which can then be characterized into activity types (sedentary, walking, and running) and correlated with metrics of cognitive variability. By identifying the patterns within a person's location data from GPS, we can observe how people interact with and traverse their environments through metrics such as distance and visit frequency over time. By taking advantage of a phone's built-in microphone, there is a potential to better understand the environmental context in which a person is thinking and behaving from moment to moment.

Mobile Assessment in Neuropsychology

In the field of neuropsychology, traditional pencil and paper tests have only just begun to be translated to computerized versions of these tests (Figure 1), and utilization rates for computerized assessment remain below 10% [63]. A review of studies comparing computerized neuropsychological tests with their analog versions found reliability was similar if not improved on computerized versions. However, normative data for pencil and paper tests could not be used interchangeably with the digital versions, given the smaller sample size for digital tests and differences in the sample characteristics [21].

Figure 1. Digital equivalents of traditional neuropsychological tests. In the mindLAMP app, the traditional Trailmaking Test B is translated to the smartphone screen (left). The task measures the accuracy and speed of finger taps between alternating numbers and letters. Spatial Span as an analog task involves a physical board with cubes; on a smartphone, squares light up in a sequential order, followed by a blank grid where the participant taps the same sequence of squares previously shown.



Beyond replicating existing tests, computerized versions of neuropsychological tests have yet to capitalize on additional passive or process-based data that are possible with mobile technology. For example, response latencies and adaptive testing methods can be used to identify a precise and unique range of within-person cognitive performance. To date, real-time cognitive data collection via ESM and EMA has yet to be fully embraced by the field of neuropsychology, but it has the potential to inform diagnosis and intervention efforts by capturing patients' fluctuation in cognition [64]. Although digital cognitive assessment is growing exponentially with the use of computers and mobile devices, some have argued that the field of neuropsychology is in the midst of *a technology crisis*, and they suggest that high-frequency data capture is one of the key methods the field should be leveraging [65-67].

A review by Moore et al [68] identified 12 studies that used self-administered mobile cognitive assessments ranging from 1 to 5 times per day over 1-14 days, across various age groups and populations. This review reported high adherence rates (80% on average) and strong construct validity, mostly in the domains of attention, working memory, and executive functioning. Notably, very few of these studies reviewed combined smartphone-based cognitive assessment with real-time measurement of other temporally dynamic contextual variables, and a number of studies only sampled cognition 1 time per day or less. For this methodology to assess potential links between variability in cognitive performance as it relates to context and state variables, a high frequency of measurement is key.

Review of High-Frequency Measurement Mobile Assessment Studies

To highlight the existing literature that has deployed high-frequency assessment, 14 studies were selected that used mobile paradigm-based assessments of cognition at a frequency of 2 or more times per day for at least one week (2). The majority of studies included in this study were conducted between 2014 and 2019. Out of the 14 studies, 6 included a nonclinical sample of young or middle-aged adults, 2 centered specifically on healthy older adults, and 1 in children. A total of 5 of the 14 studies included samples of substance users or

those at risk of smoking or alcohol abuse. Others featured clinical samples, including adults with depression and Parkinson disease. Eight out of the 14 studies, typically the more recent, used a smartphone for mobile assessment, and 5 of the 14 studies used a PDA device, one used a Nokia flip phone, and the other used a smartwatch. The majority of studies used working memory paradigms as the primary cognitive outcome measure, although other cognitive domains such as verbal and visual memory, attention, processing speed, and motor speed were also measured alone or alongside working memory in these studies. Of the 14 studies, 12 deployed repeat mobile tests for 1-2 weeks; other durations included 4, 6, and 24 weeks. A total of 6 studies prompted cognitive assessments 2-3 times per day, 6 studies prompted 4-6 times per day, and 2 studies prompted at an even higher frequency. In addition to cognitive performance, 4 out of 14 studies reported data collection related to momentary affect or mood, 3 out of 14 studies recorded reports of recent activity types or social settings, and 3 out of 14 studies examined time of day or fatigue in relation to cognitive performance. Generally, studies found high levels of concordant validity between mobile cognitive assessments and traditional in-laboratory measures.

Approximately half of the studies mentioned in Table 2 examined contextual or internal state-based variables in relation to cognitive performance; their findings on the association between context and cognition were mixed. Smartphone semantic reasoning and memory scores were greater after recent intellectually stimulating activity [32], slower smartphone-recorded reaction times were associated with greater mental fatigue [69], and increased error rates were seen in phone-based working memory and attention [70]. On the other hand, 2 studies conducted almost 15 years apart found no within-person associations between momentary mood ratings and working memory [71] and reaction time performance [36]. It should be noted that studies of repeat mobile cognitive assessment have used a variety of mobile platforms, different populations, and different ways of measuring the same contextual variables (eg, scales or dimensions of mood); as such, findings require replication using standardized methods.

Several studies commented on the psychometrics of reliable within-person fluctuations and the need to understand the drivers of these patterns of fluctuations. Dirk and Schmiedek [33] examined the psychometrics of repeated mobile assessment of cognition in children and found that moment-to-moment and day-to-day performance had reliable amounts of intraindividual variability, which was indicative of individual differences and patterns of response. Furthermore, more recent work by Sliwinski et al [15] suggested that within-person fluctuation in processing speed and working memory across the course of a day was reflective of meaningful, but unknown, contextual moderators of the intraindividual variability observed. This indicates the need to explore and identify the unknown contextual variables that influence variability in within-person cognitive functioning. Furthermore, nomothetic approaches to analyzing within-person contextually based modifiers of cognition may fail to produce reliable or meaningful findings, as the influence of particular contextual factors (eg, time of day) may be dependent on individual differences. In addition to statistical analysis that uses aggregate measures, idiographic approaches [78] could be used to understand the temporal dynamics between context, state, and cognitive function of a unique individual, which may be reliably different from another individual. Through the use of idiographic analyses such as individualized time-lagged modeling and network analyses, there is an opportunity to develop both personally targeted and ecologically generalizable interventions [31].

Table 2. Mobile assessment of cognition.

Reference	Sample	Assessment tool	Length (weeks)	Daily frequency	Cognitive domain	State or context variables	Result
Allard et al [32]	60 older adults	PDA[a]	1	2 times/day	Semantic reasoning and memory	Location, social setting, and recent activities or behaviors	Cognitive performance improved following intellectually stimulating activities
Cormack et al [71]	30 adults with depression	Apple watch	6	3 times/day	Working memory	Mood	High adherence, moderate concordance, and no relationship between momentary mood and cognition trajectories
Dagum [72]	27 young adults	Smartphone	1	Continuous	Working memory, executive function, and languages	Not specified	Digital biomarkers (eg, taps and swipes) highly correlated with traditional in-laboratory neuropsychological test scores
Dirk and Schmiedek [33]	110 participants aged 8-11 years	Smartphone	4	3 times/day	Working memory	Motivation, affect, sleep, and physical activity or accelerometer	Greater working memory variability measured by phone correlated with lower performance on in-laboratory cognitive and academic tests
Lipmeister et al [73]	44 patients with Parkinson disease; 35 controls	Smartphone	24	6 times/day	Motor speed	Motor symptoms	Phone tests of motor speed correlated with questionnaire measures and differentiated patients from controls
Pal et al [74]	12 meth addicts; 20 controls	Laboratory computer and smartphone	2	2 times/day	Working memory	Not specified	N-Back and Stop Signal on iPhone correlated with laboratory-based tests; speech detection on Stroop task did not work; no between-group differences
Price et al [69]	21 young adults	Smartphone	2	3 times/day	Working memory, attention, and processing speed	Mental fatigue	Fatigue ratings positively correlated with longer reaction times on attention task
Salthouse and Berish [36]	420 adults	PDA	1	15 times/day	Reaction time	Time of day and mood	Large within-person variability; no significant relation between time of probe or mood and reaction time
Sliwinski et al [15]	219 adults	Smartphone	2	5 times/day	Processing speed and working memory tasks	Not specified	High construct validity, reliability, and within-person variance
Schweitzer et al [75]	114 older adults	Smartphone	1	5 times/day	Memory and executive function	Physical environment and social interaction	High adherence and concordance with traditional neuropsychological test scores
Schuster et al [76]	39 high-risk smoker young adults	PDA	1	5-7 times/day	Working memory	Not specified	High feasibility or compliance and construct validity

Reference	Sample	Assessment tool	Length (weeks)	Daily frequency	Cognitive domain	State or context variables	Result
Tiplady et al [70]	38 adults who frequently consumed alcohol	Cell phone	2	2 times/day	Attention and working memory tasks	Alcohol consumption	Greater errors on phone- and laboratory-based tasks after alcohol consumption
Waters et al [77]	22 smokers; 22 controls	PDA	1	4 times/day	Working memory	State anxiety	High adherence and high reliability
Waters et al [34]	119 smokers	PDA	1	4 times/day	Attentional bias	Not specified	Between-subject craving and laboratory attentional bias associated with PDA Stroop attentional bias

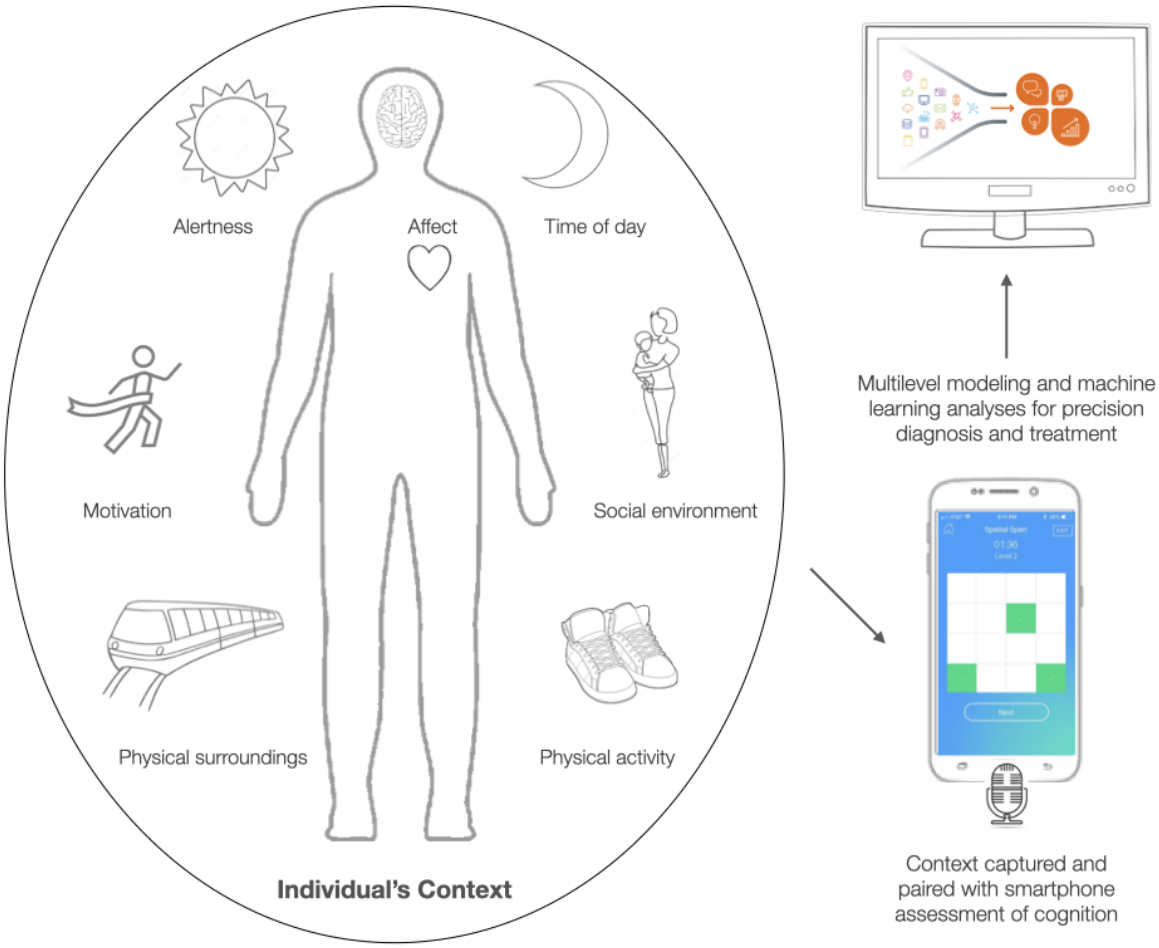
[a]PDA: personal digital assistant.

Implementation and Future Directions

Although new innovative forms of smartphone assessment are becoming increasingly common, there are several likely reasons for the dearth of research, to date, using mobile technology to study cognition and context. The first is the financial barrier to developing app-based assessment tools; in addition to this, there are significant time demands and logistical challenges of establishing the feasibility, reliability, and validity of these new instruments. In traditional neuropsychological assessment, the objective has been to measure cognitive function in a quiet setting with few distractions to obtain the best possible performance. However, when measuring cognitive function with a mobile device, there is an opportunity for ample variability and distractions. This contextual *noise* can, in fact, be a strength, reflecting a more ecologically valid environment. However, to form an accurate picture of a person's functioning across contexts, a large volume of data is needed. Analysis of acquired mobile assessment data will need to take into account the high frequency of measurements within individuals and between groups to produce clinically useful normative comparison data [79]. This brings forth the challenge of localizing the driving factors of within-person variance while also quantifying practice effects. Some have raised concerns about the use of computerized neuropsychological assessment due to the lack of standardization and normative data, which when applied to patient care could ultimately lead to poor clinical decisions and errors in diagnosis. There have been several calls to the field to address this concern by focusing on resources and investigating the psychometrics of mobile cognitive assessment [29,61,79] and centralizing the development of a toolbox of standardized mobile neuropsychological assessments that can be validated more efficiently through a network of clinical researchers across the field [80].

Inherent in pioneering mobile health technology is the challenge of developing measurement tools that are usable and study designs that are informative. However, unlike ever before, there is immense potential to study familiar neuropsychological constructs outside the laboratory setting and in the environments in which they naturally operate. To move beyond proof-of-concept validation of ambulatory assessments, the field is faced with the challenge of creating new paradigms that are not mere replications of existing neuropsychological measures but are designed to be independently administered across time, contexts, and devices. For example, moving beyond the gold standard neuropsychological measures of cognition, some studies have explored how smartphone activity, such as swipes and taps, could serve as a proxy for working memory, episodic memory, executive function, language, and intelligence in a pilot study of healthy adults [72]. The accessibility and availability of smartphones sets the stage for the emergence of new scalable research efforts that could easily recruit participants from both healthy and clinical samples and record vast quantities of new data. The results shown in Figure 2 are the ways in which some of the key ecologically situated variables can be measured using smartphones and connected to digital neuropsychological tasks performed on digital devices. The momentary and state-related variables can be measured passively (eg, GPS and microphone) or actively (eg, surveys). Together, data streams can be combined and analyzed using computational and machine learning methods to identify relationships between cognitive performance and contextual momentary factors.

Figure 2. Model of mobile assessment of intraindividual variability in cognition. Internal state-driven variables include affect motivation and alertness. External contextual variables include time of day, social environment, physical surroundings, and physical activity. Taken together, these factors give rise to fluctuations in cognitive performance. This can be captured in real time using a game-like smartphone assessment of cognition alongside sensing tools such as a smartphone microphone and GPS, which seamlessly capture information about one's environment. Advanced statistical methods can be used to analyze data to find patterns of intraindividual variability in cognition in real-world contexts.



Ultimately, the goal of mobile assessment methods is to refine the understanding of clinical impairments in cognition and give rise to preventative interventions for mitigating cognitive dysfunction and decline. New work is being done to develop a smartphone app designed to improve clinical care based on patients' and clinicians', in addition to researchers', needs. The objective of this app development is to create an open platform that can be used across a variety of research and clinical settings [18]. Unlike other existing mobile health apps developed for a specific population or function, the mindLAMP (Learn, Assess, Manage, Prevent) has been created to allow for customizable surveys, sensors, cognitive tests, and schedules of notification-prompted assessment. The app seeks to integrate active assessments such as surveys and cognitive games with passive sensing data (eg, GPS, pedometer, and microphone) and even phone metadata, such as the number of times other social media or communication apps are used on the phone. With the rich dimensionality of smartphone data from apps such as these, multilevel modeling and machine learning analyses would be logical directions to take analysis to elucidate the causal links and predictive relationships between cognitive performance and state-based and environmental factors.

Like any new method, repeat assessment of cognition and context via mobile technology must be carefully employed. Given the complexity and sheer data volume of temporarily dense, longitudinal, and dynamic smartphone and sensor data, it will be important for theory-driven hypotheses to guide this work instead of only data-driven models that will likely identify spurious correlations. Moreover, given the nature of these new data, care must be taken to ensure its privacy protections and ethical uses [81]. Looking back at the history of genetics and neuroimaging, it is clear that the greatest progress with this new spade will emerge from interdisciplinary collaborations. As health care moves toward personalized medicine, the field also has an opportunity to move toward personalized assessment of cognition. Although population-level screening tests and individualized in-office neuropsychological evaluations will remain important, they cannot fully capture the social, physical, and environmental variance that each individual experiences. Fortunately, mobile technologies such as smartphones can collect such data and thus present the opportunity for a paradigm shift with personal devices capturing personal measurements to deliver personal cognitive profiles and treatment directions.

Conflicts of Interest

None declared.

References

1. Nesselroade JR. Intraindividual variability and short-term change. Commentary. Gerontology 2004;50(1):44-47. [doi: 10.1159/000074389] [Medline: 14654727]
2. Fiske DW, Rice L. Intra-individual response variability. Psychol Bull 1955;52(3):217-250. [doi: 10.1037/h0045276] [Medline: 14371891]
3. Sonuga-Barke EJ, Castellanos FX. Spontaneous attentional fluctuations in impaired states and pathological conditions: a neurobiological hypothesis. Neurosci Biobehav Rev 2007;31(7):977-986. [doi: 10.1016/j.neubiorev.2007.02.005] [Medline: 17445893]
4. Kofler M, Alderson R, Raiker J, Bolden J, Sarver D, Rapport M. DigiNole. 2014. Working Memory and Intraindividual Variability as Neurocognitive Indicators in ADHA URL: http://diginole.lib.fsu.edu/islandora/object/fsu%3A209935/ [accessed 2020-05-15]
5. Roalf DR, Quarmley M, Mechanic-Hamilton D, Wolk DA, Arnold SE, Moberg PJ, Alzheimer's Disease Neuroimaging Initiative. Within-individual variability: an index for subtle change in neurocognition in mild cognitive impairment. J Alzheimers Dis 2016 Aug 10;54(1):325-335 [FREE Full text] [doi: 10.3233/JAD-160259] [Medline: 27567827]
6. Christ BU, Combrinck MI, Thomas KG. Both reaction time and accuracy measures of intraindividual variability predict cognitive performance in Alzheimer's disease. Front Hum Neurosci 2018;12:124 [FREE Full text] [doi: 10.3389/fnhum.2018.00124] [Medline: 29686610]
7. Hultsch DF, MacDonald SW, Hunter MA, Levy-Bencheton J, Strauss E. Intraindividual variability in cognitive performance in older adults: comparison of adults with mild dementia, adults with arthritis, and healthy adults. Neuropsychology 2000;14(4):588-598. [doi: 10.1037/0894-4105.14.4.588]
8. Hilborn JV, Strauss E, Hultsch DF, Hunter MA. Intraindividual variability across cognitive domains: investigation of dispersion levels and performance profiles in older adults. J Clin Exp Neuropsychol 2009 May;31(4):412-424. [doi: 10.1080/13803390802232659] [Medline: 18720183]
9. Hultsch DF, MacDonald SW, Dixon RA. Variability in reaction time performance of younger and older adults. J Gerontol B Psychol Sci Soc Sci 2002 Mar;57(2):P101-P115. [doi: 10.1093/geronb/57.2.p101] [Medline: 11867658]
10. Ram N, Gerstorf D. Time-structured and net intraindividual variability: tools for examining the development of dynamic characteristics and processes. Psychol Aging 2009 Dec;24(4):778-791 [FREE Full text] [doi: 10.1037/a0017915] [Medline: 20025395]
11. Braver TS. The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn Sci 2012 Feb;16(2):106-113 [FREE Full text] [doi: 10.1016/j.tics.2011.12.010] [Medline: 22245618]
12. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. J Nerv Ment Dis 1987 Sep;175(9):526-536. [doi: 10.1097/00005053-198709000-00004] [Medline: 3655778]
13. Sliwinski MJ. Measurement-burst designs for social health research. Social Pers Psych Compass 2008 Jan;2(1):245-261. [doi: 10.1111/j.1751-9004.2007.00043.x]
14. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
15. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
16. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
17. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
18. Torous J, Wisniewski H, Carpenter E. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: An interdisciplinary and collaborative approach. J Technol Bahav Sci 2019:- epub ahead of print. [doi: 10.3410/f.735622208.793567547]
19. Sandstrom GM, Lathia N, Mascolo C, Rentfrow PJ. Putting mood in context: using smartphones to examine how people feel in different locations. J Res Personal 2017 Aug;69:96-101. [doi: 10.1016/j.jrp.2016.06.004]
20. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: challenges and opportunities at the intersection of science and software. Clin Neuropsychol 2019 Feb;33(2):271-286. [doi: 10.1080/13854046.2018.1535662] [Medline: 30614374]
21. Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. Clin Neuropsychol 2013;27(8):1328-1361 [FREE Full text] [doi: 10.1080/13854046.2013.834971] [Medline: 24041037]
22. Stuss DT, Murphy KJ, Binns MA, Alexander MP. Staying on the job: the frontal lobes control individual performance variability. Brain 2003 Nov;126(Pt 11):2363-2380. [doi: 10.1093/brain/awg237] [Medline: 12876148]

23. Kelly AC, Uddin LQ, Biswal BB, Castellanos FX, Milham MP. Competition between functional brain networks mediates behavioral variability. Neuroimage 2008 Jan 1;39(1):527-537. [doi: 10.1016/j.neuroimage.2007.08.008] [Medline: 17919929]
24. Dinstein I, Heeger DJ, Behrmann M. Neural variability: friend or foe? Trends Cogn Sci 2015 Jun;19(6):322-328. [doi: 10.1016/j.tics.2015.04.005] [Medline: 25979849]
25. Vasquez BP, Binns MA, Anderson ND. Response time consistency is an indicator of executive control rather than global cognitive ability. J Int Neuropsychol Soc 2018 May;24(5):456-465. [doi: 10.1017/S1355617717001266] [Medline: 29208077]
26. Rabbitt P, Osman P, Moore B, Stollery B. There are stable individual differences in performance variability, both from moment to moment and from day to day. Q J Exp Psychol A 2001 Nov;54(4):981-1003. [doi: 10.1080/713756013] [Medline: 11765745]
27. Hertzog C, Dixon RA, Hultsch DF. Intraindividual change in text recall of the elderly. Brain Lang 1992 Apr;42(3):248-269. [doi: 10.1016/0093-934x(92)90100-s] [Medline: 1606486]
28. Troyer AK, Vandermorris S, Murphy KJ. Intraindividual variability in performance on associative memory tasks is elevated in amnestic mild cognitive impairment. Neuropsychologia 2016 Sep;90:110-116. [doi: 10.1016/j.neuropsychologia.2016.06.011] [Medline: 27297728]
29. Diehl M, Hooker K, Sliwinski M. Handbook of Intraindividual Variability Across the Life Span. New York, USA: Routledge; 2014.
30. MacDonald SW, Cervenka S, Farde L, Nyberg L, Bäckman L. Extrastriatal dopamine D2 receptor binding modulates intraindividual variability in episodic recognition and executive functioning. Neuropsychologia 2009 Sep;47(11):2299-2304. [doi: 10.1016/j.neuropsychologia.2009.01.016] [Medline: 19524093]
31. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. Proc Natl Acad Sci U S A 2018 Jul 3;115(27):E6106-E6115 [FREE Full text] [doi: 10.1073/pnas.1711978115] [Medline: 29915059]
32. Allard M, Husky M, Catheline G, Pelletier A, Dilharreguy B, Amieva H, et al. Mobile technologies in the early detection of cognitive decline. PLoS One 2014;9(12):e112197 [FREE Full text] [doi: 10.1371/journal.pone.0112197] [Medline: 25536290]
33. Dirk J, Schmiedek F. Fluctuations in elementary school children's working memory performance in the school context. J Educ Psychol 2016 Jul;108(5):722-739. [doi: 10.1037/edu0000076]
34. Waters AJ, Szeto EH, Wetter DW, Cinciripini PM, Robinson JD, Li Y. Cognition and craving during smoking cessation: an ecological momentary assessment study. Nicotine Tob Res 2014 May;16(Suppl 2):S111-S118 [FREE Full text] [doi: 10.1093/ntr/ntt108] [Medline: 23901053]
35. Bielak A, Mogle J, Sliwinski M. What did you do today? Variability in daily activities is related to variability in daily cognitive performance. J Gerontol B Psychol Sci Soc Sci 2019 Jun 14;74(5):764-771. [doi: 10.1093/geronb/gbx145] [Medline: 29240950]
36. Salthouse TA, Berish DE. Correlates of within-person (across-occasion) variability in reaction time. Neuropsychology 2005 Jan;19(1):77-87. [doi: 10.1037/0894-4105.19.1.77] [Medline: 15656765]
37. Jefferies LN, Smilek D, Eich E, Enns JT. Emotional valence and arousal interact in attentional control. Psychol Sci 2008 Mar;19(3):290-295. [doi: 10.1111/j.1467-9280.2008.02082.x] [Medline: 18315803]
38. Ellis HC, Thomas RL, Rodriguez IA. Emotional mood states and memory: elaborative encoding, semantic processing, and cognitive effort. J Exp Psychol Learn Mem Cogn 1984 Jul;10(3):470-482. [doi: 10.1037//0278-7393.10.3.470] [Medline: 6235313]
39. Brose A, Schmiedek F, Lövdén M, Lindenberger U. Daily variability in working memory is coupled with negative affect: the role of attention and motivation. Emotion 2012 Jun;12(3):605-617. [doi: 10.1037/a0024436] [Medline: 21787075]
40. Brose A, Lövdén M, Schmiedek F. Daily fluctuations in positive affect positively co-vary with working memory performance. Emotion 2014 Feb;14(1):1-6. [doi: 10.1037/a0035210] [Medline: 24364855]
41. Sliwinski MJ, Smyth JM, Hofer SM, Stawski RS. Intraindividual coupling of daily stress and cognition. Psychol Aging 2006 Sep;21(3):545-557 [FREE Full text] [doi: 10.1037/0882-7974.21.3.545] [Medline: 16953716]
42. Krawczyk DC, Gazzaley A, D'Esposito M. Reward modulation of prefrontal and visual association cortex during an incentive working memory task. Brain Res 2007 Apr 13;1141:168-177. [doi: 10.1016/j.brainres.2007.01.052] [Medline: 17320835]
43. Yeo GB, Neal A. A multilevel analysis of effort, practice, and performance: effects; of ability, conscientiousness, and goal orientation. J Appl Psychol 2004 Apr;89(2):231-247. [doi: 10.1037/0021-9010.89.2.231] [Medline: 15065972]
44. van der Heijden KB, de Sonneville LM, Althaus M. Time-of-day effects on cognition in preadolescents: a trails study. Chronobiol Int 2010 Oct;27(9-10):1870-1894. [doi: 10.3109/07420528.2010.516047] [Medline: 20969529]
45. Manly T, Lewis GH, Robertson IH, Watson PC, Datta AK. Coffee in the cornflakes: time-of-day as a modulator of executive response control. Neuropsychologia 2002;40(1):1-6. [doi: 10.1016/s0028-3932(01)00086-0] [Medline: 11595257]
46. West R, Murphy KJ, Armilio ML, Craik F, Stuss DT. Effects of time of day on age differences in working memory. J Gerontol B Psychol Sci Soc Sci 2002 Jan 01;57(1):P3-P10 [FREE Full text] [doi: 10.1093/geronb/57.1.P3]
47. Lange EB. Disruption of attention by irrelevant stimuli in serial recall. J Mem Lang 2005 Oct;53(4):513-531. [doi: 10.1016/j.jml.2005.07.002]

48. Bell R, Buchner A. Equivalent irrelevant-sound effects for old and young adults. Mem Cognit 2007 Mar;35(2):352-364. [doi: 10.3758/bf03193456] [Medline: 17645176]
49. Ljungberg JK, Neely G. Stress, subjective experience and cognitive performance during exposure to noise and vibration. J Environ Psychol 2007 Mar;27(1):44-54. [doi: 10.1016/j.jenvp.2006.12.003]
50. Whitbourne SB, Neupert SD, Lachman ME. Daily physical activity: relation to everyday memory in adulthood. J Appl Gerontol 2008 Jun;27(3):331-349 [FREE Full text] [doi: 10.1177/0733464807312175] [Medline: 25067882]
51. Phillips CB, Edwards JD, Andel R, Kilpatrick M. Daily physical activity and cognitive function variability in older adults. J Aging Phys Act 2016 Apr;24(2):256-267. [doi: 10.1123/japa.2014-0222] [Medline: 26371890]
52. Katzell RA, Thompson DE. An integrative model of work attitudes, motivation, and performance. Hum Perform 1990 Jun;3(2):63-85. [doi: 10.1207/s15327043hup0302_1]
53. Bigler ED. Symptom validity testing, effort, and neuropsychological assessment. J Int Neuropsychol Soc 2012 Jul;18(4):632-640. [doi: 10.1017/s1355617712000252] [Medline: 23057080]
54. Pessoa L. How do emotion and motivation direct executive control? Trends Cogn Sci 2009 Apr;13(4):160-166 [FREE Full text] [doi: 10.1016/j.tics.2009.01.006] [Medline: 19285913]
55. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: circadian rhythms in human cognition. Cogn Neuropsychol 2007 Oct;24(7):755-789. [doi: 10.1080/02643290701754158] [Medline: 18066734]
56. Banbury SP, Macken WJ, Tremblay S, Jones DM. Auditory distraction and short-term memory: phenomena and practical implications. Hum Factors 2001;43(1):12-29. [doi: 10.1518/001872001775992462] [Medline: 11474757]
57. Cowan N. Attention and Memory: An Integrated Framework. London, UK: Oxford University Press; 1998.
58. Sibley BA, Beilock SL. Exercise and working memory: an individual differences investigation. J Sport Exerc Psychol 2007 Dec;29(6):783-791. [doi: 10.1123/jsep.29.6.783] [Medline: 18089904]
59. Firth J, Cotter J, Elliott R, French P, Yung AR. A systematic review and meta-analysis of exercise interventions in schizophrenia patients. Psychol Med 2015 May;45(7):1343-1361. [doi: 10.1017/S0033291714003110] [Medline: 25650668]
60. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
61. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
62. Wilcockson TD, Ellis DA, Shaw H. Determining typical smartphone usage: what data do we need? Cyberpsychol Behav Soc Netw 2018 Jun;21(6):395-398. [doi: 10.1089/cyber.2017.0652] [Medline: 29781716]
63. Rabin LA, Spadaccini AT, Brodale DL, Grant KS, Elbulok-Charcape MM, Barr WB. Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. Prof Psychol Res Pract 2014 Oct;45(5):368-377. [doi: 10.1037/a0037987]
64. Casaletto KB, Heaton RK. Neuropsychological assessment: past and future. J Int Neuropsychol Soc 2017 Oct;23(9-10):778-790 [FREE Full text] [doi: 10.1017/S1355617717001060] [Medline: 29198281]
65. Miller J, Barr W. The technology crisis in neuropsychology. Arch Clin Neuropsychol 2017 Aug 1;32(5):541-554. [doi: 10.1093/arclin/acx050] [Medline: 28541383]
66. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019 Jan;3(1):igy038 [FREE Full text] [doi: 10.1093/geroni/igy038] [Medline: 30619948]
67. Sternin A, Burns A, Owen AM. Thirty-five years of computerized cognitive assessment of aging-where are we now? Diagnostics (Basel) 2019 Sep 6;9(3):114 [FREE Full text] [doi: 10.3390/diagnostics9030114] [Medline: 31489940]
68. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 2017 Dec;26(4):e1562 [FREE Full text] [doi: 10.1002/mpr.1562] [Medline: 28370881]
69. Price E, Moore G, Galway L, Linden M. Validation of a smartphone-based approach to in situ cognitive fatigue assessment. JMIR Mhealth Uhealth 2017 Aug 17;5(8):e125 [FREE Full text] [doi: 10.2196/mhealth.6333] [Medline: 28818818]
70. Tiplady B, Oshinowo B, Thomson J, Drummond GB. Alcohol and cognitive function: assessment in everyday life and laboratory settings using mobile phones. Alcohol Clin Exp Res 2009 Dec;33(12):2094-2102. [doi: 10.1111/j.1530-0277.2009.01049.x] [Medline: 19740132]
71. Cormack F, McCue M, Taptiklis N, Skirrow C, Glazer E, Panagopoulos E, et al. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. JMIR Ment Health 2019 Nov 18;6(11):e12814 [FREE Full text] [doi: 10.2196/12814] [Medline: 31738172]
72. Dagum P. Digital biomarkers of cognitive function. NPJ Digit Med 2018;1:10 [FREE Full text] [doi: 10.1038/s41746-018-0018-4] [Medline: 31304295]
73. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]

74. Pal R, Mendelson J, Clavier O, Baggott MJ, Coyle JY, Galloway GP. Development and testing of a smartphone-based cognitive/neuropsychological evaluation system for substance abusers. J Psychoactive Drugs 2016;48(4):288-294. [doi: 10.1080/02791072.2016.1191093] [Medline: 27260123]
75. Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. Int J Methods Psychiatr Res 2017 Sep;26(3) [FREE Full text] [doi: 10.1002/mpr.1521] [Medline: 27539327]
76. Schuster RM, Mermelstein RJ, Hedeker D. Acceptability and feasibility of a visual working memory task in an ecological momentary assessment paradigm. Psychol Assess 2015 Dec;27(4):1463-1470 [FREE Full text] [doi: 10.1037/pas0000138] [Medline: 25894710]
77. Waters AJ, Li Y. Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. Psychopharmacology (Berl) 2008 Mar;197(1):25-35. [doi: 10.1007/s00213-007-1006-6] [Medline: 18064440]
78. Fisher AJ, Reeves JW, Lawyer G, Medaglia JD, Rubel JA. Exploring the idiographic dynamics of mood and anxiety via network analysis. J Abnorm Psychol 2017 Nov;126(8):1044-1056. [doi: 10.1037/abn0000311] [Medline: 29154565]
79. Cohen AS. Advancing ambulatory biobehavioral technologies beyond 'proof of concept': introduction to the special section. Psychol Assess 2019 Mar;31(3):277-284. [doi: 10.1037/pas0000694] [Medline: 30802113]
80. Gershon R, Weiner M, Rentz D. Mobile Toolbox: Digital Cognition. MobileToolbox for Monitoring Cognitive Functioning (MTB) URL: https://www.mobiletoolbox.org/mobiletoolbox [accessed 2020-05-15]
81. Kane RL, Parsons TD. The Role of Technology in Clinical Neuropsychology. Oxford, UK: Oxford University Press; 2017.

Abbreviations

EMA: ecological momentary assessment
ESM: experience sampling method
PDA: personal digital assistant

Edited by G Eysenbach; submitted 13.04.19; peer-reviewed by L Germine, R Moore, D Bakker, S Jungmann; comments to author 03.10.19; revised version received 02.03.20; accepted 26.03.20; published 23.07.20.

Please cite as:
Weizenbaum E, Torous J, Fulford D
Cognition in Context: Understanding the Everyday Predictors of Cognitive Performance in a New Era of Measurement
JMIR Mhealth Uhealth 2020;8(7):e14328
URL: https://mhealth.jmir.org/2020/7/e14328
doi:10.2196/14328
PMID:32706680

Original Paper

Ecological Momentary Assessment Within a Digital Health Intervention for Reminiscence in Persons With Dementia and Caregivers: User Engagement Study

Courtney Potts[1], BSc, MSc; Raymond Bond[1], BCompSc, PhD; Assumpta Ryan[2], BSc, MEd, PhD; Maurice Mulvenna[1], BSc, MPhil, PhD; Claire McCauley[2], BPharm, MSc, PhD; Elizabeth Laird[2], BSc, MSc, PhD; Deborah Goode[2], BA, MSc, PhD

[1]Ulster University, School of Computing, Jordanstown, United Kingdom

[2]Ulster University, Institute of Nursing and Health Research, Derry, United Kingdom

Corresponding Author:
Maurice Mulvenna, BSc, MPhil, PhD
Ulster University
School of Computing
Shore Road
Jordanstown
United Kingdom
Phone: 44 28 9036 8602
Email: md.mulvenna@ulster.ac.uk

Abstract

Background: User-interaction event logs provide rich and large data sets that can provide valuable insights into how people engage with technology. Approaches such as ecological momentary assessment (EMA) can be used to gather accurate real-time data in an individual's natural environment by asking questions at any given instant.

Objective: The purpose of this study was to evaluate user engagement and responses to EMA questions using InspireD, an app used for reminiscence by persons with dementia and their caregivers. Research findings can be used to inform EMA use within digital health interventions.

Methods: A feasibility trial was conducted in which participants (n=56) used the InspireD app over a 12-week period. Participants were a mean age of 73 (SD 13) and were either persons with dementia (n=28) or their caregivers (n=28). Questions, which they could either answer or choose to dismiss, were presented to participants at various instants after reminiscence with personal or generic photos, videos, and music. Presentation and dismissal rates for questions were compared by hour of the day and by trial week to investigate user engagement.

Results: Overall engagement was high, with 69.1% of questions answered when presented. Questions that were presented in the evening had the lowest dismissal rate; the dismissal rate for questions presented at 9 PM was significantly lower than the dismissal rate for questions presented at 11 AM (9 PM: 10%; 11 AM: 50%; χ^2_1=21.4, P<.001). Questions asked following reminiscence with personal media, especially those asked after personal photos, were less likely to be answered compared to those asked after other media. In contrast, questions asked after the user had listened to generic media, in particular those asked after generic music, were much more likely to be answered.

Conclusions: The main limitation of our study was the lack of generalizability of results to a larger population given the quasi-experimental design and older demographic where half of participants were persons with dementia; however, this study shows that older people are willing to participate and engage in EMA. Based on this study, we propose a series of recommendations for app design to increase user engagement with EMA. These include presenting questions no more than once per day, after 8 PM in the evening, and only if the user is not trying to complete a task within the app.

(JMIR Mhealth Uhealth 2020;8(7):e17120) doi:10.2196/17120

KEYWORDS

ecological momentary assessment; EMA; app; behaviour analytics; event logging; dementia; carers; reminiscence; reminiscing; mHealth

Introduction

Background

Digital health and well-being products such as health apps are becoming increasingly popular given that mobile technology is ubiquitous in daily life. In addition to the user data that are recorded during use of these platforms, all user interactions and events can be logged to represent usage. Such user interaction and event logs provide rich and large data sets that can reveal valuable insights into how people engage with technology. This case study explored user engagement with ecological momentary assessment (EMA) in an app designed for persons with dementia and their caregivers.

Dementia includes a group of symptoms associated with ongoing cognitive decline and is highly prevalent with approximately 50 million cases worldwide [1]. Within the United Kingdom, 1 in 6 people over the age of 80 have dementia, and the incidence of dementia is projected to double by the year 2050 [2]. Dementia also has a large economic impact; it is estimated to cost £26 billion (approximately US $33 billion) a year in the UK alone [3]. Currently, there is no treatment that can prevent, cure, or slow the progression of dementia. Pharmacologic treatments such as antipsychotic medications have been used to treat the symptoms of dementia but with limited success and unwanted side effects [4]; therefore, nonpharmacologic interventions such as reminiscence are increasingly considered in dementia care. Reminiscence has been defined as recall of events in a person's life either individually or with others [5]. The process of reminiscence can involve the use of prompts such as photographs, music, and videos to trigger memories that have a special meaning for a person. Engaging in reminiscence increases sociability, confirms personal identity, and encourages feelings of self-worth [6,7]. A recent review [8] found evidence that reminiscence for those living with dementia helped to enhance quality of life in care homes and benefited those who felt depressed in an individual setting.

Many different techniques (such as EMA or experience sampling methodology) can be used to gather accurate data on daily living. These methodologies are used to capture real-time data in an individual's natural environment through repeated sampling [9]. This can include psychometric scales, open-ended questions, or anything else that can be used to assess an individual's condition at any given place and time. These approaches provide a high degree of ecological validity as they study people as they go about their day-to-day lives [9]. Since EMA requires that participants respond to questions at a given moment, it avoids recall bias which makes it a useful tool for those with memory impairment (such as persons with dementia). Traditionally, EMA used paper diary techniques, but now, devices such smartphones or tablets can be used to record digital data. Recently, a study [10] used EMA to identify major areas of concern for caregivers of persons with Alzheimer disease; the study's overall goal was to provide support and information for caregivers in their home. Another study used experience sampling methodology to examine the day-to-day effects of caregiving on dementia caregivers which could be used to tailor interventions to their individual needs [11]; however, there has been a paucity of research to date examining the use of EMA to sample data from both persons with dementia and their caregivers.

Previous InspireD Study

In our previous work, we developed an app for reminiscence which incorporated EMA for the Individual Specific Reminiscence in Dementia (InspireD) feasibility study. The study used a quasi-experimental design and investigated the use of an iPad app to allow people living with mild to moderate dementia and their family caregivers to reminisce (Multimedia Appendix 1). In phase one, a group of volunteers which included caregivers and persons with dementia cocreated, refined, and developed the app for InspireD. The app allowed users to electronically collect and store personal and generic media in the form of music, photographs, and videos. Phase two involved implementation which included training and use of the app with persons with dementia and their caregivers. Participants used the app at home for a period of 12 weeks. The primary outcome measure was the impact of reminiscence on mutuality [12] defined as the positive quality of the relationship between caregiver and care receiver. The secondary outcome measures included well-being measured using the 5-item World Health Organization Well-Being Index [13] and quality of the relationship between the person living with dementia and their caregiver using Quality of the Carer-Patient Relationship [14]. The third and final phase included individual interviews with participants to explore individual views on the intervention.

Reminiscing made up the largest proportion of total app interactions for persons with dementia (71%) and their caregivers (47%) [15]. Across both groups, there were more interactions with photographs than interactions with music and video [15]. The app was primarily used for reminiscence using personal multimedia content as opposed to generic photos and videos. [15]. The most popular times to use the app were around 11 AM, 3 PM, and 8 PM corresponding to postbreakfast, postlunch, and postevening mealtimes [15]. On average, a person living with dementia used the app about once per week over the 12-week trial period [15].

Participants living with dementia attained statistically significant increases in mutuality, quality of caregiver and patient relationship, and subjective well-being from the beginning to end of the study [16]. Additionally, unsupervised machine learning was used to identify behavioral clusters that characterized different user engagement with the InspireD app which were cross compared with qualitative data following interviews after the trial period [17].

Objectives

In this study, we sought to address four research questions: (1) What is the temporal engagement with EMA questions over hours of the day and across the trial period? (2) How differently do persons with dementia engage in responding to questions compared to how caregivers engage? (3) How do persons with dementia and their caregivers engage in responding to questions after reminiscing with video, audio, and photos? (4) How do persons with dementia and their caregivers engage in responding to questions after reminiscing with personal media compared

to how they respond after reminiscing with generic media? The aim of this study was to evaluate engagement with EMA questions while using an app with the overall goal to inform EMA use within digital health interventions.

Methods

Study Design

A user engagement study was conducted in which participants used a digital health intervention, an app for reminiscence, at home for a trial period of 12 weeks. EMA questions on mutuality were presented to users at various points during the trial period while they were reminiscing using the app.

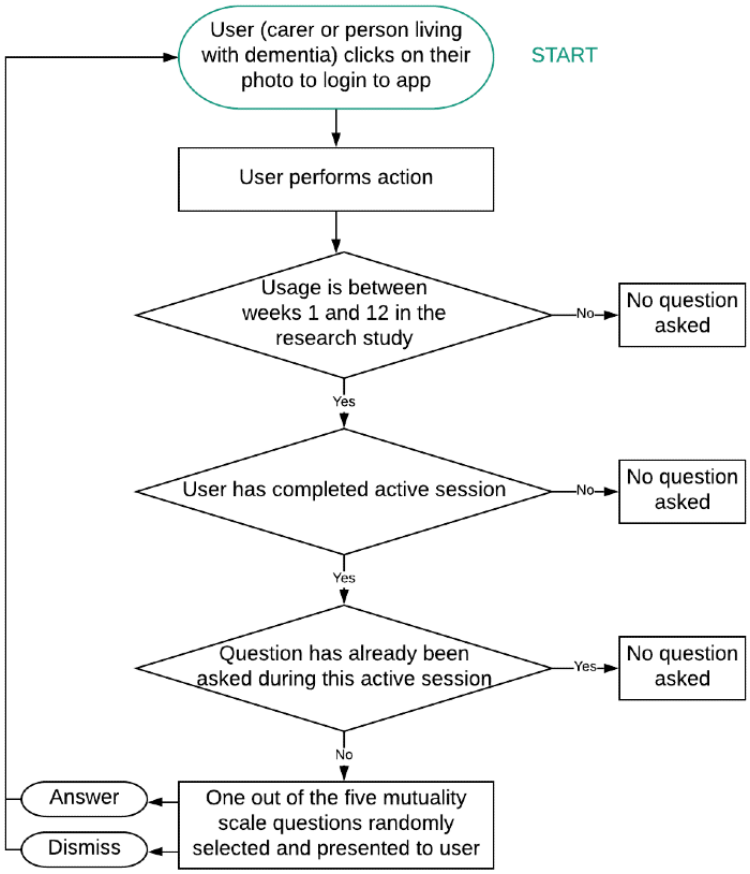
Participants

The study received ethical approval in March 2016 (16/NI/0035) in line with regional and national Health Service Trust research governance. Ethical considerations were ensuring voluntary participation by supporting separate informed consent for the persons with dementia and their caregivers, secure handling and storage of data in line with University policy, and reminding participants of their right to withdraw from the study at any time. In this study, for each person with dementia who participated, a family caregiver also participated. Each dyad (person with dementia and their caregiver) was given an iPad and asked to use the app at home for a 12-week period. Participants were encouraged verbally and in writing to use the app at least 3 times per week during the trial period. Information technology training was provided on 3 occasions; 2 sessions were provided before the trial period began, and 1 session was provided in week 6 of the trial period. The purpose of the training was to provide guidance on how to use the app, how to upload media, and to provide general app support. The app recorded user event logs locally on each iPad using an SQLite database (a public domain structured query language database). User event log data were later collected in person from the iPads using a portable storage device. Specific activities and activity types were logged: entry (logging in), administrative activities (adding or deleting photos, videos, or music), reminiscing (viewing photos, videos, or listening to music), queries at a given instant (EMA questions), and exiting (logging out). The app allowed individuals to upload their own photos, videos, and music or to access media online. For the purposes of this study, generic media was defined as photos, videos, and music that were accessed online through the app, for example, a photograph of street where the individual lived as a child. Personal media was defined as media that was uploaded, such as an old family photograph.

EMA

The EMA questions were a subset of items derived from the Mutuality Scale [12]. Over the 12-week trial period, the EMA questions were presented at random while individuals were using the app; therefore, questions were only displayed when the app was already in use. Hence, no push notifications were sent when the app was not in use to encourage completion of EMA questions. While the user was carrying out an action within the app (Multimedia Appendix 2), a series of checks were carried out before an EMA question was displayed (Figure 1). The user could either choose to answer the question using 5-point Likert scale (a great deal, quite a bit, some, a little, not at all) or choose to dismiss the question. This study did not analyze the responses that were provided, only whether the user chose to answer or dismiss the question.

Figure 1. Logic flowchart for ecological momentary assessment questions presented to people living with dementia and their caregivers (carers). If these conditions are met, then a question is presented to the user who answers or dismisses the question. Subsequently, the loop begins again.



Data Analysis

R studio (version 3.6.0) and the R programming language were used for data wrangling and statistical programming. For the first part of the analysis, the user log data were filtered to only contain data related to queries at a given instant in order to look at overall engagement with questions. Over the course of the 12-week trial period, a total of 832 questions were presented. Of these questions, 77 were asked during training days and so were excluded from this analysis, leaving 755 questions asked during the trial period.

Data were aggregated by time of day (hourly), and by week (weeks 1 to 12) to obtain question presentation counts and dismissal rates for each user type (persons with dementia or caregivers). Data were aggregated over the 12-week trial period to obtain total question presentation counts and dismissal rates for each user type (persons with dementia, or caregivers), media format (photo, video, music), media source (personal or generic), and the last logged activity (log-in, add or delete photo, exit app). Some media could not be categorized into generic or personal media and were, therefore, only included in the aggregate media category (all). For last logged activity, 124 out of 628 (16%) EMA questions were filtered out of the data since the previous recorded event had been an EMA question and not an action that could trigger a question (Multimedia Appendix 2).

Spearman rank correlation coefficients were used for association analysis between the number of questions presented and the number of interactions with the app, number of app interactions and the number of questions presented across trial weeks, dismissal rate and time of day, dismissal rate and trial week, number of questions presented and the number of questions dismissed per hour, where $P<.05$ was considered statistically significant. Chi-square tests were used to compare the proportion of questions answered versus dismissed across hours of the day, across trial weeks and for other logged activities—log-in, add or delete photo, exit app. Pairwise chi-square tests (n=36) adjusted for multiple testing using Bonferroni correction were used to compare dismissal rates for each question following reminiscence with all photos, videos and music, and then segregated into personal and generic photos, videos and music ($\alpha=.05/36$; therefore $P<.001$ was considered statistically significant).

Results

Participants

A total of 30 dyads (person with dementia and their caregiver) were recruited to the study; however, usage data was found to be corrupted in the iPad software used by 2 participants with dementia and their caregivers. Therefore, only 28 sets of tracking data (n=56 participants) were analyzed in this study. The characteristics of the participants are shown in Table 1.

Table 1. Characteristics of participants in the InspireD study.

Characteristics	All participants (N=56)	Persons with dementia (n=28)	Family caregivers (n=28)
Age (years)			
mean (SD)	73 (13)	79 (12.1)	67 (13)
range	31-94	61-94	31-91
Gender, n (%)			
Male	24 (43)	18 (64)	6 (21)
Female	32 (57)	10 (36)	22 (79)
Previous information technology experience, n (%)			
Little or none	33 (59)	23 (82)	10 (36)
Some	19 (34)	4 (14)	15 (53)
A lot	4 (7)	1 (4)	3 (11)

General Engagement

Five different EMA questions from the Mutuality Scale [12] were presented to persons with dementia and their caregivers (Table 2). There was a significant correlation between the number of questions presented and the number of interactions with the app (ρ=0.86, *P*<.001), with roughly 1 question asked for every 10 app interactions. The overall dismissal rate for questions asked during the trial period excluding training days was 30.9%. Hence, 522 out of 755 (69.1%) questions were answered. Persons with dementia used the app more in the trial period than caregivers did [15], but despite this had a lower dismissal rate for questions (121/451, 26.8%) compared to that of their caregivers (112/304, 36.8%).

A breakdown of presentation and dismissal for each question is shown in Table 2 along with chi-square test results comparing the proportion of questions answered versus dismissed. The dismissal rates were significantly different from the answer rates for each question (Table 2).

Table 2. Ecological momentary assessment questions presented and dismissed.

Questions		Presented, n (%)	Dismissed, n (%)	Chi-square (*df*)	*P* value
1	How attached are you to {partners name}?	197 (26.1)	48 (20.6)	101.5 (1)	<.001
2	How much do the two of you laugh together?	131 (17.4)	39 (16.7)	30.9 (1)	<.001
3	How much do you confide in {partners name}?	116 (15.4)	38 (16.7)	17.7 (1)	<.001
4	How much do you enjoy sharing past experiences with {partners name}?	165 (21.9)	54 (23.2)	32.7 (1)	<.001
5	How much do you like to sit and talk with {partners name}?	146 (19.3)	53 (22.8)	16.6 (1)	<.001

Engagement by Time of Day and by Week

The number of questions presented per hour and per week are shown in Figure 2. The fewest number of questions were presented between midnight and 8 AM (Figure 2). Most questions were presented late morning (between 10 AM and noon), late afternoon (2 PM to 4 PM), and after dinner (7 PM to 9 PM). A high number of questions were presented in weeks 1 and 6, which is probably due to increased use of the app post–information technology training and was consistent with increased app usage in weeks 1 and 6. There was a significant association between the number of app interactions and the number of questions presented across trial weeks (ρ=0.84, *P*=.001).

The hourly and weekly dismissal rates are shown in Figure 3. When looking at time of day, 11 AM, 4 PM, and 7 PM had the highest dismissal rates (Figure 3). These periods also had a number of questions presented (Figure 2). Questions asked at 9 PM and 10 PM had the lowest dismissal rate; users were more likely to answer questions that were presented in the evening (Figure 3). For example, the dismissal rate at 9 PM was significantly lower than the dismissal rate at 11 AM (χ^2_1=21.4, *P*<.001); questions asked at 9 PM (10% dismissal rate) were 5 times more likely to be answered than questions asked at 11 AM (50% dismissal rate). The dismissal rate for questions was high at the start and end of the trial and was lowest in the middle of the trial (Figure 3). For example, the dismissal rate in trial week 8 was significantly lower than that in week 2 (week 8: 14.1%; week 2: 54.8%; χ^2_1=19.2, *P*<.001).

Dismissal rates by time of day and by week for each user type are shown in Figure 4. When looking at time of day, user types demonstrated a similar pattern of dismissals (Figure 4). There was a strong correlation between dismissal rate and time of day between persons with dementia and their caregivers (*P*<.001, ρ=0.81). Weekly dismissal rates differed between user types (Figure 4). There was no correlation between dismissal rate and trial week between persons with dementia and their caregivers (*P*=.09, ρ=0.51).

There was a strong correlation between the number of questions presented and the number of questions dismissed per hour for caregivers (P<.001, ρ=0.84) and persons with dementia (P<.001, ρ=0.83). From investigating the residuals, there were significantly less questions dismissed at the hours of 2 PM , 9 PM, and 10 PM for both caregivers and persons with dementia. For both user types, questions in the evening were more likely to be answered than dismissed.

Figure 2. Total number of questions presented across hours of day (left) and trial weeks (right) for each user type.

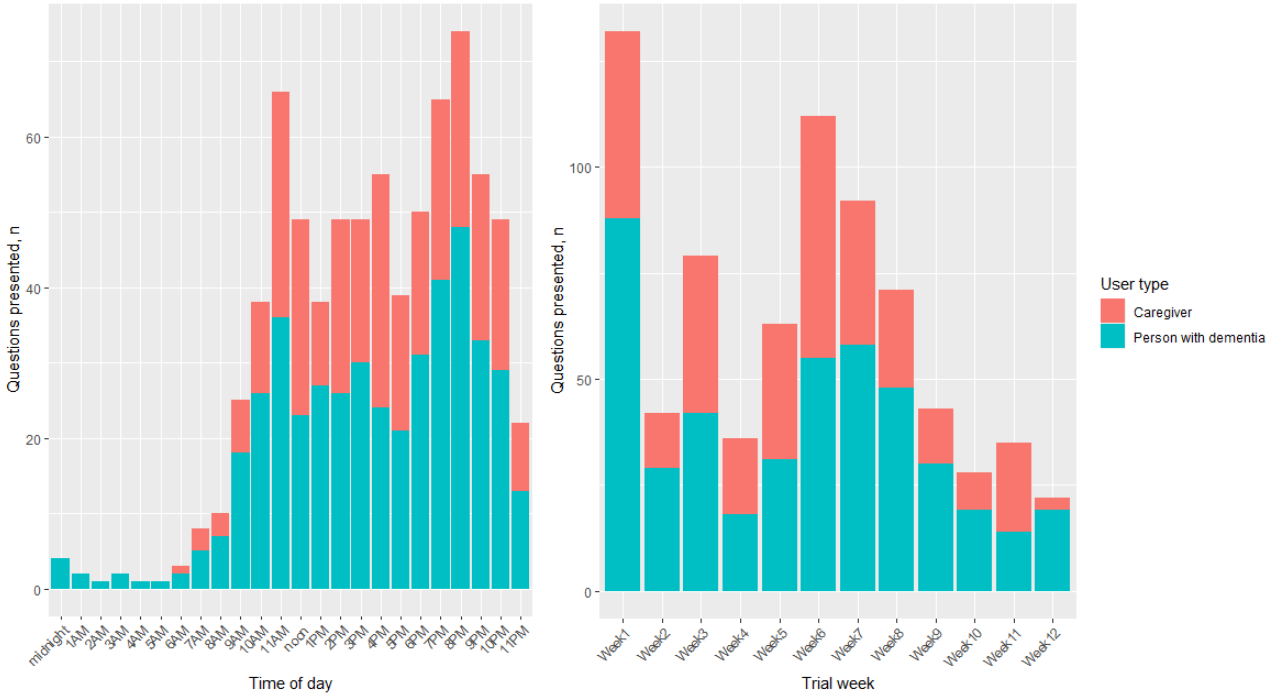


Figure 3. Dismissal rate for ecological momentary assessment questions across hours of day (left) and trial weeks (right).

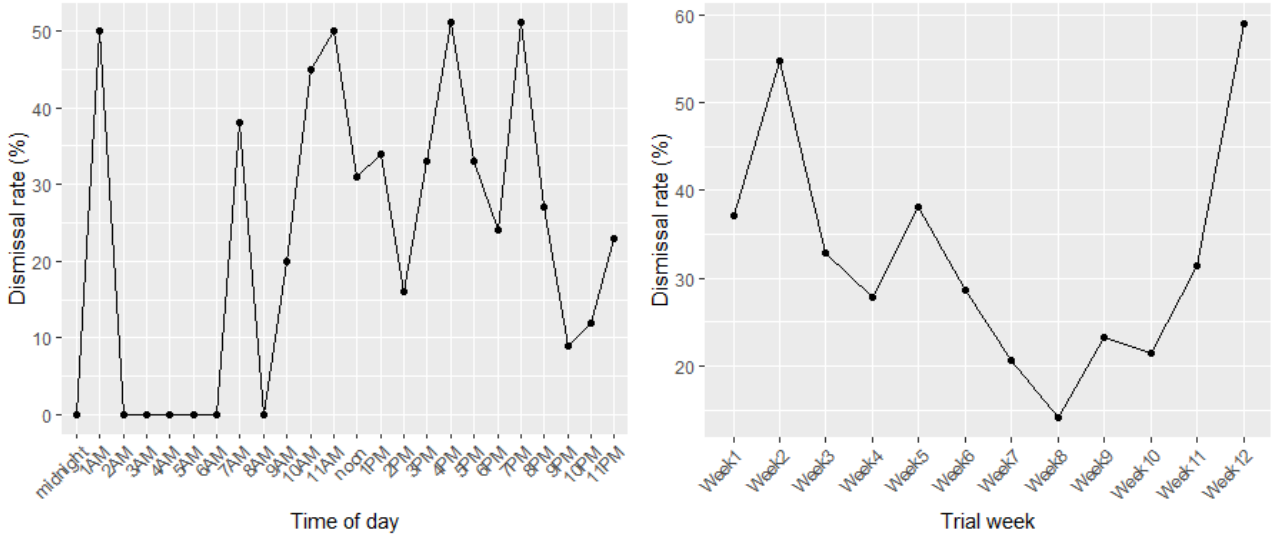
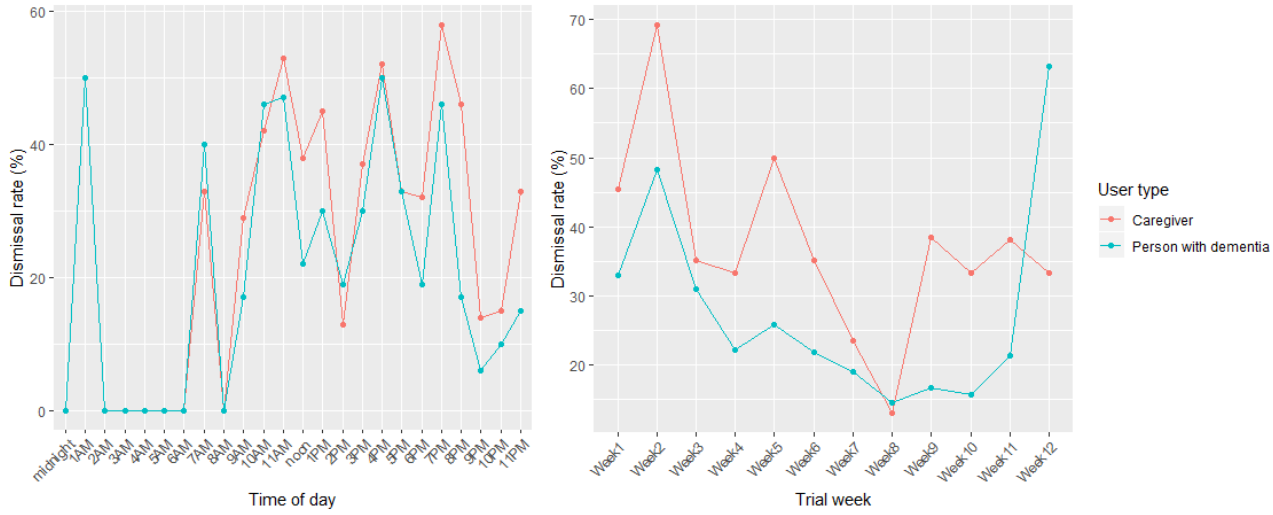


Figure 4. Dismissal rate for ecological momentary assessment questions across hours of day (left) and trial weeks (right) for each user type.



Engagement Following Reminiscence Media Types

The number presented and dismissal rates of questions following reminiscence within photo, video, and music are shown in Table 3. The highest number of questions were presented after users viewed photos, and the fewest after listening to music. Personal media was used in the app more than generic media was used, thus more questions were presented following reminiscence with personal media.

As shown in Table 3, chi-square tests comparing generic to personal media source indicated significant differences when comparing the proportion of questions of dismissed with the proportion of questions answered for generic photos versus personal photos (χ^2_1=19.4, *P*<.001) and generic music versus personal music (χ^2_1=15.3, *P*<.001). There was no significant difference between proportion of questions of dismissed and answered between generic videos and personal videos (χ^2_1=0.2, *P*=.67).

Significantly more questions were dismissed after viewing any type of photo than after listening to any type of music (all photo: 34.4%; all music: 19.3%; χ^2_1=10.9, *P*<.001). The dismissal rate for questions asked after viewing personal photos (41/122, 33.6%) was significantly higher than the rates of those asked after generic music (χ^2_1=73.7, *P*<.001), personal music (χ^2_1=19.2, *P*<.001), and generic photos (χ^2_1=19.4, *P*<.001). The dismissal rate for questions asked after interacting with generic music (4/73, 5.5%) was significantly lower than those of personal music (χ^2_1=15.3, *P*<.001), personal videos (χ^2_1=12.1, *P*=.02), or personal photos (χ^2_1=73.7, *P*<.001). These results show that questions asked following reminiscence with personal media, especially those asked after personal photos, were less likely to be answered compared to those asked after other media. In contrast, questions asked after the user had listened to generic media, in particular those asked after generic music, were much more likely to be answered.

Table 3. Question presentation and dismissal following reminiscence with different media.

Media type (format and source, questions presented)	Questions dismissed, n (%)	Comparison between generic and personal	
		Chi-square (*df*)	*P* value
Photo		19.4 (1)	<.001
Generic (n=14)	2 (15.4)		
Personal (n=122)	41 (33.6)		
All (n=256)	88 (34.4)		
Video		0.2 (1)	.67
Generic (n=2)	0 (0.0)		
Personal (n=11)	5 (45.5)		
All (n=57)	16 (28.1)		
Music		15.3 (1)	<.001
Generic (n=73)	4 (5.5)		
Personal (n=39)	39 (35.9)		
All (n=176)	34 (19.3)		

Engagement Following Other App Activities

Questions were also presented to users as they were completing other logged activities. There were significantly more questions answered than were dismissed when a person with dementia logged into the app (answered: 15/20, 75.0%; dismissed: 5/20, 25.0%; χ^2_1=8.1, *P*=.004). There was no significant difference between questions answered and questions dismissed when caregivers logged in (answered: 9/14, 64.3%; dismissed: 5/14, 35.7%; χ^2_1=1.3, *P*=.26), when caregivers added or deleted photos (answered: 21/43, 48.8%; dismissed: 22/43, 51.2%; χ^2_1=0, *P*>.999), or when caregivers or persons with dementia exited the app (answered: 35/61, 57.4%; dismissed: 26/61, 42.6%; χ^2_1=2.1, *P*=.15).

Discussion

Principal Findings

This study set out to explore engagement with EMA using a case study of a reminiscence app for persons with dementia and their caregivers. Overall engagement with EMA was high, with 69.1% questions answered when they were presented. Other studies [18] have reported between 55% and 87% engagement with EMA questions in the general population. Previous studies [19,20] that have used EMA with older adults have generally relied on sending user notifications or alarms as reminders to complete questions. In these cases, the most commonly reported reasons for not answering EMA questions included being prompted at inconvenient times or when busy, notification sound was not heard, or phone was not nearby at the time [19,20]. In our study, questions were asked only when the users where already using the app which eliminated some of the previously reported reasons for EMA nonadherence.

When trying to increase engagement with EMA within a digital intervention, one important factor to consider is the frequency of administering questions. For example, if EMA questions are presented multiple times in one day it may become burdensome for the user and increase the chances of dismissal. In our study, persons with dementia and their caregivers were less likely to dismiss EMA questions asked between 9 PM and 10 PM. This is similar to what we found in a previous study [21] exploring the temporal behavior of users completing EMA in a maternal health app, where users were more likely to complete mental health scales around 8 PM or 9 PM. Additionally, dismissal rates around 9 AM, 2 PM, and 6 PM which corresponded to postbreakfast, postlunch, and postevening mealtimes were also low suggesting users were more likely to engage in questions during postprandial reminiscence.

Dismissal rates for EMA after reminiscence with generic media were very low when compared to personal media. It has been shown that reminiscence with personal media is more likely to have positive psychosocial benefits compared to generic reminiscence material for persons with dementia [22]. This would suggest that personal media provides a more meaningful reminiscence experience, and thus, could explain why users were less likely to answer questions during this time; however, our results suggest that asking questions following reminiscence with personal photos compared to other media sources was likely to result in question dismissal. We also found that users were less likely to dismiss EMA questions following reminiscence with generic music compared to video or photos. Cognitive function has been shown to be better in persons with dementia following music therapy [23,24]. Therefore, we suggest that participants were more likely to answer questions after listening to music as this was less cognitively demanding compared to viewing photos or video. Music therapy has been shown to have other benefits for persons with dementia, improving behavioral and psychological symptoms and decreasing agitation [25-27]. This could also explain why participants were more likely to answer questions after listening to music.

When caregivers were adding or deleting photos, the dismissal rate was high. Questions asked while users were trying to complete a task such as this could be a hindrance, therefore lead to a higher dismissal rate. Overall, both users were more likely to answer questions if they were asked less frequently so this should be taken into consideration going forward.

Future work should look at the type of EMA questions being asked, for example it may be more beneficial to ask questions such as “Do you find reminiscence therapy helpful?” or “What is your favorite way to reminisce?” rather than questions with Likert scale responses. Alternatively, EMA could assess feedback on the user experience as this would provide co-design opportunities in the use phase since most cocreation activities focus on initial design. These questions could help inform app design and prrovide valuable insight into user experience.

Ultimately, to increase engagement with EMA, it is important to secure individuals as regular app users. There are several approaches which can be used to secure people as full adopters of an app, such as user notifications and prompts to encourage app usage. These notifications could be sent when individuals are likely to engage with EMA, such as in the evening at 9 PM. Personalization can also help to increase engagement. In our study, EMA questions from the mutuality scale incorporated the name of caregiver or person living with dementia, for example, “How attached are you to [partners name]?”. Future work should continue to utilize these personalized aspects such as using individual names in personalized push notifications or in EMA questions.

Limitations

The main limitation of our study was lack of generalizability to a larger population given the quasi-experimental design and older demographic where half of the participants were persons with dementia. Due to the relatively small sample size, it was not possible to attribute any findings to gender differences. Ideally, sample size should be calculated for a randomized controlled trial to ensure adequate power; however, the sample size for this study (n=56) was deemed sufficient to meet the objectives of a feasibility study.

Another limitation was the recording of log data. User events were logged in real time and were later collected in person from the iPads using portable storage devices. As a result, we could not control for lost local data due to operating system failures,

app crashes, bugs, and updates. This was the case with some of the log data which was recorded. For 16% of all EMA questions asked, the event which was logged previously was also an EMA question instead of one of the actions (Multimedia Appendix 2) that should trigger a question being asked. To allow for remote event logging and to minimize data loss, it is important to follow best practices for the collection and storage of user log data, such as storing data in the cloud. This would enable the analysis of logs over the duration of the study. Remote logging could also facilitate the use of adaptive features to motivate persons with dementia and caregivers by sending personalized notifications and motivational messages when app usage is low.

Conclusions

This study explored engagement with EMA questions using an app for reminiscence where the users were caregivers and persons with dementia. Our results show that older people are willing to participate and engage in EMA; however, more can be done to increase engagement. Notwithstanding the limitations, based on this study we propose a set of recommendations for the use of EMA to optimize user engagement within a digital intervention. EMA questions should add value and help to validate the use of the digital health app in line with the study objectives. It is important that EMA questions do not distract or interfere with the overall purpose of the app which was to allow persons with dementia and their caregivers to reminisce. To avoid overprompting the user, no more than one question should be presented per day and ideally in the evening after 8 PM as this is when people were most likely to engage. If the user is trying to complete a task within the app such as adding content then a question should not be presented. These recommendations can be broadly applied to EMA use in similar settings. Future work will be carried out to study the engagement on a larger scale with more participants, which will further support these recommendations. Future work will also involve studying app usage patterns, for example, if a user engages in an EMA question, when will they next engage, and can we predict engagement based on user log analysis.

Acknowledgments

This work was funded by the Public Health Agency Research and Development Office and Atlantic Philanthropies (COM/5016/14). The authors would like to thank the Reminiscence Network Northern Ireland and persons with dementia and their caregivers for their input in the cocreation and design of the InspireD app and the Trust's Cognitive Rehabilitation Team for facilitating recruitment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the InspireD app.
[PNG File , 210 KB - mhealth_v8i7e17120_app1.png]

Multimedia Appendix 2
List of actions completed by app users which could prompt an EMA question.
[PNG File , 69 KB - mhealth_v8i7e17120_app2.png]

References

1. Dementia. World Health Organisation. 2019 Sep 19. URL: https://www.who.int/en/news-room/fact-sheets/detail/dementia [accessed 2019-10-16]
2. Prince M, Knapp M, Guerchet M, McCrone P, Prina M, Comas-Herrera A. Dementia UK: Second edition – Overview. UK Alzheimer Soc London.: Alzheimer's Society; 2014 Sep. URL: http://eprints.lse.ac.uk/59437/1/Dementia_UK_Second_edition_-_Overview.pdf [accessed 2019-10-16]
3. Lewis F, Schaffer SK, Sussex J, O'Neill P, Cockcroft L. The Trajectory of Dementia in the UK - Making a Difference. Office of Health Economics. 2014. URL: https://www.alzheimersresearchuk.org/wp-content/uploads/2015/01/OHE-report-Full.pdf [accessed 2019-10-16]
4. Gonzalez J, Mayordomo T, Torres M, Sales A, Meléndez JC. Reminiscence and dementia: a therapeutic intervention. Int Psychogeriatr 2015 Oct;27(10):1731-1737. [doi: 10.1017/S1041610215000344] [Medline: 25765779]
5. Spector A, Orrell M, Davies S, Woods RT. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2000 Mar 01(4):CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120] [Medline: 11034700]
6. Gibson F. The Past in the Present: Using Reminiscence in Health and Social Care. Baltimore, Maryland: Health Professions Press; 2004.
7. Yu F, Mathiason MA, Johnson K, Gaugler JE, Klassen D. Memory matters in dementia: efficacy of a mobile reminiscing therapy app. Alzheimers Dement (N Y) 2019;5:644-651 [FREE Full text] [doi: 10.1016/j.trci.2019.09.002] [Medline: 31720365]

8. Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2018 Mar 01;3:CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120.pub3] [Medline: 29493789]
9. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
10. Lazzari C. Ecological momentary assessments and interventions in Alzheimer's caregiving. Curr Alzheimer Res 2018;15(11):1027-1031. [doi: 10.2174/1567205015666180702111523] [Medline: 29962343]
11. Pihet S, Moses Passini C, Eicher M. Good and bad days: fluctuations in the burden of informal dementia caregivers, an experience sampling study. Nurs Res 2017;66(6):421-431. [doi: 10.1097/NNR.0000000000000243] [Medline: 29095373]
12. Archbold PG, Stewart BJ, Greenlick MR, Harvath T. Mutuality and preparedness as predictors of caregiver role strain. Res Nurs Health 1990 Dec;13(6):375-384. [Medline: 2270302]
13. Bech P, Olsen LR, Kjoller M, Rasmussen NK. Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale. Int J Methods Psychiatr Res 2003;12(2):85-91. [Medline: 12830302]
14. Spruytte N, Van AC, Lammertyn F, Storms G. The quality of the caregiving relationship in informal care for older adults with dementia and chronic psychiatric patients. Psychol Psychother 2002 Sep;75(Pt 3):295-311. [Medline: 12396755]
15. Mulvenna M, Gibson A, McCauley C, Ryan A, Bond R, Laird L, et al. Behavioural usage analysis of a reminiscing app for people living with dementia and their carers. : Association for Computing Machinery; 2017 Presented at: Proceedings of the European Conference on Cognitive Ergonomics 2017; 2017; Umeå, Sweden p. 35-38 URL: https://dl.acm.org/doi/10.1145/3121283.3121289
16. Laird EA, Ryan A, McCauley C, Bond RB, Mulvenna MD, Curran KJ, et al. Using mobile technology to provide personalized reminiscence for people living with dementia and their carers: appraisal of outcomes from a quasi-experimental study. JMIR Ment Health 2018 Sep 11;5(3):e57 [FREE Full text] [doi: 10.2196/mental.9684] [Medline: 30206053]
17. McCauley CO, Bond RB, Ryan A, Mulvenna MD, Laird L, Gibson A, et al. Evaluating user engagement with a reminiscence app using cross-comparative analysis of user event logs and qualitative data. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):543-551. [doi: 10.1089/cyber.2019.0076] [Medline: 31403855]
18. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. J Psychiatr Res 2009 Jul;43(11):987-996. [doi: 10.1016/j.jpsychires.2009.01.014] [Medline: 19272611]
19. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. J Technol Hum Serv 2016;34(2):209-223 [FREE Full text] [doi: 10.1080/15228835.2016.1170649] [Medline: 27683018]
20. Liu H, Lou VWQ. Developing a smartphone-based ecological momentary assessment protocol to collect biopsychosocial data with community-dwelling late-middle-aged and older adults. Transl Behav Med 2019 Jul 16;9(4):711-719. [doi: 10.1093/tbm/iby096] [Medline: 30395340]
21. Bond R, Moorhead A, Mulvenna M, O'Neill S, Potts C, Murphy N. Exploring temporal behaviour of app users completing ecological momentary assessments using mental health scales and mood logs. Behav Inf Technol 2019 Aug 08;38(10):1016-1027. [doi: 10.1080/0144929x.2019.1648553]
22. Subramaniam P, Woods B. The impact of individual reminiscence therapy for people with dementia: systematic review. Expert Rev Neurother 2012 May;12(5):545-555. [doi: 10.1586/ern.12.35] [Medline: 22550983]
23. Bruer RA, Spitznagel E, Cloninger CR. The temporal limits of cognitive change from music therapy in elderly persons with dementia or dementia-like cognitive impairment: a randomized controlled trial. J Music Ther 2007;44(4):308-328. [doi: 10.1093/jmt/44.4.308] [Medline: 17997623]
24. Gómez Gallego M, Gómez García J. Music therapy and Alzheimer's disease: Cognitive, psychological, and behavioural effects. Neurologia 2017 Jun;32(5):300-308 [FREE Full text] [doi: 10.1016/j.nrl.2015.12.003] [Medline: 26896913]
25. Raglio A, Bellelli G, Traficante D, Gianotti M, Ubezio MC, Villani D, et al. Efficacy of music therapy in the treatment of behavioral and psychiatric symptoms of dementia. Alzheimer Dis Assoc Disord 2008;22(2):158-162. [doi: 10.1097/WAD.0b013e3181630b6f] [Medline: 18525288]
26. Sung H, Chang AM. Use of preferred music to decrease agitated behaviours in older people with dementia: a review of the literature. J Clin Nurs 2005 Oct;14(9):1133-1140. [doi: 10.1111/j.1365-2702.2005.01218.x] [Medline: 16164531]
27. McDermott O, Orrell M, Ridder HM. The importance of music for people with dementia: the perspectives of people with dementia, family carers, staff and music therapists. Aging Ment Health 2014;18(6):706-716. [doi: 10.1080/13607863.2013.875124] [Medline: 24410398]

Abbreviations

EMA: ecological momentary assessment
InspireD: individual specific reminiscence in dementia

Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by N Jacobson, H Rai; comments to author 06.01.20; revised version received 02.03.20; accepted 14.05.20; published 06.07.20.

Please cite as:
Potts C, Bond R, Ryan A, Mulvenna M, McCauley C, Laird E, Goode D
Ecological Momentary Assessment Within a Digital Health Intervention for Reminiscence in Persons With Dementia and Caregivers: User Engagement Study
JMIR Mhealth Uhealth 2020;8(7):e17120
URL: https://mhealth.jmir.org/2020/7/e17120
doi:10.2196/17120
PMID:32420890

Original Paper

Psychometric Properties of the Korean Version of the PsyMate Scale Using a Smartphone App: Ecological Momentary Assessment Study

Yong Sook Yang[1], MSN; Gi Wook Ryu[1], MPH; Philippe A E G Delespaul[2], PhD; Mona Choi[1], PhD

[1]Mo-Im Kim Nursing Research Institute, Yonsei University College of Nursing, Seoul, Republic of Korea

[2]Department of Psychiatry and Neuropsychology, Faculty of Health Medicine and Lifesciences, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Mona Choi, PhD
Mo-Im Kim Nursing Research Institute
Yonsei University College of Nursing
50 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3341
Email: monachoi@yuhs.ac

Abstract

Background: Ecological momentary assessment (EMA) is a method for capturing the changes in the variables in daily life with increased accuracy and decreased recall bias. The PsyMate scale assesses momentary moods in daily life and can be used in various settings.

Objective: The aim of this study was to develop a Korean version of the PsyMate (K-PsyMate) scale and evaluate its psychometric properties by using the EMA method in patients with moyamoya disease (MMD) in South Korea.

Methods: Patients with MMD aged over 18 years were recruited from July 2018 to January 2019 at the inpatient and outpatient departments of a university hospital in South Korea. The K-PsyMate scale comprising 13 items was developed following a translation/back translation approach of the English version and loaded onto a mobile app. Participants were instructed to enter their moods 4 times a day for 7 consecutive days. Content validity index, factor analysis, and Pearson's correlation were performed for validity analysis. For reliability, intraclass correlation coefficients between the first and last measurements were estimated by mean rating, absolute agreement, and a 2-way mixed-effects model. Usability was analyzed through a descriptive analysis, 2-tailed t test, and analysis of variance, and the results were confirmed by Mann-Whitney U test and the Kruskal-Wallis test, as the dependent variable was not normally distributed.

Results: In total, 1929 assessments from 93 patients were analyzed. The mean age of the participants was 40.59 (SD 10.06) years, and 66 (71%) of the 93 participants were women. Content validity was excellent as content validity index was 0.99, and 2 factors, negative affect and positive affect, were derived by an exploratory factor analysis. The correlations between the subdomains of the K-PsyMate scale and Hospital Anxiety and Depression Scale were significant (P<.001). The agreement between the first and last measurements was poor to moderate according to the obtained intraclass correlation coefficient values. Usability was evaluated by 67 (72%) out of the 93 participants. The participants rated the accuracy of assessing their momentary moods on the app at 4.13 (SD 0.97), easiness in understanding questions, operating, and inputting answers at 4.12 (SD 0.88), and interruption by the survey alarms at 2.48 (SD 1.02) out of 5.

Conclusions: The K-PsyMate scale has good validity but poor to moderate agreement, which reflects the characteristics of the EMA data collected in real and natural living environments without control. The findings of our study show that the K-PsyMate scale uploaded in a mobile app can be a valid and reliable tool for evaluating the momentary mood of patients with MMD because using a mobile app is convenient and patients are familiar with their own smartphones, which they use in their daily lives.

(JMIR Mhealth Uhealth 2020;8(7):e17926) doi:10.2196/17926

KEYWORDS

psychometrics; reliability; validity; PsyMate scale; ecological momentary assessment

Introduction

Moyamoya disease (MMD), also called as spontaneous occlusion of the circle of Willis, is a chronic progressive cerebrovascular disease with no known cause and is characterized by fine collateral vessel networks that resemble hazy cloud-like puffs of smoke in the brain [1,2]. MMD occurs worldwide with high prevalence in East Asian countries such as Japan and Korea [3]. Unlike MMD in children, MMD in adults is associated with potential rupture or ischemia of the blood vessels in the brain, similar to that reported in other cerebrovascular diseases [4,5], and this can lead to high mortality [6]. The clinical features of patients with MMD often appear as mental deterioration, headache, and speech impairment [6], which are affected by mood changes. Adults with MMD also show high levels of anxiety and depression [7] and their moods change over time during the course of life [8]. Negative mood and mood changes in adults with MMD may affect their motivation for proper health behaviors and eventually affect the health outcomes of patients with similar cerebrovascular diseases [7]. Therefore, monitoring the moods of adults with MMD can be critical for both patients and health care personnel.

Measuring moods may help people better understand their emotional states. The measuring tool to self-assess moods should be understandable and comfortable to use; at the same time, the tool should be able to assess the mood accurately and capture the fluctuations and changes in the mood reliably since mood varies according to context and time [9,10]. To avoid mood changes caused by the tool itself, the participants should be allowed to respond to the questionnaire in a flexible and friendly atmosphere of their choice.

Traditionally, mood has been assessed by retrospective measures, including pen and paper tests [11], but these methods lead to recall bias and risk of reduced accuracy [9,10]. However, momentary assessments such as ecological momentary assessment (EMA) or experience sampling methods allow increased accuracy and reduced bias because these assessments are performed in real-life environments and through repeated assessments at multiple times [9-11]. The EMA method is known for its effectiveness in capturing changes [12] and understanding daily changes in psychological states such as mood and stress [13,14]. It is broadly used to assess the psychological aspects of the participants with [15-17] and without mental health problems [18]. The increased prevalence of smartphones and advanced mobile technology has made people more familiar with the use of smartphones in their daily lives, and EMA methods using smartphones are available [19-22]. The prevalence of smartphone use in Korean adults is about 95% [23], and about 96.4% of the population under the age of 70 years in Korea can access internet services [24]. In a study conducted with healthy participants in Korea, the EMA method was evaluated as comfortable to use and easy to understand when the mood was assessed using a mobile app installed in their own smartphones [22].

Various measurement tools using mobile technology have been developed to measure moods in daily life. The PsyMate scale was developed to assess momentary moods relating to the daily life of the participants [25], and it has been used in various settings and populations [26,27]. In this study, we aimed to develop a Korean version of the PsyMate (K-PsyMate) scale, upload it on a mobile app to measure the momentary moods of Korean adults with MMD, and examine the psychometric properties of this scale.

Methods

Participants

Patients with MMD diagnosed via magnetic resonance angiography and aged over 18 years were recruited from July 2018 to January 2019 at the outpatient department or the inpatient wards of a university hospital in Seoul, South Korea. Patients who could not communicate owing to neurological problems were excluded by a physician and the physician's assistant who were involved in the treatment of patients with MMD. The Korean version of the Mini-Mental State Examination [28] was used to assess the cognitive impairment of the participants after acquiring the permission to use it from the provider. Participants with a score under 24 in the Korean version of the Mini-Mental State Examination were excluded. Participants who used operating systems other than the Android system were also excluded, since the developed mobile app was available only for the Android operating system as described in the previous study [22]. Informed consent was obtained from all the participants when they enrolled in the study. This study was approved by the Institutional Review Board of the Yonsei University Health System (No. 4-2018-0385).

Measurements

Participants' Characteristics and Baseline Mood

All participants were asked to provide their demographic and disease-specific information at the baseline measurement through a questionnaire that was developed by our research team. Demographics included age, gender, marital status, people living with, job status, monthly household income, education level, and smoking and drinking history. Detailed information on the participants' demographics has been presented in our previous work [29]. Regarding the baseline mood, the Hospital Anxiety and Depression Scale (HADS) [30] was administered after obtaining permission to use it. The HADS is known as a globally reliable and valid scale for assessing mental health in hospital settings [31]. It consists of 14 items, 7 for anxiety and 7 for depression, and it is measured on a 4-point scale ranging from 0 to 3. A higher value indicates a higher level of anxiety and depression. In this study, we administered the Korean version of HADS [32], for which Cronbach α was .89 and .86 for anxiety and depression subscales, respectively [32]. The HADS was assessed to compare the momentary negative affect and positive affect that were measured by the K-PsyMate scale, as described in a previous study [25].

K-PsyMate Scale for Measuring Momentary Mood

The mood scale of the K-PsyMate was used for the EMA section of this study. This scale consists of 13 items, of which 4 items measure positive affect and 9 items assess negative affect. The positive affect comprised feeling "cheerful," "satisfied," "relaxed," and "globally well," and the negative affect

comprised feeling "lonely," "guilty," "worried," "down," "threatened," "insecure," "irritated," "frightened," and "suspicious." These items were rated on a 7-point scale, with "1" indicating "not at all" and "7" indicating "very much." Momentary mood was measured while considering the environmental context, taking into account what the participants were doing, where they were, and with whom they were when responding to the questionnaire.

Usability Measure

The usability of the EMA method on the mobile phone was evaluated by the participants at the end of the survey. A 10-item questionnaire was uploaded in the app: 2 items for accuracy, 3 items for easiness, 2 items for enjoyment, 2 items for intention, and 1 item for interruption. It was adapted from a previous study [33] with the permission from the developer and scored on a 5-point scale (1=strongly disagree to 5=strongly agree). An open-ended question was included for the participants to offer any additional comment or opinion. The Cronbach α of the scale in this study was .74.

Procedure

Translation of the PsyMate

We freely downloaded the English version of the PsyMate scale [34] with permission from the copyright holder. To develop the K-PsyMate scale, we used the translation/back translation approach. We employed a committee approach for translation, in which we consulted 3 nursing professors who were fluent in both English and Korean to translate the English version of PsyMate [25]. Through this method, we identified language and cultural differences in selecting words. The translated PsyMate items were prepared for back translation during which we consulted 3 different nursing professors who were fluent in both English and Korean. The research team compared the back-translated PsyMate with the original English version to confirm that the translated Korean version was accurate. After confirming the translated Korean version through back translation, the research team finalized the K-PsyMate draft. This draft was then evaluated by a linguist of the Korean language prior to determining the final version. Eventually, the K-PsyMate version was finalized by the research team.

Data Collection

While obtaining informed consent from each patient, members of the research team obtained baseline data on demographic and disease-specific information by using the structured questionnaire and on mood status by using the HADS. They helped each participant install the mobile app on their mobile phone and register themselves, and they provided time for practice. Participants selected the EMA sampling period to assess momentary moods according to their convenience. They were guided by an alarm to respond 4 times a day for 7 consecutive days (4 times × 7 days = 28 times/person) at 60-minute blocked semirandom intervals. Notifications were sent in the morning between 8 AM and 9 AM, early afternoon between noon and 1 pm, evening between 5 PM and 6 PM, and at night between 9 PM and 10 PM. A reminder for each scheduled measurement was sent to the participants who did not input the response within 45 minutes after the first notification. The contact numbers and the email addresses of the members of the research team were given to the participants to contact the team if they had any questions or if they faced any difficulties during the assessment period. Finally, patients were requested to complete a 10-item survey at the end of the EMA session on the usability of the EMA method using a mobile app, including an open-ended question.

Statistical Analysis

Data were analyzed using the SPSS v.25.0 (IBM Corp). We used descriptive statistics to analyze the baseline data of the participants' characteristics, the HADS scores, and the positive affect and negative affect of the K-PsyMate scale measured by the EMA method. We applied a 2-tailed *t* test and analysis of variance to examine the differences between groups according to the general characteristics. Regarding validity tests, we used the content validity index (CVI), which calculated scale-level CVI (S-CVI) and each item-level CVI (I-CVI). To assess CVI, each item was appraised by an expert panel consisting of 6 nursing professors. The CVI for each item was scored using a 4-point scale, where 1=not relevant, 2=somewhat relevant, 3=quite relevant, and 4=highly relevant. S-CVI and I-CVI were evaluated with the cut-off of S-CVI level at 0.90 and I-CVI at 0.78 [35]. We performed exploratory factor analysis (EFA) to examine construct validity using principal component analysis and varimax rotation. We also used Pearson's correlation (*r*) to assess the association between the HADS and PsyMate in the previous study [25]. We calculated intraclass correlation coefficients (ICCs) for evaluating the test-retest reliability of the measurement tool. Reliability reflects both the degree of correlation and agreement between the measurements, and ICCs are an index that show the degree of correlation and agreement between the measurements [36]. Among the 10 different forms of ICC [37,38], we calculated the ICCs of each item between the first and the last measurement by mean rating, absolute agreement, and a 2-way mixed-effects model, since we tested absolute agreement of the mean of multiple measurements from a nonrandomly selected population [36]. The obtained ICC values of each item were interpreted by its 95% CI ranges [36]. The ICC estimates in the 95% CI were evaluated by considering ICC<0.50, 0.50≤ICC<0.75, 0.75≤ICC<0.90, and ICC≥0.90 as poor, moderate, good, and excellent reliability, respectively [39]. For usability evaluation, we performed descriptive analysis, 2-tailed *t* test, and analysis of variance, and we confirmed the results by Mann-Whitney *U* test and the Kruskal-Wallis test, as the dependent variable was not normally distributed.

Results

Participants

We recruited patients with MMD from the inpatient and outpatient departments of a tertiary hospital in South Korea. We approached 119 patients of which 109 (91.6%) agreed to participate, and they were enrolled using the mobile app. Among the 109 patients, 98 (89.9%) participated in the EMA session. The data of 93 participants who provided more than 3 measurements considered as EMA data [40] and who provided 1929 responses (74.1%) out of the 2604 scheduled assessments were analyzed. The mean age of the participants was 40.59 (SD

10.06) years, and 66 of the 93 participants (71%) were women. The HADS anxiety and HADS depression scores were 7.17 (SD 3.38) and 7.14 (SD 3.51), respectively. The overall mean negative affect and positive affect scores were 2.15 (SD 1.12) and 4.70 (SD 1.31) out of 7, respectively.

Validity of the K-PsyMate Scale

We calculated S-CVI and I-CVI as evaluated by the expert panel. The S-CVI was 0.99 (I-CVI range 0.83-1.00), which represented an excellent level of content validity. We also calculated the S-CVIs of 2 factors of negative affect and positive affect according to the original study [25], and they were 1.00 and 0.96, respectively. The I-CVIs are presented in Table 1.

EFA was performed based on 1929 completed assessments from 93 participants. We confirmed the presence of latent factors in the observed items with excellent Kaiser-Meyer-Olkin of 0.93 and with significant Bartlett's sphericity measures (P<.05). Two factors were derived with the eigenvalue greater than 1 with a cumulative percentage of explained variance of 71.1%. All items were assigned to either the negative affect (Factor 1) or positive affect (Factor 2) factor. These 2 factors were negatively related (r=–0.67). The results of the EFA are shown in Table 2.

The subdomains of negative affect and positive affect of the K-PsyMate scale and the HADS anxiety and depression were significantly correlated with coefficients (r) ranging between 0.33 and 0.68. The Cronbach α of the HADS anxiety and depression in this study were .84 and .79, respectively. The correlation coefficients between the K-PsyMate and the HADS subscales are shown in Table 3.

Table 1. Item-level content validity index of the PsyMate items (n=1929).

Item number	Items	Item-level content validity index
1	Feeling cheerful	1.00
2	Feeling satisfied	1.00
3	Feeling relaxed	1.00
4	Feeling globally well	0.83
5	Feeling lonely	1.00
6	Feeling guilty	1.00
7	Feeling worried	1.00
8	Feeling down	1.00
9	Feeling threatened	1.00
10	Feeling insecure	1.00
11	Feeling irritated	1.00
12	Feeling frightened	1.00
13	Feeling suspicious	1.00

Table 2. Factor loadings from EFA[a] of the K-PsyMate scale (n=1929).[b]

Item number	Items	Communality	Factor 1: Negative affect	Factor 2: Positive affect
2	Feeling satisfied	0.89	–0.23	*0.91*
3	Feeling relaxed	0.87	–0.26	*0.90*
1	Feeling cheerful	0.81	–0.21	*0.87*
4	Feeling globally well	0.80	–0.32	*0.84*
12	Feeling frightened	0.75	*0.83*	–0.23
10	Feeling insecure	0.75	*0.82*	–0.27
13	Feeling suspicious	0.70	*0.82*	–0.19
9	Feeling threatened	0.69	*0.82*	–0.15
6	Feeling guilty	0.63	*0.78*	–0.16
7	Feeling worried	0.70	*0.77*	–0.33
8	Feeling down	0.65	*0.68*	–0.44
11	Feeling irritated	0.62	*0.62*	–0.48
5	Feeling lonely	0.39	*0.52*	–0.35

[a]EFA: exploratory factor analysis.

[b]Values shown in italics show factor loadings >0.50, indicating that the value fits for the given factor.

Table 3. Correlations between the K-PsyMate scale and the HADS[a] (n=93).

HADS subscales	K-PsyMate subscales	
	Negative affect	Positive affect
HADS anxiety		
r	0.465[b]	–0.348[b]
P value	<.001	<.001
HADS depression		
r	0.446[b]	–0.328[b]
P value	<.001	<.001

[a]HADS: hospital anxiety and depression scale.

[b]The correlation is significant at a significance level of <.05 (2-tailed).

Reliability

The calculated ICC values between the first and the last measurements ranged from 0.52 to 0.80. Given the range of the 95% CI of each item, all 4 items of the positive affect factor and 6 items of the negative affect factor showed poor level of agreement. Three items of “down,” “insecure,” and “worried” in the negative affect factors were between 0.50 and 0.75 of the 95% CIs, which was considered as moderate agreement. The ICC value and the 95% CI of each item are shown in Table 4.

Table 4. ICC[a] and 95% CI of the K-PsyMate scale (n=93).

Items	First measure, mean (SD)	Last measure, mean (SD)	ICC	95% CI
Positive affect				
Satisfied	4.37 (1.26)	4.68 (1.41)	0.67	0.497-0.779
Cheerful	4.43 (1.32)	4.76 (1.31)	0.53	0.293-0.689
Globally well	4.57 (1.40)	4.65 (1.38)	0.53	0.297-0.691
Relaxed	4.57 (1.42)	4.74 (1.45)	0.52	0.277-0.682
Negative affect				
Down	2.74 (1.61)	2.43 (1.58)	0.80	0.694-0.866
Insecure	2.88 (1.74)	2.40 (1.63)	0.73	0.599-0.824
Worried	3.38 (1.78)	2.71 (1.67)	0.67	0.502-0.781
Frightened	2.55 (1.62)	2.19 (1.55)	0.66	0.490-0.776
Lonely	3.01 (1.66)	2.38 (1.62)	0.65	0.477-0.770
Suspicious	2.33 (1.50)	2.01 (1.41)	0.63	0.435-0.752
Guilty	2.25 (1.46)	2.04 (1.41)	0.61	0.415-0.743
Threatened	1.99 (1.33)	1.80 (1.18)	0.60	0.393-0.733
Irritated	2.72 (1.70)	2.26 (1.49)	0.54	0.327-0.704

[a]ICC: intraclass correlation coefficient.

Usability of the EMA Method by Using a Mobile App

Usability was evaluated by 67 (72%) out of the 93 participants. The mean age was 42.0 (SD 10.43) years (range, 22-67 years), and 50 (74%) of the 67 participants were females. Of the 67 participants, 26 (39%) were in their thirties, whereas only 4 (6%) were in their sixties (Multimedia Appendix 1). Participants rated the accuracy of inputting their momentary mood on the app at 4.13 (SD 0.97) out of 5. Easiness in understanding questions and in operating and in inputting answers was rated at 4.12 (SD 0.88) out of 5. The enjoyment in participating in the EMA app was rated at 3.65 (SD 0.91) out of 5. The participants rated interruption by the survey alarms at 2.48 (SD 1.02) out of 5. The intention to participate or recommend others to participate in another study using EMA apps was reported as 3.25 (SD 0.95) out of 5 (Table 5).

No difference in the usability was found in the subgroup analysis by gender, age, and marital status of the participants (Multimedia Appendix 1). Of the 67 participants, 20 (30%) wrote answers to the open-ended question in the app. Six (9%) out of 67 participants reported that they experienced discomfort owing to the small checkbox and 1 (1%) experienced network disconnection while inputting responses into the app. One participant (1%) expressed that it was a burden to answer 4 times a day.

Table 5. Usability of the EMA[a] method by using a mobile app (n=67).

Category of rating	Mean (SD) of the ratings	Possible range for the ratings
Accuracy in assessing momentary mood	4.13 (0.97)	1-5
Easiness in understanding questions, operating, and inputting answers	4.12 (0.88)	1-5
Enjoyment of participating via the EMA app	3.65 (0.91)	1-5
Interruption by the survey alarms	2.48 (1.02)	1-5
Intention to participate or recommend others to participate in another study using EMA apps	3.25 (0.95)	1-5

[a]EMA: ecological momentary assessment.

Discussion

Principal Results

In this study, we evaluated the psychometric properties of the K-PsyMate scale to measure moods by using the EMA method for adults with MMD since their moods are prone to change when they consciously strive to overcome negative feelings and stress, similar to patients with an unruptured cerebral aneurysm [5]. The content validity of the K-PsyMate scale evaluated by S-CVI was excellent with good I-CVI in all items determined by Polit et al [35]. The EFA results agreed with those of a previous evaluation of the PsyMate [25], which also had 2 factors: positive affect with 4 items and negative affect with 9 items. We found that the statistically significant correlations between the K-PsyMate scale and HADS were also in accordance with those reported in the previous study, which

had significant correlations ranging from 0.4 to 0.7 when compared with the HADS scores [25].

In this study, the agreement between the measurements evaluated by the ICC values of each item was poor to moderate. This poor to moderate level of agreement suggests that the measurements may have been affected by context variables since participants were requested to measure in an uncontrolled natural real-life environment [9,29]. This finding is in accordance with the results of a study on the Dutch version of the PsyMate, which presented sensitivity to the change of mood over time and applicability for patients in an ambulatory mental health setting in the Netherlands [25]. The poor ICCs may reflect the dynamics of affect fluctuating and changing over time [41] and the variability in the anxious and depressive mood of patients with MMD [7,42].

Considering the characteristics of the EMA data, we tried a 2-level confirmatory factor analysis [43] as an alternative method, wherein latent variables of negative affect and positive affect were assessed at the individual level. However, it did not achieve convergence, and the maximum log pseudo-likelihood stayed the same at –45854.379 from the 20th iteration. Convergence rates for 2-level confirmatory factor analysis models are known to be predicted by all possible interactions between the reliability condition (poor to moderate level by ICCs in this study), number of clusters (93 individual clusters in this study), and cluster size (ranging from 3 to 28 in this study) [44]. Multilevel confirmatory factor analysis cannot be applied to this study in which individuals form groups of interest with small cluster size and with low ICCs between measurements in single-level analysis.

The result of the usability evaluation in this study suggested that the EMA app has a good feasibility with accuracy in assessing mood and without much interruption and it may be applicable, regardless of gender or age, as no statistical difference of usability was found in the subgroup analysis.

Limitations

This study has the following limitations. First, we conducted the study by using a limited sample size. However, the psychometric properties of the scale could be analyzed by including 1929 measures obtained through the EMA method. Second, we recruited participants who only used the Android operating system, for which the mobile app was developed [22]. Hence, patients who used other operating systems such as the iPhone operating system were not included in this study. Nevertheless, the study is still meaningful in that we contacted about 1.2% of the target population with MMD in South Korea when considering the overall prevalence of 16.1 patients per 100,000 in 2011 [3]. Third, in this study, we included the responses of the participants who answered more than three times in the total duration of the study to evaluate the ability to capture the changes. This implies that our study results may vary according to participants who answered less than three times or did not answer at all. A qualitative study on participants' experience with the EMA method or research on psychometric properties of the K-PsyMate scale for different populations such as patients with other disorders and symptoms will be an interesting topic for future study.

Conclusion

This study provides evidence that the K-PsyMate scale has good validity but poor to moderate level of agreement between the measurements. The poor to moderate level of agreement evaluated by ICCs may reflect the characteristics of the EMA data collected in real and natural living situations and environments. Our study shows that the K-PsyMate scale uploaded in a mobile app can be a valid and reliable tool for evaluating the momentary mood of patients with MMD because using a mobile app is convenient for patients as they are familiar with their own smartphones, which they use in their daily lives.

Acknowledgments

This study was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2017R1D1A1B03030706).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the participants and subgroup analysis of the usability of the ecological momentary assessment method by using a mobile app (n=67).
[PDF File (Adobe PDF File), 385 KB - mhealth_v8i7e17926_app1.pdf]

References

1. Baba T, Houkin K, Kuroda S. Novel epidemiological features of moyamoya disease. J Neurol Neurosurg Psychiatry 2008 Aug;79(8):900-904. [doi: 10.1136/jnnp.2007.130666] [Medline: 18077479]
2. Festa JR, Schwarz LR, Pliskin N, Cullum CM, Lacritz L, Charbel FT, et al. Neurocognitive dysfunction in adult moyamoya disease. J Neurol 2010 May;257(5):806-815. [doi: 10.1007/s00415-009-5424-8] [Medline: 20033200]
3. Kim JS. Moyamoya Disease: Epidemiology, Clinical Features, and Diagnosis. J Stroke 2016 Jan;18(1):2-11 [FREE Full text] [doi: 10.5853/jos.2015.01627] [Medline: 26846755]

4. Koivunen R, Harno H, Tatlisumak T, Putaala J. Depression, anxiety, and cognitive functioning after intracerebral hemorrhage. Acta Neurol Scand 2015 Sep;132(3):179-184. [doi: 10.1111/ane.12367] [Medline: 25639837]
5. Pala A, Pawlikowski A, Brand C, Schmitz B, Wirtz CR, König R, et al. Quality of Life After Treatment of Unruptured Intracranial Aneurysms. World Neurosurg 2019 Jan;121:e54-e59. [doi: 10.1016/j.wneu.2018.09.010] [Medline: 30244183]
6. Ahn IM, Park D, Hann HJ, Kim KH, Kim HJ, Ahn HS. Incidence, prevalence, and survival of moyamoya disease in Korea: a nationwide, population-based study. Stroke 2014 Apr;45(4):1090-1095. [doi: 10.1161/STROKEAHA.113.004273] [Medline: 24595588]
7. Liu C, Yi X, Li T, Xu L, Hu M, Zhang S, et al. Associations of depression, anxiety and PTSD with neurological disability and cognitive impairment in survivors of moyamoya disease. Psychol Health Med 2019 Jan;24(1):43-50. [doi: 10.1080/13548506.2018.1467024] [Medline: 29678126]
8. Ryu GW, Yang YS, Choi M, Shim KW. Lived experiences of adult patients with moyamoya disease: A qualitative case study. Jpn J Nurs Sci 2020 Mar 09:e12332. [doi: 10.1111/jjns.12332] [Medline: 32153137]
9. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
10. Carpenter RW, Wycoff AM, Trull TJ. Ambulatory Assessment. Assessment 2016 Aug;23(4):414-424 [FREE Full text] [doi: 10.1177/1073191116632341] [Medline: 26887808]
11. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
12. Bolger N, Laurenceau J. Intensive longitudinal methods: an introduction to diary experience sampling research. New York: Guilford Press; 2013.
13. Cristóbal-Narváez P, Sheinbaum T, Ballespí S, Mitjavila M, Myin-Germeys I, Kwapil TR, et al. Impact of Adverse Childhood Experiences on Psychotic-Like Symptoms and Stress Reactivity in Daily Life in Nonclinical Young Adults. PLoS One 2016;11(4):e0153557. [doi: 10.1371/journal.pone.0153557] [Medline: 27082442]
14. Wichers M, Simons CJP, Kramer IMA, Hartmann JA, Lothmann C, Myin-Germeys I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011 Oct;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
15. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? Systematic Review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
16. Walz LC, Nauta MH, Aan Het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: a systematic review. J Anxiety Disord 2014 Dec;28(8):925-937. [doi: 10.1016/j.janxdis.2014.09.022] [Medline: 25445083]
17. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
18. Yang YS, Ryu GW, Choi M. Methodological Strategies for Ecological Momentary Assessment to Evaluate Mood and Stress in Adult Patients Using Mobile Phones: Systematic Review. JMIR Mhealth Uhealth 2019 Apr 01;7(4):e11215 [FREE Full text] [doi: 10.2196/11215] [Medline: 30932866]
19. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
20. Freedman MJ, Lester KM, McNamara C, Milby JB, Schumacher JE. Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. J Subst Abuse Treat 2006 Mar;30(2):105-111. [doi: 10.1016/j.jsat.2005.10.005] [Medline: 16490673]
21. Swendeman D, Ramanathan N, Baetscher L, Medich M, Scheffler A, Comulada WS, et al. Smartphone self-monitoring to support self-management among people living with HIV: perceived benefits and theory of change from a mixed-methods randomized pilot study. J Acquir Immune Defic Syndr 2015 May 01;69 Suppl 1:S80-S91 [FREE Full text] [doi: 10.1097/QAI.0000000000000570] [Medline: 25867783]
22. Yang YS, Ryu GW, Han I, Oh S, Choi M. Ecological Momentary Assessment Using Smartphone-Based Mobile Application for Affect and Stress Assessment. Healthc Inform Res 2018 Oct;24(4):381-386 [FREE Full text] [doi: 10.4258/hir.2018.24.4.381] [Medline: 30443427]
23. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally? In emerging economies, technology use still much more common among young people and the well-educated. Pew Research Center: Global Attitudes and Trends.: Pew Research; 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-05-20]
24. e-Country indicators. National government monitoring indicators. Seoul, Republic of Korea: National Indicator System URL: http://www.index.go.kr/main.do [accessed 2020-05-21]

25. Verhagen SJW, Berben JA, Leue C, Marsman A, Delespaul PAEG, van Os J, et al. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. PLoS One 2017;12(10):e0186294 [FREE Full text] [doi: 10.1371/journal.pone.0186294] [Medline: 29023568]
26. Lenaert B, Colombi M, van Heugten C, Rasquin S, Kasanova Z, Ponds R. Exploring the feasibility and usability of the experience sampling method to examine the daily lives of patients with acquired brain injury. Neuropsychol Rehabil 2019 Jun;29(5):754-766. [doi: 10.1080/09602011.2017.1330214] [Medline: 28562164]
27. Myin-Germeys I, Birchwood M, Kwapil T. From environment to therapy in psychosis: a real-world momentary assessment approach. Schizophr Bull 2011 Mar;37(2):244-247 [FREE Full text] [doi: 10.1093/schbul/sbq164] [Medline: 21224492]
28. Kang Y, Na D, Hahn S. A validity study on the Korean mini-mental state examination (K-MMSE) in dementia patients. J Kor Neurol Assoc. 1997. URL: http://jkna.org/upload/pdf/9702006.pdf [accessed 2020-04-05]
29. Yang YS, Ryu GW, Park CG, Yeom I, Shim KW, Choi M. Mood and Stress Evaluation of Adult Patients With Moyamoya Disease in Korea: Ecological Momentary Assessment Method Using a Mobile Phone App. JMIR Mhealth Uhealth 2020 May 25;8(5):e17034 [FREE Full text] [doi: 10.2196/17034] [Medline: 32449687]
30. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
31. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
32. Oh S, Min K, Park D. A study on the standardization of the Hospital Anxiety and Depression Scale for Koreans - A comparison of normal, depressed and anxious groups. J Korean Neuropsychiatr Assoc. 1999. URL: http://journal.kisep.com/pdf/002/1999/0021999026.pdf [accessed 2020-03-04]
33. Jones KK, Zenk SN, McDonald A, Corte C. Experiences of African-American Women with Smartphone-Based Ecological Momentary Assessment. Public Health Nurs 2016 Jul;33(4):371-380. [doi: 10.1111/phn.12239] [Medline: 26530781]
34. Start with the PsyMate™ yourself. PsyMate: Insight into daily life. URL: http://www.psymate.eu [accessed 2020-04-01]
35. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
36. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
37. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]
38. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
39. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Philadelphia, United States: F. A. Davis Company; 2015.
40. Hoffman L. Longitudinal analysis: Modeling within-person fluctuation and change. Abingdon, United Kingdom: Routledge; 1 edition; Jan 30, 2015.
41. Crowe E, Daly M, Delaney L, Carroll S, Malone KM. The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in Major Depression. Psychiatry Res 2019 Sep;279:98-108. [doi: 10.1016/j.psychres.2018.02.032] [Medline: 29661498]
42. Yang YS, Ryu GW, Yeom I, Shim KW, Choi M. Stress and mood of adults with moyamoya disease: A cross-sectional study. Nurs Health Sci (in press) 2020 Apr 26. [doi: 10.1111/nhs.12729] [Medline: 32336006]
43. Geldhof GJ, Preacher KJ, Zyphur MJ. Reliability estimation in a multilevel confirmatory factor analysis framework. Psychol Methods 2014 Mar;19(1):72-91. [doi: 10.1037/a0032138] [Medline: 23646988]
44. Muthén B, Asparouhov T. Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. Handbook of advanced multilevel analysis. New York: Taylor & Francis; 2011. URL: https://www.statmodel.com/download/multilevelVersion2.pdf [accessed 2020-03-06]

Abbreviations

CVI: content validity index
EFA: exploratory factor analysis
EMA: ecological momentary assessment
HADS: hospital anxiety and depression scale
ICC: intraclass correlation coefficient
I-CVI: item-level content validity index
K-PsyMate: Korean version of the PsyMate
MMD: moyamoya disease
S-CVI: scale-level content validity index

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by A Krishnakumar, R Adhikary; comments to author 31.03.20; revised version received 21.05.20; accepted 27.05.20; published 21.07.20.

Please cite as:
Yang YS, Ryu GW, Delespaul PAEG, Choi M
Psychometric Properties of the Korean Version of the PsyMate Scale Using a Smartphone App: Ecological Momentary Assessment Study
JMIR Mhealth Uhealth 2020;8(7):e17926
URL: http://mhealth.jmir.org/2020/7/e17926/
doi:10.2196/17926
PMID:32706707

Original Paper

Health Observation App for COVID-19 Symptom Tracking Integrated With Personal Health Records: Proof of Concept and Practical Use Study

Keiichi Yamamoto[1], PhD; Tsubasa Takahashi[2], MSc; Miwa Urasaki[3], BSc; Yoichi Nagayasu[4], MSc; Tomonari Shimamoto[5], PhD; Yukiko Tateyama[5], MPH, PhD; Keiichi Matsuzaki[5], MD, PhD; Daisuke Kobayashi[5], MD, PhD; Satoshi Kubo[6], MSc; Shigeyuki Mito[7], BA; Tatsuya Abe[4], BSc; Hideo Matsuura[8], MD, PhD; Taku Iwami[5], MD, PhD

[1]Information Technology Center, Wakayama Medical University, Wakayama, Japan

[2]beyondS Inc, Tokyo, Japan

[3]Wakayama City Public Health Center, Wakayama City, Japan

[4]Healthtech Laboratory, Inc, Kyoto, Japan

[5]Kyoto University Health Service, Kyoto, Japan

[6]Suzuka University of Medical Science, Mie, Japan

[7]TMI Associates, Tokyo, Japan

[8]Wakayama City Public Health Center, Wakayama, Japan

Corresponding Author:
Keiichi Yamamoto, PhD
Information Technology Center
Wakayama Medical University
811-1 Kimiidera
Wakayama, 641-8509
Japan
Phone: 81 73 441 0901
Email: kyamamo@wakayama-med.ac.jp

Abstract

Background: As a counter-cluster measure to prevent the spread of the infectious novel coronavirus disease (COVID-19), an efficient system for health observation outside the hospital is urgently required. Personal health records (PHRs) are suitable for the daily management of physical conditions. Importantly, there are no major differences between the items collected by daily health observation via PHR and the observation of items related to COVID-19. Until now, observations related to COVID-19 have been performed exclusively based on disease-specific items. Therefore, we hypothesize that PHRs would be suitable as a symptom-tracking tool for COVID-19. To this end, we integrated health observation items specific to COVID-19 with an existing PHR-based app.

Objective: This study is conducted as a proof-of-concept study in a real-world setting to develop a PHR-based COVID-19 symptom-tracking app and to demonstrate the practical use of health observations for COVID-19 using a smartphone or tablet app integrated with PHRs.

Methods: We applied the PHR-based health observation app within an active epidemiological investigation conducted by Wakayama City Public Health Center. At the public health center, a list is made of individuals who have been in close contact with known infected cases (health observers). Email addresses are used by the app when a health observer sends data to the public health center. Each health observer downloads the app and installs it on their smartphone. Self-observed health data are entered daily into the app. These data are then sent via the app by email at a designated time. Localized epidemiological officers can visualize the collected data using a spreadsheet macro and, thus, monitor the health condition of all health observers.

Results: We used the app as part of an active epidemiological investigation executed at a public health center. During the investigation, 72 close contacts were discovered. Among them, 57 had adopted the use of the health observation app. Before the introduction of the app, all health observers would have been interviewed by telephone, a slow process that took four epidemiological officers more than 2 hours. After the introduction of the app, a single epidemiological officer can carry out health observations. The app was distributed for free beginning in early March, and by mid-May, it had been used by more than 20,280 users and 400

facilities and organizations across Japan. Currently, health observation of COVID-19 is socially recognized and has become one of the requirements for resuming social activities.

Conclusions: Health observation by PHRs for the purpose of improving health management can also be effectively applied as a measure against large-scale infectious diseases. Individual habits of improving awareness of personal health and the use of PHRs for daily health management are powerful armaments against the rapid spread of infectious diseases. Ultimately, similar actions may help to prevent the spread of COVID-19.

(JMIR Mhealth Uhealth 2020;8(7):e19902) doi:10.2196/19902

KEYWORDS

public health informatics; public health administration; emerging infectious disease; preventive medicine; mobile apps; contact tracing

Introduction

The coronavirus disease (COVID-19) is widespread around the world. In clusters of the virus, there is likely to be a continuous generation of outbreaks, which may lead to larger-scale outbreaks [1]. In Japan, between 10% and 20% of all patients who are infected produce secondary infections [2]. Countermeasures aimed at preventing the spread of infection have centered on accurate detection and appropriate response to these clusters [3,4]. Accordingly, an effective and straightforward system for observing health status outside traditional health care settings would be useful as a counter-cluster measure for preventing the spread of COVID-19.

In Japan, COVID-19 is a designated infectious disease as defined by the relevant infectious disease control laws [3]. If a patient is suspected to have a COVID-19 infection based on clinical characteristics and diagnosis is confirmed by a polymerase chain reaction (PCR) test, the confirmed patient is hospitalized and isolated. An active epidemiological investigation is then conducted by public health centers, which are public institutions established by local governments based on the Japanese Community Health Act [5]. At public health centers, retrospective contact tracing [6-8] of confirmed patients and prospective symptom tracking [9,10] of their "close contacts," people with whom they have been in recent contact, are conducted. First, to identify the source of infection and make lists of close contacts for the 14 days before the onset of symptoms, a retrospective behavioral investigation is conducted based on whether the confirmed patient had participated in events characterized by the "three Cs" (closed spaces with poor ventilation, crowded places with many people nearby, or close-contact settings such as close-range conversations) [11,12], as well as their travel history. Second, to prevent secondary infections, close contacts are monitored for fever, dyspnea, coughing, and other symptoms for 14 days from the most recent date of exposure to a confirmed patient. To relieve pressure on the medical system, further measures have been implemented to care for mild or asymptomatic patients at home or in isolated accommodations [13-15].

Health observations conducted outside of traditional hospital settings can be difficult. In one example, patients on a large cruise ship were confirmed to have COVID-19 [16-19]. After quarantine, the remainder of the passengers with negative PCR tests disembarked [20,21]. The Ministry of Health, Labor and Welfare (MHLW) issued a report on health follow-up, and the public health centers in the nearby residential area conducted a health observation [22]. Such an investigation is time-consuming because it is based on telephone interviews or similar formats [23]. In another example, a hospital doctor was confirmed to be infected, and 13 other individuals were confirmed to have been infected by that doctor within 10 days, mainly from close contact [24,25]. Such heavy investigation of close contacts prompted concerns that if the infection spread further, the public health center would be overloaded [23].

In Japan, which has become a super-aged society, it is critical to fill the gap between average life expectancy and healthy life expectancy [26]. Personal health records (PHRs) are expected to extend healthy life expectancy [27]. Based on their personal judgement, individuals record their medical, nursing, caregiving, and health-care–related data (ie, person-generated data [28]). Until now, this process has been considered suitable for daily management of an individual's physical condition [29-32]. The PHR allows the user to check their own health status (eg, to be aware of their physical activity by measuring the daily number of steps or to prevent overeating by monitoring their daily weight measurements). Importantly for disease control, the PHR can be used to detect early signs of infection through regular measurements of body temperature. Management of an individual's physical condition and self-care can theoretically be made easier by recording these data.

The Kyoto University Data Health Study Group, under the supervision of the Kyoto University Health Service, conducts research on shared lifelong PHR data based on annual health checkups for students and has created standardized models to promote the use of PHRs [33]. There are no major differences between the observation items collected by daily health observation for PHRs and the disease-specific observation items for COVID-19. We hypothesized that by expanding the PHR to collect observation items specific to COVID-19, it would be possible to efficiently observe the health of individuals outside of hospital settings, including the close contacts of confirmed infected people, as well as patients who are mildly symptomatic or asymptomatic at home or in isolated accommodations.

"K-note" (*Kenko-Nikki*; "health diary") is a PHR-related app developed by Healthtech Laboratory, Inc, a Kyoto University–originated venture company in the Kyoto University Incubation Program [34]. The Kyoto University Data Health Study Group members and other volunteers, including researchers at Wakayama Medical University, gathered data

and added health observation functions specific for COVID-19 to the PHR data.

The objective of this study is to determine whether PHRs could be used for efficient health observation of emerging infectious diseases among individuals outside a traditional hospital setting. In addition, we sought to demonstrate the practical use of a smartphone and tablet app that supports PHR-based health observation by integrating monitoring functions specific to COVID-19. We explored the development, use, and efficiency of this app relative to conventional methods in the setting of an actual active epidemiological investigation of COVID-19.

Methods

Basic Concept and Features of K-Note

The PHR smartphone and tablet app "K-note" was developed to manage various data based on an individual's input as an alternative to conventional paper-based health checkup information. It is used to record day-to-day health information such as the number of steps taken, body weight, and blood pressure, as well as medical information such as medications and vaccination history. Its principal purpose is to improve individual lifestyles and, ultimately, extend healthy life expectancy.

Since July 2019, K-note has been distributed free of charge to study participants as part of a proof-of-concept study of health promotion using PHRs. We recently added a COVID-19 health observation function to the PHR app and made the software available to the public free of charge.

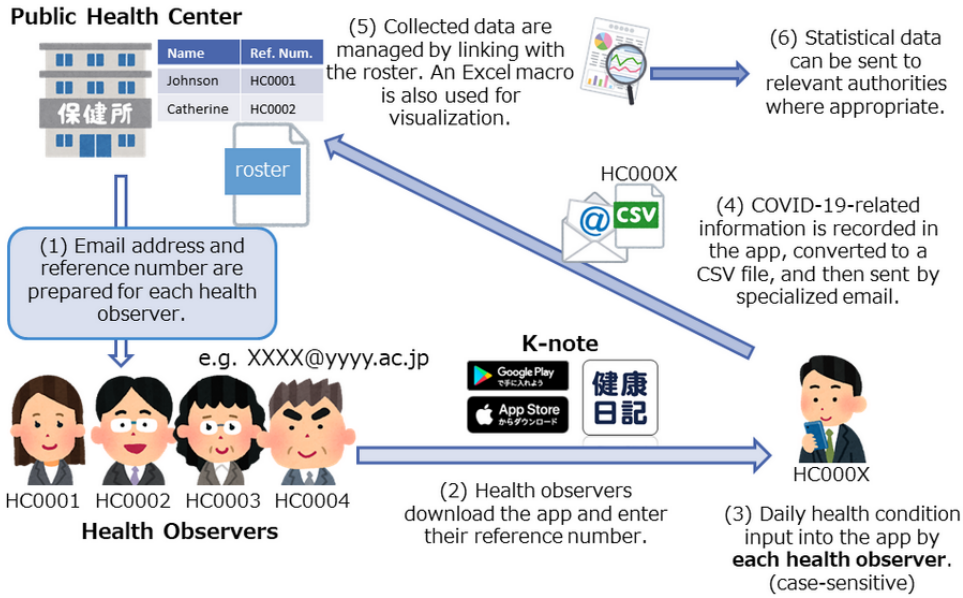
Initially, the integration of health observation data specific to COVID-19 in the K-note app was intended to streamline health observation within schools and companies of groups suspected of being infected. However, following an incidence of COVID-19 infection, at the request of Wakayama City Public Health Center, the app was applied in the context of an active epidemiological investigation [24]. To improve the efficiency of health observation work at the public health center, we referred to the survey items of the National Institute of Infectious Diseases [35] and created a Microsoft Excel (Microsoft Corporation) macro for data visualization.

In general, personal information protection and network security in Japan is strict, and there are often severe restrictions on access to webpages and social networking services, as well as personal computer settings. Because K-note is intended to be used for health observation related to COVID-19, it was designed to be usable in various information and communication technology environments, including use by local governments. The basic concepts underlying development are as follows: only the app and the readily available Excel software package are required, operation can start immediately, and due consideration is given to the protection of personal information. The benefits of the app include its availability as a free download (ie, App Store, Google Play). Data specific to COVID-19 can be recorded and managed with other health observation data, and can be sent by email to the specified destination with one click. Individuals can easily review their own health observation data, and the epidemiological officer can easily visualize the comma-separated values (CSV) data they have received using the Excel macro for data visualization. Data are managed only within the smartphone or tablet and are not made available elsewhere unless the individual who has been in close contact with an infected individual and is to be observed (health observer) specifically authorizes data sharing. K-note uses the health observation form to show the state of the health observer as "not infected" (ordinary health follow-up, close contact, etc) or "diagnosed with infection" (mild or asymptomatic), and users can choose from these options. Here, we show an example of "not infected" (ordinary health follow-up, a close contact of a known infected individual).

Overview of Health Observation for Close Contacts at the Public Health Center Using K-Note

At the public health center, a list of health observers is made; a reference number, used in management, is assigned at this time. A special email address is also created that is used when the health observer sends data to the public health center. Every day by the designated time, the health observer uses the integrated email address to send their data related to COVID-19, such as body temperature and symptoms, which were recorded by the app. When the information has been received by the epidemiological officer at the public health center, they can visualize the collected CSV data using an Excel macro, and this can be used to monitor the health condition of all health observers. If appropriate, statistical data can be sent to authorities such as the MHLW and quarantine stations (Figure 1).

Figure 1. Overview of health observation of close contacts at the public health center using K-note. CSV: comma-separated values.



Health observation at the public health center using K-note is performed in three phases: preparation, health observation by K-note, and data visualization at the public health center.

Phase 1: Preparation

As part of an active epidemiological investigation, the condition of confirmed patients is investigated at the public health center, and a list of health observers, who are to be observed for 14 days, is made. The epidemiological officer at the public health center acquires a reference number for each health observer. The ”roster“ sheet of the Excel macro for data visualization serves as the epidemiological officer management’s ledger of health observers.

Because unique email addresses are used to transmit data to the public health center, the epidemiological officer can monitor each health observer separately, which may facilitate the investigation. Although this is the system we have used, it may be useful to create the email address differently according to the operations of the public health center (eg, by creating one for each epidemiological officer at the public health center or one for each group of close contacts; Figure 2).

Figure 2. The “roster” sheet and the basic setting screen of K-note. PHR: personal health record.

Phase 2: Health Observation Using K-Note

The health observer records their daily health conditions such as body temperature and symptoms for 14 days from the start of the observation (Figure 3).

After data input, when the health observer presses the "send data page" button, the "send data page" appears. Health observers may input up to three email addresses. When the health observer presses the "send by email" button, the observation data is automatically converted into CSV format and attached to the email. By a designated time every day, the app is used to send the health observer's recorded observation data to the specified email addresses (Figures 4 and 5).

Figure 3. Health observation data input screen in K-note. Data entry items are body temperature, coughing, dyspnea, rhinorrhea, sore throat, fatigue, diarrhea, headache, other symptoms, medication, confirmation of whether a test has been taken, and test results.

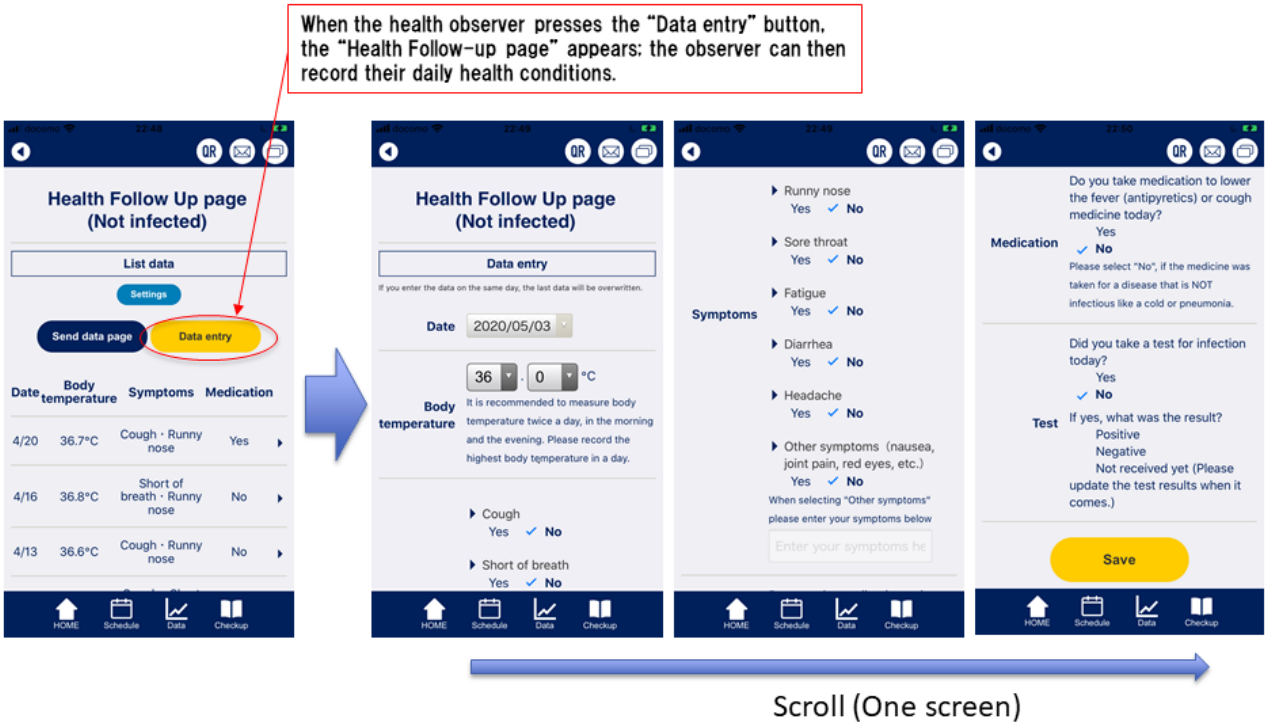
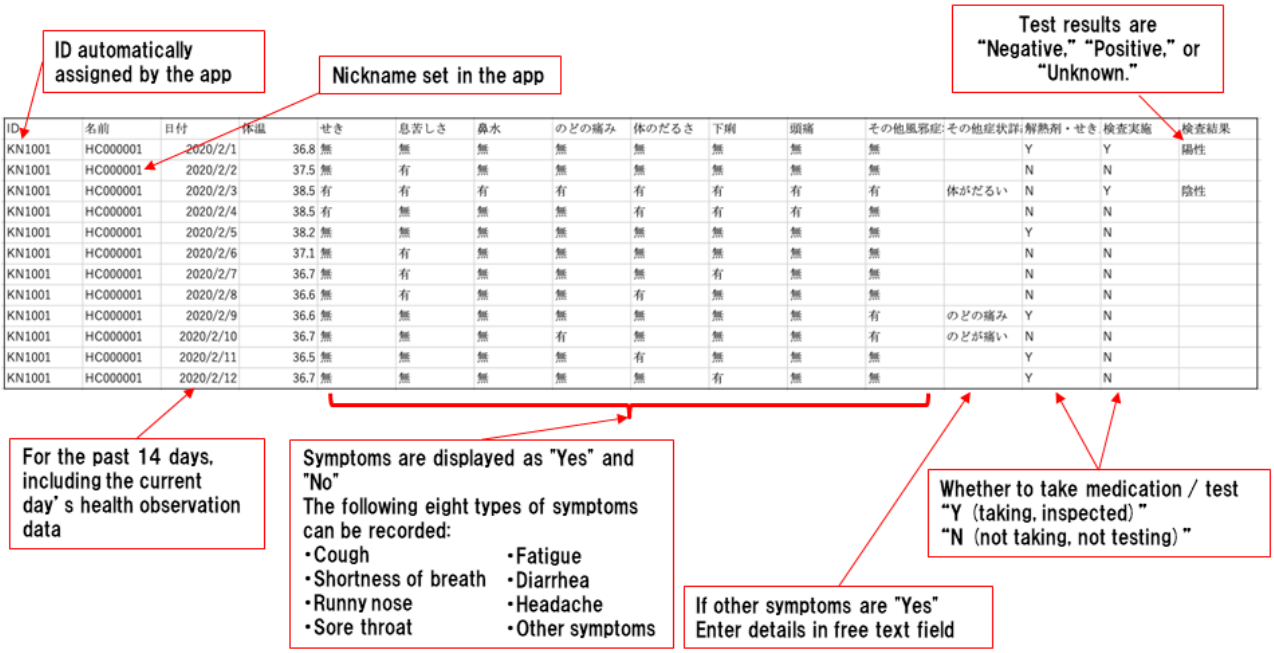


Figure 4. Email sending screens. CSV: comma-separated values.

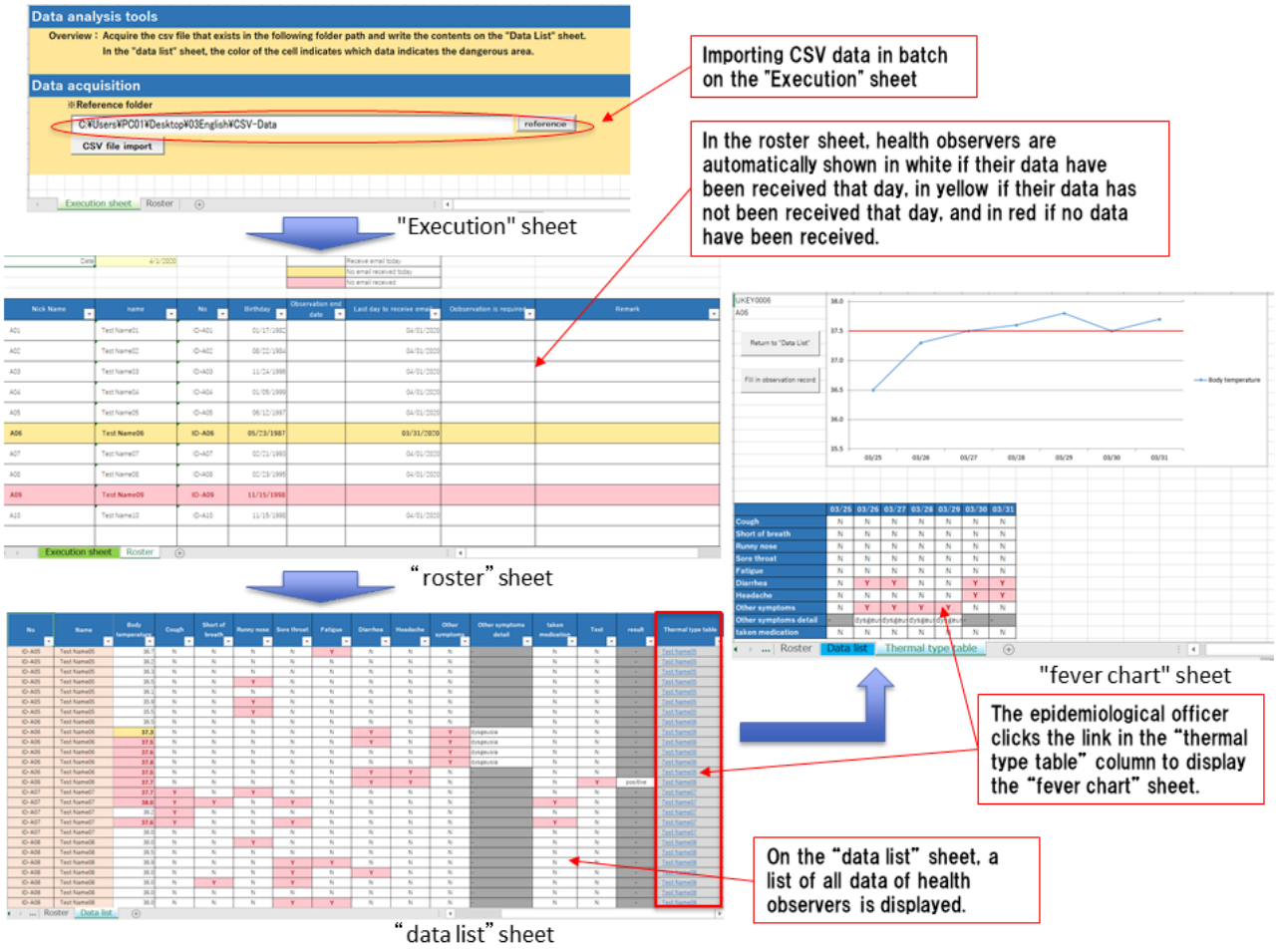
Figure 5. Transmitted comma-separated values data image.



Phase 3: Data Collection by the Epidemiological Officer at the Public Health Center

The epidemiological officer downloads the CSV data of all health observers received on that day and stores them in a single folder. By launching the Excel macro for data aggregation and batch importing the CSV data on the ”Execution“ sheet, all health observation data received on that day will be displayed on the ”data list“ sheet, and the “roster” sheet is updated. In the ”roster“ sheet, health observers whose health observation data were received that day are displayed in white, those whose data was not received on the day are displayed in yellow, and those whose data has never been received are automatically displayed in red. With this function, the epidemiological officer can easily distinguish the reporting status of all health observers. It is possible to adapt this system according to the situation (eg, it may be sufficient to contact only the health observers who are displayed in yellow or red). On the ”data list“ sheet, all the data for all the health observers are displayed. Excel functions can be used to rearrange the health observers and display only a particular day. Regarding body temperature, yellow is automatically used to distinguish health observers whose reported body temperature was ≥37°C, and red is used to distinguish those with temperatures ≥37.5°C. Coughs or other symptoms are automatically displayed in red. If the epidemiological officer checks the “data list” sheet and a health observer requires more detailed observation, they may click the link in the ”Thermal type table“ column to display the ”fever chart“ sheet. Here, body temperature is displayed graphically over time, and 37.5°C is indicated by a red line so that it is possible to see at a glance whether there are days when the temperature exceeded the standard value. Symptoms are also listed over time and displayed in red (Figure 6).

Figure 6. Screen images of the “execution,” “roster,” “data list,” and “fever chart” sheets. CSV: comma-separated values.



Results

Active Epidemiological Investigation Executed by the Public Health Center

We started developing the health observation app on February 22, 2020, and registered the initial version in the app store on February 27. On March 5, a patient with COVID-19 at a workplace in Wakayama City was confirmed, and the use of the initial version of the app was started at the Wakayama City Public Health Center on March 6. We immediately customized the health observation app according to the actual epidemiological investigation and released an updated version on March 15. Health observation had been performed for 14 days starting from the day of contact (exposure) with the confirmed patient. In principle, the investigation would be tracked daily until all health observers were contacted. For those who were exposed before March 5 and were found to be infected by health observation, observation will be completed without waiting for the final day. During the active epidemiological

investigation, the Wakayama City Public Health Center discovered 72 health observers who had close contact with a confirmed case. Among them, 57 (79.2%) adopted the use of the health observation app and 14 used telephone investigations; the observation period of the remaining health observers ended on March 5 (Table 1).

Table 1. Results of health observation from March 6, 2020.

Date in 2020	Day of week	Investigation using the app		Investigation by phone	
		Health observers, n	Spontaneous email senders, n	Health observers, n	Those who could be contacted by one phone call, n
Total		729	632	171	154
March 6	Friday	57	48	14	14
March 7	Saturday	57	54	14	14
March 8	Sunday	57	56	14	13
March 9	Monday	57	54	14	14
March 10	Tuesday	57	49	14	13
March 11	Wednesday	56	44	14	14
March 12	Thursday	53	45	12	8
March 13	Friday	53	45	11	8
March 14	Saturday	51	44	11	10
March 15	Sunday	50	39	11	10
March 16	Monday	50	44	11	11
March 17	Tuesday	49	40	11	11
March 18	Wednesday	46	40	10	7
March 19	Thursday	36	30	10	7

During the app-based investigation, the 6 health observers who did not send an email on March 19, 2020, were contacted by phone by an epidemiological officer to inform them of the end of the observation; their health condition was checked at that time. In total 57 of 72 health observers (79.2%) chose the app. The spontaneous email transmission rate was 86.7% (n=632/729).

Before the app was introduced, all health observers were interviewed by telephone, a process that took four epidemiological officers more than 2 hours. This suggested that, should the virus spread further, it would be difficult to maintain investigations by telephone with limited staff over 14 days. Additionally, the comprehensiveness of the investigation may be compromised because health observers might be absent or epidemiological officers might not be able to make a phone call because of other work. Furthermore, the health observer also had to wait in a place where they could answer the telephone during the survey period, placing a heavy burden on both the health observer and the epidemiological officers. With the introduction of K-note, it became possible to automatically manage information collectively; the health observers sent their data at a designated time every day, and a single epidemiological officer could carry out health observations alone. In addition, the arrival status of observation data and the health condition of the health observer were readily visualized, which improved the efficiency and comprehensiveness of the investigation. To improve efficiency, the app was modified 6 times, and the Excel macro 3 times, by mid-May.

Usability Evaluation of K-Note

As of May 12, 2020, K-note has been used by more than 20,280 individuals and 400 facilities or organizations, centered on companies, schools, hospitals, and local governments in Japan. The breakdown by facility was as follows: 166 (41.5%) companies, 90 (22.6%) schools, 51 (12.7%) hospitals, and 47 (11.7%) from local government. In addition, it has been introduced in 24 of 47 prefectures and is used in 35 public health centers (Figures 7 and 8).

Figure 7. The number of app users between March 1 and May 12, 2020.

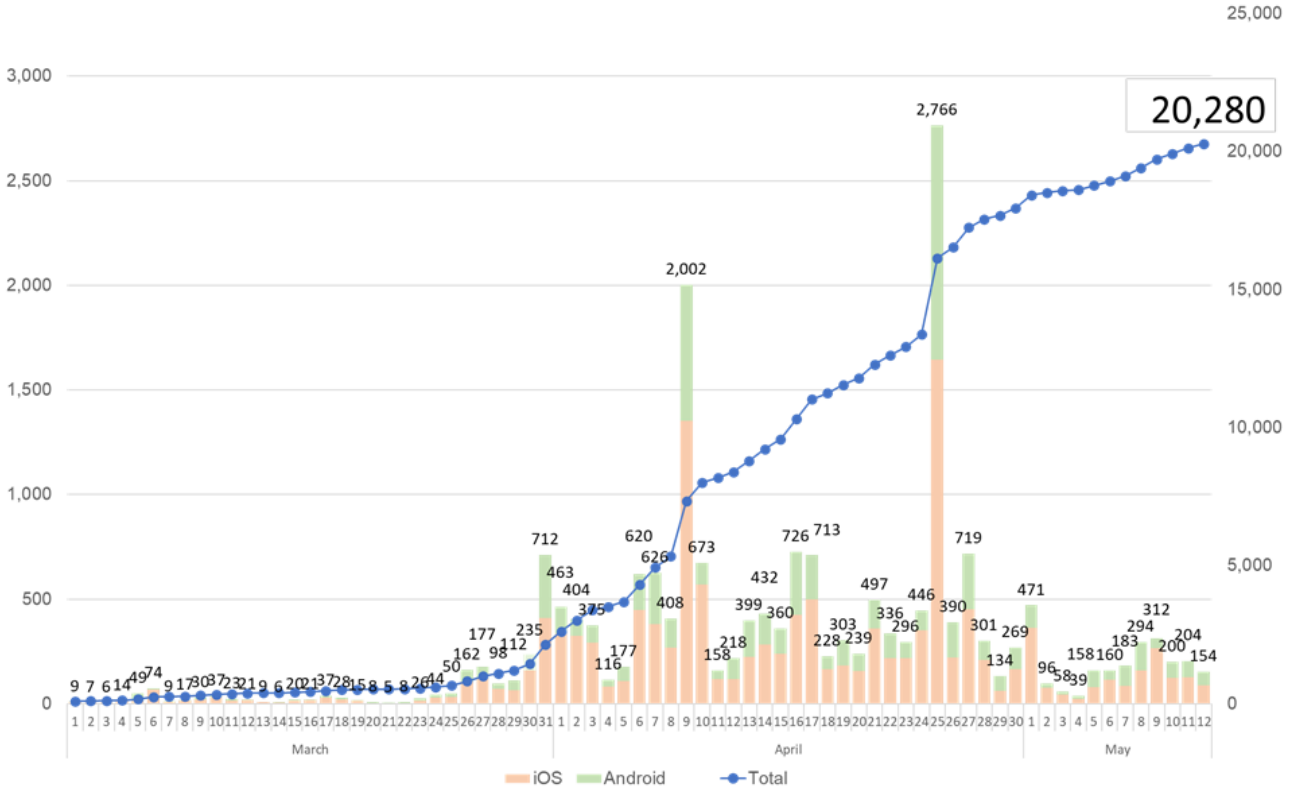
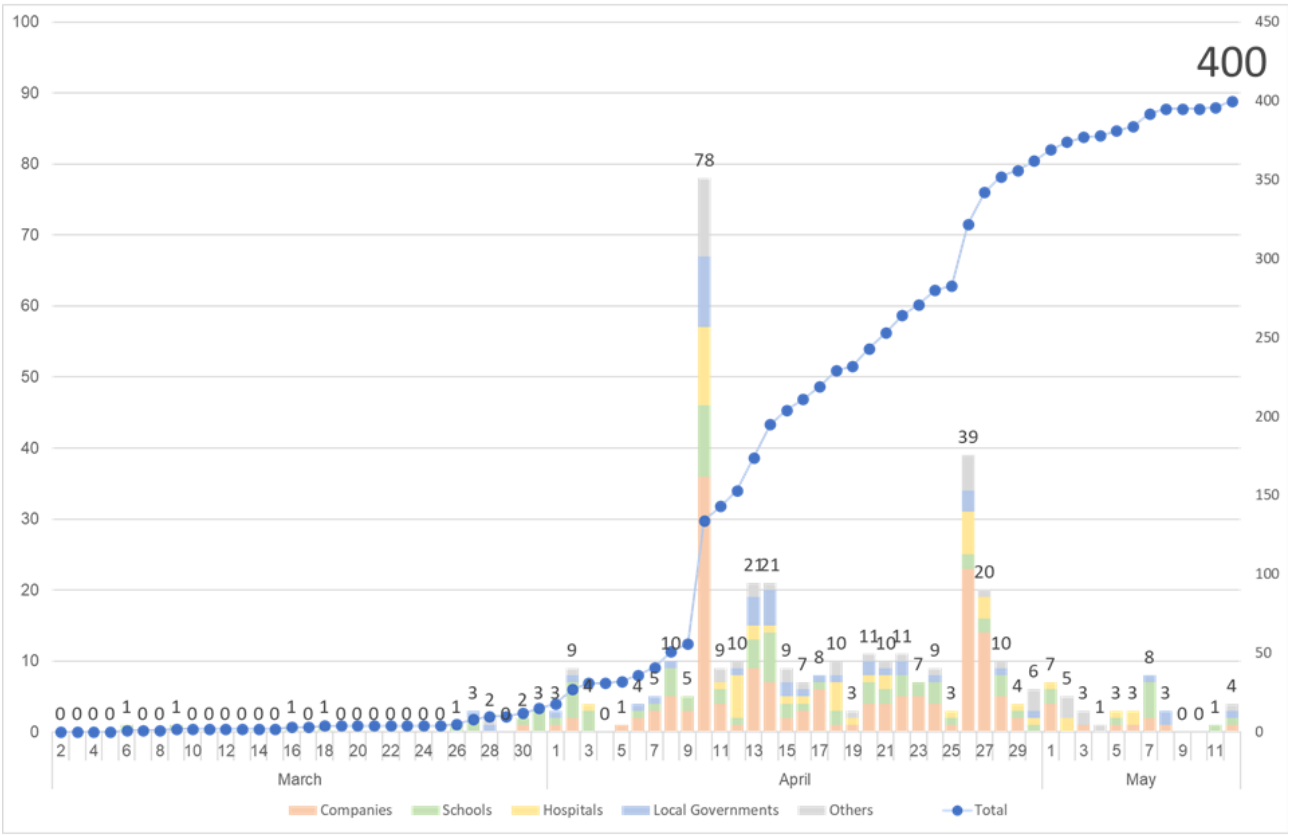


Figure 8. The number of facilities and organizations between March 1 and May 12, 2020.



At the time the app was released, only email support was implemented, but with the increase in the number of inquiries, a help desk officially opened on March 28. By mid-May, the total number of inquiries to the help desk was 99. The breakdown of inquiries was: 18 (18%) “Can’t send emails from the app,” 17 (17%) “App functions and usage,” 20 (20%) “Can’t

download Excel macro," 26 (26%) "Excel macro function and usage," and 18 (18%) "Other inquiries." In addition, online operation briefing sessions for the app were held eight times by mid-May for a total of 152 participants.

To evaluate the usability of the app, we interviewed the help desk manager about the context of use, effectiveness, efficiency, and satisfaction according to the definition of usability in ISO 9241-11:2018 [36]. The results are shown in the following sections.

Context of Use

The use of apps started with local governments and gradually spread to schools and companies. Initially, a tool for health observation was required as a COVID-19 countermeasure, and many facilities decided to introduce our app because of the motivation to respond promptly rather than seeking efficiency and effectiveness. Many users expressed their gratitude for an Excel-based system that could be introduced immediately, even if it was a bit inconvenient, because they did not know the specific method for health observation and there was no alternative system. As social awareness of the need for health observation increased, the number of users from large universities and companies increased. Since mid-May, the number of inquiries by health observers has increased due to the state of emergency being lifted in Japan and the subsequent movement toward resuming social activities [37].

Effectiveness

From the viewpoint of effectiveness, the ability to record daily health condition was of great significance as a countermeasure against COVID-19. Because it was possible to look back when some problem occurred, the app could be used for health observation not only in active epidemiological investigation at public health centers but also for health observation before and after events such as educational practice, face-to-face lectures, and physical education classes.

Efficiency

Because of the rapid spread of COVID-19, no one was demanding optimal efficiency. Still, we must acknowledge that our method for exchange of CSV data by email and manual data visualization by Excel macro was not efficient. In addition, there were many inquiries from local governments and companies indicating that, for security reasons, it was not possible to use web forms for the online briefing sessions or to download files (including the Excel) macro on the internet.

In the flood of various information technologies, the following two points are presumed to have driven increased demand for our app.

1. Among members of the general public, there was not a habit of taking daily measurements of body temperature and monitoring changes over time to lead to behavioral changes without a doctor's guidance; however, there was no alternative system that could also manage personal consent and security.
2. Even the local government had no experience with exchanging health information electronically with the general public.

Satisfaction

From the viewpoint of satisfaction, the administrator of a facility or organization was able to fulfill their social responsibility by introducing our app as the infection of COVID-19 spread. Health observers also had a sense of security that someone was observing their health condition. Therefore, it could be presumed that satisfaction was high. Recently, health observation became one of the requirements for resuming social activities in Japan.

Discussion

Principal Findings

We extended PHRs and developed a PHR-based health observation app for counter- cluster measures of COVID-19. As a result of applying the new app in an active epidemiological investigation at a public health center, we believe that the efficiency and completeness of the investigation have been improved, preventing the spread of infection. The app is used nationwide in Japan, mainly by companies, schools, hospitals, and local governments for health observation for the prevention of COVID-19. Except for active epidemiological investigations at public health centers, confirmed patients may not have been identified by self-reporting from schools or companies alone. However, the app has contributed to the early detection of COVID-19 or voluntary self-quarantine at home by individuals with suspected symptoms, and the use of the app can facilitate reopening of school and corporate activities. Our findings suggest that health monitoring of PHRs is useful for carrying out efficient health observation of emerging infectious diseases for individuals outside of a traditional hospital setting.

A factor contributing to the success of our app is that we assumed that it would be used for health observation of small groups; accordingly, we designed the system configuration so that it could be introduced immediately without a cloud system; the user only requires the app itself and Excel. Many local governments in Japan have strict control of security and protection of personal information, and are often restricted from displaying external webpages and using social networking services from their office local area networks. A system for storing health observation data in the cloud and sharing data between the health observer and the public health center would require security-related deliberation and entail large-scale initiatives for entire cities or prefectures. On the other hand, our method can be implemented immediately and at low cost in a comparatively simple ICT environment in which emails can be exchanged between the health observer and the public health center using just the app and Excel. It is therefore possible for each public health center to make a decision about whether to introduce the system. Additionally, because it was based on the COVID-19 health observation items of the National Institute of Infectious Diseases and reflected the opinions of preventive medicine specialists and epidemiological officers at public health centers, the app was practical.

The basis of this study was the idea that PHRs, which are used to promote health via daily observation, can be used as a counter-cluster measure against large-scale infectious diseases. In the case of highly infectious diseases, infection is often found in people in close contact with a known infected patient, but

they often do not have symptoms at the start of the observation. For this reason, it is extremely difficult to undertake comprehensive health observations outside the hospital, including of asymptomatic individuals or those with mild symptoms, those under isolation, those who are suspected to be infected based on airport quarantine, and those within active epidemiologic investigations. It is obviously too late to start preparations after an infectious disease begins to spread, and ideally, a system would be established during normal times. However, it is socially difficult to invest in and maintain costly measures against large-scale infectious diseases that occur once every few years or even several decades. Although there are observation items specific to infectious disease control, they are just an extension of daily health observation. We believe that the idea of converting daily health promotion activities by PHR into countermeasures against large-scale infectious diseases in emergencies could enable a swift response to these illnesses.

The popularization of PHRs is key to making daily health observations the basis of counter-cluster measures against large-scale infections. It is important for promoting the significance of the habit of self-management of health data and as a tool for individuals to maintain their own health throughout society. As a countermeasure against COVID-19, many people have undertaken health observation using our app in Japan. We would like to further promote the habit of PHR-mediated health observation as a social infrastructure even after the end of COVID-19. To do that, more people must become aware of the benefits of maintaining PHRs. Taking advantage of the experience of developing a health observation app for COVID-19, we would like to develop various additional apps of the same kind, not only as a measure against emerging infectious diseases but also for health management for public marathons, in club activities in educational institutions, and after large-scale disasters.

Limitations

For research, detailed quantitative and qualitative evaluation items should be determined, and data should be collected at the time that the research plan is created. However, due to the urgency of preventing the spread of COVID-19, app development and an introduction to public health centers were given priority. In this case, we only performed qualitative evaluation of the burden reduction effect by interviewing epidemiological officers and help desk managers. Another limitation is that some members of society do not have access to smartphones, tablets, or Wi-Fi connections. Although the process is explained well here, some people may find the method difficult to use if they are not familiar with information and communications technology. We distribute to these health observers an Excel survey sheet that can be read into the data visualization Excel macro. In addition, a proxy input function may be necessary. The current Excel macro for data visualization was developed for the health observation of individuals who have been in close contact with known infected patients at public health centers. We believe that this is suitable for surveys of up to 100 close contacts per confirmed patient. If the number of people exceeds 100, a more sophisticated data visualization system may be required. It would also be possible to develop a retrospective behavioral investigation support app for confirmed patients and close contacts using smartphones and PHR behavioral history information, but such an app has not been developed due to concerns about protecting personal information [38,39]. Currently, contact-tracing apps are being developed all over the world [6-8]. In Japan, as a counter-cluster measure, retrospective behavioral investigation of confirmed patients, explorations of close contacts, and prospective symptom tracking of close contacts are executed as active epidemiological investigations, but the amount of work required is enormous. If contact-tracing and symptom-tracking apps [9,10,40] can be used in combination in an active epidemiological investigation, it may be possible to greatly reduce the workload.

As a measure against future large-scale infectious diseases after the termination of COVID-19, it would be useful to develop an infectious disease control support network based on PHRs that could seamlessly respond to signs of infectious disease spread active case findings of behavioral investigation of confirmed patients and close contacts, health observation of out-of-hospital close contacts and mild or asymptomatic patients, and provide statistics to local governments and national headquarters.

Conclusions

We developed a health observation app integrated with PHRs as a COVID-19 counter-cluster measure. The app greatly reduced the follow-up burden of individuals who had close contact with known cases of confirmed COVID-19 infection. The relatively low-tech nature of the app and Excel combination meant that it was easily accessible, especially in Japan. The system uses emails rather than the cloud, an approach that is arguably more compatible with business and privacy practices specific to Japan. Health observation by PHRs for the purpose of improving health management and extending healthy life expectancy by using individuals' own health information is also effective as a countermeasure against large-scale infectious diseases. We believe that raising the awareness of each person about their own health and the value of using PHRs for daily health management is a powerful weapon against the rapidly expanding spread of infectious diseases. We hope that our study will help prevent the spread of COVID-19 infections and future large-scale infectious diseases around the world.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research Grant Number 19K12867, the Kyoto University Incubation Program, and the Kyoto University Health Service and HealthTech Laboratory Joint Research Fund for "Research on construction and promotion of standardized model for shared lifelong PHR based on university health checkup." These funding sources did not influence the conduct of this work.

The authors acknowledge University President Kazuhisa Miyashita, Prof Tatsuya Takeshita of Wakayama Medical University, and Masahiko Sakaguchi of Osaka Electro-Communication University for their kind support, as well as Hisakazu Tamaki, Yusuke Shimokawahara, Atsushi Matsuoka, and Masahiro Akita of Union Think for excellent software development.

Authors' Contributions

KY and TI were responsible for the conception and design. TT, YN, SK, and TA contributed to the app and data visualization tool development. TS, YT, KM, and DK contributed to the design and provided knowledge about preventive medicine. MU and HM contributed to the design and provided knowledge and executed active epidemiological investigation at the public health center. SM provided the legal knowledge. All authors read and approved the final manuscript.

Conflicts of Interest

TI is an outside director of HealthTech Laboratory and receives grants from the Kyoto University Health Service and HealthTech Laboratory Joint Research Fund. TS and YT are employed by the Kyoto University Health Service and HealthTech Laboratory Joint Research Fund. The other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
Excel macro and demo data (demo_data_visualization.zip).
[ZIP File (Zip Archive), 109 KB - mhealth_v8i7e19902_app1.zip]

References

1. Considerations in the investigation of cases and clusters of COVID-19. World Health Organization. 2020 Apr 02. URL: https://www.who.int/publications-detail/considerations-in-the-investigation-of-cases-and-clusters-of-covid-19 [accessed 2020-06-20]
2. Nishiura H, Oshitani H, Kobayashi T, et al, MHLW COVID-19 Response Team. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19) [preprint]. medRxiv 2020 Apr 16. [doi: 10.1101/2020.02.28.20029272]
3. Act on the Prevention of Infectious Diseases and Medical Care for Patients with Infectious Diseases. Act No. 114. Ministry of Health, Labor and Welfare. URL: http://www.japaneselawtranslation.go.jp/law/detail/?vm=04&re=01&id=2830 [accessed 2020-06-20]
4. Guidelines for active epidemiological investigation in patients with novel coronavirus infection (provisional version). National Institute of Infectious Diseases. URL: https://www.niid.go.jp/niid/en/2019-ncov-e/2484-idsc/9472-2019-ncov-02-en.html [accessed 2020-06-20]
5. Katsuda N, Hinohara Y, Tomita K, Hamajima N. Structure and roles of public health centers (hokenjo) in Japan. Nagoya J Med Sci 2011 Feb;73(1-2):59-68. [Medline: 21614938]
6. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
7. Ekong I, Chukwu E, Chukwu M. COVID-19 mobile positioning data contact tracing and patient privacy regulations: exploratory search of global response strategies and the use of digital tools in Nigeria. JMIR Mhealth Uhealth 2020 Apr 27;8(4):e19139 [FREE Full text] [doi: 10.2196/19139] [Medline: 32310817]
8. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 contact tracing and data protection can go together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
9. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 May 11. [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
10. Leong AY, Makowsky MJ. Quality of blood pressure tracking apps for the iPhone: content analysis and evaluation of adherence with home blood pressure measurement best practices. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e10809 [FREE Full text] [doi: 10.2196/10809] [Medline: 30977739]
11. Avoid the "Three Cs"!. Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/content/10900000/000619576.pdf [accessed 2020-06-20]
12. Suzuki K. COVID-19 strategy: the Japan model. The Japan Times. 2020 Apr 28. URL: https://www.japantimes.co.jp/opinion/2020/04/28/commentary/japan-commentary/covid-19-strategy-japan-model/#.XtuZp-fgo2w [accessed 2020-06-20]
13. About preparation for accommodation and home treatment for people with mild illness of new-type coronavirus infectious disease. Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/content/000618525.pdf [accessed 2020-06-20]
14. Tokyo negotiating with hotels to accommodate coronavirus patients with mild symptoms. The Mainichi. 2020 Apr 03. URL: https://mainichi.jp/english/articles/20200403/p2a/00m/0na/014000c [accessed 2020-06-20]
15. Japan to secure more hotel rooms for mild cases of COVID-19. The Mainichi. 2020 Apr 25. URL: https://mainichi.jp/english/articles/20200425/p2g/00m/0na/027000c [accessed 2020-06-20]
16. Field briefing: Diamond Princess COVID-19 cases. National Institute of Infectious Diseases. 2020 Feb 29. URL: https://www.niid.go.jp/niid/en/2019-ncov-e/9407-covid-dp-fe-01.html [accessed 2020-06-20]

17. Jimi H, Hashimoto G. Challenges of COVID-19 outbreak on the cruise ship Diamond Princess docked at Yokohama, Japan: a real-world story. GHM 2020;2(2):63-65. [doi: 10.35772/ghm.2020.01038]
18. Nakazawa E, Ino H, Akabayashi A. Chronology of COVID-19 cases on the Diamond Princess cruise ship and ethical considerations: a report from Japan. Disaster Med Public Health Prep 2020 Mar 24:1-8 [FREE Full text] [doi: 10.1017/dmp.2020.50] [Medline: 32207674]
19. Yamahata Y, Shibata A. Preparation for quarantine on the cruise ship Diamond Princess in Japan due to COVID-19. JMIR Public Health Surveill 2020 May 11;6(2):e18821 [FREE Full text] [doi: 10.2196/18821] [Medline: 32365046]
20. Disembarkation operation at the end of quarantine period. Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/content/10906000/000596286.pdf [accessed 2020-06-20]
21. Coronavirus: passengers leave Diamond Princess amid criticism of Japan. BBC News. 2020 Feb 20. URL: https://tinyurl.com/y9h28geu [accessed 2020-06-20]
22. Health follow-up for the disembarkation of the Diamond Princess. Ministry of Health, Labor and Welfare. URL: https://www.mhlw.go.jp/content/10900000/000599379.pdf [accessed 2020-06-20]
23. Shimura K, Mukuda K, Lee Y. Public health center workers in Tokyo exhausted as coronavirus cases keep rising. The Mainichi. 2020 Apr 21. URL: https://mainichi.jp/english/articles/20200421/p2a/00m/0in/018000c [accessed 2020-06-20]
24. Novel coronavirus infection situation in Wakayama. Wakayama Prefecture. URL: https://www.pref.wakayama.lg.jp/prefg/041200/d00203387.html?SLANG=ja&TLANG=en&XMODE=0&XCHARSET=utf-8&XJSID=0 [accessed 2020-06-20]
25. Denyer S, Kashiwagi A. A region in Japan launched its own coronavirus fight. It's now called a 'model' in local action. The Washington Post. 2020 Mar 22. URL: https://www.washingtonpost.com/world/asia_pacific/japan-coronavirus-wakayama/2020/03/22/02da83bc-65f5-11ea-8a8e-5c5336b32760_story.html [accessed 2020-06-20]
26. Tokudome S, Hashimoto S, Igata A. Life expectancy and healthy life expectancy of Japan: the fastest graying society in the world. BMC Res Notes 2016 Oct 28;9(1):482 [FREE Full text] [doi: 10.1186/s13104-016-2281-2] [Medline: 27793196]
27. Kajiwara T. Universal health coverage and long-term care in Japan. Ministry of Health, Labor and Welfare. URL: https://www.who.int/ageing/health-systems/icope/icope-consultation/Day01-ToruKajiwara.pdf [accessed 2020-06-20]
28. Gray K, Martin-Sanchez FJ, Lopez-Campos GH, Almalki M, Merolli M. Person-generated data in self-quantification. A health informatics research program. Methods Inf Med 2017 Jan 09;56(1):40-45. [doi: 10.3414/ME15-02-0006] [Medline: 27782291]
29. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014 Mar 05;2(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3017] [Medline: 25098942]
30. Dameff C, Clay B, Longhurst CA. Personal health records: more promising in the smartphone era? JAMA 2019 Jan 29;321(4):339-340. [doi: 10.1001/jama.2018.20434] [Medline: 30633300]
31. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
32. General Incorporated Association PHR Promotion Committee. URL: https://phr.or.jp/ [accessed 2020-06-20]
33. The Kyoto University Data Health Study group. Kyoto University Health Service. URL: https://dh-lab.org/ [accessed 2020-06-20]
34. K-note. Healthtech Laboratory, Inc. URL: https://www.htech-lab.co.jp/covid19/ [accessed 2020-06-20]
35. Survey items. National Institute of Infectious Diseases. URL: https://www.niid.go.jp/niid/images/epi/corona/nCoV_survey2004020.xlsx [accessed 2020-06-20]
36. ISO 9241-11:2018(en) ergonomics of human-system interaction — part 11: usability: definitions and concepts. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2020-06-20]
37. COVID-19 press conference by the Prime Minister regarding the novel coronavirus. Prime Minister of Japan and his Cabinet. 2020 May 25. URL: https://japan.kantei.go.jp/98_abe/statement/202005/_00003.html [accessed 2020-06-20]
38. Lin L, Ping CK. Singapore built a coronavirus app, but it hasn't worked so far. Wall Street J 2020 Apr 22.
39. Exposure notifications: using technology to help public health authorities fight COVID-19. Google COVID-19 Information & Resources. URL: https://www.google.com/covid19/exposurenotifications/ [accessed 2020-06-20]
40. Lalloo C, Hundert A, Harris L, Pham Q, Campbell F, Chorney J, et al. Capturing daily disease experiences of adolescents with chronic pain: mHealth-mediated symptom tracking. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e11838 [FREE Full text] [doi: 10.2196/11838] [Medline: 30664472]

Abbreviations

COVID-19: coronavirus disease
CSV: comma-separated values
MHLW: Ministry of Health, Labor and Welfare
PCR: polymerase chain reaction
PHR: personal health record

Edited by G Strudwick, G Eysenbach; submitted 14.05.20; peer-reviewed by D Estrin, C Dameff; comments to author 03.06.20; revised version received 20.06.20; accepted 21.06.20; published 06.07.20.

Please cite as:
Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, Matsuzaki K, Kobayashi D, Kubo S, Mito S, Abe T, Matsuura H, Iwami T
Health Observation App for COVID-19 Symptom Tracking Integrated With Personal Health Records: Proof of Concept and Practical Use Study
JMIR Mhealth Uhealth 2020;8(7):e19902
URL: https://mhealth.jmir.org/2020/7/e19902
doi:10.2196/19902
PMID:32568728

Original Paper

Barriers to and Facilitators of the Prescription of mHealth Apps in Australian General Practice: Qualitative Study

Oyungerel Byambasuren[1], MD, MMedRes; Elaine Beller[1], MAppStat; Tammy Hoffmann[1], PhD; Paul Glasziou[1], PhD

Institute for Evidence-Based Healthcare, Bond University, Robina, Australia

Corresponding Author:
Oyungerel Byambasuren, MD, MMedRes
Institute for Evidence-Based Healthcare
Bond University
14 University Dr
Robina, 4226 QLD
Australia
Phone: 61 432097753
Email: obyambas@bond.edu.au

Abstract

Background: The ubiquity of smartphones and health apps make them a potential self-management tool for patients that could be prescribed by medical professionals. However, little is known about how Australian general practitioners and their patients view the possibility of prescribing mobile health (mHealth) apps as a nondrug intervention.

Objective: This study aimed to determine barriers and facilitators to prescribing mHealth apps in Australian general practice from the perspective of general practitioners and their patients.

Methods: We conducted semistructured interviews in Australian general practice settings with purposively sampled general practitioners and patients. The audio-recorded interviews were transcribed, coded, and thematically analyzed by two researchers.

Results: Interview participants included 20 general practitioners and 15 adult patients. General practitioners' perceived barriers to prescribing apps included a generational difference in the digital propensity for providers and patients; lack of knowledge of prescribable apps and trustworthy sources to access them; the time commitment required of providers and patients to learn and use the apps; and concerns about privacy, safety, and trustworthiness of health apps. General practitioners perceived facilitators as trustworthy sources to access prescribable apps and information, and younger generation and widespread smartphone ownership. For patients, the main barriers were older age and usability of mHealth apps. Patients were not concerned about privacy and data safety issues regarding health app use. Facilitators for patients included the ubiquity of smartphones and apps, especially for the younger generation and recommendation of apps by doctors. We identified evidence of effectiveness as an independent theme from both the provider and patient perspectives.

Conclusions: mHealth app prescription appears to be feasible in general practice. The barriers and facilitators identified by the providers and patients overlapped, though privacy was of less concern to patients. The involvement of health professionals and patients is vital for the successful integration of effective, evidence-based mHealth apps with clinical practice.

(JMIR Mhealth Uhealth 2020;8(7):e17447) doi:10.2196/17447

KEYWORDS

mobile apps; mHealth; apps; app prescription; general practice

Introduction

The number of smartphones and mobile health (mHealth) apps has risen globally, with Australians at the forefront with smartphone ownership near 90% of the population [1]. In addition to fitness and wellness, mHealth apps are primarily created for and can benefit patients in managing chronic diseases [2]. More than half of US consumers have downloaded at least one mHealth app [3]. Despite the high initial uptake of apps, user retention rates can be low, and the duration of app usage can be short [4,5]. However, according to AppScript, an app prescription platform, prescribed mHealth apps have a higher retention rate than nonprescribed apps [2].

Health professionals prescribe mHealth apps in their practice in varying degrees [6-9]. Although relevant professional organizations provide some guidance on how to prescribe

mHealth technology in clinical practice, health professionals are often left to navigate this new area on their own [10-12]. A systematic review by Gagnon et al [13] identified about 180 individual barriers and facilitators to adoption of mHealth by health professionals, about third of which reflect factors directly relevant to their knowledge, attitude toward, and acceptance of mHealth. However, these findings were not specific to general practice.

In Australia, the Royal Australian College of General Practitioners (RACGP) offers limited guidance on mHealth apps. The college also has been collecting basic data on providers' app usage as part of their Annual Technology survey [14]. However, the survey has not explored potential barriers to mHealth app adoption in-depth. It is essential to explore the issues around app prescription in general practice in order to devise effective interventions to overcome the barriers perceived by the practitioners. Therefore, the objectives of this study were to determine the barriers to and facilitators of the prescription of mHealth apps in Australian general practice from the perspectives of general practitioners and their patients.

Methods

Participants

We conducted one-to-one semistructured interviews with 20 Australian general practitioners (GPs; 8 via telephone and 12 face-to-face) and 15 patients (all face-to-face) in South East Queensland (Australia) general practices between July and December 2017. We recruited the participants using purposive sampling to ensure a diverse range of years of experience and age. Recruitment was done mainly through snowballing via colleagues, organizational contacts, and via initial participants. Participation in the interviews was voluntary and written informed consent was obtained from each participant before the interview. GPs were interviewed in their consultation rooms or over the phone. Patients were interviewed in the waiting area of the clinic if privacy was ensured.

Procedure

We chose semistructured interviews as they allow for flexibility to explore a new subject yet are structured enough to achieve the study aim. The interview questions were developed by the authors, piloted with three academic GPs, and revised before the study. The questions were designed to explore participants' attitudes toward smartphone health apps, their thoughts on the possibility of prescribing health apps, and perceived potential barriers and facilitators to prescribing apps in general practice consultations (Figure 1). Toward the end of the interview, participants were shown (or in the case of phone interviews, apps were named) 9 examples of popular Health & Fitness and Medical apps from the major app stores and 9 examples of tested and effected apps identified through our earlier study on the evidence supporting health and medical apps in order to gauge their general familiarity with mHealth apps [15] (Figure 2). No financial or other incentives were offered to participants.

Figure 1. Guide map for the interview.

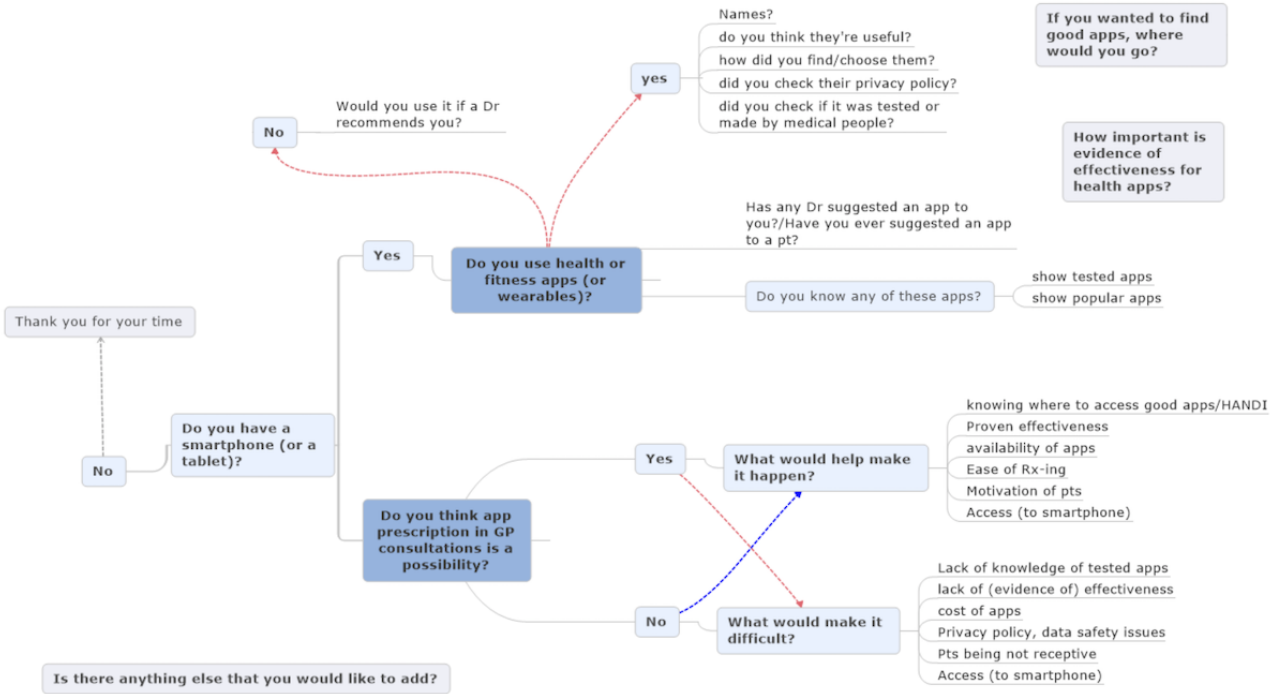


Figure 2. Example apps are shown to interviewees to identify their familiarity with evidence-based vs. popular mHealth apps.

Data Analysis

We planned to interview 15-20 GPs and a similar number of patients. Data saturation was fully achieved, ie, no new content emerged after 3 consecutive interviews in both cohorts by the time we interviewed 20 GPs and 15 patients, therefore validating that our sample size was sufficient. All interviews were conducted, audio-recorded, and transcribed verbatim by the first author (OB).

We employed a thematic analysis methodology described by Braun and Clarke [16]. The six phases of analysis were familiarization with the data, coding, generating initial themes, reviewing themes, defining and naming themes, and report writing. Two researchers (OB, RS) familiarized themselves with the interview transcripts and independently coded them. Partially inductive thematic analysis was carried out, and generated themes were reviewed in consultation with the third author (TH). The results are written up by the first author (OB).

We used the Consolidated Criteria for Reporting Qualitative Research checklist to report the details of our study [17], which is provided as Multimedia Appendix 1. Ethics approval was obtained from the Bond University Human Research Ethics Committee (16016).

Results

We interviewed 20 general practitioners and 15 patients, the demographics of whom are shown in Table 1. Ten additional patients declined to be interviewed, citing that they have never used mHealth apps or do not know anything about them. None of the providers we approached declined to be interviewed. Interviews averaged about four minutes for patients and twelve minutes for GPs.

Table 1. Participant characteristics.

Group	Total, n (%)
General practitioners (n=20)	
Years in practice	
<10	6 (30)
10-20	6 (30)
>20	8 (40)
Gender (female)	12 (60)
Patients (n=15)	
Age (years)	
18-35	5 (33.3)
36-50	5 (33.3)
Over 50	5 (33.3)
Gender (female)	10 (66.6)

General Practitioners' Perspectives

The majority of the GPs reported using health apps personally, professionally, and recommending health apps to patients. However, most GPs suggest apps mostly in generic terms in areas such as mindfulness and period-tracking without naming specific apps and let patients make the final choice because they do not know any specific vetted and safe apps.

The GPs were more familiar with popular apps such as MyFitnessPal, Headspace, and Smiling Mind than evidence-based apps. None of the GPs was familiar with five of the evidence-based apps (SuperBetter, MyMealMate, AccupedoPro, Vegethon, Tat). Sources for finding health apps to recommend varied from trusted websites and medical publications to word-of-mouth and app stores.

About half of the participants were unfamiliar with the Handbook of Non-Drug Interventions (HANDI) web resource from the RACGP [18], but upon being introduced to it, they all agreed that HANDI would be a trustworthy resource for evidence-based apps, consistent with those GPs who were already familiar with HANDI.

Barriers

We identified four themes around barriers to app prescription in general practice. The most prominent barrier cited was a generational difference in the overall digital propensity for both doctors and their patients.

> *Most patients who would benefit from them are elderly and they don't do apps. They don't do smartphones. That's the number one barrier. And most of our patients who are over 70 will be in that category. Most I would say.* [GP7]

> *Well, it's probably my age group more than anything. it's just I'm not as familiar with and as happy around technology as the newer doctors… I do feel it's important to try to learn it because it's part of the future… so it's that technology I think is the main barrier as far as being used to it. I think we need to but it's just hard to do. Can we get away with it? I guess that's what's happening to a lot of older doctors is seeing if we really need to do it or not.* [GP5]

GPs with more than 20 years of experience appeared more likely to find mHealth apps a "gimmick" and less likely to consider using them in their practice.

> *I think it would be [possible to prescribe apps]. The trouble is I think they're a bit of a gimmick. I mean we've always had accessories to health. When I was a young doctor people would bring in their calendars which showed when their next periods are. Now they pull out their phones… you find the phone is slower than the calendar because the calendar… is there and you'd be able to see it visually, whereas a phone, they're pulling it out… trying to find it miss it… then the reception is not good and it takes ages for it to download. So I don't really find them a step forward. They probably are for the person using them at home. But in the consultation, they're often not a step forward I find.* [GP9]

GPs who primarily work with elderly patients would not consider using apps as potential interventions. The "cut-off age" for recommending apps to patients was around 40-60 years.

> *For somebody like me, there will be obstacles because I don't really use apps. So, if you're not comfortable with that sort of things yourself, you have to overcome that to recommend it... Even when I've learned about them in an education session, I always forget what they are, because I don't use them… In fact, the only times I ever recommend apps at all are for young people… they don't have to be young-young, even in their 30s, a young professional… But I never recommend to people older than that. So, there is obviously a generational issue.* [GP3]

However, exceptions to the age-based generalization were commonly mentioned, making individual digital propensity a more influential factor than age in deciding to prescribe apps.

> *I got some elderly patients that don't use an iPhone too often or an app. But there's a lot of savvy oldies there too. It probably depends on the patient, their interests.* [GP11]

GPs also recognize that age is a transient barrier as the younger cohort of patients will age and become patients with chronic diseases.

> *It is a very good idea and something that can be very useful. But I never know which ones are good to use and which population would be good to use them. I'm sure it will come especially as right now the younger population are the ones who are really into the smartphones and they'll get older and have chronic medical conditions in the next 20 years I'm sure it will there's going to be a big space for these apps.* [GP16]

Another barrier was a lack of knowledge of prescribable apps and a lack of trustworthy sources to access them.

> *Probably the only other barrier is knowing which apps. And keeping on top of all the apps they become available, how much they cost and all those kinds of things.* [GP20]

> *I do. I think it's something I'm cautious of simply because if it's not something that I know a lot about then I'm a bit more concerned, you know, I don't really wanna recommend something I don't know the full workings of, especially if I'm gonna ask them to buy. There's so much of me asking them to go on medication until I'm confident that money is money worth spent and the benefit outweighs the cost, then I'm not really willing to do that.* [GP2]

> *I think for education it's really valuable. Ones that I don't use enough of and they all want more information. Otherwise, they'll just Google some unreliable search, and so if I've got a good place to go to that's evidence-based, that's good.* [GP18]

Another barrier was the time commitment required to learn about apps and integrating them into consultations as well as the time commitment required of the patients to use the recommended app. Before they were willing to expend time to adopt apps into their clinical practice, GPs described needing to be convinced of the benefits of using apps.

> *And it's time-consuming to learn about these things. It's hard enough just keeping up with what medicine is doing without this app and that app etc.* [GP1]

> *You're so busy doing in medicine there's not a lot of time to go out there and research what might need to be done to create an app or even the ones that are out there. We're so rushed for time. You're competing with lots of other demands for your time and energy as well. So that's a big limiting factor.* [GP8]

> *The patients' motivation would be a big thing. And the time involved in using it would be another big thing for someone who might be busy, for example. I don't think a lot of people have a lot of time to invest in this type of thing. So, I think your time, availability, and the motivation behind a patient.* [GP15]

> *I think for me, it's just when you are consulting, and you're busy. To modify the practice of what I'm doing, I have to have a pretty good reason for doing it.* [GP6]

Another barrier identified was the privacy, safety, and trustworthiness of health apps. The GPs perceived issues around privacy and data safety of health apps as the ultimate responsibility of the patients, since their complete and ongoing safety cannot be guaranteed. Some GPs were aware of how the industry attempts to influence health care.

> *I personally manage my privacy very, very, very carefully. But I think I leave that to the individual patient... and I think that if they're going to be using apps, my perception is that they already have made their own personal decisions upon privacy.* [GP4]

> *I think that's every day now. I mean the number of times you get on Google and then they already go okay well you've got this many children, and I know that already… that's the world we live in. I don't think we're going to stop that by not using an app when we're on the internet all the time. So, if it's demographic information and I think that's being collected by a lot of people, not just an app, I suppose.* [GP20]

> *Also, I guess I'm also very wary of who's paying for it. So, I guess my general approach to most things is to think if you're going to pay for it, then hopefully you are bearing the load of what if it's worth it. If someone else is paying then there's some hidden agenda there, whether it's a drug company or someone else who's gonna benefit of you having their app on your phone and again that comes down to my reticence to recommend something I really don't know who's designed it and what's the purpose of it and who's benefiting the most.* [GP2]

Facilitators

We identified two main themes around facilitators of mHealth app prescription in general practice. One was a trustworthy source for prescribable apps and information for GPs. The RACGP was the most preferred trustworthy source to vet, endorse, and provide prescribable apps as the majority of the GPs were RACGP members.

> *Maybe some sort of database of trusted apps that would be recommended for certain conditions or treatment strategies. Having a nice little summary of things that could be used, because of the sheer number of what's on the market, it's hard to really make it part of my regular day-to-day routine... Whereas if some organization was to put together a list of you know tried and tested and reliable apps then it would be much easier to say "look, you're young you'll you've got the time and the patience for it, let's try an app for this problem, and this is the one we trust.* [GP15]

> *I think we probably need to have them reliably approved and researched by our college. I probably wouldn't be happy to recommend any without the endorsement of the college.* [GP11]

> *We need more education on which ones to prescribe and which ones not. We have the NPS which helps us with prescribing medications. So, if there were an organization/body involved with educating GPs on which apps are good and useful and provide the right information and are easy to use, that would be really helpful.* [GP16]

GPs recognized that near-universal ownership of smartphones, the ubiquity of apps, and younger tech-savvy patients are enablers of app prescription, as this facilitates information accessibility and can sometimes provide alternatives that are more convenient and lower in cost.

> *I work in an area where there's a lot of young people, and most of these people are … generally pretty switched on. and … I can't imagine a situation where I'm not being able to recommend an app so somebody. This hasn't happened yet.* [GP12]

> *Apps are quite a neat way of showing somebody all that information without having a book to give them. It's often a very low-cost solution if your alternatives are costly. And it's very accessible... you don't need to wait for in-hours care, you can be 10 p.m. at night and do some of the work, whether it's treatment or information and knowledge that you're sharing that can all be done at a time that suits a patient. so that's the kind of value of apps I suppose over other resources.* [GP20]

Patient Perspective

Two-thirds of the patients have used or use mHealth apps personally. The most commonly used app types were fitness, wellness, and women's health-related apps (period, ovulation, and pregnancy tracking). Those who did not use apps said the reason was they do not have any need or significant reason to

do so. People often chose apps through the recommendations of friends or family, directly from app stores, or through subscribed services such as the local gym. From the list of popular and evidence-based apps, they recognized some of the most popular apps, such as MyFitnessPal and 5K runner, but none of the evidence-based apps.

Barriers

From the patients' perspective, a perceived barrier to using mHealth apps in general practice consultations was older patients.

> *I don't think there'd be any problem. But if we're talking about the elderly, they're not really very computer-savvy. So, they might find it difficult. I have a number of friends, even older than me; they don't wanna use smartphones. It's too much trouble. They just wanna make a phone call and get a text. That's it.* [Pt8]

> *I guess only with the older generation not having smartphones. They would not use apps. So that would really be the only problem. Yeah, I wouldn't see any other [issue].* [P2]

Another barrier was poor usability of apps.

> *I think the ease-of-use has to be paramount. Ease of use is gotta be the big issue. For me, that was the problem. I can't speak for other people. I mean I'm pretty good with technology, but I just found it very tedious…And I just found it was annoying to put data in. That was the issue with it. So, when it becomes difficult, I didn't do it.* [Pt12]

Data safety and privacy issues related to health app use were not a barrier described by the patients interviewed. Patients were more concerned about the loss of financial information than health-related information.

> *I don't worry about it too much, privacy stuff, cause I don't have much to hide. Maybe if someone got on to your phone, they could see your personal information, they usually have like a passcode for private apps anyway.* [P13]

> *Me personally, no. data safety… it is what it is. I think there are measures in place to keep it safe. Other people probably don't have that opinion. I think honestly it is safe enough.* [P15]

> *I'm more worried about financial stuff than health that would create a financial loss. if somebody finds out what my blood pressure is, what's the big deal, right?* [Pt8]

> *I don't worry about it. well, I'm not really putting anything into an app that is that [important]…I mean I don't know who else in the world cares the day my period came or how big the baby is, I haven't really put in banking details or anything. I haven't used any paid apps.* [Pt6]

Facilitators

From the patients' perspective, the ubiquity of smartphones and apps and young patients were perceived as facilitators of app prescription.

> *I imagine younger people wouldn't have any problems.* [Pt11]

> *I think it's quite a common thing now. Everyone has smartphones nowadays. So, it's an easy way to access it.* [Pt1]

Most patients expressed that they think mHealth apps can be beneficial, especially when recommended by their doctor and that app prescription by GPs is possible, helpful, and welcome because it would eliminate the challenge of finding health apps for themselves.

> *Great idea if they're prescribed from a doctor. I don't think getting them off the net without advice would be a good thing. I think it needs to be advised or prescribed by someone who knows, and then from there, it can only be a good thing, yeah.* [Pt15]

> *Yeah. If my doctor recommended one, I'd probably go with that rather than trying to go from the recommendations on iPhone, you know, the stars, from the app stores. I mean, they are helpful too what other users have found, cause a lot of apps crash and have problems if they're not maintained and upgraded, but yeah, if a doctor was recommending one, I'd probably use it.* [Pt9]

> *If there are multiple people being like, "this is really good. This has helped me with this…" then I'll actually go have a look. If I like it for more than a week, then I'll just continue using it. I went through a lot [of apps] before finding the one that wasn't an effort to use, one that was just easy. I went through probably ten to find one I actually liked, which is a bit annoying. It kind of turns you off that. So, if there was one or two that all doctors recommended, then people would probably more likely to use them.* [Pt1]

A theme identified from both the GPs and patient perspectives that could not be categorized as a barrier or facilitator was evidence of app effectiveness. GPs expressed they would not want to prescribe apps not supported by evidence, yet they also feel that simple apps do not need high levels of evidence.

> *…as it applies to anything in medicine, I think it would need a reasonable degree of efficacy to run with it. You know, you can't just have an app for the sake of having an app.* [GP1]

> *… [evidence is] pretty important to officially recommend. Some of it is common sense. Like something to log your blood pressure and mood diary is self-explanatory and makes sense. But for some more complex health apps, you sort of wanna know if there's good evidence so that it's reasonably well made. Especially if you're going to spend money on it.* [GP14]

Most patients viewed the evidence of app effectiveness to be important in the same way as the standards required of pharmaceutical interventions. Some were not concerned about the effectiveness and preferred personal evaluations and the freedom to make the ultimate decision.

> *Very important. It's like anything in health; it's like medication, if a doctor is gonna recommend something, they have to know it works. Cause it might not be suitable for somebody, and if that person is to use something that's not suitable, then that could have a bad effect instead of a good one.* [Pt13]

> *Definitely, yea. I would want one that has references, the app that I use have medical references they've sourced the information. Otherwise, anyone could be sitting home and writing an app.* [Pt6]

> *I'm not too worried about that. I just get on and try it, if it works for me then that's great. If it doesn't, then I'll just delete it if it's not gonna do what I needed it to do for me… I mean, as long as it's our choice to use them or not. I mean up to us, they can make a recommendation, but if we find it's not suitable or it doesn't work, and if it does then great if it makes our life easier, it's a fast and busy world, so if they think it can help, it's great.* [Pt10]

Discussion

In this study of the barriers to and facilitators of mHealth app prescription in Australian general practice, all patients and GPs recognized that mHealth apps could be beneficial, and app prescription is achievable. From the GPs' perspective, uptake is hindered by barriers around a generational difference in the digital propensity for both GPs and patients, lack of knowledge of prescribable apps and lack of trustworthy source to access them, time commitments required of the GPs and patients, and privacy, safety, and trustworthiness of health apps.

Both patients and GPs cited the old age of patients as a barrier to app prescription, although also offered examples of exceptions to this age-related division of digital propensity. Annual mobile consumer surveys showing that older age groups have seen the highest increase in smartphone ownership in Australia [19]. Doctors and patients also believed that the ubiquity of smartphones and apps, and young patients are facilitating factors as Australians approach "peak smartphone"—the peak level of usage before the vast majority of consumers to begin actively limiting their phone use [20]. Almost all the interviewed GPs reported using apps personally and professionally; however, they do not recommend specific apps to patients. Instead, they remind patients of the availability of mHealth apps in the general area of focus (eg, mindfulness) should they wish to use them.

The evidence base (weak or strong) for mHealth apps emerged as an important theme as a barrier or facilitator. Patients viewed doctors as a trustworthy source of health apps, and the GPs acknowledged that they needed a trustworthy source for prescribable apps as they have neither the knowledge nor time to find such apps themselves. A national-level professional organization such as the RACGP is well-placed to address this barrier. One of the many resources they provide for GPs is HANDI—a database of effective nondrug interventions, which includes several mHealth apps [18]. However, half of the GPs interviewed in this study were not familiar with or commonly used HANDI, but all agreed that RACGP is the most trusted source for them to access professional and practice-related information and guidance. The majority of participants recognized a few of the most popular apps from app stores and fewer apps from the evidence-based apps list, emphasizing the need for information dissemination about evidence-based apps.

Several comparable studies have explored barriers and facilitators to novel technology adoption in medical practice. Many studies report a lack of education and training as one of the most common barriers that face health care providers in adopting new technology in health care [13,21,22], and our study echoes this. Furthermore, the potential to increase doctors' time strain and workload are universally common factors of poor uptake of new health technologies [23-25]. Brandt et al found that the experiences of GPs with eHealth support for lifestyle changes were an essential factor in recommending it for their patients [26]. Our findings also emphasize the digital propensity of the health care providers and patients would make a big difference in uptake of mHealth apps. Building on this finding and educating and supporting GPs so they understand the value of new technology such as the potential to save consultation time and keep patients connected and motivated between consultations can help mitigate against these barriers and help them recommend apps to their patients with confidence.

Recent research suggests that individuals with poor self-reported health and low rates of physical activity were the least likely to report downloading and using these health tools [27]; however, patients adhere better to prescribed apps [2] beyond the typical one week of usage [4,5]. Thus, the official prescription of apps by trusted medical practitioners might help increase the uptake of effective health apps among such patients who would benefit the most. Future studies should measure the real-world adherence of the patients following app prescription by health professionals.

The present qualitative study appears to be the first of its kind to explore the perspectives of GPs and patients regarding mHealth app prescription in Australian general practice. The barriers identified in this study were added to the mHealth section of the RACGP Annual Technology survey in 2017 to explore further how they would rank among a national sample of GPs [28]. The question about barriers to app prescription gathered over six hundred responses and the top barriers reflected the central theme identified in this study, which was lack of knowledge of effective apps and lack of trustworthy source to access them, further validating the findings of this study.

Limitations of this study include a small sample size that skewed towards relatively healthy patients from high socioeconomic areas and GPs from the metropolitan area. Although we attempted to mitigate this limitation by purposively sampling participants from a variety of age and work experience, future studies should target patients with long-term medical conditions, those from rural and remote areas, and low socioeconomic areas.

Other limitations include lack of triangulation of the results, member checking, a reflection of possible interviewer bias about the potential of apps, and not piloting the interview questions with patients.

Conclusions

mHealth app prescription appears to be feasible in general practice. The barriers and facilitators we identified from both GPs and patients widely overlapped. The involvement of all stakeholders of consumer mobile technology, medical professionals, and patients is vital in the successful integration of mHealth apps with clinical practice. The findings of this study will inform the development of effective interventions to overcome the identified barriers and help the adoption of health apps to general practice to patients' benefit.

Acknowledgments

We would like to thank Rebecca Sims for the qualitative analysis assistance. This research was supported by the Australian Government Research Training Program scholarship.

Conflicts of Interest

PG and TH are members of the HANDI project committee at the RACGP.

Multimedia Appendix 1
COREQ checklist.
[PDF File (Adobe PDF File), 150 KB - mhealth_v8i7e17447_app1.pdf]

References

1. GSMA Intelligence. The Mobile Economy: Asia pacific. The Mobile Economy: Asia Pacific. 2019. URL: https://www.gsma.com/r/mobileeconomy/asiapacific/ [accessed 2018-01-01]
2. Aitken M, Lyle J. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance, IMS Institute for Healthcare Informatics. 2015. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf?la=en&hash=B3ACFA8ADDB143F29EAC0C33D533BC5D7AABD689 [accessed 2015-10-01]
3. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101-e103 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
4. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
5. 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. Ann Intern Med 2014 Nov 18;161(10_Supplement):S5. [doi: 10.7326/m13-3005]
6. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Aug 24;30(4):439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
7. El Amrani L, Oude Engberink A, Ninot G, Hayot M, Carbonnel F. Connected Health Devices for Health Care in French General Medicine Practice: Cross-Sectional Study. JMIR Mhealth Uhealth 2017 Dec 21;5(12):e193 [FREE Full text] [doi: 10.2196/mhealth.7427] [Medline: 29269336]
8. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: A mixed method study. Health Policy and Technology 2017 Sep;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
9. 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]
10. National IFHE. Evidence Standards Framework for Digital Health Technologies. Evidence Standards Framework for Digital Health Technologies. 2019. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2019-04-01]
11. American MA. Integration of Mobile Health Applications and Devices into Practice. Integration of Mobile Health Applications and Devices into Practice. 2016. URL: https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/about-ama/councils/Council%20Reports/council-on-medical-service/interim-2016-council-on-medical-service-report-6.pdf [accessed 2017-01-01]
12. American PA. App Evaluation Model. App Evaluation Model. 2017. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2018-01-01]
13. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan 15;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

14. The RACOGP. Views and attitudes towards technological innovation in general practice Survey report 2016. Views and attitudes towards technological innovation in general practice: Survey report 2016. Melbourne, Vic: RACGP; 2017. URL: https://www.racgp.org.au/FSDEDEV/media/documents/Running%20a%20practice/Technology/Member%20engagement/RACGP-Technology-Survey-2016-Results.pdf [accessed 2017-01-01]
15. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018 May 9;1(1):12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
16. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
17. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
18. The Royal Australian College of General Practitioners. Handbook of Non-Drug Interventions. Handbook of Non-Drug Interventions. URL: https://www.racgp.org.au/your-practice/guidelines/handi/about/the-handi-project/ [accessed 2018-01-01]
19. Deloitte TTL. Mobile Consumer Survey 2017: Smart everything, everywhere. Mobile Consumer Survey 2017: Smart everything, everywhere. URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/deloitte-au-tmt-mobile-consumer-survey-2017-211117-final.pdf [accessed 2018-01-01]
20. Deloitte TTL. Mobile Consumer Survey 2018 : Behaviour unlimited. Mobile Consumer Survey 2018 : Behaviour unlimited. URL: http://images.content.deloitte.com.au/Web/DELOITTEAUSTRALIA/%7B51598b8c-1c72-4ceb-8272-3a052f322992%7D_20181103-tel-inbound-mobile-consumer-survey-2018-report-2.pdf?elqTrackId=34b48a629a3e44ec8407b5f742142f4e&elqaid=790&elqat=2 [accessed 2018-01-01]
21. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and Facilitators to eHealth Use in Daily Practice: Perspectives of Patients and Professionals in Dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
22. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving Implementation of eMental Health for Mood Disorders in Routine Practice: Systematic Review of Barriers and Facilitating Factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
23. Ayre J, Bonner C, Bramwell S, McClelland S, Jayaballa R, Maberly G, et al. Factors for Supporting Primary Care Physician Engagement With Patient Apps for Type 2 Diabetes Self-Management That Link to Primary Care: Interview Study. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11885 [FREE Full text] [doi: 10.2196/11885] [Medline: 30664468]
24. Thies K, Anderson D, Cramer B. Lack of Adoption of a Mobile App to Support Patient Self-Management of Diabetes and Hypertension in a Federally Qualified Health Center: Interview Analysis of Staff and Patients in a Failed Randomized Trial. JMIR Hum Factors 2017 Oct 03;4(4):e24 [FREE Full text] [doi: 10.2196/humanfactors.7709] [Medline: 28974481]
25. Thomson K, Brouwers C, Damman OC, de Bruijne MC, Timmermans DR, Melles M. How Health Care Professionals Evaluate a Digital Intervention to Improve Medication Adherence: Qualitative Exploratory Study. JMIR Hum Factors 2018 Feb 20;5(1):e7 [FREE Full text] [doi: 10.2196/humanfactors.8948] [Medline: 29463494]
26. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General Practitioners' Perspective on eHealth and Lifestyle Change: Qualitative Interview Study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
27. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health App Use Among US Mobile Phone Users: Analysis of Trends by Chronic Disease Status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
28. Byambasuren O, Beller E, Glasziou P. Current Knowledge and Adoption of Mobile Health Apps Among Australian General Practitioners: Survey Study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]

Abbreviations

GP: general practitioner
HANDI: Handbook of Non-Drug Interventions
mHealth: mobile health
RACGP: Royal Australian College of General Practitioners

Edited by G Eysenbach; submitted 12.12.19; peer-reviewed by E Sadeghi-Demneh, XF Zhong, A Vallée, K Drixler, M Semwal, M Nomali, T Muto, C Jacob; comments to author 06.03.20; revised version received 31.05.20; accepted 03.06.20; published 30.07.20.

Please cite as:
Byambasuren O, Beller E, Hoffmann T, Glasziou P
Barriers to and Facilitators of the Prescription of mHealth Apps in Australian General Practice: Qualitative Study
JMIR Mhealth Uhealth 2020;8(7):e17447
URL: http://mhealth.jmir.org/2020/7/e17447/
doi:10.2196/17447
PMID:32729839

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org