JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 9 (2021), Issue 12 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Reviews

Original Papers

Review

Tools for Evaluating the Content, Efficacy, and Usability of Mobile Health Apps According to the Consensus-Based Standards for the Selection of Health Measurement Instruments: Systematic Review

Antonio Muro-Culebras[1,2*], PhD; Adrian Escriche-Escuder[1,2*], MSc; Jaime Martin-Martin[1,2], PhD; Cristina Roldán-Jiménez[1,2], PhD; Irene De-Torres[3], PhD; Maria Ruiz-Muñoz[1,2], PhD; Manuel Gonzalez-Sanchez[1,2], PhD; Fermin Mayoral-Cleries[4], PhD; Attila Biró[5], MSc; Wen Tang[6], PhD; Borjanka Nikolova[7], MSc; Alfredo Salvatore[8], MSc; Antonio Ignacio Cuesta-Vargas[1,2,9], PhD

[1]Grupo Clinimetría (F-14), University of Málaga, Málaga, Spain
[2]Instituto de Investigación Biomédica de Málaga (IBIMA), Malaga, Spain
[3]Physical Medicine and Rehabilitation Unit, Regional University Hospital of Malaga, Málaga, Spain
[4]Mental Health Unit, Regional University Hospital of Malaga, Málaga, Spain
[5]ITWare, Budapest, Hungary
[6]Faculty of Science and Technology, Bournemouth University, Bournemouth, United Kingdom
[7]Arthaus, Skopje, Republic of North Macedonia
[8]SensorID Snc, Boiano, Italy
[9]School of Clinical Science, Faculty of Health Science, Queensland University Technology, Queensland, Australia
[*]these authors contributed equally

Corresponding Author:
Antonio Ignacio Cuesta-Vargas, PhD
Grupo Clinimetría (F-14), University of Málaga
C/ Arquitecto Francisco Peñalosa 3
Málaga, 29071
Spain
Phone: 34 951952852
Email: acuesta@uma.es

Abstract

Background: There are several mobile health (mHealth) apps in mobile app stores. These apps enter the business-to-customer market with limited controls. Both, apps that users use autonomously and those designed to be recommended by practitioners require an end-user validation to minimize the risk of using apps that are ineffective or harmful. Prior studies have reviewed the most relevant aspects in a tool designed for assessing mHealth app quality, and different options have been developed for this purpose. However, the psychometric properties of the mHealth quality measurement tools, that is, the validity and reliability of the tools for their purpose, also need to be studied. The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative has developed tools for selecting the most suitable measurement instrument for health outcomes, and one of the main fields of study was their psychometric properties.

Objective: This study aims to address and psychometrically analyze, following the COSMIN guideline, the quality of the tools that are used to measure the quality of mHealth apps.

Methods: From February 1, 2019, to December 31, 2019, 2 reviewers searched PubMed and Embase databases, identifying mHealth app quality measurement tools and all the validation studies associated with each of them. For inclusion, the studies had to be meant to validate a tool designed to assess mHealth apps. Studies that used these tools for the assessment of mHealth apps but did not include any psychometric validation were excluded. The measurement tools were analyzed according to the 10 psychometric properties described in the COSMIN guideline. The dimensions and items analyzed in each tool were also analyzed.

Results: The initial search showed 3372 articles. Only 10 finally met the inclusion criteria and were chosen for analysis in this review, analyzing 8 measurement tools. Of these tools, 4 validated ≥5 psychometric properties defined in the COSMIN guideline.

Although some of the tools only measure the usability dimension, other tools provide information such as engagement, esthetics, or functionality. Furthermore, 2 measurement tools, Mobile App Rating Scale and mHealth Apps Usability Questionnaire, have a user version, as well as a professional version.

Conclusions: The Health Information Technology Usability Evaluation Scale and the *Measurement Scales for Perceived Usefulness and Perceived Ease of Use* were the most validated tools, but they were very focused on usability. The Mobile App Rating Scale showed a moderate number of validated psychometric properties, measures a significant number of quality dimensions, and has been validated in a large number of mHealth apps, and its use is widespread. It is suggested that the continuation of the validation of this tool in other psychometric properties could provide an appropriate option for evaluating the quality of mHealth apps.

(JMIR Mhealth Uhealth 2021;9(12):e15433) doi:10.2196/15433

KEYWORDS

mobile health; mHealth; eHealth; mobile apps; assessment; rating; smartphone; questionnaire design; mobile phone

Introduction

Background

Nowadays, in the age of digital content, people, regardless of age group, have access to mobile and smart devices (eg, phones, tablets, and smart televisions), or special devices with the possibility of internet connection. In dedicated (iOS and Android) app stores (Apple App Store and Google Play Store), there are thousands of apps with a vast number of functions, and this number is increasing every day. According to a new report by Grand View Research, Inc [1], the mobile health (mHealth) app market size is expected to reach US $149.3 billion by 2028 and is expected to register a compound annual growth rate of 17.7% over the forecast period. In these app catalogs, mHealth apps are a very important field, and there has been a growing interest from users in the last few decades. Some studies report that up to 34% of mobile phone owners have at least one health app installed on their device [2]. The World Health Organization [3] described the term mHealth as *the use of mobile wireless technologies for health*, being a subset of eHealth, which is described as *the use of information and communications technology in support of health and health-related fields*. The World Health Organization [3] also highlighted the relevance of digital health interventions to address health needs, remarking that they should always be used as an aid and an improvement for health systems, not as a substitute.

The variety of features in apps available on different platforms or cross-platforms is wide. Therefore, many of the mHealth apps are (1) simply a catalog of recommendations; some of them work as a (2) follow-up tool, complementing an intervention program, whereas other mHealth apps are (3) connected to dedicated sensors to offer information about health signals or health status.

Most of these mHealth apps enter the market with limited filters or controls that usually do not consider aspects such as the veracity of its content and their effectiveness as relevant [4]. According to a previous study, only a small percentage of available apps in some health fields have referred to medical professional involvement in their development or content [5]. Both, apps that users use autonomously and those that can be directly recommended by clinicians to their patients require a prior study that investigates their evidence to minimize the risk of using the apps that do not work or that may even cause harm [6,7]. The development of these studies and analyses have been described in academic contexts, but its execution is not always easy in commercial apps [5]. Some experts attribute this fact to a much slower pace of academic research than that of app development, which can result in long delays in the diffusion of apps in commercial markets and among users [8]. Thus, many of the apps are only rated by the general subjective perception of users with vague rating tools, such as numerical or star-based scores from 1 to 5.

Many attempts have been made to develop effective and practical validation tools to measure the quality of mHealth apps. The quality-based concept has been interpreted in different ways according to each author and field, evaluating or resulting in different components [9]. Some of the first attempts used existing generic tools, such as the System Usability Scale (SUS) [10], to measure the usability, that is, the ease of use, of systems. This tool was developed in 1986 to allow a quick and basic measurement of the usability of any system and is still used in many studies despite being 30 years old. The increase in the use of new technologies, such as smartphones and health apps, brought the need to develop new types of mHealth apps with specific measurement tools that adapt the existing ones.

Previous studies have reviewed which methods have been used to assess the quality of mHealth apps [11,12], determining which aspects could be the most relevant in a tool designed for this purpose [2]. In general, quality evaluation methodologies can be divided into 2 categories: (1) methodologies based on the downloaded app content (using a predefined list of requirements that the app should contain, assessing the inclusion of evidence-based content, and assessing the usability of predefined app functions) and (2) methodologies that are content-independent and the app does not need to be downloaded (using app market or website assessment tools, users' reviews and ratings, and other assessment methods such as the analysis of the app description or a medical professional involvement in the app) [11]. Similarly, a previous review suggested that the essential contents to be evaluated can be grouped into 4 categories: (1) content analysis (coding and qualitatively evaluating the app content); (2) usability testing (evaluating whether the app works correctly and its ease of use); (3) observational studies (that can be used to assess app use and

satisfaction and to predict its usefulness in certain contexts); and (4) efficacy testing (assessing whether the app achieves meaningful effects in previously determined outcomes) [2]. Other areas of interest in this quality evaluation may include exploring the technical functions of the app, the management of security and privacy of user data, and how the developer will use these or developer transparency [2]. Because the measurement tools do not measure all the dimensions and properties of an mHealth app, a prior analysis is necessary to select the appropriate tool for each purpose.

In addition to the inclusion of the previously proposed aspects, it is important to determine the validity of these instruments, that is, the ability to properly assess what they intend to assess. Therefore, it is relevant for any measurement tool to study its psychometric properties. Psychometric properties are different concepts related to the validity and reliability of the instruments, each helping us to determine whether a tool adequately does what it was designed to do according to essential aspects. The development of psychometrics has provided the possibility of knowing the existence of individual differences in the use of measurement tools and their quantification [13,14]. In 2005, the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) started to develop practical tools for selecting the most suitable measurement instrument in research and clinical practice to improve the selection of outcome measurement instruments for health outcomes. One of the main fields of study was the psychometric properties of the assessment tools. Psychometric analysis, using a wide variety of terminology, has been a source of controversy and confusion for decades. The progress of the COSMIN initiative has improved this aspect through the development of the COSMIN Taxonomy of Measurement Properties. This tool aims to standardize the psychometric criteria necessary to validate patient-reported outcome measures [15,16]. These guidelines were not specifically designed for digital health. However, it is necessary to bring the tools of psychometrics closer to this field to evaluate the measurement instruments used to choose mHealth apps in a clinical context. Therefore, it is essential to integrate this analysis in the field of digital health in their original format and possibly later in a more specific version adapted for this field. In this context, a wide analysis of mHealth assessment tools, following this guide, seems to be an appropriate method for assessing their quality and suitability, bringing scientific consensus closer to this field of health.

Objective

The purpose of this systematic review is to address and psychometrically analyze, following the COSMIN guideline, the quality of the tools that are currently used to measure the quality of mHealth apps.

Methods

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [17].

Data Sources and Searches

PubMed and Embase databases were processed by 2 reviewers (AE-E and AM-C) from February 2019 to December 31, 2019. The following search term combinations were used in PubMed: (*mHealth* OR *mobileapp** OR *healthtechnology*) AND (*scale* OR *checklist* OR *score*) AND *app*. The search was extended to all fields. Embase was searched using the following search string: (*health*/*exp* OR *health*) AND *app* AND (*quality*/*exp* OR *quality*). Complementary searches were performed on the reference lists of the reviews and included articles. The outcome selection process was as follows: when a tool was identified, the instrument was specifically searched in the database search engines, to find all the validation studies associated with this tool.

For inclusion, the studies had to be meant to validate a tool (scale, score, index, or questionnaire) designed to evaluate the quality of systems and used to assess mHealth apps.

Studies that included the application of these tools for the evaluation of mHealth apps but that did not include any type of psychometric validation were excluded. In addition, studies that contained self-written questionnaires where authors focused on measuring usability or content of apps without any validation of their use were also excluded because of the lack of reliability.

Study Selection

The search results were screened by title and abstract by 2 independent authors (AE-E and AM-C). Whenever the information contained in the title and abstract was insufficient, the full text was examined to decide. Full texts of all potentially eligible studies were independently screened by the same reviewers to identify those that met the abovementioned selection criteria. Disagreements were agreed upon by a third reviewer (AIC-V). Finally, the measurement tools included in the selected studies were identified and retrieved.

Data Extraction

Data from the selected studies were extracted by the same independent reviewers using an extraction form. The extracted information included the original language, cross-cultural adaptations available, number of dimensions, number of items, and the fulfillment of the 10 psychometric characteristics described in COSMIN (internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity, responsiveness, and interpretability). Any discrepancies identified were discussed and resolved by bringing in a third reviewer (AIC-V) whenever a consensus could not be reached.

Quality Assessment: Psychometric Characteristics (COSMIN Analysis)

The psychometric characteristics of each of the retrieved tools were analyzed following the COSMIN guidelines [15] based on the COSMIN Taxonomy of Measurement Properties to assess their methodological quality. The COSMIN guideline also offers updated values to consider psychometric properties as sufficient [15]. The definitions provided by COSMIN for all psychometric characteristics are presented in Table 1. We examined 10 psychometric characteristics. First, the content validity of each

tool was assessed. According to COSMIN, content validity is considered to be the most important measurement property because it evaluates whether an outcome measurement instrument is relevant, comprehensive, and comprehensible with respect to the construct of interest and target population [15]. Second, the internal structure of the outcome measures was evaluated using structural validity, internal consistency, and cross-cultural validity. Third, the remaining measurement properties were evaluated (reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness). Finally, the interpretability and feasibility of each measurement tool were evaluated.

Table 1. Consensus-based Standards for the Selection of Health Measurement Instruments definitions of domains, measurement properties, and aspects of measurement properties [18].

Domain	Measurement property	Aspect of a measurement property	Definition
Reliability	N/A[a]	N/A	The degree to which the measurement is free from measurement error
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: for example, using different sets of items from the same HR-PROs[b] (internal consistency) over time (test-retest), by different persons on the same occasion (interrater), or by the same persons (ie, raters or responders) on different occasions (intrarater)
	Internal consistency	N/A	The degree of the interrelatedness among the items
	Reliability	N/A	The proportion of the total variance in the measurements which is due to *true*[c] differences between patients
	Measurement error	N/A	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Validity			The degree to which an HR-PRO instrument measures the construct it purports to measure
	Content validity		The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
		Face validity	The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of an HR-PRO instrument are consistent with the hypotheses (for instance, with regard to internal relationships, relationships to the scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
		Hypotheses testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
	Criterion validity	N/A	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a *gold standard*
Responsiveness			The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Responsiveness	N/A	Idem responsiveness
Interpretability[d]	N/A	N/A	Interpretability is the degree to which one can assign qualitative meaning, that is, clinical or commonly understood connotations, to an instrument's quantitative scores or change in scores

[a]N/A: not applicable.

[b]HR-PRO: health-related patient-reported outcome.

[c]The word *true* must be seen in the context of the classical test theory, which states that any observation is composed of 2 components: a true score and an error associated with the observation. *True* is the average score that would be obtained if the scale were given an infinite number of times. It refers only to the consistency of the score and not to its accuracy.

[d]Interpretability is not considered a measurement property but is an important characteristic of a measurement instrument.

The information obtained from the validation of the psychometric criteria was used in 2 ways. First, the number of properties described in the COSMIN guideline that were validated was quantified. Second, the meaning of the values obtained for each property was analyzed.

Measured Dimensions

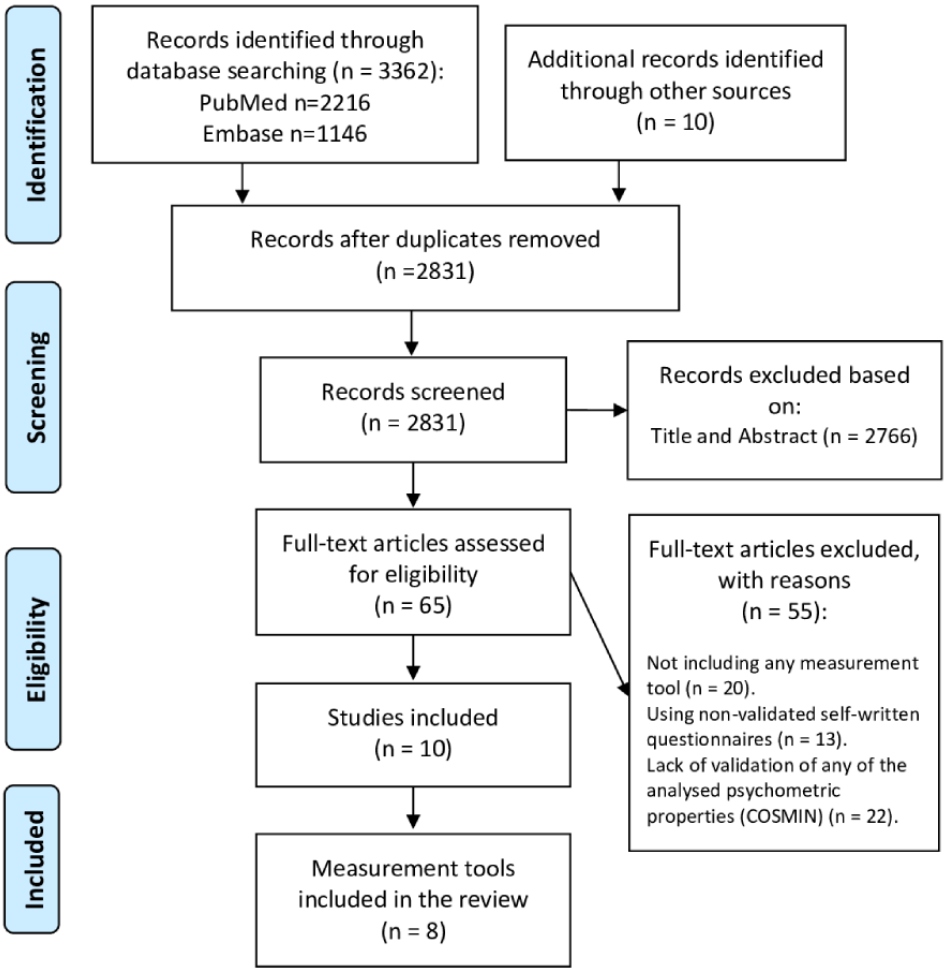
The number and content of the measured dimensions were identified to facilitate the characterization of each measurement tool. In addition, the number of items was retrieved to determine tool length.

Results

Selection of Studies

The literature research identified 3372 articles, of which 2831 remained after deleting duplicates. From these, 65 studies were selected as potentially eligible after reading the title and abstract (full texts of the studies were retrieved in the case of doubt). We excluded 20 studies because they did not include any measurement tool, and 13 studies were excluded because of the use of nonvalidated self-written questionnaires. After analyzing the psychometric characteristics, 22 studies were excluded because of the lack of validation of any of the psychometric properties recommended by the COSMIN guideline. Therefore, only 10 studies were finally included in the review (Figure 1), including the development process or analysis of 8 measurement tools, with some of these different versions of the same original tool. Therefore, 2 tools (Mobile App Rating Scale [MARS] and mHealth Apps Usability Questionnaire [MAUQ]) have different versions for users and professionals. In addition, the MAUQ provides different versions for interactive or stand-alone mHealth apps.

Figure 1. Flowchart of the selection process. COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments.



Measurement Tools

The oldest scale identified was the SUS, which was developed in 1986, whereas the newest was the MAUQ, which was developed in 2016. In total, 8 tools were identified, some of which were different versions of the same original tool.

Usability

As described in the Introduction section, the SUS tool was initially developed to evaluate the usability of engineering and electronic office systems. Nowadays, it is used to evaluate many products and services, such as software, webpages, or mobile apps. It focuses on measuring usability using a Likert scale of 10 elements [10]. It has been adapted to multiple languages,

such as Portuguese, Spanish, French, German, Persian, and Malay, and it is considered a highly reliable tool [19-22].

Similarly, 3 other tools included were specifically designed to assess usability [23-25]. First, the tool *Measurement Scales for Perceived Usefulness and Perceived Ease of Use* was developed in 1989 to measure the usability of computer systems [23]. As its name suggests, this tool incorporates 2 scales to evaluate the perception of 2 aspects of usability: usefulness and ease of use. Second, the Health Information Technology Usability Evaluation Scale (Health-ITUES) questionnaire was one of the first questionnaires to focus specifically on health areas [24]. Initial attempts to develop and value Health-ITUES were conducted using a web-based communication system that supported nurse staffing and scheduling [26]. However, its use in an mHealth app was not validated until a few years ago [27]. Finally, the MAUQ questionnaire exclusively focuses on the usability aspects of apps [25]. This questionnaire provides 4 versions, depending on whether it is used by a health professional or by a patient and whether it is intended to analyze an interactive or stand-alone app (interactive app for patients, interactive app for health care providers, stand-alone app for patients, and stand-alone app for health care providers). However, only the patient's versions have been validated and are therefore included in the analysis of this review, as shown in Table 2.

Table 2. Main characteristics of the mobile health apps quality measurement tools included in this review.

Measurement tool	Validation	Year	Language	Cross-cultural adaptation available	Dimensions of the measurement tool	Number of items
Mobile App Rating Scale	60 mental health apps	2015	English	Italian and Spanish	5: engagement; functionality; esthetics; information quality; and subjective app quality	23
iSYScore index	257 health apps	2016	Spanish	__[a]	3: popularity and interest; trust and quality; and usefulness	14
User version of the Mobile App Rating Scale	2 health apps	2016	English	—	5: engagement; functionality; esthetics; information; and subjective app quality	20
Health information technology usability evaluation scale	100 adults who tested the use of a mobile health app	2015	English	—	4: quality of work life; perceived usefulness; perceived ease of use; and user control	20
Measurement scales for perceived usefulness and perceived ease of use	2 studies: 112 users and 2 systems and 40 users and 2 systems	1998	English	—	2: perceived usefulness and perceived ease of use	12
The mHealth app usability questionnaire for interactive mHealth apps (patient version)	2 health apps	2019	English	—	3: Ease of use and satisfaction; System information arrangement; Usefulness	21
The mHealth app usability questionnaire for stand-alone mHealth apps (patient version)	2 health apps	2019	English	—	3: ease of use; interface and satisfaction; and usefulness	18
System Usability Scale	3 studies: 20 people, 206 studies using System Usability Scale, and 9000 System Usability Scale questionnaires	1986	English	Portuguese, Spanish, French, German, Persian, and Malay	1: usability	10

[a]Not analyzed.

Overall Quality

iSYScore, developed in 2015, was initially designed to measure the reliability and overall quality of mHealth apps, not only their usability [28]. It was validated according to 3 main aspects: popularity and interest, trust and quality, and usefulness [28]. Another important tool was the MARS developed in 2015 [29]. This scale has been adapted to Spanish and Italian languages [30]. It is worth noting that, in addition to its original version, this tool has a specific user version—user version of MARS (uMARS)—developed in 2016 [31]. Both the MARS versions (user and professional versions) focus on measuring other components of the quality of mHealth apps and not just usability, and they are widely used methods for measuring the quality of mHealth apps in different contexts [32-39]. The number of items and the different dimensions assessed by each mHealth measurement tool, as well as the main characteristics, are listed in Table 2.

Quality Assessment: Psychometric Characteristics (COSMIN Analysis)

The main characteristics and the results of the quality analysis according to the psychometric properties of the measurement tools (COSMIN analysis) are summarized in Multimedia Appendix 1. Most of the studies did not assess or report all the properties recommended by COSMIN. Therefore, the information reflected in this review refers to the values of the properties reported in the original studies consulted.

Discussion

Principal Findings

The aim of this study is to review the literature and collect and analyze the tools used to assess the quality of mHealth apps. As described earlier, an objective criterion (analysis of psychometric characteristics through the COSMIN guideline) was used along with a subjective criterion (assessment of the adequacy of the number of dimensions and items evaluated by the tools). The main finding of this review was the generalized lack of validation of the psychometric characteristics of the available tools, including those most commonly used. In addition, there is no robust set of outcome measures for understanding the different dimensions of mHealth apps. For overall quality, the MARS scale seems to be a potentially valid tool to establish a standardized use of it.

Validation of Psychometric Properties and Dimensions Included: 2 Characteristics to Consider When Choosing an Appropriate Tool

Usability Measurement

Regarding psychometric validation, 5 of the mHealth measurement tools included in this review met 4 or more of the properties recommended in the COSMIN guideline. The most validated tools are the *Measurement Scales for Perceived Usefulness and Perceived Ease of Use* and the Health-ITUES, with validation of 6 out of the 10 psychometric properties, and the SUS and the 2 versions of the MAUQ, with the validation of 4 properties each. However, these tools mainly focus on usability. Although usability is a critical aspect of an app that is expected to be used regularly, quality assessment cannot focus solely on this feature. This is worth highlighting because depending on whether the professional intends to assess only this specific dimension or requires further examination of mHealth apps, the choice of measurement tool based solely on the amount of validated psychometric properties may not be sufficient. The *Measurement Scales for Perceived Usefulness and Perceived Ease of Use* and the SUS are frequently used for evaluating mHealth apps; however, they are not mobile-specific. In contrast, the Health-ITUES and the MAUQ were explicitly designed for smartphones, as they allow one to evaluate the specific properties of this type of technology. The main limitation of Health-ITUES is that a unique mHealth app developed for community-dwelling adults living with HIV was used to validate the psychometric characteristics.

General Assessment

Considering the variety of dimensions evaluated, both MARS scales (professional and user versions) seem to be the tools that allow the most detailed measurement, assessing 5 aspects of the mHealth apps: engagement, functionality, esthetics, information quality, and subjective app quality. The evaluation of these additional characteristics of the apps allows for an in-depth analysis. Similarly, the MARS scale contains items related to the theoretical background, target population, or technical aspects of security or privacy of user data. However, these are not considered in the final score of the measurement instrument. Both user and professional versions have validated 3 essential psychometric properties with adequate results using around 60 mental health apps (content validity, evaluated by an expert panel to select the questionnaire items; internal consistency, Cronbach α=.90; and reliability, intraclass correlation=0.79 and 0.70 for professional and user versions, respectively). One of the strengths of this tool is the availability of a specific version for users (uMARS) with validation of the same psychometric characteristics (and with similar results) as the standard version. In addition, the MARS scale has been cross-culturally adapted and validated for different languages and is currently being adapted to other languages. There is abundant literature on the use of MARS for the evaluation of several mHealth apps, making it one of the most studied tools for assessing the quality of this kind of app.

The iSYScore tool is another method for measuring the overall quality of mHealth apps [28]. This tool has 2 significant disadvantages: it has insufficient validation (only content validity by an expert panel) and is only available in the Spanish language, so its use is severely limited.

Availability of Different Versions of the Tools: Are They All Equally Studied?

Overview

Most of these tools focus on professional use. They seek to measure the usability or the overall quality of mHealth apps in an expert way [10,23,24,28]. However, allowing users of a specific population to evaluate the quality of apps designed for their use can allow one to analyze quality from another point of view, by dividing the beliefs and expectations of professionals and users. This requires the development of tools specifically designed for users or, at least, validated versions. Only 2 of the tools included in this review have a user version. The MARS [29] uses uMARS as the user version [31]. The MAUQ has 2 versions: the stand-alone mHealth app for patients and the interactive mHealth app for patients [25]. Therefore, one of the main strengths of the MAUQ tool is the availability of 4 versions of the questionnaire: interactive apps for patients, interactive apps for providers, stand-alone apps for patients, and stand-alone apps for providers. This fact allows greater versatility in the use of the questionnaire, which is specific for each type of user and app. However, only the 2 user versions have been tested and validated using 2 health apps, so the validity and reliability of the professional versions should still be studied.

Lack of Reliability of Self-created Questionnaires

The objective of this review was to use psychometrically validated measurement tools to assess mHealth apps. However, in the scientific literature, there are mainly 2 options used by authors in studies for this evaluation of mHealth apps. First, a large number of authors choose to use self-created questionnaires explicitly designed to evaluate the characteristics of their specific apps. Because it can be personalized, this option allows greater flexibility than other generic tools. However, most of these tools are used in studies that do not focus on their validation by recruiting small samples of participants that do not allow a reliable analysis of the psychometric characteristics. Consequently, although personalized questionnaires are frequently used to determine the usability and quality of app contents, their lack of validation and the lack of knowledge about their reliability raise questions about their suitability for use. Therefore, this type of tool was not included in this review. Second, many authors choose to use previously validated tools to maximize the reliability of the results obtained, despite the possible loss of personalization derived from the use of generic tools. Traditionally, there has been widespread use of tools designed for the technological environment in health, such as the SUS scale, which is still used today. However, the use of tools not specifically designed for mobile environments limits the analysis of specific characteristics of this type of technology. For this reason, it is essential to standardize the use of validated measurement tools designed explicitly for mobiles.

Developer Transparency and Data Privacy and Security

This review shows a lack of evaluation of relevant aspects of mHealth apps, such as developer transparency and policies regarding user data privacy and security [1]. Although some tools such as MARS incorporate items with some of these aspects, the fact that these are not included in the final scores of the available tools demonstrates the need to focus on this point.

Strengths and Limitations

This is the first review to analyze the psychometric characteristics of the assessment tools of mHealth apps adhering to the COSMIN criteria. The main weakness is the difficulty in performing an optimal search and establishing adequate selection criteria because of the great heterogeneity in the tools and studies available in the literature. Another weakness is the possible risk of bias due to the possibility of losing a tool published in unreviewed databases. The implemented systematic methodology minimized biases derived from this situation.

Future studies should focus on creating and validating new tools or improving the validation of the most commonly used tools.

Conclusions

In conclusion, although there is growing evidence about the use of tools to assess the quality and content of mHealth apps, the availability of specific, highly validated tools for mobile apps is still an unexplored topic in the market. There is no robust scorecard to understand the different dimensions of mHealth apps. The COSMIN guideline allows clinicians and scientific consensus to be brought closer to the field of digital health. According to the psychometric properties, the Health-ITUES scale and the *Measurement Scales for Perceived Usefulness and Perceived Ease of Use* were the most validated tools. However, the validation is specific to a single app from a field, and its design is focused on evaluating its usability aspects. The MARS tool obtained adequate outcomes in a moderate number of psychometric characteristics, and it has been validated in a large number of mHealth apps. Its current use is widespread and evaluates different aspects of the app quality, as well as its usability.

This review suggests that the continuation of the validation of this tool in other psychometric properties might provide an appropriate option for evaluating the quality of mHealth apps that is requested by the market in the long term to quickly identify relevant apps.

Acknowledgments

This study is part of a project that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie actions (grant 823871 [Multi-dimensional Intervention Support Architecture for Gamified eHealth and mHealth Products]).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of the quality analysis of the measurement tools according to their psychometric properties (Consensus-based Standards for the Selection of Health Measurement Instruments).
[DOCX File , 16 KB - mhealth_v9i12e15433_app1.docx]

References

1. mHealth apps market size, share and trends analysis report by type (Fitness, Medical), by region (North America, APAC, Europe, MEA, Latin America), and segment forecasts, 2021 - 2028. Grand View Research. 2021. URL: https://www.grandviewresearch.com/industry-analysis/mhealth-app-market [accessed 2021-02-07]

2. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
3. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019:1-206.
4. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
5. O'Neill S, Brady RR. Colorectal smartphone apps: opportunities and risks. Colorectal Dis 2012 Sep;14(9):530-534. [doi: 10.1111/j.1463-1318.2012.03088.x] [Medline: 22646729]
6. 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]
7. O'Neil A, Cocker F, Rarau P, Baptista S, Cassimatis M, Barr Taylor C, et al. Using digital interventions to improve the cardiometabolic health of populations: a meta-review of reporting quality. J Am Med Inform Assoc 2017 Jul 01;24(4):867-879 [FREE Full text] [doi: 10.1093/jamia/ocw166] [Medline: 28339628]
8. 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]
9. Nouri R, Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
10. Brooke J. SUS: A quick and dirty usability scale. In: Usability Evaluation in Industry. 1st ed. London: CRC Press; 1996:189-194.
11. 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]
12. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. Br Med J 2002 Mar 9;324(7337):569-573 [FREE Full text] [Medline: 11884320]
13. Cooper SR, Gonthier C, Barch DM, Braver TS. The role of psychometrics in individual differences research in cognition: a case study of the AX-CPT. Front Psychol 2017 Sep 04;8:1482 [FREE Full text] [doi: 10.3389/fpsyg.2017.01482] [Medline: 28928690]
14. Kohan N. The importance of psychometric research. Rev Latinoam Psicol 2002 Jan 01;34(3):229-240 [FREE Full text]
15. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HC, et al. Cosmin guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1147-1157 [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
16. Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. Cosmin methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018 May;27(5):1159-1170 [FREE Full text] [doi: 10.1007/s11136-018-1829-0] [Medline: 29550964]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The cosmin study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
19. Marzuki MF, Yaacob NA, Yaacob NM. Translation, cross-cultural adaptation, and validation of the Malay version of the system usability scale questionnaire for the assessment of mobile apps. JMIR Hum Factors 2018 May 14;5(2):e10308 [FREE Full text] [doi: 10.2196/10308] [Medline: 29759955]
20. Dianat I, Ghanbari Z, AsghariJafarabadi M. Psychometric properties of the Persian language version of the system usability scale. Health Promot Perspect 2014;4(1):82-89 [FREE Full text] [doi: 10.5681/hpp.2014.011] [Medline: 25097841]
21. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP. European Portuguese validation of the system usability scale (SUS). Procedia Comput Sci 2015;67:293-300. [doi: 10.1016/j.procs.2015.09.273]
22. Brooke J. SUS: a retrospective. J Usability Stud 2013 Jan;8:29-40 [FREE Full text]
23. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
24. Yen P, Wantland D, Bakken S. Development of a customizable health it usability evaluation scale. AMIA Annu Symp Proc 2010 Nov 13;2010:917-921 [FREE Full text] [Medline: 21347112]
25. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]

26. Yen P, Sousa KH, Bakken S. Examining construct and predictive validity of the health-it usability evaluation scale: confirmatory factor analysis and structural equation modeling results. J Am Med Inform Assoc 2014 Oct;21(e2):241-248 [FREE Full text] [doi: 10.1136/amiajnl-2013-001811] [Medline: 24567081]
27. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
28. Grau I, Kostov B, Gallego JA, Grajales III F, Fernández-Luque L, Sisó-Almirall A. Assessment method for mobile health applications in Spanish: the iSYScore index. Semergen 2016 Nov;42(8):575-583. [doi: 10.1016/j.semerg.2015.12.001] [Medline: 26879598]
29. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
30. Payo RM, Álvarez MM, Díaz MB, Izquierdo MC, Stoyanov SR, Suárez EL. Spanish adaptation and validation of the mobile application rating scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
31. 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]
32. Salehinejad S, Kalhori SR, Gohari SH, Bahaadinbeigy K, Fatehi F. A review and content analysis of national apps for COVID-19 management using Mobile Application Rating Scale (MARS). Inform Health Soc Care 2021 Mar 02;46(1):42-55. [doi: 10.1080/17538157.2020.1837838] [Medline: 33164594]
33. 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]
34. Jovičić S, Siodmiak J, Watson ID, European Federation of Clinical Chemistry Laboratory Medicine Working Group on Patient Focused Laboratory Medicine. Quality evaluation of smartphone applications for laboratory medicine. Clin Chem Lab Med 2019 Feb 25;57(3):388-397. [doi: 10.1515/cclm-2018-0710] [Medline: 30496130]
35. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul;15(3):e12768 [FREE Full text] [doi: 10.1111/mcn.12768] [Medline: 30569549]
36. Gamwell KL, Kollin SR, Gibler RC, Bedree H, Bieniak KH, Jagpal A, et al. Systematic evaluation of commercially available pain management apps examining behavior change techniques. Pain 2021 Mar 01;162(3):856-865. [doi: 10.1097/j.pain.0000000000002090] [Medline: 33003110]
37. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
38. Ho L, Macnab A, Matsubara Y, Peterson K, Tsang B, Stothers L. Rating of pelvic floor muscle training mobile applications for treatment of urinary incontinence in women. Urology 2021 Apr;150:92-98. [doi: 10.1016/j.urology.2020.08.040] [Medline: 32890617]
39. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. JMIR Mhealth Uhealth 2019 Nov 07;7(11):e13236 [FREE Full text] [doi: 10.2196/13236] [Medline: 31697245]

Abbreviations

COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments
Health-ITUES: Health Information Technology Usability Evaluation Scale
MARS: Mobile App Rating Scale
MAUQ: mHealth Apps Usability Questionnaire
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SUS: System Usability Scale
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 10.07.19; peer-reviewed by A Paglialonga, B Greer, R Ramezani, PY Yen, R Bipeta; comments to author 26.10.20; revised version received 13.11.20; accepted 16.04.21; published 01.12.21.

Please cite as:
Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, Roldán-Jiménez C, De-Torres I, Ruiz-Muñoz M, Gonzalez-Sanchez M, Mayoral-Cleries F, Biró A, Tang W, Nikolova B, Salvatore A, Cuesta-Vargas AI
Tools for Evaluating the Content, Efficacy, and Usability of Mobile Health Apps According to the Consensus-Based Standards for the Selection of Health Measurement Instruments: Systematic Review
JMIR Mhealth Uhealth 2021;9(12):e15433
URL: https://mhealth.jmir.org/2021/12/e15433
doi:10.2196/15433
PMID:34855618

Review

Behavioral Theories and Motivational Features Underlying eHealth Interventions for Adolescent Antiretroviral Adherence: Systematic Review

Alemitu Mequanint Bezabih[1], MSc; Kathrin Gerling[1], PhD; Workeabeba Abebe[2], MD, MPH; Vero Vanden Abeele[1], PhD

[1]Department of Computer Science, e-Media Research Lab, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Department of Pediatrics and Child Health, Addis Ababa University, Addis Ababa, Ethiopia

Corresponding Author:
Alemitu Mequanint Bezabih, MSc
Department of Computer Science, e-Media Research Lab
Katholieke Universiteit Leuven
Andreas Vesaliusstraat 13
Leuven, 3000
Belgium
Phone: 32 0485949954
Email: alem.bezabih@kuleuven.be

Abstract

Background: eHealth systems provide new opportunities for the delivery of antiretroviral therapy (ART) adherence interventions for adolescents. They may be more effective if grounded in health behavior theories and behavior change techniques (BCTs). Prior reviews have examined the effectiveness, feasibility, and acceptability of these eHealth systems. However, studies have not systematically explored the use of health behavior theories and BCTs in the design of these applications.

Objective: The purpose of this review was to explore whether health behavior theories and BCTs were considered to ground designs of eHealth systems supporting adolescents' (10-24 years) ART adherence. More specifically, we examined which specific theories and BCTs were applied, and how these BCTs were implemented as design features. Additionally, we investigated the quality and effect of eHealth systems.

Methods: A systematic search was performed on IEEE Xplore, ACM, ScienceDirect, PubMed, Scopus, and Web of Science databases from 2000 to 2020. Theory use and BCTs were coded using the Theory Coding Scheme and the Behavior Change Technique Taxonomy version 1 (BCTTv1), respectively. Design features were identified using the lenses of motivational design for mobile health (mHealth). The number of BCTs and design features for each eHealth system and their prevalence across all systems were assessed.

Results: This review identified 16 eHealth systems aiming to support ART adherence among adolescents. System types include SMS text message reminders (n=6), phone call reminders (n=3), combined SMS text message and phone call reminders (n=1), electronic adherence monitoring devices (n=3), smartphone apps (n=1), smartphone serious games (n=1), gamified smartphone apps (n=1), leveraging existing social media (n=2), web-based applications (n=1), videoconferencing (n=1), and desktop applications (n=1). Nine were grounded in theory, of which 3 used theories extensively. The impact of adolescent developmental changes on ART adherence was not made explicit. A total of 42 different BCTs and 24 motivational design features were used across systems. Ten systems reported positive effects on 1 or more outcomes; however, of these ten systems, only 3 reported exclusively positive effects on all the outcomes they measured. As much as 6 out of 16 reported purely no effect in all the outcomes measured.

Conclusions: Basic applications (SMS text messaging and phone calls) were most frequent, although more advanced systems such as mobile apps and games are also emerging. This review indicated gaps in the use of theory and BCTs, and particularly the impact of developmental changes on ART adherence was not adequately considered. Together with adopting a developmental orientation, future eHealth systems should effectively leverage health theories and consider developing more advanced systems that open the door to using BCTs more comprehensively. Overall, the impact of eHealth systems on adolescent ART adherence and its mediators is promising, but conclusive evidence on effect still needs to be provided.

(JMIR Mhealth Uhealth 2021;9(12):e25129) doi:10.2196/25129

KEYWORDS

HIV; adolescents; ART adherence; eHealth; health theories; behavior change techniques; motivational design principles

Introduction

Background

HIV disproportionately affects adolescents worldwide (an extended definition covering ages 10-24 [1] is used here). AIDS is the first cause of death among adolescents in Africa and the second worldwide [2]. There is no cure for HIV yet. However, antiretroviral therapy (ART) is an effective measure to control HIV if properly adhered to. Unfortunately, suboptimal ART adherence is common among adolescents with HIV. While there is an overall improvement in other age groups in the epidemic control of HIV, when compared with other age categories [3], adolescents are characterized by (1) higher treatment dropout rate [4,5], (2) lower viral suppression achievement [6], (3) rising AIDS-related illnesses [5], and (4) a smaller decrease in AIDS-related deaths.

While adolescents share several barriers to ART adherence with adults [7], there are also challenges unique to adolescents that further complicate adherence [8-12]. Some of these challenges emerge from the unique developmental changes associated with adolescence, including biological, cognitive, and psychosocial changes [13-19]. Adolescence is a period of cognitive maturation; however, this is a gradual process, and (younger) adolescents are still limited in formal and hypothetical thinking [17]. In periods of stress, even older adolescents may regress to simplistic preformal reasoning. Consequently, they may not foresee the long-term importance of ART adherence and underestimate the severity of the HIV condition and the susceptibility of facing nonadherence consequences [20]. This in turn can lead to risk-taking behaviors. Additionally, adolescence is a period of becoming autonomous in which control over their own lives is paramount [21]. Therefore, threats to personal agency, for example, impositions from health professionals, requesting to comply with ART, are likely met with psychological reactance, an aversive response, possibly resulting in nonadherence. Adolescence is also a developmental period characterized by an orientation toward the peer group and the need to conform [22]. Because of the need to maintain appearances and fit in with their peers, and possibly engage in sexual relations [8], adolescents may choose to hide or even completely deny having HIV, especially if the disease is asymptomatic. As an unfortunate consequence, nonadherence to ART and death rates among adolescents living with HIV are disconcerting.

Prior Work

In the last decade, the availability of eHealth systems has dramatically increased [23]. eHealth refers to health services, information, and support that are delivered or enhanced through web-based technologies and related software applications [24], including SMS text messaging, web-based applications, social media, mobile apps, and games. eHealth interventions are also increasingly being used to improve ART adherence, and integrated into HIV self-management and service delivery [25-28]. Overall, eHealth is considered a promising approach to deliver effective interventions for ART treatment, for both adult [29-32] and adolescent groups [27,33-35].

ART adherence is a complex health behavior determined by multiple sociobehavioral factors, and for adolescents, further complicated by unique developmental changes at the biological, social, and psychological level. Prior reviews of eHealth ART adherence systems for adolescents have examined the effectiveness, feasibility, and acceptability [33-36] and found that such systems have generally encouraging impact. Similar findings have also been synthesized for other chronic health conditions [34,37-39]. However, to the best of the authors' knowledge, no reviews have systematically explored the extent to which health behavior theories ground the intervention, guide the selection of behavior change techniques (BCTs), or inform the design features in the app itself. This is surprising, as several studies suggest that grounding the design of eHealth systems in theory is associated with increased effectiveness [40-42]. Specific to HIV, recent systematic reviews also revealed that, for adults, designing eHealth systems based on behavioral theory is associated with efficacy in improving adherence to HIV medication [30].

Health behavior theories may provide a comprehensive understanding of adolescent's adherence behavior and its determinant factors [43]. A number of theories are already applied to model HIV medication adherence [44], including the Health Belief Model [45], Information–Motivation–Behavioral Skills Model [46,47], Social Cognitive Theory [48], Theory of Reasoned Action [49], Theory of Planned Behavior [50], and Transtheoretical Model [51]. Therefore, health behavior theories can provide insights on achieving behavior change, and hence could support the design of eHealth systems that specifically consider adolescent adherence behavior.

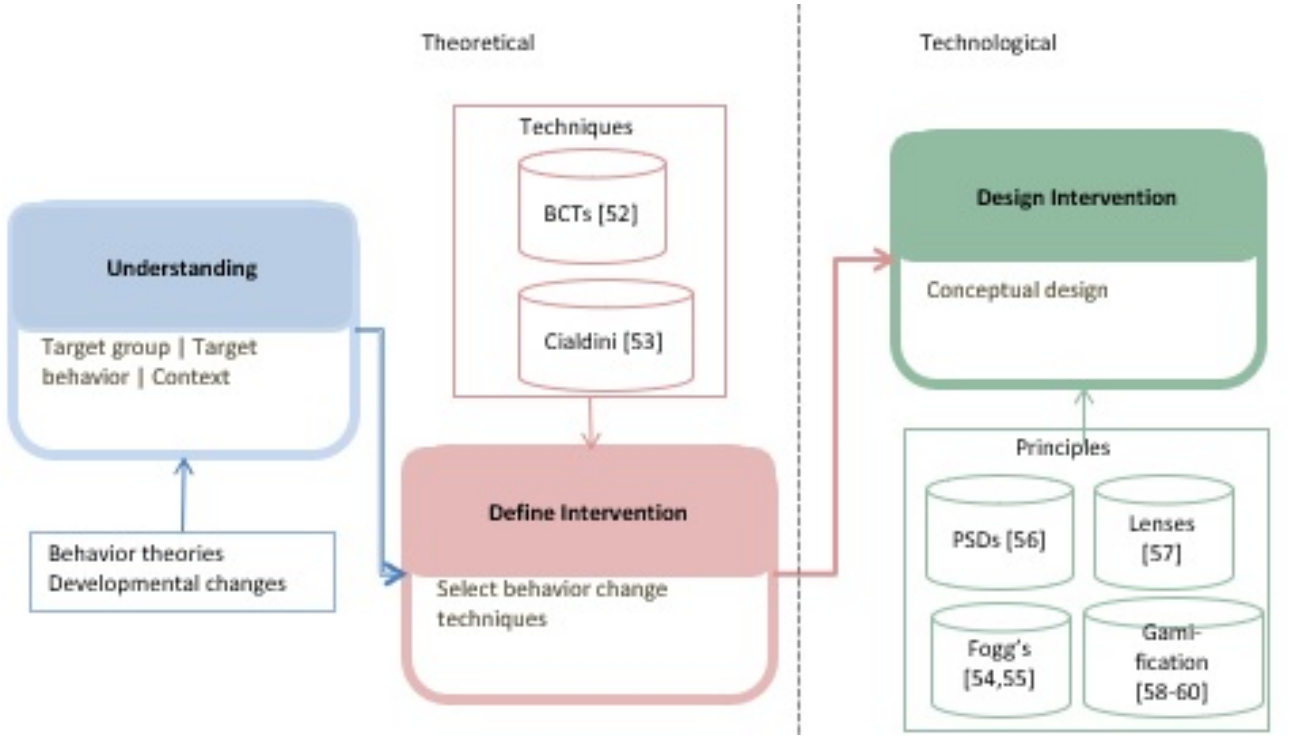
Derived from the aforementioned theories and models, several behavior change strategies exist that may influence, motivate, or persuade people to adhere to healthy behaviors. These strategies are organized into taxonomies such as the Behavior Change Technique Taxonomy version 1 (BCTTv1) of Michie et al [52] or Cialdini's influence techniques [53]. Systematic reviews also revealed that using such specific techniques facilitate behavior change, and are associated with efficacy in improving adherence to HIV medication [30]. Note that while these frameworks are applied to clinical practice, they are still technology agnostic. They apply to human-delivered interventions, first and foremost.

Hence, for eHealth interventions, such BCTs still need translation into designed features of a delivery platform. Also here, frameworks exist to guide researchers, such as Fogg's persuasive principles [54,55], persuasive system design principles of Oinas-Kukkonen and Harjumaa [56], lenses of motivational design for mHealth features by Geuens et al [57], or taxonomies of gamification elements [58-60]. Different from the BCTs, these necessitate and describe features of a technology-enabled intervention. Despite their applied nature,

here too, research has shown the need for design considerations to be based on health behavior change theories [40,42,61].

In sum, designing an eHealth intervention implies a 3-stage process: (1) understanding health behaviors through insight from appropriate *theories*, (2) defining appropriate BCTs as elements of the intervention, and (3) a translation into designed *features* for a chosen eHealth platform [42,57,62]. An illustration of this pipeline is given in Figure 1. Ideally, step 3 (designing features) is proceeded by steps 1 (understanding through theory) and 2 (specifying BCTs) [57]. However, this is not a mandate and researchers/app designers may jump to step 3 directly [42]. Considering this design pipeline perspective, it is therefore interesting to investigate whether theories matter for quality and impact of eHealth interventions for ART adherence of adolescents as well [42].

Figure 1. A stepwise design process of e-Health behavioral interventions. BCT: behavior change technique; PSD: Persuasive Systems Design.



Research Objectives

The purpose of this systematic review is to assess the extent to which studies implement the different steps of this pipeline (health behavior *theory* > *BCT* > designed *features*) and how this is related to the quality and impact of eHealth systems to promote adolescents' ART adherence. Therefore, we explore the research questions (RQs) in Textbox 1.

Textbox 1. Research questions (RQs).

- RQ1. Are eHealth interventions addressing antiretroviral therapy (ART) adherence among adolescents with HIV grounded in health behavior theory?
  - Which theories are used?
  - How often were theories used?
  - Do theories used specifically address developmental changes related to adolescence?
- RQ2. Do the eHealth interventions addressing ART adherence among adolescents rely on behavior change techniques (BCTs)?
  - Which BCTs are used?
  - How many BCTs are used by interventions?
  - Are BCTs linked to behavior change theories?
- RQ3. How are eHealth interventions addressing ART adherence among adolescents implementing the BCTs as design features?
  - What platforms are chosen?
  - Which features are designed?
  - How many design features are used in the different platforms?
- RQ4. What are the quality and impact of the eHealth interventions and how do they relate to grounding in theories related to health behavior and behavior change?
  - What is the evidence quality of the interventions?
  - What is the impact of the interventions?
  - What is the relation between grounding in theory and impact of the interventions?

Methods

Overview

The process of formulating the RQs and the search strategy was guided by the PICO(S) (Patient/Population/Problem, Intervention, Comparison, Outcome, Study design) concept of Cochrane Collaboration [63]. We used the PRIMA guidelines as a basis for conducting and reporting this systematic review [64].

Search Strategy

The search query was crafted as a combination of the PICO text words and then applied on the databases, limited to metadata fields search (title, abstract, and keyword). We restricted the search on eHealth interventions pertaining to ART adherence among adolescents, excluding those regarding other HIV care services such as HIV testing. Search terms were structured as per the syntax of each database (Multimedia Appendix 1). The entire search query was refined via several tests and peer reviews. To expand the search into an interdisciplinary space, electronic databases relevant to technology and medical fields (IEEE Xplore, ACM, ScienceDirect, PubMed, Scopus, and Web of Science) were searched on April 25, 2019, including all papers up to this date. A secondary search was conducted on all databases on January 22, 2021, to check for new relevant citations. The search sting was formulated as follows:

[HIV OR “Human Immunodeficiency Virus” OR “HIV/AIDS” OR “Acquired Immunodeficiency Syndrome” OR “HIV-positive” OR “HIV+” OR “living with HIV”] AND [adolescent OR teen* OR young OR youth] AND [ARV OR antiretroviral OR “Antiretroviral Therapy” OR “HIV treatment” OR “HIV care”] AND [“eHealth” OR “e-health” OR “electronic health” OR “digital health” OR telemedicine OR “tele-medicine” OR technology OR “computer-based” OR “web” OR “web-based” OR Internet OR online OR “social media” OR “social networking” OR “mHealth” OR “m-health” OR “mobile health” OR “mobile phone” OR “cell phone” OR “cellular phone” OR smartphone OR “text message” OR SMS OR “short message service” OR “app” OR “application” OR game OR videogame OR gamif* OR “play”] AND [adherence OR attrition OR dropout OR drop-out OR completers OR “lost to follow-up” OR withdrawal OR nonresponse OR non-response OR “completion” OR “did not complete” OR retention OR loss OR compliance OR concordance]

Eligibility Criteria

This review focused on eHealth interventions designed for adolescents to support ART adherence. We developed and applied the inclusion/exclusion criteria listed in Textbox 2.

Textbox 2. Inclusion and exclusion criteria.

Inclusion Criteria

- eHealth interventions designed for adolescents (age average between 10 and 24 years) to improve HIV medication adherence.
- Studies clearly describing the intervention content and characteristics.
- Interventions empirically evaluating antiretroviral therapy (ART) adherence, measured as one of the following:
  - Primary adherence behavior–related outcomes (eg, change in knowledge and attitude about HIV and ART, self-efficacy in taking ART medication, social support).
  - ART adherence (directly measured):
    - Objective measure (eg, using real-time electronic adherence monitoring, pill count).
    - Subjective measure (eg, self-report, caregiver report).
  - Biological outcomes (eg, CD4 count, viral load suppression).
- Peer-reviewed articles published in English (journal or conference).

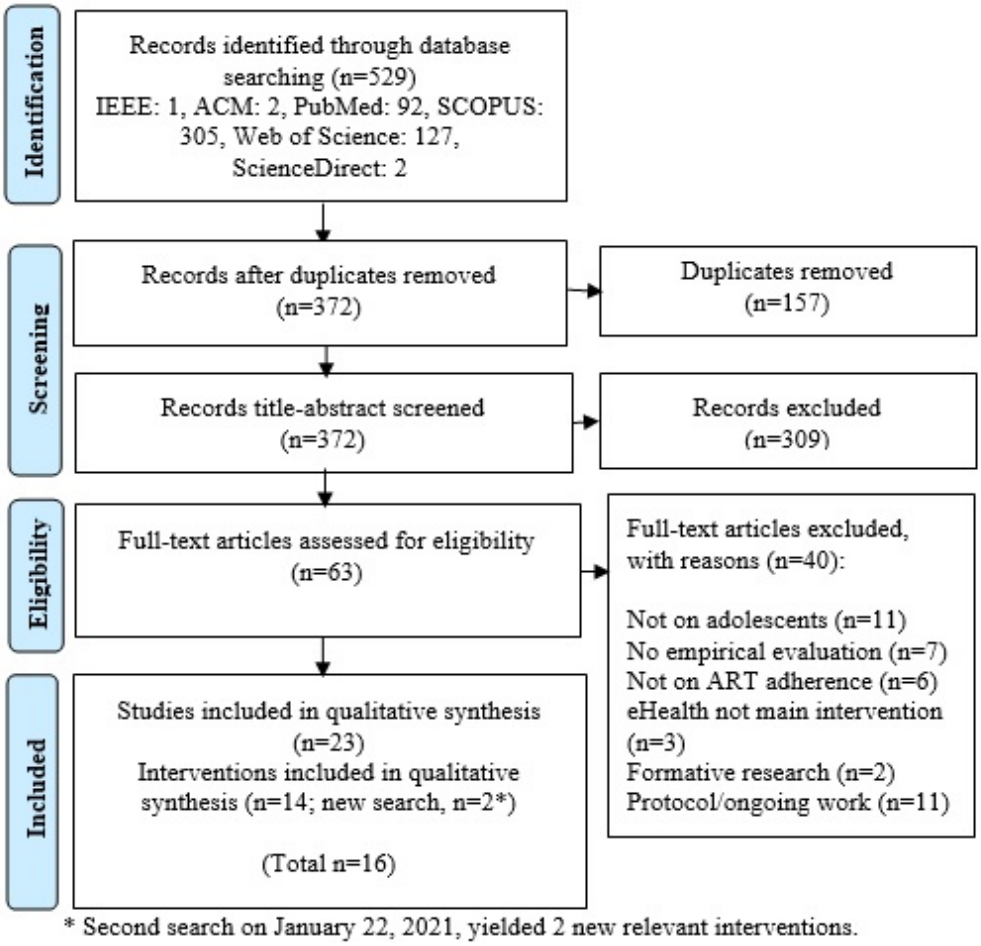
Exclusion Criteria

- Interventions designed for other age groups or health professionals (or mixed but no age subgroup analysis).
- Abstracts, reviews, protocols (however, we kept papers describing an already included eHealth system to get extra information about the intervention), ongoing works, books, book chapters.
- Interventions that only evaluated user experience, acceptability, or feasibility but not ART adherence.
- Interventions that focused on other HIV care services such as reducing risky behaviors, promoting HIV testing, or pretreatment care.

Screening and Inclusion

A multistep screening process (Figure 2) was applied to get to the final included interventions. As a first step, 529 papers were obtained, and duplicates (n=157) were removed. Next, 372 unique papers were screened by applying the criteria in Textbox 2 on both title and abstract. After this screening, 63 papers remained. In case of doubt on title/abstract information, papers were simply added to the next phase. In the fourth step, the 63 full papers were read and screened against the inclusion/exclusion criteria, resulting in further removal of 40 papers, eventually retaining 23 papers describing 14 different interventions. When available, various papers on an intervention (protocol, feasibility study, user experience study, pilot test, and randomized controlled trial [RCT]) were all kept to the end for further reference of intervention details. In April 2021, a second database search was performed to capture any new studies, resulting in 165 new unique citations. Following the same screening process, 11 papers were screened (after excluding 154) via title/abstract screening. Full-text reading of these papers resulted in 2 papers describing 2 eHealth systems for final inclusion. Totally, out of 694 papers from both searches, 16 different eHealth apps were included and analyzed.

Figure 2. Screening and inclusion process.



Data Extraction and Analysis

We coded relevant characteristics including intervention name, publication year, country, delivery technology, study design, population, sample size, follow-up duration, intervention summary, theory use and justification of theory selection, consideration of developmental changes related to adolescence, outcomes, and effectiveness.

Coding of Behavior Theories, Behavior Change Techniques, and Designed Features

Behavior Change Theories

We used the Theory Coding Scheme (TCS) developed by Michie and Prestwich [65] to describe the theory basis of each intervention and to assess to what extent theories were utilized. This checklist has 19 items to verify whether an intervention mentions a theory or model, to what extent a theory or model is used in designing intervention features, and whether the theory is tested in pre/poststudy. Item 1 verifies whether a particular theory is mentioned in the studies, whereas items 2 and 4-11 measure to what extent the theory is used during designing interventions. Item 3 checks whether more than 1 theory is applied. Item 12a verifies whether the impact on adherence mediators was measured. The remaining items were left uncoded as they deal with reliability of methods in measuring and testing theory, and theory refinement, which is beyond the scope of this study. Each item was coded as 1 if applied and 0 if not present. As a last step and composite measure, the included eHealth systems were graded on the extent of theory usage as no, low, medium, or high usage. Studies were labeled as follows: "no," if there was no mention of theory, or they mentioned a theory but there was no grounding of the intervention on it (TCS item 1); "low," if it was explicitly stated that theory concepts were used to design the intervention (1 or more of the TCS items 2, 5, 8, and 11); "medium," if all intervention techniques originated from theory concepts (TCS item 7 or 9); "high," when all theory concepts were addressed, or used to select participants or tailor interventions (1 or more of TCS items 4, 6, and 10).

Consideration of Developmental Changes

We investigated studies for the presence of discussions concerning how developmental theory concepts influence ART adherence among adolescents. This examination was particularly inspired by TCS item 1 which says "Theory/model of behavior mentioned—Models/theories that specify relations among variables, in order to explain or predict behavior are mentioned, even if the intervention is not based on this theory" [65]. Hence, we examined the presence of explicit discussions on how developmental changes (biological, social, psychological) would "explain or predict" (to use the exact words) ART adherence among adolescents.

Behavior Change Techniques

For coding BCTs from intervention descriptions, BCTTv1 [52] was used. BCTTv1 is a well-established taxonomy with an

extensive list of theoretical methods of behavior change, containing 93 BCTs grouped into 16 clusters. It has been widely applied to specify intervention techniques in various behavioral domains such as physical activity [66], alcohol use [67], and medication adherence [68]; as well as to identify the presence of BCTs in existing interventions [41,62]. To examine the use of BCTs in interventions, the prevalence of individual BCT across all interventions was calculated. Similarly, we also calculated the total number of BCTs per intervention. Considering the complexity of coding BCTs, coders completed a certified online training on BCT taxonomy (BCTTv1) [69].

Designed Features

Because BCTs are only descriptions of behavior change from a psychological perspective, in other words they are technology agnostic, we additionally coded the designed features, using the lenses of motivational design for mHealth developed by Geuens et al [57]. Geuens et al [57] explained how theoretical concepts of behavior change can be translated into design principles, by including also implementation instantiations of these principles through examples of mobile app features. These lenses of motivational designs provide implementation-level descriptions of design principles, also encompassing persuasive principles of [54-56]. The prevalence of individual design features among systems and the number of features per each system are calculated.

Impact

Because interventions differ substantially in measuring their impact, *each outcome measure* was coded (mediators, adherence, or health outcomes) as positive effect (+), no effect (0), or negative effect (–). Hence, we did not exclude any outcome measures or limit ourselves to direct measures of adherence such as medication intake or viral load. This inclusive approach is justified by the Information–Motivation–Behavioral Skills Model [47] of ART adherence, as it accommodates a broad range of outcome measures [47,70], and is widely adopted in ART adherence research [71-73]. This model conceptualizes ART adherence as a function of 3 mediators: information about HIV and ART, motivation to take medication, and behavioral skills required for taking medication. An increase in any of these constructs is theorized to result in improved adherence behavior which in turn produces favorable health outcomes.

Methodological Quality Assessment

Finally, we evaluated the strength of evidence for the methodological quality of included interventions with respect to the review questions, using the quality assessment method used in Johnson et al [74], with a minor modification: we discerned pre/posttest studies from single-subject and case studies (Table 1). This method was first used to assess the quality of evidence for the impact of computer games and serious games on learning [75,76], and later for their impact on health and well-being [74]. It has 5 criteria against which every included intervention is scored from 1 to 3. Adding up each of the 5 marks gives a possible maximum score of 15. A subsample of interventions (4/15, 26.67%) was coded independently by 2 coders (first and last authors). Interrater reliability was calculated using intraclass correlation coefficient with 2-way mixed effects and absolute agreement. The score was 0.89, showing a good agreement between the 2 coders. The quality measure here refers to the quality of clinical validation test (study design, eg, RCT, pre/post; sample size), not the quality in terms of design and development of eHealth systems (ie, integration of theory, BCTs, design elements).

Table 1. Evidence quality assessment method.

Criterion	Min score	Max score
How appropriate is the research design for addressing the question, or subquestions of this review: randomized controlled trails (3), quasi-experimental study (2.5), pre/posttest design (2), case study, single subject-experimental design (1)?	1	3
How appropriate are the methods and analysis?	1	3
How generalizable are the findings of the study to the target population with respect to the size and representativeness of sample?	1	3
How relevant is the particular focus of the study (including conceptual focus, context, sample, and measures) for addressing the question or subquestions of this review?	1	3
To what extent can the study findings be trusted in answering the study question(s)?	1	3

Results

Overview

Of the 16 included systems, most were built in the United States (n=12), 2 in Nigeria, 1 in Uganda, and 1 in Argentina. Some studies (n=5) intentionally included participants with poor adherence performance. Half of the systems employed pre–post designs, whereas the other half used RCTs for clinical validation. Table 2 shows the list of included studies, and the summary of interventions is provided in Multimedia Appendix 2.

Table 2. Included studies.

No.	Intervention studies[a]	References
1	Whiteley et al	[77,78]
2	Tanner et al	[79,80]
3	Stankievich et al	[81]
4	Spratt et al	[82]
5	Shegog et al	[83]
6	Belzer et al	[84-86]
7	Saberi et al	[87]
8	Linnemayr et al	[88,89]
9	Puccio et al	[90]
10	Naar-King et al	[91,92]
11	Hightow-Weidman et al	[93]
12	Dowshen et al	[94,95]
13	Garofalo et al	[96]
14	Dworkin et al	[97-99]
15	Dulli et al	[100]
16	Abiodun et al	[101]

[a]For interventions with multiple studies, only the first author of one of the papers is used.

RQ1: Are eHealth Interventions Grounded in Behavior Theory?

Which Theories Were Commonly Applied to Inform Behavior Change?

A total of 10 different theories were mentioned (Table 3), of which 3 theories appeared commonly, namely, Social Cognitive Theory, the Information–Motivation–Behavioral Skills Model, and Motivational Interviewing. The complete theory coding sheet of included studies using the TCS is provided in Multimedia Appendix 3.

Table 3. Theories used by the included interventions.

Theory	Frequency	Interventions
Social Cognitive Theory	4	Tanner et al [79,80], Shegog et al [83], Hightow-Weidman et al [93], Garofalo et al [96]
Information–Motivation–Behavioral Skills Model	3	Whiteley et al [77,78], Linnemayr et al [88,89], Dworkin et al [97-99]
Motivational Interviewing	3	Spratt et al [82], Shegog et al [83], Naar-King et al [91,92]
Fogg Behavior Model	1	Hightow-Weidman et al [93]
Empowerment Theory	1	Tanner et al [79,80]
Transtheoretical Model	1	Spratt et al [82]
Stress, Appraisal, and Coping Theory	1	Belzer et al [84-86]
Narrative Communication (Storytelling)	1	Hightow-Weidman et al [93]
Ecological Momentary Intervention	1	Dowshen et al [94,95]
Social Action Theory	1	Puccio et al [90]

Were Behavior Change Theories Used Extensively?

Of the 16 eHealth systems, 9 were grounded in theory. However, the extent of theory utilization (the extent to which interventions address particular theory-relevant constructs [65]) varies substantially. Four interventions were supported by more than 1 theory. Of these, 1 (Hightow-Weidman et al [93]) was guided by 3 theories, whereas 3 interventions (Tanner et al [79,80], Spratt et al [82], and Shegog et al [83]) combined 2 theories. Table 3 provides an overview of the specific theories used per intervention. The remaining 7 interventions were not grounded in theory. While 3 of them mentioned a particular theory and its constructs in relation to adherence, they were not utilizing

it to inform the design of the intervention. The remaining 4 did not refer to theory at all.

Based on the degree of application of theories in designing interventions, that is, based on the TCS [65] (also see Multimedia Appendix 3), we grouped studies into 4 usage categories as stated in the "Methods" section: no, low, medium, and high (Table 4). As much as 7 of 16 interventions have no use of theory, 5 have low theory usage, 1 used theory moderately, and 3 used theories extensively.

Table 4. The extent of theory usage of the included interventions, based on the Theory Coding Scheme [67].

Theory usage category	Theory Coding Scheme items	Interventions
No		
No theory mentioned	Not applicable	Stankievich et al [81], Saberi et al [87], Dulli et al [100], Abiodun et al [101]
Mentioned theory	1	Puccio et al [90], Linnemayr et al [88,89], Dowshen et al [94,95]
Low	2, 5, 8, 11	Spratt et al [82], Shegog et al [83], Belzer et al [84-86], Naar-King et al [91,92], Garofalo et al [96]
Medium	7, 9	Whiteley et al [77,78]
High	4, 6, 10	Tanner et al [79,80], Hightow-Weidman et al [93], Dworkin et al [97-99]

Were Developmental Changes Related to Adolescence Considered?

Given the importance of using suitable theoretical foundation for interventions that address adolescents (see the "Introduction" section), we explored whether the studies included in our review adequately address this aspect in how they report their work. Hence, we entirely examined them for information concerning the influence of developmental changes (biological, social, psychological) on ART adherence among adolescents. For example, we explored whether the studies considered the question of which theories could work better for adolescents (ie, which theory would provide better coverage of factors associated with developmental changes). Our findings suggest that none of the theory-informed eHealth interventions explicitly discussed theory selection from a developmental perspective: none of the studies provided a discussion of why or how the respective theories were appropriate for adolescents, suggesting lack of consideration of developmental changes. However, they mentioned previous use of those health theories in behavior change research including in adolescent populations (see Multimedia Appendix 2), although they did not include explicit reasoning for their choice. This may suggest that, in general, developmental changes related to adolescence are currently not included in the choice of theoretical foundation for interventions to increase ART adherence among adolescents in a transparent way, leaving room for future work that explicitly draws from suitable theory to achieve better outcomes as suggested by [102]. Additionally, we searched for evidence that developmental changes were accounted for in the intervention design—whether any specific design features were related to these changes—but no such explicit association was found.

RQ2: Are eHealth Interventions Using Behavior Change Techniques?

Which Behavior Change Techniques Were Most Common?

Across the 16 eHealth systems reviewed, a total of 42 BCTs were identified (Multimedia Appendix 4). The most popular technique was "Prompting/cueing," which was used by 11 interventions. The second most frequent technique was "Social support (unspecified)," used in 10 interventions. The third was "Problem solving," used in 8 interventions. Next, "Monitoring of behavior by others without feedback" was used 7 times, and "Information about health consequences," "Credible source," "Social support (emotional)," and "Instructions on how to perform behavior" were each used 5 times. "Demonstration of the behavior" has been used 4 times and the remaining techniques appeared 3 times or less.

How Many BCTs Were Used Per Intervention?

The median number of BCTs used was 5. The intervention with the highest number of techniques was by Dworkin et al [97-99] utilizing 17 different ones. The lowest number of BCTs was found in Abiodun et al [101], using 1 technique. A complete list of details on how many techniques each intervention has applied is provided in Multimedia Appendix 4.

Are Behavior Change Techniques Linked to Behavior Change Theories?

Most theory-based studies (7/9: Shegog et al [83], Belzer et al [84-86], Garofalo et al [96], Whiteley et al [77,78], Tanner et al [79,80], Hightow-Weidman et al [93], Dworkin et al [97-99]) linked at least one intervention technique to 1 theory (TCS item 8) and at least one theory-relevant construct to an intervention technique (TCS item 11). However, only half (4/9: Whiteley et al [77,78], Tanner et al [79,80], Hightow-Weidman et al [93], Dworkin et al [97-99]) linked all theory-relevant constructs to at least one intervention technique (TCS item 10), and only 5 out of 9 (Whiteley et al [77,78], Tanner et al [79,80], Hightow-Weidman et al [93], Garofalo et al [96], Dworkin et al [97-99]) linked all intervention techniques to at least one theory-relevant construct (TCS item 7). Two (Shegog et al [83] and Naar-King et al [91,92]) studies used theory constructs to tailor intervention techniques to recipients (TCS item 6), but no study selected recipients for the intervention based on theory constructs (TCS item 4).

RQ3. How Are eHealth Interventions Designed?

What Platforms Were Used to Implement eHealth?

Many of the systems were simplistic applications from a technical perspective developed on basic phones (ie, SMS text

messaging [n=6] and phone calls [n=3]; Table 5). However, advanced applications designed for smartphones were also emerging (n= 3). These smartphone-based apps were designed in various forms—ordinary apps, serious games, and gamified apps. Already existing social media apps were also utilized. Electronic adherence monitoring devices (electronic medication containers that look like ordinary bottles or mobile phones, eg, WisePill) were also common, but mainly used in combination with other platforms and not as a standalone system. The primary purpose of these devices was to objectively measure adherence, except in 1 intervention (Spratt et al [82]) in which it was used to deliver reminders in the form of blinking lights and chime sounds. Other systems included web-based applications (desktop/laptop), remote videoconferencing, and desktop applications (Naar-King et al [91,92]). Such systems were, however, less frequent, each appearing just once.

Table 5. Type of systems in included studies.

System type	Frequency	Study
SMS text messaging	6	Stankievich et al [81], Spratt et al [82], Linnemayr et al [88,89], Dowshen et al [94,95], Garofalo et al [96], Abiodun et al [101]
Phone call	3	Spratt et al [82], Puccio et al [90], Belzer et al [84-86]
Electronic adherence monitoring device[a]	3	Whiteley et al [77,78], Spratt et al [82], Linnemayr et al [88,89]
Smartphone app	1	Dworkin et al [97-99]
Smartphone serious game	1	Whiteley et al [77,78]
Gamified smartphone app	1	Hightow-Weidman et al [93]
Social media	2	Tanner et al [79,80], Dulli et al [100]
Web-based application (desktop/laptop)	1	Shegog et al [83]
Videoconferencing	1	Saberi et al [87]
Desktop applications	1	Naar-King et al [91,92]

[a]Used in combination with others, not as a standalone intervention system.

Which Design Features Were Common?

Out of the 28 motivational design features proposed by Geuens et al [57], we found 24, with the most frequent features being "Reminders," "Personalization," and "General information"—appearing 11, 10, and 9 times, respectively (Multimedia Appendix 5). Information to educate adolescents about HIV and ART adherence was sometimes tailored to participants based on personal profiles—hence "Microtailoring." "Instructions" on how to perform certain tasks such as taking medication and talking to providers were also detected. In some interventions, generic information and instructions were from health care expert sources (ie, "Expertise"). Asking patients to manually enter information (ie, "Logging") about performance of their behavior (eg, whether they took medication or not) or outcome data (eg, viral load, CD4 count) was also common, although sometimes this was also done automatically, via "Tracking," using electronic devices such as medication adherence monitoring devices.

How Many Design Features Were Used in the Different Platforms?

The median number of design features was 4. The intervention with the highest number of design features (n=15) was Hightow-Weidman et al [93], while the interventions with the lowest number of features (n=1) were Puccio et al [90] and Abiodun et al [101] (Multimedia Appendix 5).

RQ4. What Are the Quality and Impact of the eHealth Interventions and How Do They Relate to Grounding in Theory?

Quality

We adopted Johnson et al's [74] method to categorize papers on methodological quality, computing the quality of evidence score, ranging from 5 to 15. Papers with a rating 8 or below are categorized as "weak evidence," 9-12 as "moderate evidence," and 13 and above as "strong evidence." Six interventions scored strong on quality of evidence, 4 moderate, and the remaining 7 weak (Table 6). The quality ratings only pertain to the strength of empirical evidence for outcome effects, and do not judge the overall quality of the studies.

Impact

As mentioned in the "Methods" section, outcome measures were coded following Information–Motivation–Behavioral Skills Model's conceptualization of ART adherence behavior as a function of 3 mediators: information, motivation, and behavioral skills [47]. Therefore, the measures corresponding to these parameters are knowledge on HIV and ART, personal motivation and social motivation, and ART self-efficacy, respectively. Direct measures of adherence behavior are ART medication adherence and appointment adherence. Similarly, measures of biological outcomes include viral load and CD4 count. The number of measurements for each outcome, effect type, and quality of evidence is summarized in Table 6. Of the 16 eHealth systems, 10 reported positive effect on 1 or more of the outcomes measured, yet 7 of these also reported no effect

on 1 or more of other outcomes. As much as 6 out of 16 reported purely no effect in all the outcomes measured.

Table 6. Impact of included interventions on Integrated Behavioral Model mediators of antiretroviral therapy adherence among adolescents and quality ratings.

Interventions	Theoretical grounding and motivational features			Quality and effectiveness[a]									
	Overall quality of theory integration (Table 4)	Behavior change techniques, n	Design features, n	HIV knowledge	Antiretroviral therapy (ART) knowledge	ART motivation (personal motivation)	Social support (social motivation)	ART self-efficacy	ART medication adherence	Medical appointment adherence	Viral load	CD4 count	Quality of evidence
Whiteley et al [77], [78][d,e]	Medium	15	10	+	+	0	0	0	+		+		15
Tanner et al [79], [80][e,f]	High	16	12							+	+		11
Hightow-Weidman et al [93][e]	High	15	15	+	+			+					9
Dworkin et al [97]-[99][e]	High	17	14	0				0	+				12
Stankievich et al [81][g]	No	2	3								0		6
Spratt et al [82][e,g]	Low	9	4						0			-	6
Saberi et al [87][g]	No	7	2		0		0	0					6
Linnemayr et al [88], [89][g]	No	3	2						0				13
Puccio et al [90][g]	No	3	1						0		0		5
Naar-King et al [91], [92][g]	Low	11	4						0		0		15
Dowshen et al [94], [95][g]	No	2	3						+		0	0	8
Garofalo et al [96][g]	Low	4	4						+		0		13
Shegog et al [83][g]	Low	2	5	+	+	+		+					7
Belzer et al [84]-[86][g]	Low	4	2			0		0	+		+		11
Dulli et al [100][g]	No	6	4	+			0		0				13
Abiodun et al [101][g]	No	1	1						0		+		14

[a]"+" means positive effect; "0" means no effect; "–" means negative effect.

[b]True for newly started ART; no significant effect on patients who stayed longer on ART.

[c]Higher usage of theory and motivational features, scoring at least medium and high once.

[d]Electronic monitoring alone (control) acts as a better intervention than additional signal and SMS text message reminders (intervention).

[e]Lower usage of theory and motivational features.

What Is the Relation Between Grounding in Theory and Impact of the Interventions?

The nature of this review precludes any firm conclusions. A closer look at Table 6 suggests mixed results; we cannot unambiguously conclude that more extensive grounding of design features in theory related to health behavior or behavior change is associated with better (significant) effectiveness on outcomes with good evidence quality (compare studies with footnotes c and e). However, the results suggest that 10/16 (Whiteley et al [77,78], Tanner et al [79,80], Hightow-Weidman et al [93], Dworkin et al [97-99], Dowshen et al [94,95], Garofalo et al [96], Shegog et al [83], Belzer et al [84-86], Dulli et al [100], Abiodun et al [101]) of the included eHealth apps report a positive impact on ART adherence or on its mediators.

Discussion

Summary

This systematic review examined theory usage, integration of BCTs, and motivational design features and technology platforms used in existing eHealth ART adherence interventions for adolescents, and how these aspects relate to the quality and impact of interventions. Generally, we found the impact of eHealth systems on adolescent ART adherence and its mediators promising. Moreover, most included systems attempted rooting eHealth interventions in theoretical frameworks. Nevertheless, we found a gap between the discussion of theories to root an intervention and the actual application of those theories in terms of system design. Additionally, we only came across few systems that contain a considerable number of BCTs and motivational design features. Instead, elementary designs characterize current systems. In the following paragraphs, we detail the main findings and relate them to existing theoretical and empirical work.

Principal Findings

Are eHealth Interventions Grounded in Health Behavior Change Theory?

The review shows that current eHealth systems to improve adolescents' ART adherence refer to theory only lightly. Of the 16 included eHealth systems, only 4 interventions show extensive to moderate usage of theory, while the remaining have low or no usage (Table 4). This finding contrasts with the studies that have argued that grounding the design of eHealth systems in theory is associated with increased effectiveness [40-42]. In the context of HIV in particular, designing eHealth systems based on behavioral theory is associated with efficacy in improving adherence to HIV medication [30]. This is attributed to the fact that health behavior theories provide a comprehensive understanding of ART adherence behavior and its determinant factors, to inform the design of the intervention [44,103]. Theory underutilization may result in a limited understanding of the different moderators of ART adherence among adolescents, leaving several factors unaddressed by the intervention, which in turn reduces effectiveness [104].

Moreover, our findings indicate that the context of developmental changes and their impact on ART adherence are not explicitly addressed in the included studies. This is somewhat surprising as prior research has shown that factors associated with developmental changes affect adolescents' health including medication adherence [105]. Adolescence presents a specific developmental stage marked by profound biological, psychological, and social changes [106-111] that may affect ART adherence behavior [20]. For instance, we did not find explicit mention of how the ongoing development of cognitive capacity [17], oppositional behavior associated with becoming autonomous [21], and conformity to peer pressure [22] and the synergy it may have with HIV stigma [112] can affect ART adherence among teenagers. Being informed of these developmental theory concepts might be useful in the choice of appropriate BCTs and design features. However, the absence of explicit referral to developmental theory does not automatically imply that interventions are not (implicitly) informed by developmental theories.

Are Behavior Change Techniques Guided by Theory?

Almost half of the interventions applied only 4 or less BCTs, which may render them less effective in delivering intervention content: prior research on BCTs and health behaviors, (eg, [40-42,62]) has indicated that the number of BCTs applied in an intervention influences its effect, that is, interventions that employ more BCTs were found to have a larger effect on behavior than those that apply fewer BCTs [41]. This relatively low number of BCTs usage might be the result of a less comprehensive conceptualization of adherence and its determinants, or a less comprehensive conceptualization of adolescents' developmental changes and their impact on ART adherence. For example, an intervention narrowing adherence to "consuming medication," and focusing primarily on "forgetfulness" as main barrier, might end up employing the "prompts/cues" BCT only; "reminders" as the main technique can be seen in [90,94]. A deeper look through the perspective of developmental changes (ie, a reflection on why adolescents forget) might perhaps suggest another solution; for example, building adolescents' skills on how to better integrate medications into daily life. In terms of breadth (covering a broad range of factors), reviews on barriers of adolescent's ART adherence list many factors complicating adherence [9-11] that eHealth systems need to address. Moreover, we found that BCTs which might be relevant for adolescents such as demonstration of behavior (modeling), peer comparison, or incentivization were scarce.

What Are the Prevalent Platforms and Designed Features?

We found that most included systems were limited to SMS text messaging or phone calling. Compared with more advanced information and communications technology systems, SMS text messaging and phone calls offer the advantage of being cheap to deliver. However, such systems also limit the implementation of more sophisticated features. For example, it is challenging to deliver intervention content enriched with engaging audiovisual content. SMS text messages alone might not be sufficiently engaging to adolescents, which could be a possible explanation for researchers reporting noneffectiveness [88,113,114]. A previous work noted that for adolescents "texts

that say the same thing are boring" [77], and the authors recommended dynamically changing the content of the message so that adolescents do not get tired of reading the same messages [89].

As for the designed features, "reminders (notifications)" and "personalization (adapting color schemes and skins)" were the most frequently applied followed by displaying "general information." The remaining features were scarcely used. Providing reminders for medication, enabling customization on system features, and educating adolescents about HIV and ART are still appropriate. However, interventions could be more effective if they additionally include "social" features such as principles grouped under "reward and incentives" and "social interactions" categories [57].

What Are the Quality and Impact of the eHealth Interventions?

Of the 16 eHealth systems, 10 (Whiteley et al [77,78], Tanner et al [79,80], Hightow-Weidman et al [93], Dworkin et al [97-99], Dowshen et al [94,95], Garofalo et al [96], Shegog et al [83], Belzer et al [84-86], Dulli et al [100], Abiodun et al [101]) reported positive effect on 1 or more of the outcomes measured, yet 7 (Whiteley et al [77,78], Dworkin et al [97-99], Dowshen et al [94,95], Garofalo et al [96], Belzer et al [84-86], Dulli et al [100], Abiodun et al [101])out of these 10 also reported no effect on 1 or more of other outcomes. Six out of 16 (Stankievich et al [81], Spratt et al [82], Saberi et al [87], Linnemayr et al [88,89], Puccio et al [90], Naar-King et al [91,92]) reported purely no effect in all the outcomes measured. Overall, while evidence is mixed, the impact of these systems on ART adherence and its mediators has a positive trend. However, the nature of this review and diversity of eHealth studies preclude any firm conclusions based on our study findings (Table 6). Considering the relatively small number of digital interventions identified in this review and the mixed evidence on its impact, more digital interventions for adolescents' HIV self-management should be evaluated to come to firm conclusions.

Implications and Recommendations

Overview

A clear understanding of theoretical insights enables designers to translate various BCTs into features appropriate to this specific audience; the way BCTs are implemented and presented to users matters in respect to their effectiveness. Moreover, implementation sophistication on top of applying relevant intervention content is paramount. In the paragraphs below, we detail the recommendations derived from this review.

Ground Interventions in Theories and Methods Tailored to Adolescents

Deciding on a particular theory is challenging for eHealth designers as a plethora of health theories exist, and there is not one theory specifically developed for adolescents. In this respect, protocols for developing behavioral interventions such as the IM (intervention mapping) [115] and capability, opportunity, motivation, behavior frameworks [116,117] may be of interest. According to IM, one should first consult existing literature to list reported causes of the target health problem (in this case, barriers/facilitators of ART adherence among adolescents). Next, concepts from this list should be linked to theoretical constructs (ie, the theory with better coverage of this list could be a good candidate). Moreover, use of a combination of multiple theories (such as Integrated Behavioral Model or any customized combination) might be considered. Complementing this approach, the practical study of the actual target group (eg, user-centered methods such as participatory design [118,119] with adolescents) might also help identify the most relevant determinants (eg, [120]) on which an intervention can focus [121].

Selection of Behavior Change Techniques Tailored to Adolescents

Incentivization holds particular potential to improve adherence to HIV care [122-125] among adolescents. Moreover, as peer pressure among teen population is high [19], implementing social-based techniques involving peer role modeling and comparisons among peers might be effective. In this respect, Hightow-Weidman et al [93] reported that participants enjoyed "social-media-like" discussion boards that prompt peers to daily discuss on HIV matters and share experiences. However, any use of social media with adolescents should be done with utmost care, discussed further below.

Additionally, methods that help adolescents analyze barriers they face, and help them generate strategies to overcome these issues might be crucial. This is already addressed by some interventions that seek to develop problem-solving skills, for example, phone conversations with a professional adherence counselor [87] or a similar discussion with a humanlike character animated in computer software [91].

Implement Sophisticated Features With More Advanced Technology Platforms

More advanced and engaging design features such as gamification and social connections are needed in eHealth systems. Although evidence on the efficacy of gamification specific to adolescents' ART adherence is not synthesized yet, literature on gamification and adherence indicated its potential [126,127] and positive influence on health behavior change in general [74,128-132]. First, as mentioned above, adolescents might be more attracted to immediate rewards [17] (eg, points) and incentives associated with adherence achievements (eg, badges) than foreseeing future long-term health consequences. In addition to material rewards, virtual rewards in the form of gamification elements such as points, levels, leaderboards, and easter eggs might be interesting. Additionally, as adolescents are heavily influenced by peer pressure [13], integrating "social interaction" principles [57] might be important. Allowing adolescents to link and share via social media platforms, connect with others in a similar condition and demonstrating their success (social identification), know how other peers are performing (social comparison), compete or cooperate with peers on social interaction environments could be useful. However, we acknowledge that given the specific developmental characteristics of this age group and the dangers of issues related to HIV stigma (eg, concerns similar to privacy unraveling

[133,134]), any such designed features should be stigma sensitive and introduced with the greatest care.

Finally, to enable using more advanced features, implementations such as desktop applications, mobile apps, games, or including gamification elements and stigma-sensitive social media–like features might be more appropriate to engage teenagers. As such applications are emerging, implementations in these platforms are warranted in future interventions.

Limitations and Future Work

As per our knowledge, this is the first systematic review to examine theory usage, BCTs, motivational design features, and technology platforms used in existing eHealth ART adherence interventions for adolescents. To investigate our main points, we used established coding frameworks and taxonomies. Nevertheless, this study comes with its limitations. First, this review focused primarily on exploring existing eHealth ART interventions for adolescents. There is a need for a more rigorous study of what best promotes ART adherence among adolescents, that is, looking at the future. Second, although established coding taxonomies were applied, and coders were trained, coding BCTs and design principles remained challenging due to variations in intervention content descriptions. Templates for reporting eHealth interventions for behavior change might be useful here. Finally, the findings in this review depended only on explicit referral to theories used to ground intervention and design. We must acknowledge that this methodological approach does not allow us to conclude that interventions and designs are not tailored toward adolescents if done implicitly. It may be that interventions are still informed by knowledge on adolescents' developmental changes, but simply without explicit mentioning, for example, through the embodied knowledge of experts contributing to the design. Hence, future work may address this by including prespecified developmentally appropriate BCTs and incorporating these in the systematic review of eHealth interventions for ART adherence among adolescents. However, it should be noted here that, for the sake of reproducibility, design knowledge should be transparently communicated as a good practice.

Conclusions

In this review, a total of 16 eHealth interventions targeting adolescents' ART adherence were included. Overall, the impact of these systems on ART adherence and its mediators is promising, but evidence remains mixed. We observed mostly simple applications (ie, for SMS text messages and phone calls); however, advanced smartphone apps are emerging. Moreover, most interventions applied only a limited number of BCTs and designed features. While 9/16 systems were grounded in theory, overall, health theories were utilized sparsely. Moreover, we observed a dearth of approaches addressing the specific developmental changes related to adolescence and their implications for intervention design and responsiveness to BCTs. In summary, we suggest that eHealth interventions, as well as the design of specific delivery platforms, should include health theories that are appropriate to adolescents' development, and implement features that cater to this age group.

Acknowledgments

This work was supported by Group T Academy, Belgium and the Ministry of Science and Higher Education, Ethiopia via a joint scholarship offered to the first author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms structured as per the syntax of each database.
[DOCX File , 18 KB - mhealth_v9i12e25129_app1.docx]

Multimedia Appendix 2
Characteristics of included studies.
[DOCX File , 25 KB - mhealth_v9i12e25129_app2.docx]

Multimedia Appendix 3
Use of theory in the included interventions from the Theory Coding Scheme (TCS).
[DOCX File , 22 KB - mhealth_v9i12e25129_app3.docx]

Multimedia Appendix 4
Presence of behavior change techniques (BCTs) from BCT taxonomy version 1 in the included interventions.
[DOCX File , 31 KB - mhealth_v9i12e25129_app4.docx]

Multimedia Appendix 5
Presence of design principles in the included interventions from lenses of motivational principles.
[DOCX File , 25 KB - mhealth_v9i12e25129_app5.docx]

References

1. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. Lancet Child Adolesc Health 2018 Mar;2(3):223-228. [doi: 10.1016/S2352-4642(18)30022-1] [Medline: 30169257]
2. Active involvement of young people is key to ending the AIDS epidemic by 2030. URL: https://www.unaids.org/en/resources/presscentre/featurestories/2015/august/20150812_PACT [accessed 2019-08-13]
3. Slogrove AL, Sohn AH. The global epidemiology of adolescents living with HIV: time for more granular data to improve adolescent health outcomes. Curr Opin HIV AIDS 2018 May;13(3):170-178 [FREE Full text] [doi: 10.1097/COH.0000000000000449] [Medline: 29432227]
4. Kranzer K, Bradley J, Musaazi J, Nyathi M, Gunguwo H, Ndebele W, et al. Loss to follow-up among children and adolescents growing up with HIV infection: age really matters. Journal of the International AIDS Society 2017 Jul 17;20(1):21737. [doi: 10.7448/ias.20.1.21737]
5. Jerene D, Abebe W, Taye K, Ruff A, Hallstrom I. Adolescents living with HIV are at higher risk of death and loss to follow up from care: Analysis of cohort data from eight health facilities in Ethiopia. PLoS One 2019 Oct 17;14(10):e0223655 [FREE Full text] [doi: 10.1371/journal.pone.0223655] [Medline: 31622391]
6. Fokam J, Sosso SM, Yagai B, Billong SC, Djubgang Mbadie RE, Kamgaing Simo R, et al. Viral suppression in adults, adolescents and children receiving antiretroviral therapy in Cameroon: adolescents at high risk of virological failure in the era of "test and treat". AIDS Res Ther 2019 Nov 19;16(1):36 [FREE Full text] [doi: 10.1186/s12981-019-0252-0] [Medline: 31744517]
7. Bezabhe WM, Chalmers L, Bereznicki LR, Peterson GM, Bimirew MA, Kassie DM. Barriers and facilitators of adherence to antiretroviral drug therapy and retention in care among adult HIV-positive patients: a qualitative study from Ethiopia. PLoS One 2014 May 14;9(5):e97353 [FREE Full text] [doi: 10.1371/journal.pone.0097353] [Medline: 24828585]
8. Haberer J, Mellins C. Pediatric adherence to HIV antiretroviral therapy. Curr HIV/AIDS Rep 2009 Nov;6(4):194-200 [FREE Full text] [doi: 10.1007/s11904-009-0026-8] [Medline: 19849962]
9. Hudelson C, Cluver L. Factors associated with adherence to antiretroviral therapy among adolescents living with HIV/AIDS in low- and middle-income countries: a systematic review. AIDS Care 2015;27(7):805-816. [doi: 10.1080/09540121.2015.1011073] [Medline: 25702789]
10. Ammon N, Mason S, Corkery J. Factors impacting antiretroviral therapy adherence among human immunodeficiency virus-positive adolescents in Sub-Saharan Africa: a systematic review. Public Health 2018 Apr;157:20-31. [doi: 10.1016/j.puhe.2017.12.010] [Medline: 29501984]
11. MacCarthy S, Saya U, Samba C, Birungi J, Okoboi S, Linnemayr S. "How am I going to live?": exploring barriers to ART adherence among adolescents and young adults living with HIV in Uganda. BMC Public Health 2018 Oct 04;18(1):1158 [FREE Full text] [doi: 10.1186/s12889-018-6048-7] [Medline: 30286746]
12. Gray WN, Schaefer MR, Resmini-Rawlinson A, Wagoner ST. Barriers to Transition From Pediatric to Adult Care: A Systematic Review. J Pediatr Psychol 2018 Jun 01;43(5):488-502. [doi: 10.1093/jpepsy/jsx142] [Medline: 29190360]
13. Christie D, Viner R. Adolescent development. BMJ 2005 Feb 05;330(7486):301-304 [FREE Full text] [doi: 10.1136/bmj.330.7486.301] [Medline: 15695279]
14. Vijayakumar N, Op de Macks Z, Shirtcliff EA, Pfeifer JH. Puberty and the human brain: Insights into adolescent development. Neurosci Biobehav Rev 2018 Sep;92:417-436 [FREE Full text] [doi: 10.1016/j.neubiorev.2018.06.004] [Medline: 29972766]
15. Arain M, Haque M, Johal L, Mathur P, Nel W, Rais A, et al. Maturation of the adolescent brain. Neuropsychiatr Dis Treat 2013;9:449-461 [FREE Full text] [doi: 10.2147/NDT.S39776] [Medline: 23579318]
16. Steinberg L, Morris AS. Adolescent development. Annu Rev Psychol 2001;52:83-110. [doi: 10.1146/annurev.psych.52.1.83] [Medline: 11148300]
17. Steinberg L. Cognitive and affective development in adolescence. Trends Cogn Sci 2005 Feb;9(2):69-74. [doi: 10.1016/j.tics.2004.12.005] [Medline: 15668099]
18. Steinberg L. Risk taking in adolescence: what changes, and why? Ann N Y Acad Sci 2004 Jun;1021:51-58. [doi: 10.1196/annals.1308.005] [Medline: 15251873]
19. Smetana JG, Campione-Barr N, Metzger A. Adolescent development in interpersonal and societal contexts. Annu Rev Psychol 2006;57:255-284. [doi: 10.1146/annurev.psych.57.102904.190124] [Medline: 16318596]
20. Taddeo D, Egedy M, Frappier J. Adherence to treatment in adolescents. Paediatr Child Health 2008 Jan;13(1):19-24 [FREE Full text] [doi: 10.1093/pch/13.1.19] [Medline: 19119348]
21. Lowenthal E, Matesva M, Marukutira T, Bayani O, Chapman J, Tshume O, et al. Psychological Reactance is a Novel Risk Factor for Adolescent Antiretroviral Treatment Failure. AIDS Behav 2021 May 04;25(5):1474-1479. [doi: 10.1007/s10461-020-02986-z] [Medline: 32754779]
22. Gommans R, Sandstrom MJ, Stevens GW, ter Bogt TF, Cillessen AH. Popularity, likeability, and peer conformity: Four field experiments. Journal of Experimental Social Psychology 2017 Nov;73:279-289. [doi: 10.1016/j.jesp.2017.10.001]
23. 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; 2017.
24. 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]

25. 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]
26. 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]
27. Muessig KE, LeGrand S, Horvath KJ, Bauermeister JA, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441 [FREE Full text] [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
28. Hightow-Weidman LB, Muessig KE, Bauermeister JA, LeGrand S, Fiellin LE. The future of digital games for HIV prevention and care. Curr Opin HIV AIDS 2017 Sep;12(5):501-507 [FREE Full text] [doi: 10.1097/COH.0000000000000399] [Medline: 28692490]
29. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth Interventions To Support Self-Management In HIV: A Systematic Review. Open AIDS J 2017;11:119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
30. Henny KD, Wilkes AL, McDonald CM, Denson DJ, Neumann MS. A Rapid Review of eHealth Interventions Addressing the Continuum of HIV Care (2007-2017). AIDS Behav 2018 Jan;22(1):43-63 [FREE Full text] [doi: 10.1007/s10461-017-1923-2] [Medline: 28983684]
31. Wang Z, Zhu Y, Cui L, Qu B. Electronic Health Interventions to Improve Adherence to Antiretroviral Therapy in People Living With HIV: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14404 [FREE Full text] [doi: 10.2196/14404] [Medline: 31621641]
32. Purnomo J, Coote K, Mao L, Fan L, Gold J, Ahmad R, et al. Using eHealth to engage and retain priority populations in the HIV treatment and care cascade in the Asia-Pacific region: a systematic review of literature. BMC Infect Dis 2018 Feb 17;18(1):82 [FREE Full text] [doi: 10.1186/s12879-018-2972-5] [Medline: 29454322]
33. Navarra AD, Gwadz MV, Whittemore R, Bakken SR, Cleland CM, Burleson W, et al. Health Technology-Enabled Interventions for Adherence Support and Retention in Care Among US HIV-Infected Adolescents and Young Adults: An Integrative Review. AIDS Behav 2017 Nov;21(11):3154-3171 [FREE Full text] [doi: 10.1007/s10461-017-1867-6] [Medline: 28776275]
34. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text Messaging and Mobile Phone Apps as Interventions to Improve Adherence in Adolescents With Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
35. Mulawa MI, LeGrand S, Hightow-Weidman LB. eHealth to Enhance Treatment Adherence Among Youth Living with HIV. Curr HIV/AIDS Rep 2018 Aug 29;15(4):336-349 [FREE Full text] [doi: 10.1007/s11904-018-0407-y] [Medline: 29959649]
36. Badawy SM, Kuhns LM. Economic Evaluation of Text-Messaging and Smartphone-Based Interventions to Improve Medication Adherence in Adolescents with Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
37. Ramsey W, Heidelberg R, Gilbert A, Heneghan M, Badawy S, Alberts N. eHealth and mHealth interventions in pediatric cancer: A systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37 [FREE Full text] [doi: 10.1002/pon.5280] [Medline: 31692183]
38. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
39. Badawy SM, Kuhns LM. Texting and Mobile Phone App Interventions for Improving Adherence to Preventive Behavior in Adolescents: A Systematic Review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
40. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
41. 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]
42. Vollmer Dahlke D, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015 Mar 27;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
43. Glanz K, Rimer BK, Viswanath KV. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA, US: Jossey-Bass; 2008.

44. Amico KR, Mugavero M, Krousel-Wood MA, Bosworth HB, Merlin JS. Advantages to Using Social-Behavioral Models of Medication Adherence in Research and Practice. J Gen Intern Med 2018 Feb 4;33(2):207-215 [FREE Full text] [doi: 10.1007/s11606-017-4197-5] [Medline: 29204968]
45. Rosenstock IM, Strecher VJ, Becker MH. Social Learning Theory and the Health Belief Model. Health Education Quarterly 2016 Sep 04;15(2):175-183. [doi: 10.1177/109019818801500203]
46. Fisher W, Fisher J, Harman J. The Information-Motivation-Behavioral Skills Model: A general social psychological approach to understanding and promoting health behavior. In: Social Psychological Foundations of Health and Illness. Massachusetts: Blackwell Malden; 2003:82-106.
47. 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]
48. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Upper Saddle River, NJ: Prentice-Hall; 1986.
49. Fishbein M. A theory of reasoned action: some applications and implications. Nebr Symp Motiv 1980;27:65-116. [Medline: 7242751]
50. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior. Berlin, Heidelberg: Springer; 1985:11-39.
51. Prochaska J, Velicer W. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
52. 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]
53. Cialdini RB. The 6 principles of persuasion. In: Influence: Science and Practice. Boston, MA: Pearson Education; 2009:e1.
54. Fogg BJ. Persuasive technology. Ubiquity 2002 Dec;2002(December):2. [doi: 10.1145/764008.763957]
55. Fogg B. Persuasive computers: perspectives and research directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Internet. New York, NY: ACM Press/Addison-Wesley Publishing Co; 1998 Presented at: CHI98: ACM Conference on Human Factors and Computing Systems; April 18-23, 1998; Los Angeles, CA p. 225-232.
56. 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]
57. Geuens J, Geurts L, Gerling K, Croon R, Abeele V. A Dyad of Lenses for the Motivational Design of mHealth: Bridging the Gap Between Health Theory and App Design. 2019 Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); June 10-13, 2019; Xi'an, China p. 1-12. [doi: 10.1109/ichi.2019.8904839]
58. Chou Y. Yu-kai Chou: Gamification & Behavioral Design. URL: https://yukaichou.com/gamification-examples/octalysis-complete-gamification-framework/ [accessed 2019-10-03]
59. Marczewski A. 52 Gamification Mechanics and Elements. 2017. URL: https://www.gamified.uk/user-types/gamification-mechanics-elements/ [accessed 2019-10-03]
60. Tondello G, Mora A, Nacke L. Elements of Gameful Design Emerging from User Preferences. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. New York, NY, United States: Association for Computing Machinery; 2017 Oct 15 Presented at: CHI PLAY '17: The annual symposium on Computer-Human Interaction in Play; October 15 - 18, 2017; Amsterdam The Netherlands p. 129-142. [doi: 10.1145/3116595.3116627]
61. Geuens J, Swinnen T, Geurts L, Westhovens R, Croon R, Vanden AV. Worriers versus Warriors:Tailoring mHealth to Address Differences in Patients with Chronic Arthritis. In: 2020 IEEE International Conference on Healthcare Informatics (ICHI). New York, NY: IEEE; 2020 Nov 30 Presented at: Eighth IEEE International Conference on Healthcare Informatics (ICHI 2020). ICHI 2020; November 30 to December 3, 2020; Oldenburg, Germany. [doi: 10.1109/ICHI48887.2020.9374322]
62. Geuens J, Swinnen TW, Westhovens R, de Vlam K, Geurts L, Vanden Abeele V. A Review of Persuasive Principles in Mobile Apps for Chronic Arthritis Patients: Opportunities for Improvement. JMIR Mhealth Uhealth 2016 Oct 13;4(4):e118 [FREE Full text] [doi: 10.2196/mhealth.6286] [Medline: 27742604]
63. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Matthew P, et al. Cochrane Handbook for Systematic Reviews of Interventions. URL: https://training.cochrane.org/handbook/current [accessed 2019-11-28]
64. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
65. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
66. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
67. Newbury-Birch D, Coulton S, Bland M, Cassidy P, Dale V, Deluca P, et al. Alcohol screening and brief interventions for offenders in the probation setting (SIPS Trial): a pragmatic multicentre cluster randomized controlled trial. Alcohol Alcohol 2014;49(5):540-548. [doi: 10.1093/alcalc/agu046] [Medline: 25063992]

68. Bobrow K, Brennan T, Springer D, Levitt NS, Rayner B, Namane M, et al. Efficacy of a text messaging (SMS) based intervention for adults with hypertension: protocol for the StAR (SMS Text-message Adherence suppoRt trial) randomised controlled trial. BMC Public Health 2014 Jan 11;14:28 [FREE Full text] [doi: 10.1186/1471-2458-14-28] [Medline: 24410738]
69. BCT Taxonomy Training. URL: https://www.bct-taxonomy.com/ [accessed 2021-01-28]
70. Fisher JD, Amico KR, Fisher WA, Harman JJ. The information-motivation-behavioral skills model of antiretroviral adherence and its applications. Curr HIV/AIDS Rep 2008 Nov 25;5(4):193-203. [doi: 10.1007/s11904-008-0028-y] [Medline: 18838059]
71. Movahed E, Morowatisharifabad MA, Farokhzadian J, Nikooie R, Hosseinzadeh M, Askarishahi M, et al. Antiretroviral Therapy Adherence Among People Living With HIV: Directed Content Analysis Based on Information-Motivation-Behavioral Skills Model. Int Q Community Health Educ 2019 Jun 19;40(1):47-56. [doi: 10.1177/0272684x19858029]
72. Peng Z, Yu Y, Wei W, Hou Y, Sun Z, Wang Y, et al. Reliability and Validity of the LifeWindows Information-Motivation-Behavioral Skills Antiretroviral Therapy Adherence Questionnaire Among HIV+ Patients in Shanghai. Patient Prefer Adherence 2020;14:507-515 [FREE Full text] [doi: 10.2147/PPA.S234041] [Medline: 32184577]
73. Chang SJ, Choi S, Kim S, Song M. Intervention Strategies Based on Information-Motivation-Behavioral Skills Model for Health Behavior Change: A Systematic Review. Asian Nursing Research 2014 Sep;8(3):172-181. [doi: 10.1016/j.anr.2014.08.002]
74. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
75. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Computers & Education 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
76. Boyle EA, Hainey T, Connolly TM, Gray G, Earp J, Ott M, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education 2016 Mar;94:178-192. [doi: 10.1016/j.compedu.2015.11.003]
77. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A Mobile Gaming Intervention to Increase Adherence to Antiretroviral Treatment for Youth Living With HIV: Development Guided by the Information, Motivation, and Behavioral Skills Model. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e96 [FREE Full text] [doi: 10.2196/mhealth.8155] [Medline: 29685863]
78. Whiteley L, Brown LK, Mena L, Craker L, Arnold T. Enhancing health among youth living with HIV using an iPhone game. AIDS Care 2018 Jan 09;30(sup4):21-33 [FREE Full text] [doi: 10.1080/09540121.2018.1503224] [Medline: 30626196]
79. Tanner AE, Song EY, Mann-Jackson L, Alonzo J, Schafer K, Ware S, et al. Preliminary Impact of the weCare Social Media Intervention to Support Health for Young Men Who Have Sex with Men and Transgender Women with HIV. AIDS Patient Care STDS 2018 Nov;32(11):450-458 [FREE Full text] [doi: 10.1089/apc.2018.0060] [Medline: 30398955]
80. Tanner AE, Mann L, Song E, Alonzo J, Schafer K, Arellano E, et al. weCARE: A Social Media-Based Intervention Designed to Increase HIV Care Linkage, Retention, and Health Outcomes for Racially and Ethnically Diverse Young MSM. AIDS Educ Prev 2016 Jun;28(3):216-230 [FREE Full text] [doi: 10.1521/aeap.2016.28.3.216] [Medline: 27244190]
81. Stankievich E, Malanca A, Foradori I, Ivalo S, Losso M. Utility of Mobile Communication Devices as a Tool to Improve Adherence to Antiretroviral Treatment in HIV-infected Children and Young Adults in Argentina. Pediatr Infect Dis J 2018 Apr;37(4):345-348. [doi: 10.1097/INF.0000000000001807] [Medline: 29023355]
82. Spratt ES, Papa CE, Mueller M, Patel S, Killeen T, Maher E, et al. Using Technology to Improve Adherence to HIV Medications in Transitional Age Youth: Research Reviewed, Methods Tried, Lessons Learned. J Gen Med (Dover) 2017;1(1):1002 [FREE Full text] [Medline: 30345429]
83. Shegog R, Markham CM, Leonard AD, Bui TC, Paul ME. "+CLICK": pilot of a web-based training program to enhance ART adherence among HIV-positive youth. AIDS Care 2011 Sep 08;24(3):310-318. [doi: 10.1080/09540121.2011.608788] [Medline: 21899488]
84. Belzer ME, Naar-King S, Olson J, Sarr M, Thornton S, Kahana SY, Adolescent Medicine Trials Network for HIV/AIDS Interventions. The use of cell phone support for non-adherent HIV-infected youth and young adults: an initial randomized and controlled intervention trial. AIDS Behav 2014 Apr;18(4):686-696 [FREE Full text] [doi: 10.1007/s10461-013-0661-3] [Medline: 24271347]
85. Belzer ME, Kolmodin MacDonell K, Clark LF, Huang J, Olson J, Kahana SY, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Acceptability and Feasibility of a Cell Phone Support Intervention for Youth Living with HIV with Nonadherence to Antiretroviral Therapy. AIDS Patient Care STDS 2015 Jun;29(6):338-345 [FREE Full text] [doi: 10.1089/apc.2014.0282] [Medline: 25928772]
86. Sayegh CS, MacDonell KK, Clark LF, Dowshen NL, Naar S, Olson-Kennedy J, et al. The Impact of Cell Phone Support on Psychosocial Outcomes for Youth Living with HIV Nonadherent to Antiretroviral Therapy. AIDS Behav 2018 Oct 12;22(10):3357-3362 [FREE Full text] [doi: 10.1007/s10461-018-2192-4] [Medline: 29948339]
87. Saberi P, Yuan P, John M, Sheon N, Johnson MO. A pilot study to engage and counsel HIV-positive African American youth via telehealth technology. AIDS Patient Care STDS 2013 Sep;27(9):529-532 [FREE Full text] [doi: 10.1089/apc.2013.0185] [Medline: 23991691]

88. 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]
89. Rana Y, Haberer J, Huang H, Kambugu A, Mukasa B, Thirumurthy H, et al. Short message service (SMS)-based intervention to improve treatment adherence among HIV-positive youth in Uganda: focus group findings. PLoS One 2015;10(4):e0125187 [FREE Full text] [doi: 10.1371/journal.pone.0125187] [Medline: 25881059]
90. Puccio JA, Belzer M, Olson J, Martinez M, Salata C, Tucker D, et al. The use of cell phone reminder calls for assisting HIV-infected adolescents and young adults to adhere to highly active antiretroviral therapy: a pilot study. AIDS Patient Care STDS 2006 Jun;20(6):438-444. [doi: 10.1089/apc.2006.20.438] [Medline: 16789857]
91. Naar-King S, Outlaw AY, Sarr M, Parsons JT, Belzer M, Macdonell K, Adolescent Medicine Network for HIV/AIDS Interventions. Motivational Enhancement System for Adherence (MESA): pilot randomized trial of a brief computer-delivered prevention intervention for youth initiating antiretroviral treatment. J Pediatr Psychol 2013 Jul;38(6):638-648 [FREE Full text] [doi: 10.1093/jpepsy/jss132] [Medline: 23359664]
92. Outlaw AY, Naar-King S, Tanney M, Belzer ME, Aagenes A, Parsons JT, Adolescent Medicine Trials Network for HIV/AIDS Interventions. The initial feasibility of a computer-based motivational intervention for adherence for youth newly recommended to start antiretroviral treatment. AIDS Care 2014 Jan;26(1):130-135 [FREE Full text] [doi: 10.1080/09540121.2013.813624] [Medline: 23869650]
93. Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, et al. A Gamified Smartphone App to Support Engagement in Care and Medication Adherence for HIV-Positive Young Men Who Have Sex With Men (AllyQuest): Development and Pilot Study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34 [FREE Full text] [doi: 10.2196/publichealth.8923] [Medline: 29712626]
94. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: a pilot study using personalized, interactive, daily text message reminders. J Med Internet Res 2012 Apr 05;14(2):e51 [FREE Full text] [doi: 10.2196/jmir.2015] [Medline: 22481246]
95. Dowshen N, Kuhns LM, Gray C, Lee S, Garofalo R. Feasibility of interactive text message response (ITR) as a novel, real-time measure of adherence to antiretroviral therapy for HIV+ youth. AIDS Behav 2013 Jul;17(6):2237-2243. [doi: 10.1007/s10461-013-0464-6] [Medline: 23546844]
96. Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A Randomized Controlled Trial of Personalized Text Message Reminders to Promote Medication Adherence Among HIV-Positive Adolescents and Young Adults. AIDS Behav 2016 May;20(5):1049-1059 [FREE Full text] [doi: 10.1007/s10461-015-1192-x] [Medline: 26362167]
97. 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]
98. Dworkin M, Chakraborty A, Lee S, Monahan C, Hightow-Weidman L, Garofalo R, et al. A Realistic Talking Human Embodied Agent Mobile Phone Intervention to Promote HIV Medication Adherence and Retention in Care in Young HIV-Positive African American Men Who Have Sex With Men: Qualitative Study. JMIR Mhealth Uhealth 2018 Jul 31;6(7):e10211 [FREE Full text] [doi: 10.2196/10211] [Medline: 30064971]
99. Lee S, Lu Y, Chakraborty A, Dworkin M. Development of a mobile personal health guide for HIV-infected African American MSM. 2016 Oct 19 Presented at: International Conference on Intelligent Virtual Agents; September 20–23, 2016; Los Angeles, CA p. 247-450. [doi: 10.1007/978-3-319-47665-0_52]
100. Dulli L, Ridgeway K, Packer C, Murray KR, Mumuni T, Plourde KF, et al. A Social Media-Based Support Group for Youth Living With HIV in Nigeria (SMART Connections): Randomized Controlled Trial. J Med Internet Res 2020 Jun 02;22(6):e18343 [FREE Full text] [doi: 10.2196/18343] [Medline: 32484444]
101. Abiodun O, Ladi-Akinyemi B, Olu-Abiodun O, Sotunsa J, Bamidele F, Adepoju A, et al. A Single-Blind, Parallel Design RCT to Assess the Effectiveness of SMS Reminders in Improving ART Adherence Among Adolescents Living with HIV (STARTA Trial). J Adolesc Health 2021 Apr;68(4):728-736. [doi: 10.1016/j.jadohealth.2020.11.016] [Medline: 33342719]
102. Michielsen K, Chersich M, Temmerman M, Dooms T, Van Rossem R. Nothing as Practical as a Good Theory? The Theoretical Basis of HIV Prevention Interventions for Young People in Sub-Saharan Africa: A Systematic Review. AIDS Res Treat 2012;2012:345327-345318 [FREE Full text] [doi: 10.1155/2012/345327] [Medline: 22900155]
103. Simoni JM, Ronen K, Aunon FM. Health Behavior Theory to Enhance eHealth Intervention Research in HIV: Rationale and Review. Curr HIV/AIDS Rep 2018 Dec 3;15(6):423-430 [FREE Full text] [doi: 10.1007/s11904-018-0418-8] [Medline: 30511186]
104. Munro S, Lewin S, Swart T, Volmink J. A review of health behaviour theories: how useful are these for developing interventions to promote long-term medication adherence for TB and HIV/AIDS? BMC Public Health 2007 Jun 11;7(1):104 [FREE Full text] [doi: 10.1186/1471-2458-7-104] [Medline: 17561997]
105. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. The Lancet 2016 Jun;387(10036):2423-2478. [doi: 10.1016/s0140-6736(16)00579-1]

106. Leany B. Brain development and health implications in adolescents. In: Handbook of Adolescent Health Psychology. New York, NY: Springer; Mar 26, 2013:235-244.
107. Linders A. Deconstructing adolescence. In: International Handbook on Adolescent Health and Development. Berlin, Germany: Springer International Publishing; Nov 23, 2016:15-28.
108. Berenbaum S, Beltz A, Corley R. The importance of puberty for adolescent development: conceptualization and measurement. Adv Child Dev Behav 2015 Jan 1;48:53-92. [doi: 10.1016/bs.acdb.2014.11.002] [Medline: 25735941]
109. Bell B. Understanding adolescents. In: Perspectives on HCI Research with Teenagers. Cham, Switzerland: Springer International Publishing; 2016:11-27.
110. O' Donohue W, Benuto L, Tolle L. Handbook of Adolescent Health Psychology. New York, NY: Springer-Verlag; 2013.
111. Fitton D, Bell B. Working with teenagers within HCI research: Understanding teen-computer interaction. 2014 Presented at: 28th International BCS Human Computer Interaction Conference on HCI 2014; September 9-12, 2014; Southport UK. [doi: 10.14236/ewic/hci2014.31]
112. Blake Helms C, Turan JM, Atkins G, Kempf M, Clay OJ, Raper JL, et al. Interpersonal Mechanisms Contributing to the Association Between HIV-Related Internalized Stigma and Medication Adherence. AIDS Behav 2017 Jan 10;21(1):238-247 [FREE Full text] [doi: 10.1007/s10461-016-1320-2] [Medline: 26864692]
113. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills EJ, Smieja M, et al. The Cameroon Mobile Phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. PLoS One 2012 Dec 6;7(12):e46909 [FREE Full text] [doi: 10.1371/journal.pone.0046909] [Medline: 23236345]
114. Sherman EM, Niu J, Elrod S, Clauson KA, Alkhateeb F, Eckardt P. Effect of mobile text messages on antiretroviral medication adherence and patient retention in early HIV care: an open-label, randomized, single center study in south Florida. AIDS Res Ther 2020 May 13;17(1):16 [FREE Full text] [doi: 10.1186/s12981-020-00275-2] [Medline: 32404130]
115. Kok G, Schaalma H, Ruiter RAC, van Empelen P, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. J Health Psychol 2004 Jan;9(1):85-98. [doi: 10.1177/1359105304038379] [Medline: 14683571]
116. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
117. Heneghan M, Hussain T, Barrera L, Cai S, Haugen M, Duff A, et al. Applying the COM-B model to patient-reported barriers to medication adherence in pediatric acute lymphoblastic leukemia. Pediatric Blood & Cancer 2020:e28216 [FREE Full text] [doi: 10.1002/pbc.28216]
118. Fitton D, Little L, Bell B. Introduction: HCI Reaches Adolescence. In: Perspectives on HCI Research with Teenagers. Cham, Switzerland: Springer International Publishing; 2016:1-9.
119. Fitton D, Read J, Horton M. The challenge of working with teens as participants in interaction design. New York, NY: Association for Computing Machinery; 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France. [doi: 10.1145/2468356.2468394]
120. Badawy S, Thompson A, Liem R. Technology Access and Smartphone App Preferences for Medication Adherence in Adolescents and Young Adults With Sickle Cell Disease. Pediatr Blood Cancer 2016 May;63(5):848-852 [FREE Full text] [doi: 10.1002/pbc.25905] [Medline: 26844685]
121. Montaño D, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Health Behavior: Theory, Research, and Practice (5th edition). Hoboken, NJ: Jossey-Bass; 2015:95-124.
122. Bassett IV, Wilson D, Taaffe J, Freedberg KA. Financial incentives to improve progression through the HIV treatment cascade. Curr Opin HIV AIDS 2015 Nov;10(6):451-463 [FREE Full text] [doi: 10.1097/COH.0000000000000196] [Medline: 26371461]
123. Galárraga O, Enimil A, Bosomtwe D, Cao W, Barker DH. Group-based economic incentives to improve adherence to antiretroviral therapy among youth living with HIV: safety and preliminary efficacy from a pilot trial. Vulnerable Child Youth Stud 2020;15(3):257-268 [FREE Full text] [doi: 10.1080/17450128.2019.1709678] [Medline: 33281920]
124. Galárraga O, Kuo C, Mtukushe B, Maughan-Brown B, Harrison A, Hoare J. iSAY (incentives for South African youth): Stated preferences of young people living with HIV. Soc Sci Med 2020 Nov;265:113333 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113333] [Medline: 32896799]
125. El-Sadr WM, Donnell D, Beauchamp G, Hall HI, Torian LV, Zingman B, HPTN 065 Study Team. Financial Incentives for Linkage to Care and Viral Suppression Among HIV-Positive Patients: A Randomized Clinical Trial (HPTN 065). JAMA Intern Med 2017 Aug 01;177(8):1083-1092 [FREE Full text] [doi: 10.1001/jamainternmed.2017.2158] [Medline: 28628702]
126. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and Adherence to Web-Based Mental Health Interventions: A Systematic Review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
127. De CR, Geuens J, Verbert K, Abeele V. A systematic review of the effect of gamification on adherence across disciplines. 2021 Presented at: International Conference on Human-Computer Interaction; July 24–29, 2021; Virtual Event. [doi: 10.1007/978-3-030-77277-2_14]
128. King D, Greaves F, Exeter C, Darzi A. 'Gamification': Influencing health behaviours with games. Journal of the Royal Society of Medicine 2013 Mar 12;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]

129. 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]
130. Hervas R, Ruiz-Carrasco D, Mondejar T, Bravo J. Gamification mechanics for behavioral change: a systematic review and proposed taxonomy. New York, NY: Association for Computing Machinery; 2017 Presented at: PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154939]
131. 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]
132. 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 [FREE Full text] [doi: 10.1177/1460458214537511] [Medline: 24986104]
133. Peppet S. Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future. 2011 Jan 1. URL: https://scholar.law.colorado.edu/articles/177 [accessed 2021-11-10]
134. Warner M, Gutmann A, Sasse MA, Blandford A. Privacy Unraveling Around Explicit HIV Status Disclosure Fields in the Online Geosocial Hookup App Grindr. Proc. ACM Hum.-Comput. Interact 2018 Nov 01;2(CSCW):1-22. [doi: 10.1145/3274450]

Abbreviations

ART: antiretroviral therapy
BCT: behavior change technique
TCS: Theory Coding Scheme

Edited by L Buis; submitted 19.10.20; peer-reviewed by S Badawy, L Kuhns; comments to author 12.12.20; revised version received 06.02.21; accepted 17.09.21; published 10.12.21.

Please cite as:
Bezabih AM, Gerling K, Abebe W, Abeele VV
Behavioral Theories and Motivational Features Underlying eHealth Interventions for Adolescent Antiretroviral Adherence: Systematic Review
JMIR Mhealth Uhealth 2021;9(12):e25129
URL: https://mhealth.jmir.org/2021/12/e25129
doi:10.2196/25129
PMID:34890353

Review

Remote Monitoring Systems for Patients With Chronic Diseases in Primary Health Care: Systematic Review

Mariana Peyroteo[1,2], BSc, MSc; Inês Augusto Ferreira[3,4], BSc; Luís Brito Elvas[2,3], BSc; João Carlos Ferreira[2,3], PhD; Luís Velez Lapão[5,6], PhD

[1]NOVA School of Science and Technology, NOVA University of Lisbon, Setúbal, Portugal

[2]Inov Inesc Inovação, Instituto de Novas Tecnologias, Lisbon, Portugal

[3]Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisbon, Portugal

[4]School of Biology, St Leonard's Postgraduate College, The University of St Andrews, St Andrews, United Kingdom

[5]Unidade de Investigação e Desenvolvimento em Engenharia Mecanica e Industrial, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Setúbal, Portugal

[6]Comprehensive Health Research Center, NOVA Medical School, NOVA University of Lisbon, Lisboa, Portugal

Corresponding Author:
Luís Velez Lapão, PhD
Unidade de Investigação e Desenvolvimento em Engenharia Mecanica e Industrial
Department of Mechanical and Industrial Engineering
NOVA School of Science and Technology
Faculdade de Ciências e Tecnologia
Campus da Caparica
Setúbal, 2829-516
Portugal
Phone: 351 212 948 300
Email: l.lapao@fct.unl.pt

Abstract

Background: The digital age, with digital sensors, the Internet of Things (IoT), and big data tools, has opened new opportunities for improving the delivery of health care services, with remote monitoring systems playing a crucial role and improving access to patients. The versatility of these systems has been demonstrated during the current COVID-19 pandemic. Health remote monitoring systems (HRMS) present various advantages such as the reduction in patient load at hospitals and health centers. Patients that would most benefit from HRMS are those with chronic diseases, older adults, and patients that experience less severe symptoms recovering from SARS-CoV-2 viral infection.

Objective: This paper aimed to perform a systematic review of the literature of HRMS in primary health care (PHC) settings, identifying the current status of the digitalization of health processes, remote data acquisition, and interactions between health care personnel and patients.

Methods: A systematic literature review was conducted using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines to identify articles that explored interventions with HRMS in patients with chronic diseases in the PHC setting.

Results: The literature review yielded 123 publications, 18 of which met the predefined inclusion criteria. The selected articles highlighted that sensors and wearables are already being used in multiple scenarios related to chronic disease management at the PHC level. The studies focused mostly on patients with diabetes (9/26, 35%) and cardiovascular diseases (7/26, 27%). During the evaluation of the implementation of these interventions, the major difficulty that stood out was the integration of information into already existing systems in the PHC infrastructure and in changing working processes of PHC professionals (83%).

Conclusions: The PHC context integrates multidisciplinary teams and patients with often complex, chronic pathologies. Despite the theoretical framework, objective identification of problems, and involvement of stakeholders in the design and implementation processes, these interventions mostly fail to scale up. Despite the inherent limitations of conducting a systematic literature review, the small number of studies in the PHC context is a relevant limitation. This study aimed to demonstrate the importance of matching technological development to the working PHC processes in interventions regarding the use of sensors and wearables

for remote monitoring as a source of information for chronic disease management, so that information with clinical value is not lost along the way.

(JMIR Mhealth Uhealth 2021;9(12):e28285) doi:10.2196/28285

KEYWORDS

sensors; wearables; remote monitoring; digital health; primary health care; chronic diseases

Introduction

Background

Digitalization of care processes, holistic sensing supported by the Internet of Things (IoT), and artificial intelligence (AI) tools are being actively applied with benefits to the health sector and giving rise to the smart health paradigm [1]**.** This corresponds to an emerging market that was evaluated at US $143.6 billion in 2019, with an estimated annual growth rate of 16.2% from 2020 to 2027 [2]. In this transformative process, health remote monitoring systems (HRMS) are recognized as an emerging technology that use sensors and wearable devices to collect patient data. However, to provide clinical value, these systems have to be associated with clinical processes and therapeutics so that measurements can be linked with actual patient care.

The use of sensors in health is a recent but growing area of research. These sensors are usually called biosensors, as they often collect patients' vital signs. Wearable devices are devices that patients use in direct contact with the body to provide clinically relevant data for care. This continuous monitoring process defines personalized care. An IoT sensor can be used in either a discrete manner or a continuous manner [3]. A Scopus search showed that there were 96,888 publications regarding biosensors just in 2020.

Wearable devices are equipped with sensing capabilities for user mobility tracking and monitoring physical activity (for example, counting steps), heart rate, oxygen levels, and blood pressure. For instance, a wearable electrocardiogram monitor has been used by many patients with serious conditions under the prescription of their physicians [4]**.** Another example of a wearable device used by patients is a wearable blood pressure monitor, which can be integrated into a watch [5]. There is, however, the need for proper evaluation or certification to regulate the efficacy of these wearables, as some frauds have already been identified. The Food and Drug Association in the United States has established a wearables certification [6]**.** Lastly, there are also patch biosensors, which are self-adhesive patches that can collect a range of different data such as heart rate, respiratory rate, temperature, and body posture and can detect falls.

The implementation of HRMS is based on an IoT system that incorporates, stores, and communicates the information gathered by a set of wearable devices and sensors. The computer senses and records the daily physiological data of the patient by means of a data processing device, data transition, data archive, data analytics, and AI [7]. An HRMS is based on 4 main pillars: (1) development of systems to identify disease progression and prevention through remote sensors; (2) use of big data (BD) processing and analysis, which processes multiple heterogeneous data sources to integrate different patient data, aiming at providing high-quality personalized treatment; (3) development of predictive models supported by AI to be implemented on top of the processed BD, allowing the classification of patients and the discovery of behavioral patterns to enable alerts to be generated when an abnormality is registered for quicker clinical action; and (4) create an entirely remote interaction process from the hospital to the patient. The HRMS enables a data-intensive approach, in which a large amount of health data is generated, stored, and available for data mining, allowing the generation of useful knowledge. All data processed from IoT sensors and wearable devices can be input into BD analytics, which allows the generation of knowledge and alerts that can be used to monitor health [8].

The use of HRMS to record data supports the development of personal health records, involving patients in their own data collection, health monitoring, exercise, and lifestyle [9]. A single health database allows personalized care, as the health care professional can tailor the treatment according to both the patient conditions and the device readings. The term personalized care refers to the design and adaptation of clinical treatment to the characteristics, needs, and individual preferences of the patient throughout all stages: care, prevention, diagnosis, treatment, and follow-up [10]. The use of this term has been growing in recent years, as more recent technology such as genome sequencing, wearable devices, and HRMS has allowed the use of precision medicine [11]. The area of personalized medicine has rapidly grown over the last decade due to improvements in areas such as diagnostic testing, BD technologies, and others. The European Union is highly concerned about the provision of high-quality personalized medicine [12].

Communication of devices with cloud platforms allows for data to be stored in the cloud and easily accessed by doctors, allowing remote health monitoring functions. Remote patient monitoring requires machinery that collects and interprets biometric and physiological data [13]. Remote patient monitoring has many applications such as real-time illness detection, continuous monitoring of patients such as those with chronic disease or less severe conditions, or monitoring of athletes' health [13]. Recent guidelines have pointed out that the creation and adoption of person-centered integrated care for older adults is critical, as a decline in intrinsic capacity is reported among older adults. Intrinsic capacity includes the mental and physical attributes that the patient can use to perform any daily task, can be used to identify those patients who would benefit more from interventions, and can be measured by wearable devices [14].

Chronic Diseases in the Primary Health Care Context

Chronic diseases are among the most important health problems to benefit from HRMS. The role of primary health care (PHC)

centers as the first point of contact is considered a universal health coverage model. For that reason, health care data management and remote monitoring benefit from being included in PHC, enabling a comprehensive collection of data from patients and making the data available to health professionals afterwards.

Not only has research into wearable medical devices increased but also the availability of these devices to the general public. Devices such as mobile phones and smart bands, usage of which is becoming customary, are increasing the amount of data generated that can be used to improve the management of chronic diseases. More than 100,000 apps have been created to use this type of data, and the number has been doubling every 2.5 years [15]. However, often these applications cannot be integrated into the health care process, resulting in dispersed data.

PHC is suffering from larger demand from an increase in the number of patients with chronic diseases. Due to hospital overload, the use of remote monitoring systems is often considered of added value. Systematic monitoring would also allow for remote tracking of symptom progression in less severe COVID-19 patients, allowing one to closely monitor and increase the comfort of patients, but also reduce the strain on the health system. An HRMS is fundamental for supporting the collection of data necessary to improve the management of novel conditions, such as COVID-19, as it allows the collection of data useful in medical research and to identify patterns of symptoms that could indicate patients' symptom progression [16].

Furthermore, remote monitoring systems might improve health professionals' effectiveness in managing chronic diseases, as an HRMS permits the early detection of disease warning signs, which is crucial to improve survival rates of specific diseases (eg, hypertension, diabetes, chronic obstructive pulmonary disease). These systems promise to, by tracking patients' disease progression, increase patients' awareness of and engagement with chronic therapeutics.

Goal of the Study

Although several other studies are being conducted on this topic, it is paramount to analyze published studies (beyond clinical trials or pilot studies) to understand the weaknesses and opportunities that still persist in this area

This paper aimed at performing a systematic review of the literature on HRMS in the PHC setting, identifying the current status of the digitalization of the health process, regarding (1) digital monitoring of chronically ill patients, (2) early detection of acute episodes in patients with decompensated chronic pathologies, (3) the outcomes of the implementation process, and (4) patient empowerment.

As secondary outcomes, the following was also assessed: (1) the digital communication between PHC professionals and patients or caregivers, (2) the integration of the information collected by the health care information systems, (3) reduction of hospital burden, and (4) user satisfaction (patient, caregivers, and professionals).

Methods

Search Strategy and Inclusion Criteria

A systematic literature review was conducted by following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) methodology [17] and with the research question: "What is the state of the art on healthcare remote monitoring system usage for chronic patients in primary health care?"

We searched the databases of Scopus and Web of Science Core Collection (WoSCC), and the research was conducted through December 31, 2020. The results had to be articles, published between 2015 and 2020, and written in English or Portuguese. The documents collected were only about computer science, medicine, engineering, and health professions.

The search strategy was based on 6 queries, each with a different focus of research (detailed in Multimedia Appendix 1). This method allowed for the observation of the number of articles existing in both databases, considering the concept and context as well as the population under study. It is important to note that the values corresponding to the queries still have duplicate articles.

For this review, only articles were considered. Grey literature, reviews, conference papers, workshops, books, and editorials, as well as works not related to the domain, were excluded. The population included all ages, genders, and ethnic groups diagnosed with multimorbidity or at least one chronic disease. The study was considered eligible for inclusion if the intervention included one of these criteria: (1) continuous electronic recording of patient indicators (sensors or wearables) connected to a computer system integrated into PHC centers, (2) patient input devices linked to a computer system allowing the display of data in real time for analysis by PHC professionals, and (3) collection of personal electronic health or clinical data transmitted for review by a remote PHC professional.

Study Selection

The initial selection of papers was done using the title and abstract, and in some cases in which that information was insufficient, the full document was analyzed. The process was performed by 3 researchers independently: 2 performed the process, and in case of disagreement, the third resolved the disagreement.

Data Extraction and Synthesis

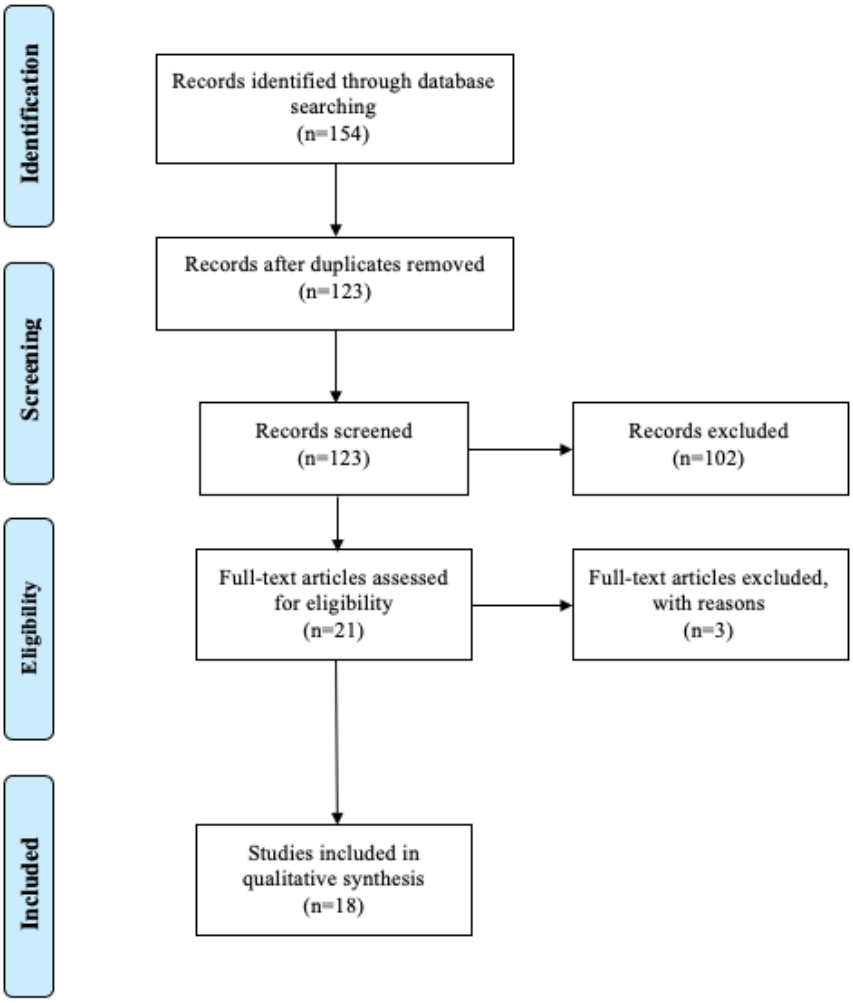
The data were managed and stored by Zotero and Microsoft Excel version 16.46 (21021202). These data were title, author, year, journal, subject area, keywords, and abstract. For data synthesis and analysis, a qualitative assessment was conducted based on the results presented in the previous section. The databases—Scopus and WoSCC—were searched systematically regarding the published work related with the concept "Healthcare Remote Monitoring Systems" or "Smart Health," with the target population "Chronic Patients," and within the context of the study "Primary Health Care."

Results

Queries and Themes

The research was conducted using queries and themes. Each query was conducted in the individual databases and with the same restrictions and filters. Figure 1 shows the PRISMA workflow diagram from the total number of articles studied.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) workflow diagram.



With query 1, Scopus and WoSCC were searched for literature regarding the concept of this study, and we found 61,230 results.

With query 2, the search included the target population of patients with a chronic disease, and 61,469 documents were found in both databases.

With query 3, which involved the context of this study, 73,550 documents were found in Scopus and WoSCC.

With the fourth query, the databases were searched for the concept of the study and the target population but without any context, and 1458 documents were found.

With query 5, queries 1, 2, and 3 were combined, and the documents collected involved all the aforementioned inclusion criteria. The merging of all the queries resulted in 154 documents. After performing a manual process to identify significant subjects based on the research questions, identify the outcomes, and remove the duplicates, 18 documents were obtained. Our systematic research took into account year, area, research question topic, and a short description. We also looked at articles published in 2020 for the COVID-19 pandemic effect.

Regarding these results, additional research was conducted (query 6) with which the purpose was to compare the results in both databases for HRMS used with patients with a chronic disease but in another specific context (in hospital care). In the search, query 1 and query 2 were combined, and hospital care was added (“Hospital” OR “Acute Care” OR “Clinical”), resulting in 726 results.

Study Characteristics

All 18 studies included in the review were selected through the use of the aforementioned specific criteria. Table 1 shows the key study characteristics regarding the year, region, disease of focus, interface, data collection methods, collection frequency, stakeholders’ involvement, and existence of pilot studies. Classification of the studies regarding these characteristics was not mutually exclusive, given that these were assigned due to presence or absence in the study.

Table 1. Study characteristics (n=18).

Characteristics	Articles, n (%)
Region	
Europe	8 (44)
Americas	7 (39)
Southeast Asia	1 (6)
Africa	1 (6)
Western Pacific	1 (6)
Disease of focus (n=26)	
Diabetes	9 (35)
Cardiovascular diseases	7 (27)
Respiratory diseases	4 (15)
Multimorbidity	4 (15)
Mental disorders	2 (8)
Interface (n=20)	
Mobile phone or telephone	10 (50)
Tablet	6 (30)
Web-based platform	4 (20)
Data collection method (n=19)	
Sensors	9 (47)
Questionnaires	5 (26)
Wearables	5 (26)
Collection frequency (n=14)	
Daily	6 (43)
Monthly	4 (29)
Weekly	3 (21)
Permanent	1 (7)
All stakeholders involved	
Yes	9 (83)
No	6 (33)
Pilot study	
Yes	8 (44)
Being developed	3 (17)
No	7 (39)

Outcomes Analysis

The results previously defined in the goal of this review are summarized in Table 2. The description of the indicators was explicit, and no requests to the authors of the articles for clarification were necessary. As mentioned before, classification of the studies regarding the outcome was not mutually exclusive, given that these were attributed due to presence or absence in the study.

Table 2. Outcome comparison.

Outcomes	References	Number of documents
Primary outcomes		
Digital monitoring of the patients' chronic diseases	[18-35]	18
Early detection of acute episodes	[18-24,27-29,31-35]	15
Outcomes of the implementation studies: effectiveness and cost-effectiveness	[18,19,22,28,30,33,34]	7
Outcomes of the implementation studies: implementation process with multidisciplinary teams	[18-21,23,24,27-31,33]	12
Patient empowerment (self-management applications)	[18,19]	2
Secondary outcomes		
Digital communication between PHC[a] professionals and patients	[18-32,34]	16
Integration of information into PHC centers	[18,20,26]	3
Reduction of hospital inflow	[18-21,23-29,31,33-35]	16
User satisfaction (patients, carers, and professionals); role of informal carers, especially to facilitate the use of technology by older adult patients	[19,21,28]	3

[a]PHC: primary health care.

Risk of Bias in Included Studies

Given the categorization of articles included in the study, there may be bias in the definition of "Remote Patient Monitoring." Thus, it is possible that some articles may have been excluded.

Discussion

Principal Findings

After assessing all the included studies, it was possible to acknowledge the growing prevalence of remote monitoring systems worldwide in recent years. Although the main goals across the different studies varied, the majority of the articles included common features for these devices. The interventions included the following features: medical condition management (n=3), diagnosis (n=1), conceptual models (n=6), reminders and alerts (n=2), self-reported monitoring (n=5), wearable remote monitoring device (n=4), health promotion and education (n=4).

Most of the articles selected included research about conceptual models regarding work methodologies or efficacy of the devices or the interventions. Of the articles, 24% (4/18), which had a specific focus on wearable remote monitoring systems for patients with a chronic disease in PHC, were studies or proposals for conceptual care models. The topics in these studies varied from changes in chronic disease management through the use of biosensors to the impact assessment of the implementation of telecare projects. From these studies, approximately 75% (3/4) were theoretical models that focused on the understanding of barriers and challenges to digital transformation, but no context-related implementation evidence was presented to support the conclusions drawn.

The research using different queries with different keywords allowed the comparison of the state of the art regarding the concept in the chronic care context. Therefore, it was possible to observe the difference in HRMS applications for patients with a chronic disease carried out in both the hospital and PHC contexts. Given that chronic disease management is a major focus of PHC services [36], it would be expected that most studies would be applied in that setting.

However, by observing and evaluating the number of articles resulting from the research between contexts (see query 5 vs query 6), it was possible to verify that the influx of studies in the hospital context (n=726) was 7 times higher than studies in the PHC context (n=100). This difference may be due to the recent restructuring (since 2014) of health service contexts or to the difficulty in implementing studies in the PHC context [37]. This can also be explained by the fact that there are many clinical conditions that are still managed from the hospital point of view that focus on symptomatic treatment.

According to World Health Statistics 2020, it was estimated that 71% of deaths worldwide were caused by noncommunicable diseases (cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes) [38]. In this literature review, it was possible to observe that the use of smart health is mostly focused on noncommunicable diseases that not only cause most deaths but also can be better controlled and prevented by controlling risk factors and monitoring of the patient's health [39]—see the "Disease of focus" information in Table 1.

Remote Monitoring and the COVID-19 Pandemic

With the COVID-19 pandemic, the role and importance of health digitalization had a special focus for the measures implemented to reduce the risk of transmission, not only because of the extreme need for adaptation of the entire population to the situation and the social and economic impact that this pandemic brought but also because of the pressure and exhaustion of health professionals (eg, burnout).

It was possible to identify 4 articles [21,25,33,34] in which the focus and role of health digitalization, through telecare, were seen as essential and urgent for the current or future pandemics and crisis situation management. The small number of articles found may suggest the need for further research and studies, to assess the reach of COVID-19 not only in other countries,

particularly those where mental health infrastructures are less developed [40], but also in other vulnerable populations (eg, children, adolescents, older adults) and in areas facing barriers to accessing health care [41].

As presented in the Results section, 67% (12/18) of the studies included in this literature review presented, in some way, an implementation process (ie, proof of concept, pilot study, or clinical trial). However, it is necessary to take into account that, in 28% (5/18) of the studies in which proposals, evaluations, or conceptual models were presented, real and scientific evidence of the study success and effectiveness was still absent.

Most studies presented as a limitation the difficulties of large-scale implementation, often due to lack of clinical context, which could be improved by the participation of health workers and patients in the design and implementation process.

Some articles that have already conducted a proof of concept in pilot studies considered extrapolating the use of remote monitoring systems to many patients simultaneously as a limitation. Furthermore, those who focused on project evaluations mentioned that the biggest weakness was large-scale implementation, not only for stakeholders but also for the system users themselves. Still, the articles that focused on evaluations were, as already mentioned, only conceptual models that still lacked proof and real-world evidence.

Despite all the studies and results analyzed, there is still a dearth of scientific consistency in the development and implementation process of these interventions and devices [42,43]. According to Gagnon et al [44], the success of an intervention linked to telecare lies in the implementation process. Thus, and as it can be observed in this systematic literature review, it is possible to confirm that the weaknesses mentioned as limitations in the reviewed articles are included in the critical success factors already described by the MOMENTUM framework [45].

Failures include the low adherence by users and, in particular, health professionals. The justifications for the failure of remote monitoring include the difficulty of integrating into existing working methods [23,27], the limited integration of data and information provided in existing communication systems [24,28,29,31], and the lack of correlation between the results of these interventions and specific and individual clinical knowledge for patients with a chronic disease [46]. There is a need for better methodological and evaluative approaches to the development and evaluation of health care improvement interventions [47].

Comparison With Prior Work

Systematic reviews have been based on studies of the (1) identification of opportunities and barriers and (2) acceptability, effectiveness, and impact of the development and implementation of new methods of chronic disease management using remote monitoring systems (eHealth, mobile health, and telehealth). As stated by Trifan et al [46], many reviews that have already been conducted focused on specific conditions and pathologies or had a general focus without a clinical context, as can be seen with this study. However, only a few review studies considered the use of sensors and wearables as a method to collect information for remote monitoring [48-57].

This study aimed to demonstrate the importance that integration and contextualization have in the development and implementation of interventions regarding the use of sensors and wearables for remote monitoring as a source of information for chronic disease management in PHC. To avoid losing information with clinical value throughout the process, it is necessary for not only technological development to be adequate and innovative but also human resources and work methodologies to be able to adapt to new changes.

The PHC context integrates multidisciplinary teams and patients with often complex chronic pathologies. As such, the implementation of new methods and processes for chronic disease management has to be phased and patient-centered and involve all stakeholders [45]. The impact on health care working processes is still not very well studied.

The technological evolution has enabled remote monitoring to grow almost exponentially [56], to solutions, artefacts, or devices that are increasingly smaller, faster, and easier to use and ready to be integrated into the clinical context. The integration of these devices could allow for cost reduction, improved patient quality of life, and early detection of acute episodes, enabling more adequate and personalized intervention and management of disease according to the needs of each patient [57]. However, integrating them in the context of care, specifically in PHC, means that the developed devices have to be adapted not only to patients' specific needs but also to health care professionals' requirements. With the involvement of health care professionals, it is possible to design a solution that takes into account not only the technological requirements behind the system but also the medical requirements, thereby contributing to the improvement of disease management for patients with complex pathologies such as chronic diseases [58].

Limitations

The limitations identified in this systematic review are the use of only 2 databases and the exclusion criteria, which may have led to the exclusion of relevant articles, as well as the time between the search and review process. In addition, most studies were conducted in a hospital setting, and only a few were conducted in a PHC setting; therefore, another limitation is the low number of studies in a PHC context.

Conclusions

This literature review identified several studies on the implementation of remote monitoring devices for patients with chronic diseases in the PHC context. These studies were mainly of cardiac, respiratory, or metabolic pathologies. Despite the opportunities observed, the limitations presented are based on difficulties in generalizing the studies and implementing them on a larger scale. This may be due to both the lack of senior managerial engagement as well as the lack of contextualization of the solutions presented, which, despite being able to prove the technology concept, are not compatible with the health professionals' working methods or with the complexity required for multimorbid patients.

It is clear that innovative technological solutions are being developed. In order to fulfil the need in the area, these technologies have to be properly selected and adapted to the

context of the patients as well as to the health care environment, meaning that more research will be necessary to improve knowledge in this field.

Acknowledgments

MP and LVL acknowledge Fundação para a Ciência e a Tecnologia (FCT-MCTES) for its financial support via the project UIDB/00667/2020 (UNIDEMI).

Authors' Contributions

MP, LVL, and JCF conceptualized the study, with input from IAF and additional input from LBE. MP created the search strategy, with input from LVL and JCF. MP conducted the review, screening, data extraction, and quality assessment and drafted the manuscript. LVL and JCF performed independent screening of abstracts and full-text papers. LVL acted as the third reviewer. MP wrote the drafts of the review paper, with significant input from LVL and JCF and additional input from IAF. All authors have read and approved the final manuscript.

Conflicts of Interest

MP received support from the Portuguese National Funds through FITEC – Programa Interface, with reference CIT INOV – INESC INOVAÇÃO. Funding was also provided from the PhD program in Industrial Management, NOVA Science and Technology Faculty.

Multimedia Appendix 1
Detailed search strategy.
[DOCX File , 14 KB - mhealth_v9i12e28285_app1.docx]

References

1. Röcker C, Ziefle M, Holzinger A. From Computer Innovation to Human Integration: Current Trends and Challenges for Pervasive Health Technologies. In: Holzinger A, Ziefle M, Röcker C, editors. Pervasive Health. Human–Computer Interaction Series. London, England: Springer; 2014.
2. Grand View Research. Global Smart Healthcare Market Size Report, 2020-2027. Grand View Research. 2020. URL: https://www.grandviewresearch.com/industry-analysis/smart-healthcare-market [accessed 2021-12-04]
3. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020 Jun 17;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
4. Guzik P, Malik M. ECG by mobile technologies. J Electrocardiol 2016;49(6):894-901. [doi: 10.1016/j.jelectrocard.2016.07.030] [Medline: 27524474]
5. Qi L, Wen Y. Research and application progress of wearable blood pressure measurement device. China Med Devices 2016;31(12):79.
6. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med 2018 Nov 6;15(11):e1002689 [FREE Full text] [doi: 10.1371/journal.pmed.1002689] [Medline: 30399149]
7. Haghi M, Thurow K, Stoll R. Wearable devices in medical Internet of Things: scientific research and commercially available devices. Healthc Inform Res 2017 Jan;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]
8. Chang A. The Role of Artificial Intelligence in Digital Health. In: Wulfovich S, Meyers A, editors. Digital Health Entrepreneurship. Health Informatics. Cham, Switzerland: Springer International Publishing; 2020:71-81.
9. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
10. Ginsburg G. Medical genomics: Gather and use genetic data in health care. Nature 2014 Apr 24;508(7497):451-453. [doi: 10.1038/508451a] [Medline: 24765668]
11. Jeong IC, Bychkov D, Searson PC. Wearable devices for precision medicine and health state monitoring. IEEE Trans Biomed Eng 2019 May;66(5):1242-1258. [doi: 10.1109/TBME.2018.2871638] [Medline: 31021744]
12. Nimmesgern E, Norstedt I, Draghia-Akli R. Enabling personalized medicine in Europe by the European Commission's funding activities. Per Med 2017 Jul;14(4):355-365 [FREE Full text] [doi: 10.2217/pme-2017-0003] [Medline: 29749834]
13. Malasinghe LP, Ramzan N, Dahal K. Remote patient monitoring: a comprehensive study. J Ambient Intell Human Comput 2017 Oct 26;10(1):57-76. [doi: 10.1007/s12652-017-0598-x]
14. de Carvalho IA, Martin FC, Cesari M, Sumi Y, Thiyagarajan JA, Beard J. Operationalising the concept of intrinsic capacity in clinical settings. World Health Organization. 2017 Nov. URL: https://www.who.int/ageing/health-systems/clinical-consortium/CCHA2017-backgroundpaper-1.pdf [accessed 2021-12-04]

15. Connected health: how digital technology is transforming health and social care. Deloitte Centre for Health Solutions. 2015. URL: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-connected-health.pdf [accessed 2021-12-04]
16. Lukas H, Xu C, Yu Y, Gao W. Emerging telemedicine tools for remote COVID-19 diagnosis, monitoring, and management. ACS Nano 2020 Dec 22;14(12):16180-16193 [FREE Full text] [doi: 10.1021/acsnano.0c08494] [Medline: 33314910]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339(jul21 1):b2535-b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
18. Porath A, Irony A, Borobick AS, Nasser S, Malachi A, Fund N, et al. Maccabi proactive Telecare Center for chronic conditions - the care of frail elderly patients. Isr J Health Policy Res 2017 Dec 11;6(1):68 [FREE Full text] [doi: 10.1186/s13584-017-0192-x] [Medline: 29228992]
19. Lygidakis C, Uwizihiwe J, Kallestrup P, Bia M, Condo J, Vögele C. Community- and mHealth-based integrated management of diabetes in primary healthcare in Rwanda (D²Rwanda): the protocol of a mixed-methods study including a cluster randomised controlled trial. BMJ Open 2019 Jul 24;9(7):e028427 [FREE Full text] [doi: 10.1136/bmjopen-2018-028427] [Medline: 31345971]
20. Salisbury C, Thomas C, O'Cathain A, Rogers A, Pope C, Yardley L, et al. TElehealth in CHronic disease: mixed-methods study to develop the TECH conceptual model for intervention design and evaluation. BMJ Open 2015 Feb 06;5(2):e006448 [FREE Full text] [doi: 10.1136/bmjopen-2014-006448] [Medline: 25659890]
21. Emerson J, Welch M, Rossman W, Carek S, Ludden T, Templin M, et al. A multidisciplinary intervention utilizing virtual communication tools to reduce health disparities: a pilot randomized controlled trial. Int J Environ Res Public Health 2015 Dec 22;13(1):ijerph13010031 [FREE Full text] [doi: 10.3390/ijerph13010031] [Medline: 26703661]
22. Oksman E, Linna M, Hörhammer I, Lammintakanen J, Talja M. Cost-effectiveness analysis for a tele-based health coaching program for chronic disease in primary care. BMC Health Serv Res 2017 Feb 15;17(1):138 [FREE Full text] [doi: 10.1186/s12913-017-2088-4] [Medline: 28202032]
23. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris M, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]
24. Diez-Canseco F, Toyama M, Ipince A, Perez-Leon S, Cavero V, Araya R, et al. Integration of a technology-based mental health screening program into routine practices of primary health care services in Peru (The Allillanchu Project): development and implementation. J Med Internet Res 2018 Mar 15;20(3):e100 [FREE Full text] [doi: 10.2196/jmir.9208] [Medline: 29588272]
25. Bovbjerg ML, Lee J, Wolff R, Bangs B, May MA. Pilot study for managing complex chronic care Medicaid patients with diabetes using a mobile health application achieves "Triple Aim" improvement in a primary care setting. Clin Diabetes 2017 Oct 16;35(4):227-231 [FREE Full text] [doi: 10.2337/cd17-0006] [Medline: 29109612]
26. Paula A, Maldonado J, Gadelha C. Healthcare telemonitoring and business dynamics: challenges and opportunities for SUS. Rev Saude Publica 2020;54:65 [FREE Full text] [doi: 10.11606/s1518-8787.2020054001996] [Medline: 32609277]
27. Lugo-Palacios DG, Hammond J, Allen T, Darley S, McDonald R, Blakeman T, et al. The impact of a combinatorial digital and organisational intervention on the management of long-term conditions in UK primary care: a non-randomised evaluation. BMC Health Serv Res 2019 Mar 12;19(1):159 [FREE Full text] [doi: 10.1186/s12913-019-3984-6] [Medline: 30866917]
28. Jindal D, Gupta P, Jha D, Ajay V, Goenka S, Jacob P, et al. Development of mWellcare: an mHealth intervention for integrated management of hypertension and diabetes in low-resource settings. Glob Health Action 2018;11(1):1517930 [FREE Full text] [doi: 10.1080/16549716.2018.1517930] [Medline: 30253691]
29. Dang S, Ruiz DI, Klepac L, Morse S, Becker P, Levy CR, et al. Key characteristics for successful adoption and implementation of home telehealth technology in Veterans Affairs home-based primary care: an exploratory study. Telemed J E Health 2019 Apr;25(4):309-318. [doi: 10.1089/tmj.2018.0009] [Medline: 29969381]
30. Alonso-Domínguez R, Gómez-Marcos MA, Patino-Alonso MC, Sánchez-Aguadero N, Agudo-Conde C, Castaño-Sánchez C, et al. Effectiveness of a multifactorial intervention based on an application for smartphones, heart-healthy walks and a nutritional workshop in patients with type 2 diabetes mellitus in primary care (EMID): study protocol for a randomised controlled trial. BMJ Open 2017 Sep 14;7(9):e016191 [FREE Full text] [doi: 10.1136/bmjopen-2017-016191] [Medline: 28912193]
31. Hans PK, Gray CS, Gill A, Tiessen J. The provider perspective: investigating the effect of the Electronic Patient-Reported Outcome (ePRO) mobile application and portal on primary care provider workflow. Prim Health Care Res Dev 2017 Sep 13;19(02):151-164. [doi: 10.1017/s1463423617000573]
32. Dang S, Olsan T, Karuza J, Cai X, Gao S, Intrator O, et al. Telehealth in home-based primary care: factors and challenges associated with integration into veteran care. J Am Geriatr Soc 2019 Sep;67(9):1928-1933. [doi: 10.1111/jgs.16045] [Medline: 31343732]
33. Ramallo-Fariña Y, García-Bello MA, García-Pérez L, Boronat M, Wägner AM, Rodríguez-Rodríguez L, INDICA Team. Effectiveness of internet-based multicomponent interventions for patients and health care professionals to improve clinical

outcomes in type 2 diabetes evaluated through the INDICA Study: multiarm cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 02;8(11):e18922 [FREE Full text] [doi: 10.2196/18922] [Medline: 33136059]
34. Hernandez-Quiles C, Bernabeu-Wittel M, Garcia-Serrano MDR, Vergara-Lopez S, Perez-de-Leon JA, Ruiz-Cantero A, et al. A multicenter randomized clinical trial to evaluate the efficacy of telemonitoring in patients with advanced heart and lung chronic failure. Study protocol for the ATLAN_TIC project. Contemp Clin Trials Commun 2020 Mar;17:100512 [FREE Full text] [doi: 10.1016/j.conctc.2019.100512] [Medline: 31938754]
35. Martín-Lesende I, Orruño E, Mateos M, Recalde E, Asua J, Reviriego E, et al. Telemonitoring in-home complex chronic patients from primary care in routine clinical practice: Impact on healthcare resources use. Eur J Gen Pract 2017 Dec 27;23(1):135-142 [FREE Full text] [doi: 10.1080/13814788.2017.1306516] [Medline: 28446045]
36. Ham C. The ten characteristics of the high-performing chronic care system. Health Econ Policy Law 2010 Jan;5(Pt 1):71-90. [doi: 10.1017/S1744133109990120] [Medline: 19732475]
37. Carlfjord S, Lindberg M, Bendtsen P, Nilsen P, Andersson A. Key factors influencing adoption of an innovation in primary health care: a qualitative study based on implementation theory. BMC Fam Pract 2010 Aug 23;11(1):60 [FREE Full text] [doi: 10.1186/1471-2296-11-60] [Medline: 20731817]
38. World health statistics 2020: monitoring health for the SDGs, sustainable development goals. World Health Organization. 2020 May 13. URL: https://www.who.int/publications/i/item/9789240005105 [accessed 2021-12-04]
39. 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]
40. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]
41. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
42. Queiroz AAFLN, Mendes IAC, de Godoy S, Velez Lapão L, Dias S. mHealth strategies related to HIV postexposure prophylaxis knowledge and access: systematic literature review, technology prospecting of patent databases, and systematic search on app stores. JMIR Mhealth Uhealth 2021 Feb 16;9(2):e23912 [FREE Full text] [doi: 10.2196/23912] [Medline: 33591289]
43. Laur CV, Agarwal P, Mukerji G, Goulbourne E, Baranek H, Pus L, et al. Building health services in a rapidly changing landscape: lessons in adaptive leadership and pivots in a COVID-19 remote monitoring program. J Med Internet Res 2021 Jan 13;23(1):e25507. [doi: 10.2196/25507]
44. Gagnon M, Duplantie J, Fortin J, Landry R. Implementing telehealth to support medical practice in rural/remote regions: what are the conditions for success? Implement Sci 2006 Aug 24;1:18 [FREE Full text] [doi: 10.1186/1748-5908-1-18] [Medline: 16930484]
45. Maia MR, Castela E, Pires A, Lapão LV. How to develop a sustainable telemedicine service? A Pediatric Telecardiology Service 20 years on - an exploratory study. BMC Health Serv Res 2019 Sep 23;19(1):681 [FREE Full text] [doi: 10.1186/s12913-019-4511-5] [Medline: 31547824]
46. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. JMIR Mhealth Uhealth 2019 Aug 23;7(8):e12649 [FREE Full text] [doi: 10.2196/12649] [Medline: 31444874]
47. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
48. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
49. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
50. Thomas JG, Bond DS. Review of innovations in digital health technology to promote weight control. Curr Diab Rep 2014;14(5):485. [doi: 10.1007/s11892-014-0485-1] [Medline: 24664797]
51. Ko PT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: 10.5664/jcsm.5288] [Medline: 26156958]
52. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Jun;18(2):163-169 [FREE Full text] [Medline: 27489456]
53. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
54. Rajagopalan A, Shah P, Zhang M, Ho R. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci 2017 Nov 12;7(11):150 [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]

55. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Feb 13;19(2):e41 [FREE Full text] [doi: 10.2196/jmir.6397] [Medline: 28193600]
56. Michalopoulos S. Over 20% of digital healthcare startups emerged during the pandemic. EURACTIV. 2021 Feb 17. URL: https://www.euractiv.com/section/coronavirus/news/over-20-of-digital-healthcare-startups-emerged-during-the-pandemic/ [accessed 2021-12-04]
57. de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Ortega Bravo M, CONNECARE-Lleida Group. Implementing mobile health-enabled integrated care for complex chronic patients: intervention effectiveness and cost-effectiveness study. JMIR Mhealth Uhealth 2021 Jan 14;9(1):e22135 [FREE Full text] [doi: 10.2196/22135] [Medline: 33443486]
58. Williams K, Markwardt S, Kearney SM, Karp JF, Kraemer KL, Park MJ, et al. Addressing implementation challenges to digital care delivery for adults with multiple chronic conditions: stakeholder feedback in a randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 01;9(2):e23498 [FREE Full text] [doi: 10.2196/23498] [Medline: 33522981]

Abbreviations

AI: artificial intelligence
BD: big data
HRMS: health remote monitoring systems
IoT: Internet of Things
PHC: primary health care
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
WoSCC: Web of Science Core Collection

Edited by L Buis; submitted 27.02.21; peer-reviewed by L Barbosa, T Saheb, T Capote, J de Batlle; comments to author 31.03.21; revised version received 25.05.21; accepted 08.10.21; published 21.12.21.

Please cite as:
Peyroteo M, Ferreira IA, Elvas LB, Ferreira JC, Lapão LV
Remote Monitoring Systems for Patients With Chronic Diseases in Primary Health Care: Systematic Review
JMIR Mhealth Uhealth 2021;9(12):e28285
URL: https://mhealth.jmir.org/2021/12/e28285
doi:10.2196/28285
PMID:34932000

Original Paper

Identifying Data Quality Dimensions for Person-Generated Wearable Device Data: Multi-Method Study

Sylvia Cho[1], PhD; Chunhua Weng[1], PhD; Michael G Kahn[2], MD, PhD; Karthik Natarajan[1], PhD

[1]Department of Biomedical Informatics, Columbia University, New York, NY, United States

[2]Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Corresponding Author:
Sylvia Cho, PhD
Department of Biomedical Informatics
Columbia University
622 West 168th Street PH20
New York, NY, 10032
United States
Phone: 1 212 305 5334
Email: sc3901@cumc.columbia.edu

Abstract

Background: There is a growing interest in using person-generated wearable device data for biomedical research, but there are also concerns regarding the quality of data such as missing or incorrect data. This emphasizes the importance of assessing data quality before conducting research. In order to perform data quality assessments, it is essential to define what data quality means for person-generated wearable device data by identifying the data quality dimensions.

Objective: This study aims to identify data quality dimensions for person-generated wearable device data for research purposes.

Methods: This study was conducted in 3 phases: literature review, survey, and focus group discussion. The literature review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline to identify factors affecting data quality and its associated data quality challenges. In addition, we conducted a survey to confirm and complement results from the literature review and to understand researchers' perceptions on data quality dimensions that were previously identified as dimensions for the secondary use of electronic health record (EHR) data. We sent the survey to researchers with experience in analyzing wearable device data. Focus group discussion sessions were conducted with domain experts to derive data quality dimensions for person-generated wearable device data. On the basis of the results from the literature review and survey, a facilitator proposed potential data quality dimensions relevant to person-generated wearable device data, and the domain experts accepted or rejected the suggested dimensions.

Results: In total, 19 studies were included in the literature review, and 3 major themes emerged: device- and technical-related, user-related, and data governance–related factors. The associated data quality problems were incomplete data, incorrect data, and heterogeneous data. A total of 20 respondents answered the survey. The major data quality challenges faced by researchers were completeness, accuracy, and plausibility. The importance ratings on data quality dimensions in an existing framework showed that the dimensions for secondary use of EHR data are applicable to person-generated wearable device data. There were 3 focus group sessions with domain experts in data quality and wearable device research. The experts concluded that intrinsic data quality features, such as conformance, completeness, and plausibility, and contextual and fitness-for-use data quality features, such as completeness (breadth and density) and temporal data granularity, are important data quality dimensions for assessing person-generated wearable device data for research purposes.

Conclusions: In this study, intrinsic and contextual and fitness-for-use data quality dimensions for person-generated wearable device data were identified. The dimensions were adapted from data quality terminologies and frameworks for the secondary use of EHR data with a few modifications. Further research on how data quality can be assessed with respect to each dimension is needed.

(JMIR Mhealth Uhealth 2021;9(12):e31618) doi:10.2196/31618

KEYWORDS

patient-generated health data; data accuracy; data quality; wearable device; fitness trackers; qualitative research

Introduction

Use of Person-Generated Wearable Device Data for Research Purposes

The growing interest in quantified self along with the routine use of consumer wearables is generating substantial amounts of person-generated wearable device data [1,2]. These passively and objectively collected data hold great potential for use in biomedical research as they capture data that occur outside the clinic, without having to rely on patient recall [3]. An example of using wearable device data for biomedical research is a study by Lim et al [4] in which consumer-grade fitness tracker data (Fitbit Charge HR) was used along with survey and electronic health record (EHR) data. In addition, wearable device data can be reused in multiple studies to answer many different research questions. The investigators of the Lim et al [4] study made their data publicly available for other researchers, expanding the opportunity to generate and validate medical evidence. McDonald et al [5] used these data to investigate the relationship between sleep time and BMI in a Chinese population. This study was conducted to confirm the results of Xu et al [6], who examined the relationship between sleep duration and BMI. One of the limitations of the Xu et al [6] study was that their data primarily consisted of Europeans, and thus the study results needed further investigation to be generalizable. McDonald et al [5] added further evidence to the association between sleep and BMI by examining the same research question using a data set comprising Asian individuals. This type of evidence generation is expected to become more widespread with the All of Us Research Program, a precision medicine initiative by the National Institutes of Health, which is collecting, integrating, and providing wearable device data (eg, Fitbit) to the public for research purposes [7]. Considering that there is a lack of publicly available data sets generated from consumer wearable devices with a large number of participants and long-term observation, the All of Us data are expected to become a promising resource for many researchers interested in analyzing wearable device data.

Significance of Data Quality Assessment

Although person-generated wearable device data are a promising new source of biomedical data, there are concerns regarding the quality of data. For example, missing data owing to users not wearing the device or incorrect data owing to device malfunction are a few data quality problems that could occur [8,9]. As these data anomalies could lead to various challenges when analyzing wearable device data, data quality assessment is a critical step that should be implemented before any analyses [8]. In this setting, data quality assessment is not only about whether the wearable device captures valid and reliable data but also whether a data set is fit-for-use for a specific research purpose, ensuring valid results [8,10]. However, the question about what data quality means, more specifically, how data quality is defined for the use of person-generated wearable device data for research purposes still remains.

Data Quality Dimensions

Data quality dimensions are criteria or aspects of data quality that are considered essential for a specific user's task and are constructs used when assessing data [11,12]. For example, the quality of data could be assessed in terms of its completeness ("Are data values present?"), conformance ("Do data values adhere to specified standards and formats?"), and plausibility ("Are data values believable?") [13,14]. Various methods have been previously used to derive the data quality dimensions for biomedical data sets. First, Weiskopf et al [13] and Johnson et al [15] used systematic reviews to derive data quality dimensions. They both abstracted data quality attributes from studies on EHR data quality and then derived broad dimensions of data quality [13,15]. Second, stakeholder meetings are another method used by Kahn et al [14]. Stakeholders reviewed the literature on data quality, publications on best practices, operational manuals, and data quality rules from several EHR-based research networks. Data quality terms were then integrated into categories through an iterative process [14]. Finally, surveys have also been used as a method to identify data quality dimensions [16,17]. For example, Huang et al [17] investigated important data quality dimensions for genome annotation by asking genomic researchers to rank the importance of 17 data quality dimensions. The strength of these empirical methods is that it captures the perspective of data users and reveals data quality dimensions that may not have been considered by data quality researchers [16,18]. This is important as data quality is a concept that depends on the data users and their research tasks.

Currently, there is a lack of studies that derive dimensions for person-generated wearable device data using empirical methods. To our knowledge, the study by Codella et al [19] is the most relevant study on data quality dimensions for person-generated wearable device data. The study [19] first reviewed the literature to identify stakeholders' concerns regarding person-generated health data (PGHD) and mapped the concerns to the corresponding data quality dimensions in the Wang and Strong [16] framework. However, the Wang and Strong [16] framework was derived by surveying business data consumers, which might not include important data quality dimensions for PGHD. Therefore, there is a great need to investigate the essential challenges and dimensions for assessing the quality of person-generated wearable device data for biomedical research because it is a growing, new data type.

Objective

The aim of this study is to identify important data quality dimensions for using person-generated wearable device data for research purposes. The focus of this study is on intrinsic (data quality features inherent to the data) and contextual and fitness-for-use data quality dimensions (features that are task-dependent). Extrinsic and operational data quality features, such as data accessibility, security, or privacy, are not the focus of this study.

Methods

Study Design

Owing to the lack of literature or experts in the data quality field for person-generated wearable device data, a multi-method approach was used to complement and validate information found by each method. A combination of literature review and

survey was used to improve reliability through constant data comparison [20]. In addition, focus group discussions were conducted to derive data quality dimensions from the collected data.

Part 1: Literature Review

The goal of the literature review was to identify (1) factors affecting the quality of person-generated wearable device data and (2) associated intrinsic data quality challenges that could potentially occur when conducting research. Studies were examined from scholarly databases using a combination of search terms related to data quality and wearable devices. One reviewer (SC) screened the titles and abstracts of the studies based on a set of selection criteria. For example, studies containing any content on the data quality of wearable device data or sensor data when used for research purposes were included, but studies on clinicians wearing devices for patient care were excluded because the focus was on person-generated data being used for research purposes. The full text was screened using the same criteria by 2 reviewers (SC and KN). Sentences on data quality challenges and factors affecting those challenges were annotated, and semantically similar challenges and factors were grouped into the same category. The categorization process was performed by 3 researchers (SC, KN, and Ipek Ensari), including the 2 reviewers (SC and KN). Details of the literature review process are described in a previously published manuscript [9].

Part 2: Survey

Survey Development

The survey was developed with a mixture of multiple-choice, open-ended, and Likert-type scale questions. The survey was iteratively refined based on feedback from 6 experts—3 in data quality, 2 in wearable devices, and 1 in survey development. The experts were recruited through the professional network of the research team, and the experts were those who actively conducted research in either data quality, wearable devices, or survey development. A web-based survey was created using Qualtrics (Qualtrics; version August 2019), which is a web-based survey software [21].

Data Collection and Analysis

The eligibility criteria for survey participation included the following: (1) an individual with experience in analyzing passively collected wearable device data for their research and (2) an individual with knowledge of data quality challenges when dealing with wearable device data. Potential survey participants were identified by searching the authors of research studies that used wearable device data and through referrals. The survey link was sent via email to the candidate respondents. In addition, a link to the survey was posted on the Observational Health Data Sciences and Informatics forum [22]. This forum was chosen because it focuses on observational health data, and individuals with diverse research backgrounds including PGHD and data quality frequently visit the forum. Participation was voluntary, and the survey was self-administered and anonymous.

Answers to multiple-choice questions were analyzed using descriptive statistics, and thematic analysis was conducted to identify themes from answers to open-ended questions. Responses to Likert-type scale questions were analyzed by comparing mean (importance of the dimensions) with SD (reliability) of the importance ratings of the dimensions. Dimensions with high mean (importance) and low SD (less variability in ratings among respondents) were determined as important.

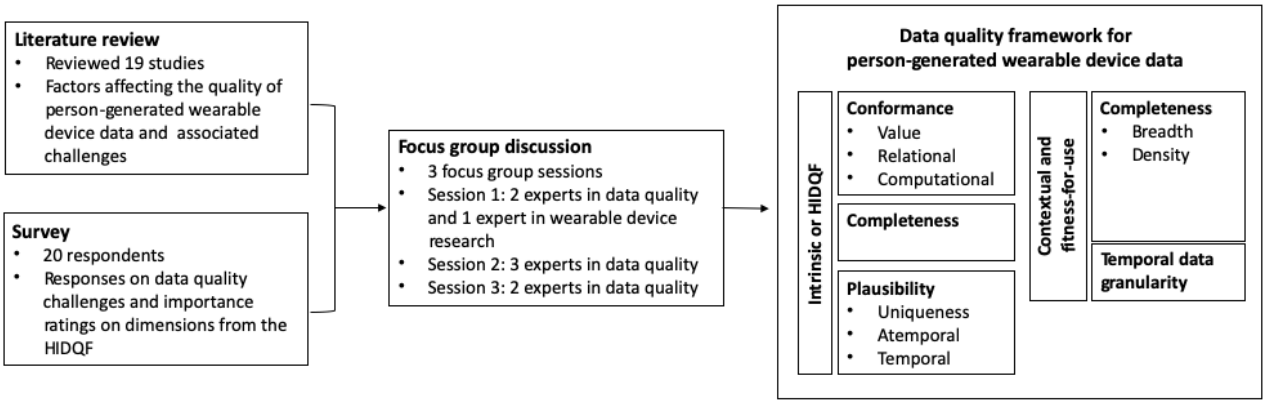
Part 3: Focus Group Discussion

Domain experts in data quality or wearable device data were recruited through a professional network of authors. The facilitator (SC) combined the results of the literature review and survey and proposed potential dimensions to domain experts. Domain experts discussed the information provided and determined whether to accept or reject the suggested data quality dimensions. The importance ratings on dimensions in the harmonized intrinsic data quality framework (HIDQF) were also used as a reference to determine its relevance to wearable device data [14]. The discussion continued until consensus was reached among the experts.

Results

An overview of the results is depicted in Figure 1 followed by further details regarding the results.

Figure 1. An overview of study processes and results. HIDQF: harmonized intrinsic data quality framework.



Part 1: Literature Review

In total, 1290 studies were retrieved and screened, resulting in 1.47% (19/1290) studies being selected for analysis. Data extracted from the studies were categorized into 3 groups of factors affecting data quality, that is, device- and technical-related, user-related, and data governance–related factors, and 3 data quality challenges, that is, incompleteness, incorrectness, and heterogeneity. Most studies have discussed device- and human-related factors that influence data quality. For example, device malfunction, network and connectivity, and users not wearing the device can lead to incomplete or missing data, whereas poor quality of sensor or algorithms and users' incorrect use of the device may lead to incorrect data. In addition, lack of data standardization, such as different data formats, measurement units, and different algorithms, for the same parameter may cause difficulty in making a direct comparison between data from different devices. The full results of the literature review have been published [9].

Part 2: Survey

Survey Design and Participant Recruitment

The survey was designed in 3 parts: (1) questions on the respondents' research background, (2) questions on the research that the participants have conducted, and (3) questions on participants' perception and knowledge of data quality. The survey included a Likert-type scale question that asked to rate the data quality dimensions from the HIDQF regarding their importance [14]. The HIDQF harmonizes 9 existing data quality terminologies and frameworks that are applicable to the secondary use of EHR data [14]. The harmonized framework involved a consensus among various stakeholders and experts in data quality; thus, it made sense to leverage the framework as a basis for the data quality dimensions of wearable devices. The full survey can be found in the link cited in the reference [23].

Emails were sent out to 100 researchers from August 2019 to September 2019. The exact number of survey recipients is unknown because the email recipients forwarded the email to other eligible individuals, and the survey was posted on a public online forum. In total, 20 responses were collected—most respondents were from the United States, but there were also a few respondents from the United Kingdom, France, and Singapore. Using 100 as a proxy for the number of eligible researchers, there was a 20% (20/100) response rate for the survey.

Background of Respondents

Table 1 shows the background of the survey respondents based on the responses collected from part 1 and part 2 of the survey.

Most respondents published 1-3 peer-reviewed articles (12/20, 60%), and 3 respondents (3/20, 15%) published >10 articles. The most common types of studies previously conducted by respondents were device validation or reliability studies (11/20, 55%), modeling to predict health state (10/20, 50%), and tracking behavioral changes (8/20, 40%). Other research types, such as pattern analysis on activity data and tracking body movement or stress, were also mentioned.

Nearly half of the respondents (9/20, 45%) used research-grade and consumer-grade devices with similar frequency, and 8 respondents (8/20, 40%) had only used consumer-grade devices. The respondents gave multiple answers regarding the brand and model of the devices they had used before. Among consumer-grade devices, the most frequently mentioned brand was Fitbit (19/20, 95%), followed by Garmin, Withings, Jawbone, and Apple Watch. Research-grade devices, especially accelerometers, such as ActiGraph, GENEactiv, and Actical, were mentioned 6 times. Other devices were mentioned, such as the Huawei Watch 2, Samsung Gear 2, and Misfit Shine 2.

Table 1. Background of respondents (N=20).

Characteristic	Value
Number of peer-reviewed articles using wearable device data, n (%)	
None	1 (5)
1 to 3	12 (60)
3 to 5	2 (10)
5 to 10	2 (10)
10 or more	3 (15)
Type of research conducted (multiple choice possible), n (%)	
Device validation or reliability studies	11 (55)
Modeling to predict health state	10 (50)
Modeling to inform treatment decisions	2 (10)
Tracking behavioral changes	8 (40)
Other	3 (15)
Type of devices used for research, n (%)	
Consumer-grade wearable	8 (40)
Research-grade wearable	3 (15)
Used both with similar frequency	9 (45)
Brand of devices used (multiple choice possible), n (%)	
Fitbit (Charge HR, Alta HR, Ultra, etc)	19 (95)
Garmin (Vivofit, Vivosmart, Fenix, etc)	6 (30)
Withings (Go, Pulse, or BP cuff)	4 (20)
Jawbone (UP)	2 (10)
Apple Watch	1 (5)
Accelerometer (ActiGraph, GENEactiv, etc)	6 (30)
Other (Huawei, Samsung gear, Misfit, etc)	14 (70)

Data Quality Challenges

In total 3 main themes and 1 minor issue were derived from the open-ended question on data quality challenges: (1) completeness, (2) accuracy, (3) plausibility, and (4) data access and semantics.

Completeness

One of the major themes was the completeness. Missing data were a concern for the respondents because of the uncertainty involved in dealing with missingness as it can have a negative effect on the analysis results. Many respondents wrote about missing data caused by various reasons, such as device error or users not wearing devices, which aligns with the results from the literature review. One respondent specifically talked about a different aspect of missing data, which is the lack of a certain variable that they needed for their research (“Lack of availability of heart rate variability”).

Accuracy

Another major theme was accuracy—*Do the data represent the true value?* Respondents talked about their doubts about whether the data correctly capture the true physiological measure they are supposed to represent. For example, steps might not be counted if one does not wear the device during exercise owing to discomfort. On the other hand, other activities, such as motorcycle rides, could falsely increase the step counts. In addition, a respondent mentioned the problem of GPS devices only recording known locations rather than the actual route, affecting distance traveled. There could also be inaccuracies in the sleep data. For instance, activities that are performed while lying on the bed (eg, using phones) could be counted as sleep mode, and sleep or wake time could be recognized inaccurately. These concerns match the challenges found in the literature review.

Plausibility

Plausibility was another major theme—*Do the data make sense?* One of the issues mentioned was that the data did not agree with their common knowledge. For example, there are problems in inconsistency between variables (“large spikes or drops in activity that are highly inconsistent with their surrounding measured values”). Respondents also stated that outliers in the data made them question the validity of that data point (“knowing whether unusual data are real”).

There were also time-related plausibility issues. For example, even though the data for 2 different variables are captured at

the same time, the recorded timestamp on the server could be different between the 2 variables because of problems with data upload ("lag between device and data server—some variables are collected at slightly different time due to problems with wifi connection, data uploading"). In addition, people traveling between different time zones may produce implausible time patterns when the device does not recognize the change in time zone ("Subjects may travel between different time zones during study period. Some devices don't recognize a different time zone and the recorded data has weird time pattern that is hard to understand"). These challenges were not explicitly mentioned in the literature but are implied by incorrect data problems.

Data Access and Semantics

There were data quality challenges related to data access and semantics. For example, the difficulty in accessing raw data and minute-level data was mentioned by a few respondents. In addition, a few respondents mentioned that interpreting the data may be a challenge because of the lack of information on context and provenance (eg, no documentation of exposures). Lack of transparency owing to consumer devices being proprietary was also mentioned. These challenges were not mentioned in our literature review study on data quality challenges because the scope of research was only on intrinsic data quality challenges, but there were studies mentioning these challenges.

Ratings of Data Quality Dimensions

Respondents' importance ratings on dimensions from the HIDQF are presented in Figures 2 and 3 [14].

Figure 2. Importance ratings on dimensions from harmonized intrinsic data quality framework.

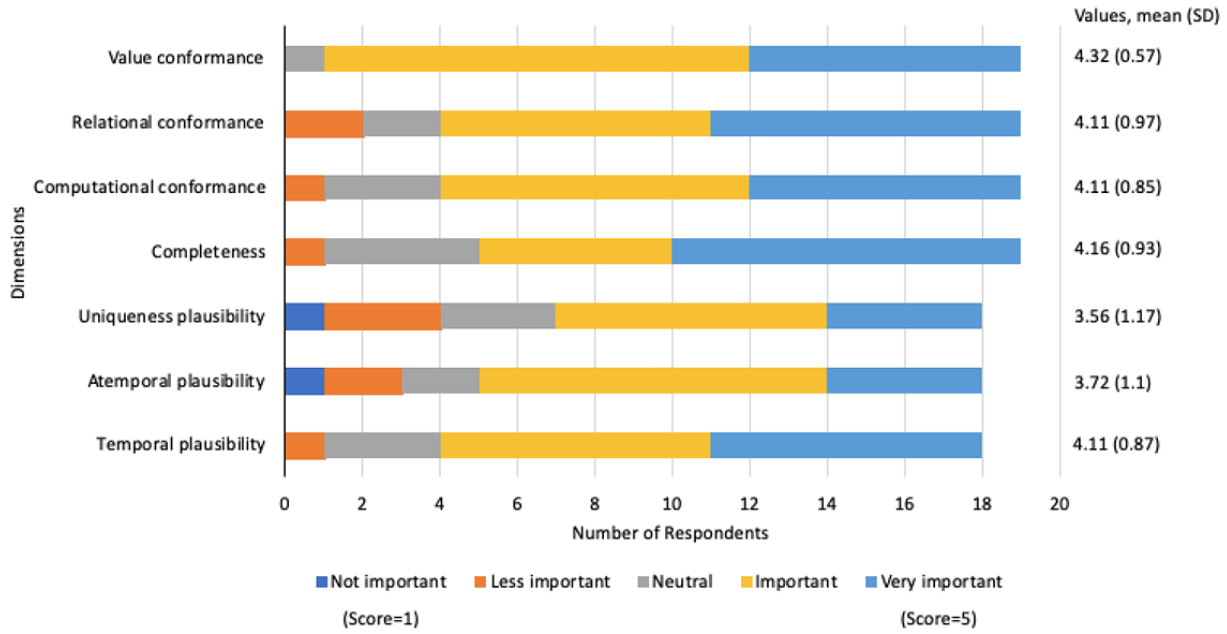
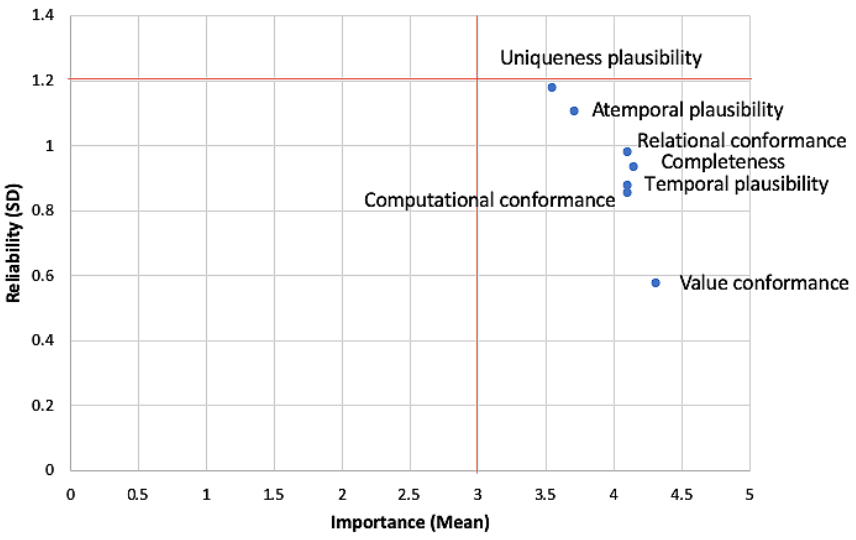


Figure 3. Importance versus reliability of ratings on data quality dimensions in the harmonized intrinsic data quality framework.



Adopting the cutoffs used in a previously published study, dimensions with mean ratings>3 were determined to be important, and ratings with SD <1.2 were considered reliable [24]. Overall, respondents considered dimensions from the HIDQF as important data quality features for wearable device data. A follow-up question on the most important dimension identified completeness as the most important dimension (n=7), followed by relational conformance (n=4), computational conformance (n=4), value conformance (n=1), temporal plausibility (n=1), and atemporal plausibility (n=1).

A few respondents answered the free-response question on additional data quality dimensions that need to be added. Various problems were mentioned, including the importance of a consistent sampling rate when dealing with multiple device data and the need for contextual information about the data set. For instance, metadata on whether the data set is raw data or processed using proprietary algorithms and whether the users brought their own device or whether it was provided was considered important information to respondents. Furthermore, information on the wearing status of users was considered important.

Part 3: Deriving Dimensions Through Focus Group Discussion

The potential data quality dimensions proposed by the facilitator (SC) are presented in Table 2 (the full version of this table can be found in Table S1 of Multimedia Appendix 1). Conformance was included as a potential data quality dimension based on the factors related to data heterogeneity found in the literature review and survey responses on the importance of data conformance. Completeness was one of the most frequently mentioned data quality challenge in both the literature review and survey. It was also selected as the most important data quality dimension by the survey respondents and thus was included in the list of potential data quality dimensions. Data quality challenges related to accuracy (data incorrectness) were frequently mentioned in both the literature and the survey. In addition, plausibility, which has a similar context with accuracy, was mentioned by survey respondents (eg, “large spikes or drops in activity that are highly inconsistent with their surrounding measured values”). Both challenges were presented to the experts for further discussion. The difficulty of accessing minute- or second-level data was mentioned as a challenge in both the literature and the survey (this is more of an extrinsic data quality challenge, which was why it was not reported in the previously published literature review study). As the objective of this study was to focus on intrinsic and contextual and fitness-for-use data quality dimensions, not extrinsic data quality dimensions, data accessibility was not included as a potential data quality dimension. Instead, the challenge of accessing minute- or second-level data was interpreted as the researchers’ need for more temporally granular data. Thus, temporal data granularity was added as a potential data quality dimension. Finally, data interpretability was proposed to domain experts based on survey responses on the need for contextual information and metadata.

Table 2. List of data quality dimensions suggested based on findings from literature review and survey.

Dimensions suggested to experts	Corresponding content from the literature review	Corresponding content from survey responses	Importance rating (only for HIDQF[a])
Conformance			
Value conformance	• Different devices may use a different measurement unit.	• "Data set not conforming to data dictionary will be hard to fix"	4.32
Relational conformance	—[b]	• "Without relational conformance you can't link one wearable device to another or to health outcomes"	4.11
Computational conformance	• Companies do not always reveal whether or when they update their device algorithms or whether or when the users install the provided software updates. • Lack of standardization: (for multi-device studies) different devices may use different algorithms, a different definition for the same parameter, different sampling rate.	• "I don't know a way to proceed with the data analyses if the computational conformance isn't met with satisfaction. it suggests that the data collected cannot be trusted."	4.11
Completeness	• Missing data due to various reasons: device malfunction, connectivity issues, nonadherence to the device, quality of skin contact of the device.	• "Missing data is a large issue for our research, especially because we are trying to identify patterns or subsequences of activity. Missing data has to either be interpolated or treated as a zero value, and either of these methods can have a large negative effect on the results of our pattern mining techniques."	4.16
Breadth completeness	—	• "Lack of availability of HRV[c]"	
Plausibility			
Uniqueness plausibility	—	—	3.56
Atemporal plausibility	—	• "Large spikes or drops in activity that are highly inconsistent with their surrounding measured values"	3.72
Temporal plausibility	• Companies do not always reveal whether or when they update their device algorithms, or whether or when the users install the provided software updates.	• "Devices might cause problem with recording different time zone or time during traveling: Subjects may travel between different time zones during study period. Some devices don't recognize a different time zone and the recorded data has weird time pattern that is hard to understand"	4.11
Temporal data granularity	• Fitbit only provides access to day-level data unless the minute-level or second-level data is requested and approved.	• "Access to minute level data."	—
Accuracy	• Poor data accuracy caused by device malfunction, unknown limitations of proprietary algorithms, user error in device use.	• "Other activities generating step counts (eg, motorcycle ride, vibration)" • "Inaccurate sleep and wake time recognition"	—
Interpretability	—	• "Trying to nail down exactly what a participant was doing when data was being collected offsite."	—

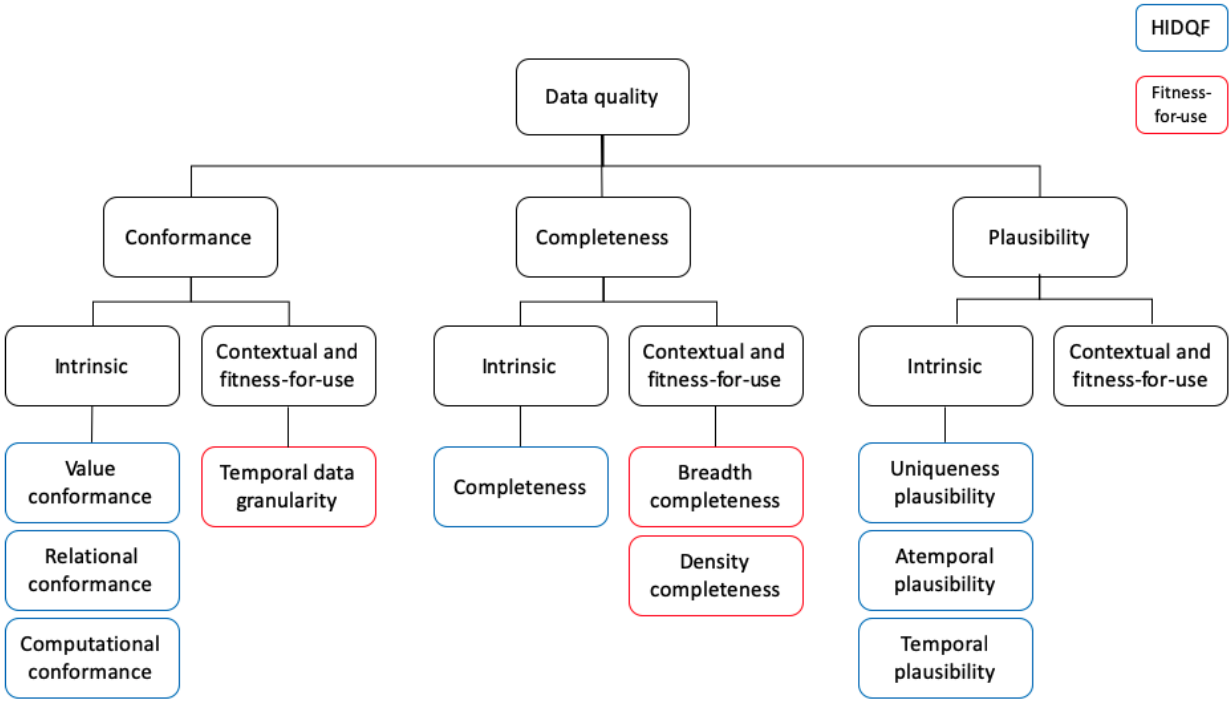
[a]HIDQF: harmonized intrinsic data quality framework.

[b]No available data.

[c]HRV: heart rate variability.

In total, 3 separate discussion sessions were conducted in January, May, and September 2020. All sessions were conducted with 2-3 domain experts and 1 facilitator. In all, 2 data quality experts and 1 wearable device expert participated in the first discussion session. To continue the discussion on the relevance of dimensions to wearable device data, the second and third discussion sessions were conducted with 3 and 2 data quality experts, respectively. The domain experts agreed that all dimensions in the HIDQF were applicable to person-generated wearable device data. In addition, it was suggested to add contextual and fitness-for-use data quality dimensions that consider data quality in the context of a given research task [16]. Although the dimensions of the HIDQF are for research purposes as well, they focused on intrinsic data quality that assesses data quality in terms of the structure and presence of the data itself, independent of research tasks [14]. Considering that our focus was on using wearable device data for research purposes, aspects of data quality that can be determined once the research task is known were considered important. The final list of dimensions is shown in Figure 4.

Figure 4. Data quality dimensions for assessing person-generated wearable device data for research purposes. HIDQF: harmonized intrinsic data quality framework.



There was substantial discussion on *completeness*. The completeness dimension in the HIDQF is defined as "Are data values present?" which measures completeness based on the presence of data without referring to research tasks [14,25]. However, determining missing data could be complicated for wearable device data when conducting research, especially for activity or step count data, because missing data could appear as null but more often as zero values [9]. Interpreting zero values is not easy because it could mean that a person was not wearing the device (true missingness) or was sedentary (a valid zero value). Zero values generated from being sedentary are not simply missing data, as they provide information on device users' physical activity [26]. As it is impossible to know the cause of zero values, researchers typically make assumptions on thresholds for the inactivity period to determine nonwear time (eg, 60 minutes of inactivity [zero step count] is considered as a user not wearing the device) [26,27]. Thus, data completeness for activity-related data can be assessed based on the measures and thresholds that researchers set to define what is or is not missing data. This was why the fitness-for-use data completeness dimensions were considered important by domain experts. There were 2 fitness-for-use completeness dimensions determined as applicable to wearable device data, which were breadth and density completeness. Breadth completeness assesses whether a data set contains all types of data that are required for a specific task. For example, to investigate the association between activity and heart rate, a data set that does not provide heart rate data would not be suitable for use. Density completeness assesses whether a data set contains sufficient amount of data in terms of density, regularity, and duration. For instance, examining the association between step count and blood pressure might require the data set to have ≥10 days of step count data per month for 2 months [28]. The 2 subcategories of completeness, which are breadth and density completeness, were adopted from Weiskopf et al [25].

There was also a significant debate on whether accuracy (*Do the data reflect the true value?*) should be included as a dimension. On one hand, accuracy was considered a dimension that can be easily understood by stakeholders and the ultimate goal of data quality. On the other hand, accuracy was viewed as a vague term that could be interpreted in many different ways. For example, inaccuracy could be an umbrella term that

incorporates invalid data, missing data, or data not conforming to data dictionaries. In addition, accuracy was considered inapplicable for assessing data quality from the secondary use of the data perspective. This is because it is impossible to know whether the data are correct or incorrect in the absence of a known truth. For instance, although the data indicated that an individual took 8 steps at 9 AM on April 5, 2020, there would be no way for a researcher to assess whether that is right or wrong when they retrospectively assess the accuracy of that value. The accuracy of the data values can only be assessed by comparing the device to a gold standard device. In reality, this is not feasible as people rarely wear more than one device in their daily lives, which restricts the ability to assess the accuracy of values in a longitudinal and continuously collected wearable device data. This was why the dimension *plausibility* (*Do the data make sense?*) was eventually included rather than *accuracy*.

Temporal data granularity was another fitness-for-use dimension considered important. As wearable device data are time-series data, the granularity of time points was deemed as an essential aspect. Temporal data granularity is about how frequently the data are documented (eg, every second, minute, or hour) and whether it fits the purpose of the research task. For example, a data set with timestamps every hour would not be suitable for research requiring data points every minute.

Other minor issues mentioned in the literature review and survey were not included as a dimension. For example, survey respondents mentioned the difficulty of interpreting data values, understanding what was really happening while data were being collected, or knowing how the data were collected. This was considered a metadata quality problem rather than a quality metric for the data. The definitions and examples of the final set of dimensions derived from focus group discussions are presented in Table 3.

Table 3. Data quality dimensions for person-generated wearable device data identified by domain experts.

Type and dimension	Definition[a]	Example
Intrinsic		
Conformance: Do data values adhere to specified standards and formats?		
Value conformance	Data values conform to internal formatting constraints, allowable values, or ranges.	• Unit of distance is "miles." • "Sleep stages" only has values "deep," "light," "rem," and "wake," which conform to the data dictionary.
Relational conformance	Assuming there are multiple tables or files, recorded data elements agree with structural constraints imposed by the physical database structures that store data values.	• Participant ID number links to other tables as required. • The wearable device identifier is appropriately linked for all observations.
Computational conformance	Computations used to create derived values from existing variables yield the intended results either within a data set or between data sets.	• Sleep duration conforms to the difference between start time and end time of sleep.
Completeness: Are data values present?	Missing data is determined based on the presence of data. Typically, absence of data is expected if the device is not worn, but this could sometimes be difficult to know retrospectively.	• There is no NA (Not Available) in the step count data.
Plausibility: Are data values believable?		
Uniqueness plausibility	Objects do not appear multiple times in settings where they should not be duplicated or cannot be distinguished within a database or when compared with an external reference.	• A single participant only has one participant ID number.
Atemporal plausibility	Observed data values, distributions, or densities agree with local or "common" knowledge or from comparisons with external sources that are deemed to be trusted or relative gold standards.	• Step count and distance values are positive. • Trends of step counts and distance agree with each other. • Step counts do not show a sudden spike during sleep or during sedentary time. • The range of heart rate values is biologically plausible. • Heart rate is higher when active compared with when sedentary.
Temporal plausibility	Time-varying variables change values as expected based on known temporal properties or across one or more external comparators or gold standards.	• Start time of sleep occurs before end time of sleep. • Aggregate step count is higher during daytime than nighttime.
Contextual and fitness-for-use		
Completeness: Are data values present fit for intended use?		
Breadth completeness	All data types required for intended use exist.	• Heart rate data are essential for studies analyzing the relationship between physical activity and heart rate.
Density completeness	Data set contains a specified number of data values or occurs regularly over a certain period.	• Heart rate should be measured at least once a day. • Sleep data should be recorded every day consecutively for a 6-week period to be considered complete.
Temporal data granularity: does the device collect data granular enough for intended use?	Granularity of time stamps are sufficient for the task at hand.	• Data values are recorded every second, which is appropriate for marathon research studies (the exact start and end time of the marathon for each runner is important for marathon-related studies).

[a]Definitions were adopted and adapted from the studies by Weiskopf et al [25] and Kahn et al [14].

Discussion

Principal Findings

In this study, data quality dimensions for person-generated wearable device data were identified using multiple methods. A literature review and survey was conducted to understand the data quality challenges of researchers and their perceptions on data quality dimensions. On the basis of this information, domain experts determined the appropriate dimensions. Experts agreed that the data quality dimensions from the HIDQF are applicable to person-generated wearable device data, and fitness-for-use dimensions were also considered important, especially for research purposes. The final data quality dimensions deemed important were intrinsic data quality dimensions, such as conformance, completeness, and plausibility, and fitness-for-use data quality dimensions, such as breadth and density completeness and temporal data granularity.

Data Quality Assessment Guidelines for Researchers

Completeness

In this study, breadth and density completeness, which are contextual and fitness-for-use data quality dimensions, were considered important for conducting research. Assessing breadth completeness is important, especially for data sets collected in a bring-your-own-device research setting [9]. This is because different brands and models that users bring may collect different data types, which means that not all individuals in the data set would have all the data types that are needed to answer a research question.

Density completeness is also an essential fitness-for-use dimension for wearable device data because the amount of data sufficient and valid for a specific research task is determined by researchers. Researchers first need to determine how wear versus nonwear of the device is defined. Typically, consumer wearables do not provide information on the wear status; thus, researchers need to make decisions based on existing data. The recorded zero step counts could be due to nonwear (missing data) or it could mean inactivity, and thus researchers need to determine thresholds to define nonwear. An alternative method to determine the wearing status could be based on the existence of heart rate data or the values of heart rate data. For example, Lim et al [4] used the confidence values of heart rate data points as surrogate measures for which *−1* indicates invalid data because the device is not worn or incorrectly worn. This approach opens up the discussion on missing data, whether it should be simply based on the absence of data values or whether the default values for missing data and their semantic meaning should be considered. This was the reason why the fitness-for-use completeness dimensions were considered important.

On the basis of decisions made on wear versus nonwear, researchers can determine the appropriate level of data density for their research. Researchers can first determine the thresholds for how much health behavior data are sufficient for a day. For example, Tang et al [9,29] systematically addressed the incompleteness of physical activity data by presenting heuristic criteria for the definition of a valid day: a day is valid (1) if the step count is above a certain threshold, (2) if the number of hours with data is above a certain threshold, (3) if there are data within 3 periods. Researchers can also define completeness based on the number of valid days needed within a certain data collection period, or how regularly the data should be present for the individual data to be included in the analysis [9]. As recently released devices have the ability to examine various data types and collect data seamlessly for years, further investigation is needed to determine how completeness is characterized in research studies.

Conformance

Value, relational, and computational conformance are all considered important dimensions for wearable device data, but there are challenges in data management and quality assessment. Value and relational conformance can only be assessed in terms of the data dictionary and relational model specific to the brand, model, and version of the device but only if this information is publicly available. In addition, computational conformance can be assessed for values that can be calculated using generic equations, such as sleep duration, which is the difference between the start and end of sleep time. However, it can be difficult to assess computational conformance for variables calculated using proprietary algorithms, as these are not disclosed to data users. Another challenge related to data conformance is the lack of a common data standard for wearable device data. A common data standard would be crucial for a data set collected from disparate devices (eg, Apple Watch and Fitbit Charge HR), such as data collected under a bring-your-own-device protocol. There is a movement in the mobile health community, called Open mHealth, to create a common data schema that explicitly states the format and data definitions for patient-generated data [30]. Adopting these standards for wearable device data might solve the discrepancy between the definition of data values among multi-device data. For example, currently there is no industry standard for defining activity intensity (eg, light, moderate, and vigorous). These challenges indicate that facilitating the use of consumer wearables for research purposes would not be feasible without the support of device companies and the research community.

Plausibility

Plausibility aligns with the needs of researchers for accurate data values. For instance, data may be deemed implausible when step counts are higher than normal, but the corresponding heart rate values are lower than usual. Typically, researchers arbitrarily come up with their own rules to assess the plausibility of data before proceeding with the analysis. However, domain knowledge and a considerable amount of experts' time are required to formulate a set of potential data quality rules. Thus, creating a knowledge base of data quality rules for person-generated wearable device data would not only save time for future researchers but also prevent the use of ad hoc data quality rules [9]. Another challenge for plausibility is that there are few known external benchmarks that can be used to validate or triangulate the data (data quality validation per the HIDQF). For example, the summary statistics of steps, active minutes, and BMI have been compared with the corresponding

values in the Centers for Disease Control and Prevention survey (eg, Behavioral Risk Factor Surveillance System) [31]. Further discussion among the researcher community would be needed to find potential methods or data sources to check the plausibility of data.

Although plausibility was chosen over accuracy as a data quality dimension, it is true that many people are concerned about whether data values are trustworthy. Even though accuracy cannot be assessed in the secondary use of data scenarios, it could be indirectly verified through the results of device validation studies [32-34]. Thus, it is important to provide metadata information on the device brand, model, and version that generated the data set as each element can change device validity [35]. However, knowing the validity and reliability of a device is insufficient to understand the accuracy of data because there are other factors that affect data quality such as incorrect device use by the user. In addition, device validation studies are generally conducted in a controlled setting for a short period.

Limitations

This study has a few limitations. First, the study focuses only on intrinsic and fitness-for-use data quality dimensions and thus does not include extrinsic data quality features, that is, features that affect the data but are not about the data values themselves (eg, security, privacy, or data accessibility). There might be contextual information or metadata that are considered important when determining the fitness-for-use of a data set. For example, some researchers might want to know the process or operational aspects of data collection (eg, Were the data collected under the bring-your-own-device policy or were devices provided?) [36]. These factors were not captured as a data quality dimension, but it is an aspect that might need to be considered when assessing the fitness-for-use of a data set. Second, the study was conducted with a small number of survey respondents and domain experts. Therefore, survey responses and experts' opinions may not be representative and comprehensive. As survey responses match the results of the literature review, it is likely that the survey was able to capture most of the data quality challenges despite the small number of respondents. Furthermore, the intrinsic data quality dimensions identified in this study leveraged the dimensions of the HIDQF. The HIDQF was determined through iterative meetings with stakeholders and data quality experts; thus, it is highly likely that most intrinsic data quality dimensions were included in our final list of dimensions. In future studies, contextual and fitness-for-use data quality dimensions could be further investigated with a larger group of stakeholders of person-generated wearable device data.

Conclusions

Person-generated wearable device data are an emerging data type for biomedical research because of the growing use of wearable devices in people's daily lives. However, there is a lack of agreement on how data quality should be assessed for person-generated wearable device data. As the first step to solve this challenge, data quality dimensions were identified specifically for person-generated wearable device data. We found that data quality dimensions for secondary use of EHR data are applicable to person-generated wearable device data. The identified dimensions will be able to provide guidance to researchers on how data quality is defined and what aspects of data quality should be assessed for person-generated wearable device data. Further research on how data quality can be assessed with regard to dimensions is needed.

Acknowledgments

The authors would like to thank all the respondents who participated in the survey to provide their knowledge and experience. They also thank Carol E Garber for her input on wearable device data. This research was supported by the National Center for Advancing Translational Sciences grant 1U01TR002062-01 and the National Institute of Health's All of Us Research Program grant 1U2COD023196-01.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Dimensions suggested by facilitator and final decision by domain experts.
[DOCX File , 18 KB - mhealth_v9i12e31618_app1.docx]

References

1. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110. [doi: 10.2196/jmir.9157] [Medline: 29567635]
2. Fawcett T. Mining the quantified self: personal knowledge discovery as a challenge for data science. Big Data 2015 Dec;3(4):249-266. [doi: 10.1089/big.2015.0049] [Medline: 27441406]
3. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018 Jul;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]

4. Lim WK, Davila S, Teo JX, Yang C, Pua CJ, Blöcker C, et al. Beyond fitness tracking: the use of consumer-grade wearable data from normal volunteers in cardiovascular and lipidomics research. PLoS Biol 2018 Feb;16(2):e2004285 [FREE Full text] [doi: 10.1371/journal.pbio.2004285] [Medline: 29485983]
5. McDonald L, Mehmud F, Ramagopalan SV. Sleep and BMI: do (Fitbit) bands aid? F1000Res 2018 Apr 27;7:511 [FREE Full text] [doi: 10.12688/f1000research.14774.2] [Medline: 30271578]
6. Xu X, Conomos MP, Manor O, Rohwer JE, Magis AT, Lovejoy JC. Habitual sleep duration and sleep duration variation are independently associated with body mass index. Int J Obes (Lond) 2018 Apr;42(4):794-800. [doi: 10.1038/ijo.2017.223] [Medline: 28895585]
7. All of us research program overview. National Institutes of Health. URL: https://allofus.nih.gov/about/about-all-us-research-program [accessed 2021-12-06]
8. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
9. Cho S, Ensari I, Weng C, Kahn MG, Natarajan K. Factors affecting the quality of person-generated wearable device data and associated challenges: rapid systematic review. JMIR Mhealth Uhealth 2021 Mar 19;9(3):e20738 [FREE Full text] [doi: 10.2196/20738] [Medline: 33739294]
10. DiazVasquez JF. DataGauge: a model-driven framework for systematically assessing the quality of clinical data for secondary use. UT School of Biomedical Informatics. 2016. URL: https://digitalcommons.library.tmc.edu/uthshis_dissertations/33/ [accessed 2021-12-06]
11. Karkouch A, Mousannif H, Al Moatassime H, Noel T. Data quality in internet of things: a state-of-the-art survey. J Netw Comput Applications 2016 Sep;73:57-81. [doi: 10.1016/j.jnca.2016.08.002]
12. Wang R, Guarascio L. Dimensions of data quality: toward quality data by design. Massachusetts Institute of Technology. 1991 Aug. URL: https://tinyurl.com/ynh4j4 [accessed 2021-12-06]
13. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 01;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
14. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016 Sep 11;4(1):1244 [FREE Full text] [doi: 10.13063/2327-9214.1244] [Medline: 27713905]
15. Johnson S, Speedie S, Simon G, Kumar V, Westra B. A data quality ontology for the secondary use of EHR data. AMIA Annu Symp Proc 2015 Nov 5;2015:1937-1946 [FREE Full text] [Medline: 26958293]
16. Wang R, Strong D. Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 2015 Dec 11;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
17. Huang H, Stvilia B, Jörgensen C, Bass HW. Prioritization of data quality dimensions and skills requirements in genome annotation work. J Am Soc Inf Sci 2011 Oct 04;63(1):195-207. [doi: 10.1002/asi.21652]
18. Liu L, Chi L. Evolutional data quality: a theory-specific view. In: Proceedings of the Seventh International Conference on Information Quality (ICIQ 2002). 2002 Presented at: Seventh International Conference on Information Quality (ICIQ 2002); Jan 2002; Cambridge, MA, USA.
19. Codella J, Partovian C, Chang H, Chen C. Data quality challenges for person-generated health and wellness data. IBM J Res Dev 2018 Jan 1;62(1):3:1-3:8. [doi: 10.1147/jrd.2017.2762218]
20. Silverman D. Doing Qualitative Research. London: SAGE Publications; 2010.
21. Qualtrics homepage. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2021-12-10]
22. OHDSI Forums. Observational Health Data Sciences and Informatics (OHDSI). URL: https://forums.ohdsi.org/ [accessed 2021-12-06]
23. Full survey on data quality challenges and dimensions of wearable device data. Columbia University. URL: https://cumc.co1.qualtrics.com/jfe/preview/SV_bNlmgCekd9FC3Gd?Q_CHL=preview&Q_SurveyVersionID=current [accessed 2021-12-05]
24. Zozus MN, Pieper C, Johnson CM, Johnson TR, Franklin A, Smith J, et al. Factors affecting accuracy of data abstracted from medical records. PLoS One 2015 Oct 20;10(10):e0138649 [FREE Full text] [doi: 10.1371/journal.pone.0138649] [Medline: 26484762]
25. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform 2013 Oct;46(5):830-836 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.010] [Medline: 23820016]
26. Mailey EL, Gothe NP, Wójcicki TR, Szabo AN, Olson EA, Mullen SP, et al. Influence of allowable interruption period on estimates of accelerometer wear time and sedentary time in older adults. J Aging Phys Act 2014 Apr;22(2):255-260 [FREE Full text] [doi: 10.1123/japa.2013-0021] [Medline: 23752299]
27. Evenson KR, Terry JW. Assessment of differing definitions of accelerometer nonwear time. Res Q Exerc Sport 2009 Jun;80(2):355-362 [FREE Full text] [doi: 10.1080/02701367.2009.10599570] [Medline: 19650401]

28. Menai M, Brouard B, Vegreville M, Chieh A, Schmidt N, Oppert J, et al. Cross-sectional and longitudinal associations of objectively-measured physical activity on blood pressure: evaluation in 37 countries. Health Promot Perspect 2017;7(4):190-196 [FREE Full text] [doi: 10.15171/hpp.2017.34] [Medline: 29085795]
29. Tang LM, Meyer J, Epstein DA, Bragg K, Engelen L, Bauman A, et al. Defining adherence: making sense of physical activity tracker data. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-22. [doi: 10.1145/3191769]
30. Estrin D, Sim I. Open mHealth architecture: an engine for health care innovation. Science 2010 Nov 05;330(6005):759-760. [doi: 10.1126/science.1196187] [Medline: 21051617]
31. Evenson KR, Wen F, Furberg RD. Assessing validity of the Fitbit indicators for U.S. public health surveillance. Am J Prev Med 2017 Dec;53(6):931-932 [FREE Full text] [doi: 10.1016/j.amepre.2017.06.005] [Medline: 28755981]
32. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. J Med Eng Technol 2016 Aug;40(6):298-306. [doi: 10.1080/03091902.2016.1193238] [Medline: 27266422]
33. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. Ann Intern Med 2017 Apr 11;166(8):610-612. [doi: 10.7326/L16-0353] [Medline: 28395305]
34. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16(1):433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
35. Collins T, Woolley SI, Oniani S, Pires IM, Garcia NM, Ledger SJ, et al. Version reporting and assessment approaches for new and updated activity and heart rate monitors. Sensors (Basel) 2019 Apr 10;19(7):1705 [FREE Full text] [doi: 10.3390/s19071705] [Medline: 30974755]
36. Reynolds MW, Bourke A, Dreyer NA. Considerations when evaluating real-world data quality in the context of fitness for purpose. Pharmacoepidemiol Drug Saf 2020 Oct;29(10):1316-1318 [FREE Full text] [doi: 10.1002/pds.5010] [Medline: 32374042]

Abbreviations

EHR: electronic health record
HIDQF: harmonized intrinsic data quality framework
PGHD: person-generated health data

Edited by L Buis; submitted 28.06.21; peer-reviewed by K Burns, T Sagi; comments to author 28.08.21; revised version received 27.09.21; accepted 11.11.21; published 23.12.21.

Please cite as:
Cho S, Weng C, Kahn MG, Natarajan K
Identifying Data Quality Dimensions for Person-Generated Wearable Device Data: Multi-Method Study
JMIR Mhealth Uhealth 2021;9(12):e31618
URL: https://mhealth.jmir.org/2021/12/e31618
doi:10.2196/31618
PMID:34941540

Original Paper

A Smartphone-Based Decision Support Tool for Predicting Patients at Risk of Chemotherapy-Induced Nausea and Vomiting: Retrospective Study on App Development Using Decision Tree Induction

Abu Saleh Mohammad Mosa[1,2,3,4], MS, PhD; Md Kamruz Zaman Rana[1,2], MS; Humayera Islam[2,4,5], MS; A K M Mosharraf Hossain[6,7], MD; Illhoi Yoo[1†], PhD

[1]Health Management and Informatics, University of Missouri School of Medicine, Columbia, MO, United States
[2]Institute for Data Science and Informatics, University of Missouri, Columbia, MO, United States
[3]Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States
[4]Center for Biomedical Informatics, University of Missouri School of Medicine, Columbia, MO, United States
[5]Institute of Statistical Research and Training, University of Dhaka, Dhaka, Bangladesh
[6]Ellis Fischel Cancer Center, University of Missouri School of Medicine, Columbia, MO, United States
[7]Department of Hematology and Medical Oncology, BayCare Health System, South Florida Baptist Hospital, Plant City, FL, United States
[†]deceased

Corresponding Author:
Abu Saleh Mohammad Mosa, MS, PhD
Health Management and Informatics
University of Missouri School of Medicine
1 Hospital Drive
Columbia, MO, 65212
United States
Phone: 1 573 882 8349
Email: mosaa@health.missouri.edu

Abstract

Background: Chemotherapy-induced nausea and vomiting (CINV) are the two most frightful and unpleasant side effects of chemotherapy. CINV is accountable for poor treatment outcomes, treatment failure, or even death. It can affect patients' overall quality of life, leading to many social, economic, and clinical consequences.

Objective: This study compared the performances of different data mining models for predicting the risk of CINV among the patients and developed a smartphone app for clinical decision support to recommend the risk of CINV at the point of care.

Methods: Data were collected by retrospective record review from the electronic medical records used at the University of Missouri Ellis Fischel Cancer Center. Patients who received chemotherapy and standard antiemetics at the oncology outpatient service from June 1, 2010, to July 31, 2012, were included in the study. There were six independent data sets of patients based on emetogenicity (low, moderate, and high) and two phases of CINV (acute and delayed). A total of 14 risk factors of CINV were chosen for data mining. For our study, we used five popular data mining algorithms: (1) naive Bayes algorithm, (2) logistic regression classifier, (3) neural network, (4) support vector machine (using sequential minimal optimization), and (5) decision tree. Performance measures, such as accuracy, sensitivity, and specificity with 10-fold cross-validation, were used for model comparisons. A smartphone app called CINV Risk Prediction Application was developed using the ResearchKit in iOS utilizing the decision tree algorithm, which conforms to the criteria of explainable, usable, and actionable artificial intelligence. The app was created using both the bulk questionnaire approach and the adaptive approach.

Results: The decision tree performed well in both phases of high emetogenic chemotherapies, with a significant margin compared to the other algorithms. The accuracy measure for the six patient groups ranged from 79.3% to 94.8%. The app was developed using the results from the decision tree because of its consistent performance and simple, explainable nature. The bulk questionnaire approach asks 14 questions in the smartphone app, while the adaptive approach can determine questions based on the previous questions' answers. The adaptive approach saves time and can be beneficial when used at the point of care.

Conclusions: This study solved a real clinical problem, and the solution can be used for personalized and precise evidence-based CINV management, leading to a better life quality for patients and reduced health care costs.

(JMIR Mhealth Uhealth 2021;9(12):e27024) doi:10.2196/27024

KEYWORDS

chemotherapy; CINV risk factors; data mining; prediction; decision trees; clinical decision support; smartphone app

Introduction

Background

Chemotherapy is a drug treatment commonly used to treat nearly every type of cancer [1]. As estimated, each year, as many as 1 million Americans receive some type of chemotherapy [2]. Cancer cells multiply at an unusually faster rate compared to healthy cells, and chemotherapy is used to kill those fast-growing cells in the body. However, chemotherapy can lead to many side effects, such as nausea, vomiting, appetite changes, anemia, hair loss, constipation, and diarrhea, among others [3-11]. Chemotherapy-induced nausea and vomiting (CINV) are the two most frightful and unpleasant side effects of chemotherapy [3,4,12-15].

CINV can lead to consequences that affect both patients and the health care system as a whole. First, CINV engenders other side effects, such as nutritional deficits, dehydration, and electrolyte imbalance, which diminishes the quality of life in cancer patients [16-20]. Second, the various side effects of CINV lead to a low-quality social life [19,21]. Third, CINV can also lead to loss of workdays, which in return increases the economic burden [19,22-24]. Fourth, CINV surges health care costs arising from CINV-related outpatient visits, hospitalization, and the cost of drugs [18,19,22-27]. Fifth, intolerance of cancer patients toward CINV can lead to discontinuation of cancer treatment, leading to poor treatment outcomes, treatment failure, or even death [12,28-30].

The management of CINV is a complex process due to two factors. The first level of complexity arises from the different impacts of the different emetogenicity levels of the chemotherapeutic agents. The emetogenicity of chemotherapy is fractionated into four emetic risk categories based on the percentage of patients who suffer from CINV without antiemetics: (1) minimal (<10%), (2) low-emetogenic chemotherapy (LEC: 10%-30%), (3) moderate-emetogenic chemotherapy (MEC: 30%-90%), and (4) high-emetogenic chemotherapy (HEC: >90%). CINV has two different pathophysiological phases (acute and delayed) that can lead to different consequences, adding a second level of complexity. The acute phase of CINV occurs within the first 24 hours of chemotherapy. Chemotherapy triggers the release of serotonin in the peripheral pathway (gastrointestinal tract), which binds to the 5-hydroxytryptamine (5-HT_3) receptors and sends a signal to the vomiting center in the medulla [31,32]. The central pathway is associated with the delayed phase of CINV that occurs after the first 24 hours of chemotherapy administration and may persist up to 1 week. This pathway is located in the brain, where chemotherapy triggers a neuropeptide release named substance P, which binds to the neurokinin-1 (NK-1) receptor in the vomiting center, causing CINV [31,32].

There are several antiemetic guidelines for the management of CINV, such as the American Society of Clinical Oncology (ASCO) guideline [33,34], the National Comprehensive Cancer Network (NCCN) guideline [35], and the guideline from the Multinational Association of Supportive Care in Cancer (MASCC) in cooperation with the European Society of Medical Oncology (ESMO) [36]. Despite the improvements in CINV management, many recent studies have reported various percentages of patients experiencing CINV with the use of antiemetics: 28% [37], 38%-52% [38], 56.1% [39], 61.2% [19], and 62% [20]. The guideline-recommended standard antiemetic prophylaxis takes only the chemotherapeutic emetogenicity into consideration for CINV management.

However, several patient-related risk factors can potentially worsen the risk of CINV, but none of the guidelines considers those factors [40]. Since physicians cannot entirely rely on the guidelines, they use their own experiences to manage CINV. Consequently, CINV management is inconsistent among physicians, since their decisions are subjective to their experiences in managing CINV [41].

The use of risk prediction algorithms for clinical decision making at the point of care would require completing and processing massive patient panels, which can be time consuming and can lead to inaccurate results [42]. In recent years, smartphones have become popular among physicians for accessing health care information at the point of care [43]. The advent of open-source frameworks, such as Apple ResearchKit, Apple CareKit, and Android frameworks (eg, PhoneGap), has opened up tremendous opportunities to capture patient-related data and deliver patient-specific clinical decision support information through smartphones. Data mining techniques are beneficial in predictive analytics on medical data [44]. Various machine learning (ML) algorithms have the potential to help build robust clinical decision support systems using clinical data. Smartphone apps integrated with robust clinical decision support developed from rigorously validated ML models and artificial intelligence (AI) can be immensely useful for clinicians and can significantly improve overall health care delivery.

Objective

The objective of this study was to develop a smartphone app for clinical decision support to predict patients' risk of CINV using patient-related risk factors. ML algorithms, such as the decision tree, naive Bayes algorithm, logistic regression classifier, neural network, and support vector machine, were applied to determine the best-performing algorithm for CINV risk prediction based on electronic medical records (EMRs). Standard performance metrics, such as accuracy, sensitivity,

and specificity, were used to compare the performance among the algorithms. This paper also illustrates the use of the ML model to develop a smartphone app and demonstrates its usage from the users' perspective. The developed app aims to help clinicians identify high-CINV-risk patients and can be integrated with antiemetic guidelines for better CINV management.

Methods

Data Sources and Population Selection

This was a retrospective study, and data were collected from the EMRs from a single center called the University of Missouri Ellis Fischel Cancer Center. The study was approved by the MU Health Sciences institutional review board. Our study included only patients who received chemotherapy and standard antiemetic prophylaxis (based on national antiemetic guidelines) at the oncology outpatient service from June 1, 2010, to July 31, 2012. However, we excluded patients with missing information and those who underwent concurrent radiotherapy or surgical procedures.

We planned to collect two independent data sets for each stage of CINV. Since acute and delayed CINV follows two different pathophysiologies, we planned to discover the patient-related risk factors for causing CINV during both phases independently. In each data set, there were three groups based on the emetogenicity level of the chemotherapy regimens. Of the four emetic risk categories, the minimal risk category of chemotherapy for causing CINV is not clinically crucial, since only less than 10% of those patients suffer from CINV. Thus, we collected data in three separate groups corresponding to three clinically meaningful categories: low, moderate, and high.

Our significant interest classes included both CINV and non-CINV cases. However, LEC led to CINV in less than 30% of patients, and the use of standard antiemetic treatment further reduced this percentage. Thus, the data set had few CINV cases compared to non-CINV cases. In addition, the number of CINV cases was higher than the non-CINV cases in the HEC group. Hence, class balancing in each data group (LEC, MEC, and HEC) was considered necessary. We addressed the class imbalance issue by making the data set's size in each class for each group approximately equal.

Variable Selection

In a previous study, we completed a systematic review by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline to identify potential patient-related variables that cause CINV [45]. Our previous study used MEDLINE to identify articles that demonstrated patient-related risk factors of CINV through clinical studies. A total of 26 patient-related risk factors were documented in that study from reviewing 49 articles [46]. For this study, we included 14 independent variables and 1 dependent variable (CINV outcome) [46]. We chose the risk factors based on the recommendations from chemotherapy experts in the MU Elis Fischel Cancer Center and our literature review. The selected variables were also easy to collect through clinical encounters, which can facilitate the usability of the prediction model at the point of care before chemotherapy.

Data Mining

Data mining or knowledge discovery in databases (KDD) can discover hidden patterns, previously unknown, and potentially useful information from data. In general, data mining algorithms are categorized into two groups: descriptive or unsupervised learning and predictive or supervised learning. In supervised learning, the class labels of the observations or tuples are known, whereas in unsupervised learning, those class labels are unknown. For this study, we developed a prediction model that falls into the supervised learning or classification category.

Classification is a supervised learning method for building classification models based on a data set (called training data) and the values in classifying attributes (called a class label). The classification model is used to predict the categorical class label. Classification is a two-step process in which the model is constructed in the first step and the accuracy of the model is determined using a data set (called test data set) in the second step. The accuracy of the classification model is the percentage of test data set tuples that are correctly classified by the model. To overcome the overfitting problem, the test data set must be independent of the training data set. In general, the classification model consists of IF-THEN rules or mathematical formulas. For our study, we used five popular data mining algorithms: (1) naive Bayes [47], (2) logistic regression classifier [48], (3) neural network (voted perceptron) [49], (4) support vector machine (using sequential minimal optimization) [50], and (e) decision tree [51-53]. There are several tools available for data mining. We used the most widely used tools, called WEKA [54]. Performance measures, such as accuracy, sensitivity, and specificity, were used for model comparisons. In addition, 10-fold cross-validation were used for model validation [55].

Smartphone App Development

ResearchKit is an open-source framework based on iOS that makes it easy to create mobile apps. It allows researchers and drug developers to tailor it to their own particular needs, whether for collecting clinical research data, recruiting patients, or obtaining informed consent. The framework allows for collecting information through electronic data capture, creating a small task to gather any specific information required for the study, and then storing the data as part of a sandbox, thereby protecting patient information. We developed our smartphone app using some modules, including a survey engine, visual consent flow, and active tasks from this framework. As the users of this app will be care providers, and no identifiable data will be stored, we did not use the visual consent flow. The smartphone app was built using the algorithm that had the most consistent performance among the ML algorithms and is also explainable, usable, and actionable AI for clinical decision support.

Results

Data Summary

In total, 6124 records were extracted based on inclusion and exclusion criteria. The number of records was 3053 and 3071 for the acute-phase and the delayed-phase data set, respectively. Table 1 presents the frequency distribution of both data sets for

combinations of three chemotherapy categories and two treatment outcomes.

Table 1. Data summary.

CINV[a] treatment group	Records, n	CINV, n (%)	No CINV, n (%)
Acute phase			
HECb[b]	1026	504 (49.12)	522 (50.88)
MEC[c]	1012	506 (50.00)	506 (50.00)
LEC[d]	1015	506 (50.15)	509 (49.85)
Total	3053	1519 (49.75)	1534 (50.25)
Delayed phase			
HEC	1166	586 (50.26)	580 (49.74)
MEC	891	447 (50.17)	444 (49.83)
LEC	1014	519 (51.18)	495 (48.82)
Total	3071	1552 (50.54)	1519 (49.46)

[a]CINV: chemotherapy-induced nausea and vomiting.

[b]HEC: high-emetogenic chemotherapy.

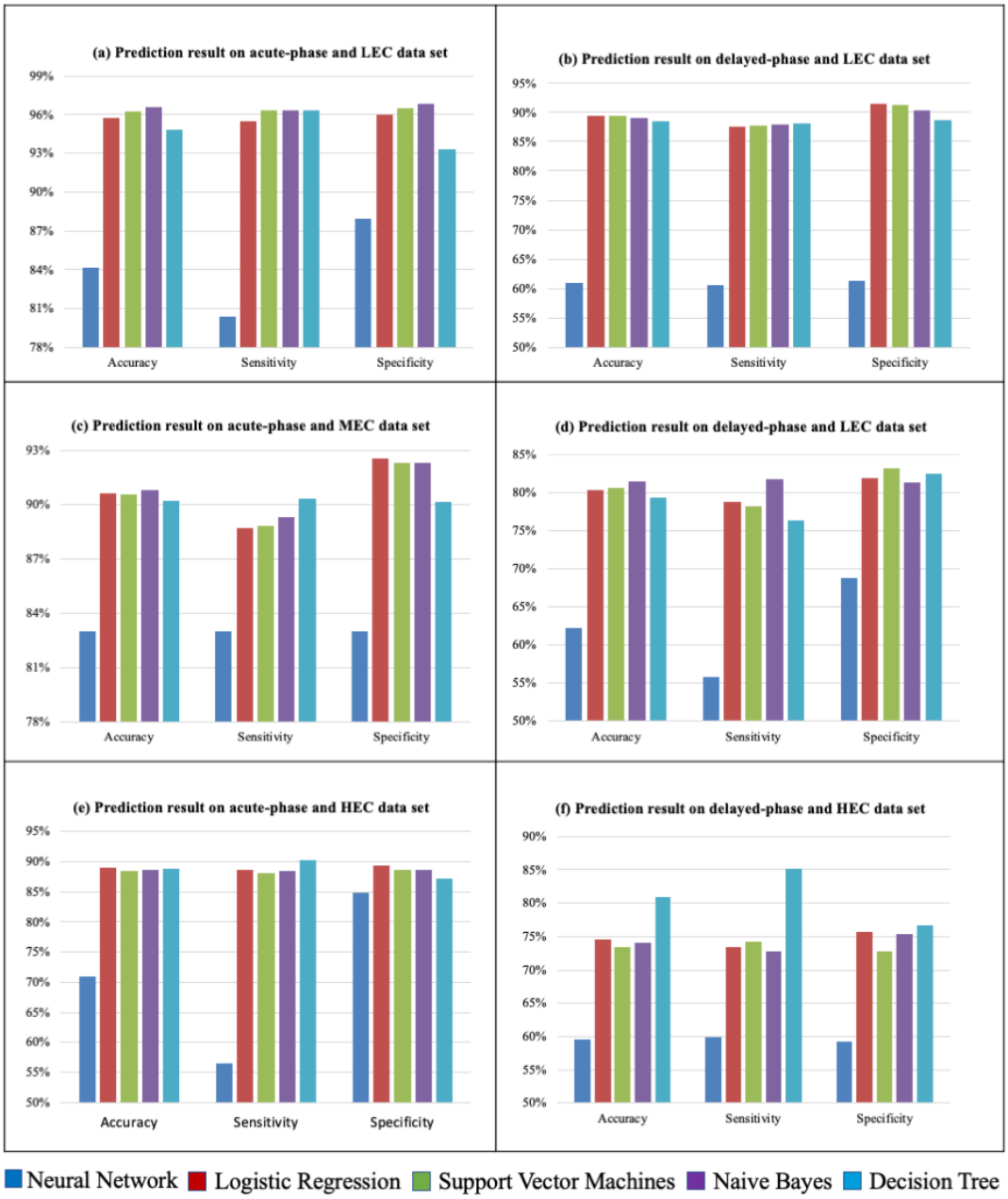
[c]MEC: moderate-emetogenic chemotherapy.

[d]LEC: low-emetogenic chemotherapy.

Data Mining Model Performance Comparison

The models' performances for all the emetogenicity levels and CINV phases (accuracy, sensitivity, specificity) are compared in Figure 1. The differences between performances of the different models were not consistent in each data set's model. The naive Bayes algorithm showed the best performance in the acute phase for LEC (the accuracy was 96.6%, sensitivity was 96.3%, and specificity was 96.8%), the acute phase for MEC (the accuracy was 90.8%, sensitivity was 89.3%, and specificity was 92.3%), and the delayed phase for MEC (the accuracy was 81.5%, sensitivity was 81.7%, and specificity was 81.3%). For the delayed phase for LEC, the support vector machine gave the best performance (the accuracy was 89.5%, sensitivity was 87.8%, and specificity was 91.3%).

Figure 1. Accuracy, sensitivity, and specificity of different ML algorithms used to predict CINV status among patients. CINV: chemotherapy-induced nausea and vomiting; HEC: high-emetogenic chemotherapy; LEC: low-emetogenic chemotherapy; MEC: moderate-emetogenic chemotherapy; ML, machine learning.



The decision tree gave the most consistent performance in both phases of HEC, with a significant margin compared to the other algorithms. Although different algorithms gave the best performance for different stages, we selected the decision tree model to develop the app for its consistent performance across measures and its simple, explainable nature. Moreover, clinical decision support integrated with explainable, usable, and actionable AI is more convenient for oncologists to understand, and thus, it can help them understand the app's background functioning.

Decision Tree Models

The six decision tree models for predicting CINV in both acute and delayed phases for each type of emetogenicity resulted in six flowcharts (Figures 2-7). Table 2 shows the description of the abbreviated form of each patient-related risk factor shown in the decision trees. We optimized the confidence factor for tree size and used the same confidence factor for all the decision trees. A threshold of >0 was used as the cutoff point. The accuracy of the six models was 94.8%, 88.5%, 90.2%, 79.3%, 88.7%, and 81%, respectively. In addition, sensitivity (correct prediction for the positive outcome of CINV) measures were 96.3%, 88.2%, 90.3%, 76.3%, 90.3%, and 85.2%, respectively,

while specificity (correct prediction for the negative outcome of CINV) measures were 93.3%, 88.7%, 90.1%, 82.4%, 87.2%, and 76.7%, respectively.

Figure 2. Decision tree. Phase: acute; emetogenicity: low. CINV: chemotherapy-induced nausea and vomiting.

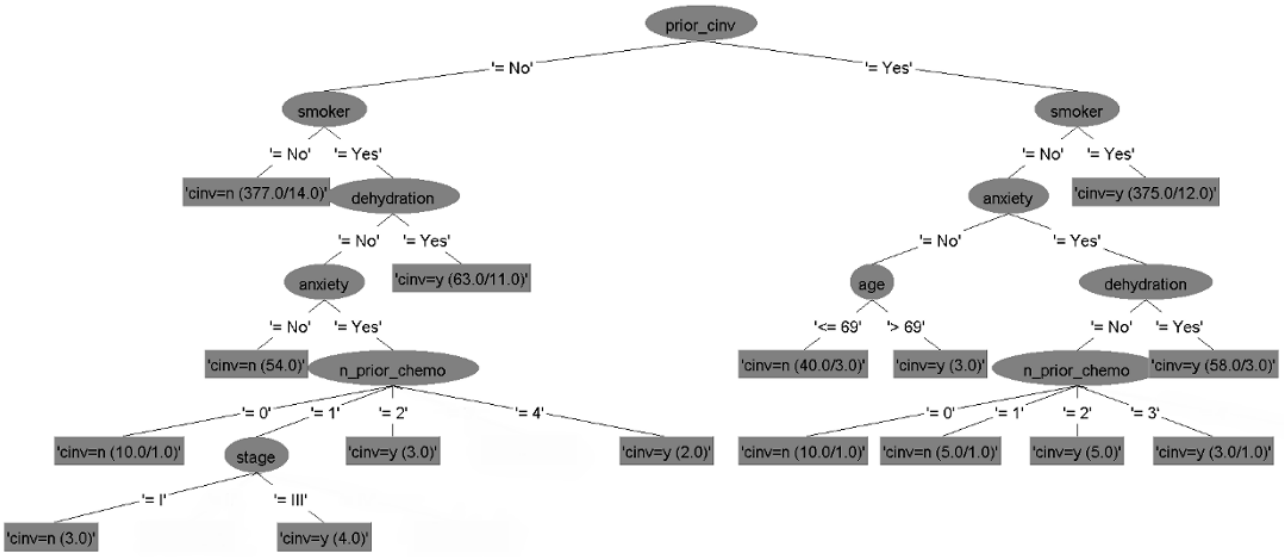


Figure 3. Decision tree. Phase: delayed; emetogenicity: low. BMI: body mass index; CINV: chemotherapy-induced nausea and vomiting.

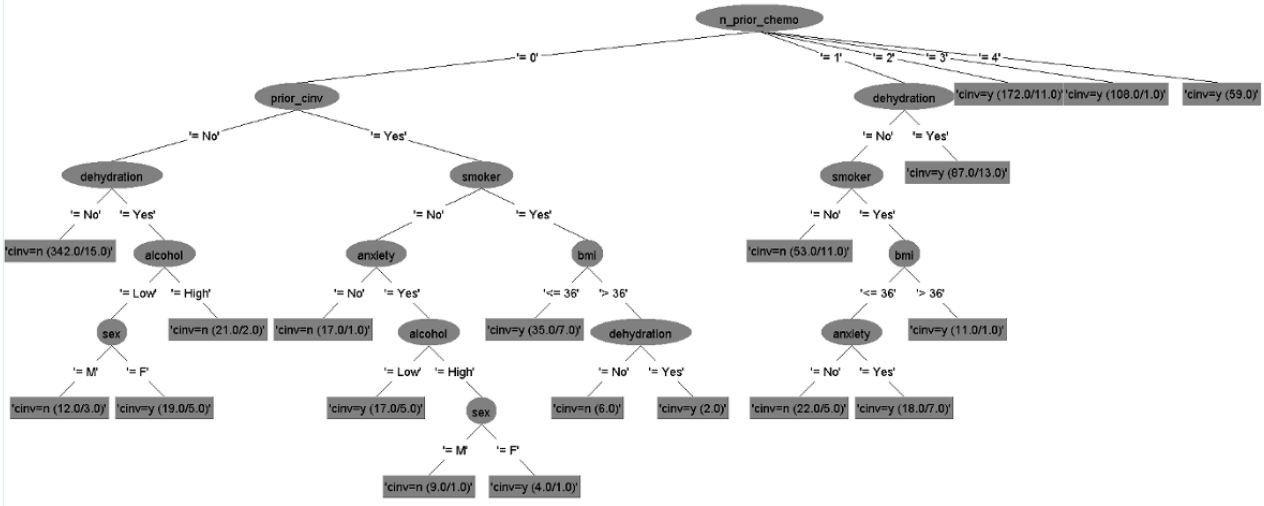


Figure 4. Decision tree. Phase: acute; emetogenicity: moderate. CINV: chemotherapy-induced nausea and vomiting.

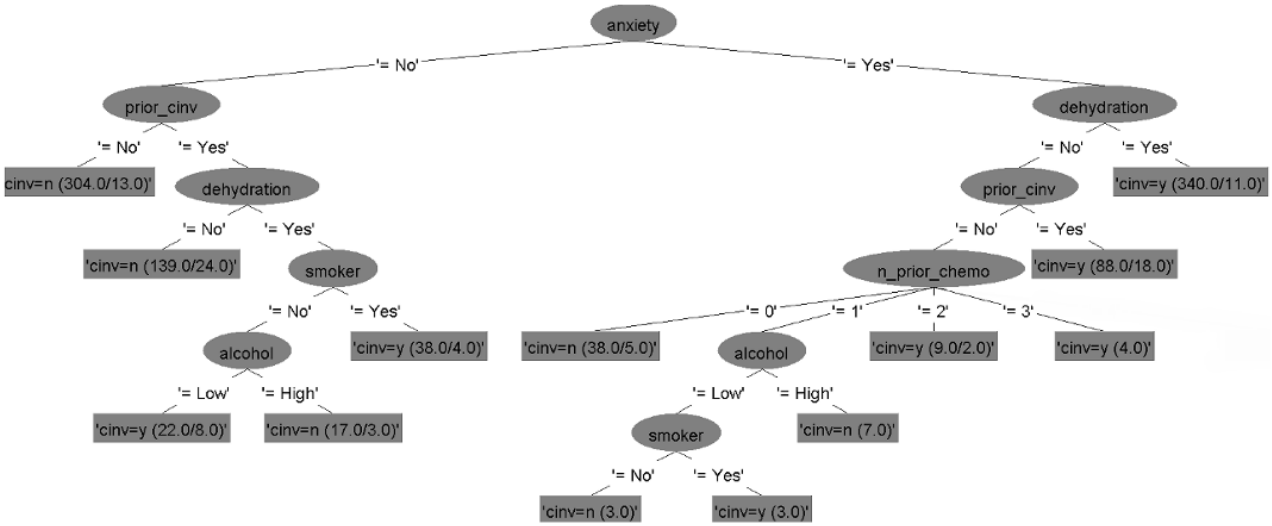


Figure 5. Decision tree. Phase: delayed; emetogenicity: moderate. CINV: chemotherapy-induced nausea and vomiting.

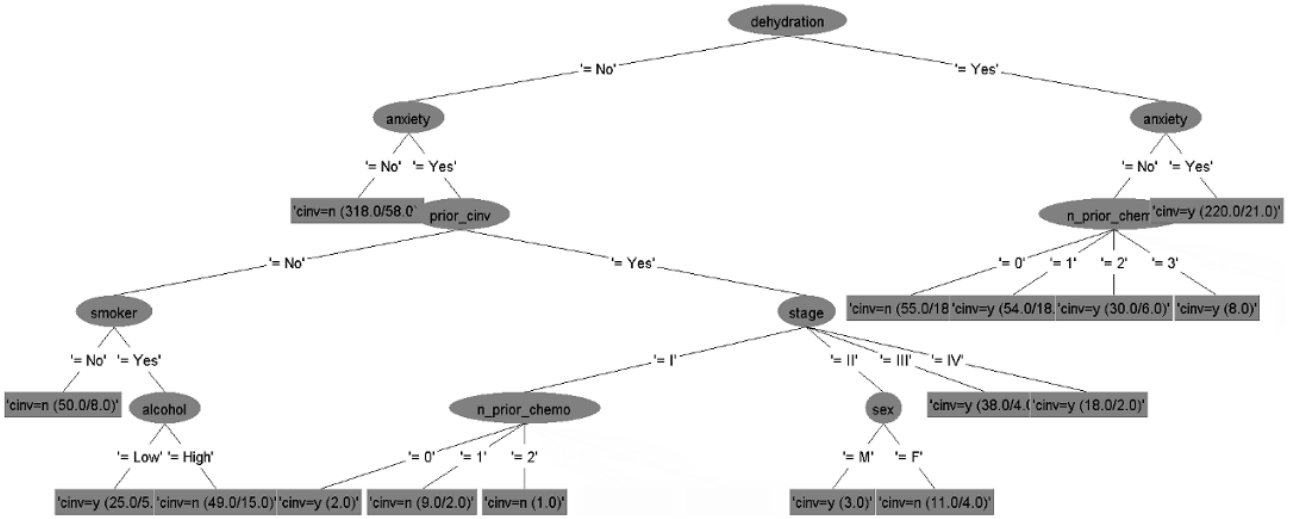


Figure 6. Decision tree. Phase: acute; emetogenicity: high. BMI: body mass index; CINV: chemotherapy-induced nausea and vomiting.

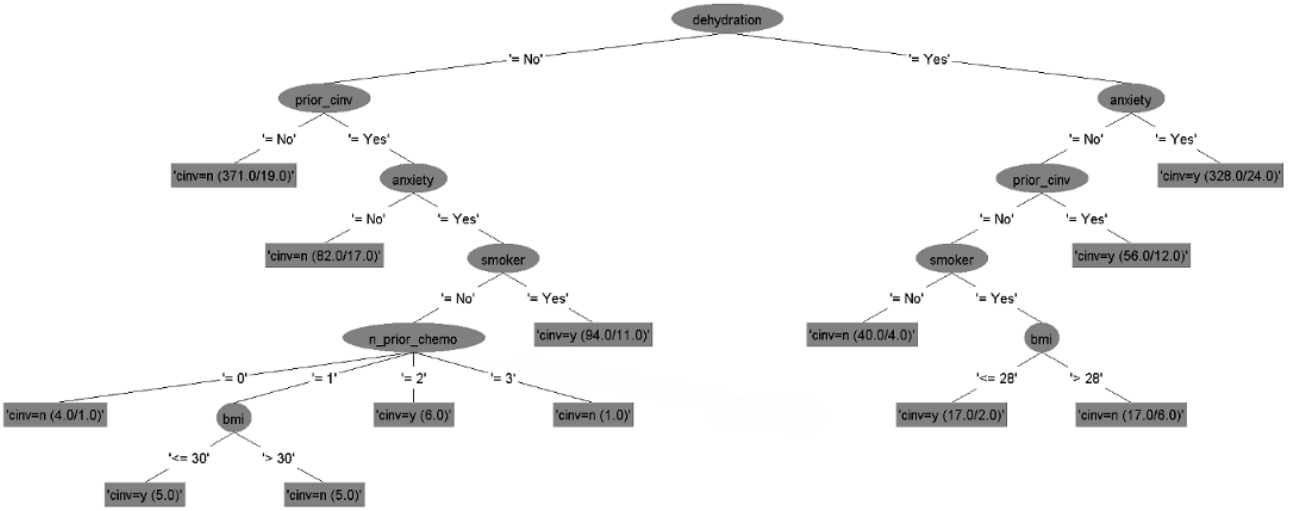


Figure 7. Decision tree. Phase: delayed; emetogenicity: high. CINV: chemotherapy-induced nausea and vomiting.

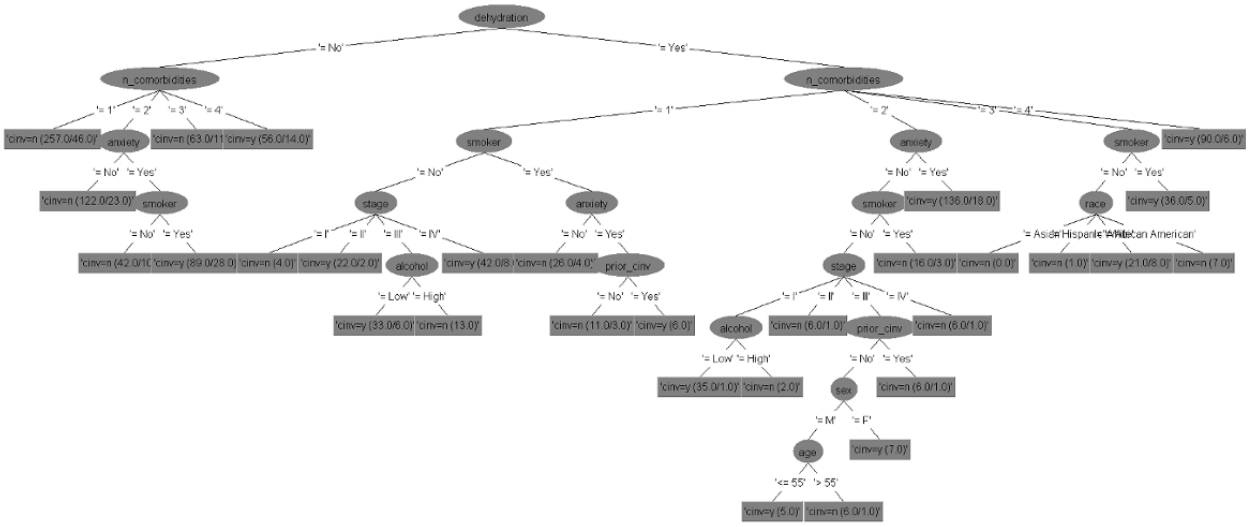


Table 2. Patient-related risk factors and their abbreviations used in the decision trees.

Risk factor abbreviation	Description
smoker	Is the patient a current smoker?
race	Race of the patient
age	Age of the patient in years
bmi	Body mass index during chemotherapy
anxiety	Did the patient have anxiety during chemotherapy?
prior_cinv	History of previous CINV[a]
n_prior_chemo	Number of prior chemotherapy regimen
n_comorbidities	Number of comorbidities
sex	Sex of the patient
alcohol	Alcohol consumption
stage	Stage of cancer
type	Type of cancer
dehydration	Did the patient have dehydration during chemotherapy?

[a]CINV: chemotherapy-induced nausea and vomiting.

Clinical Decision Support Smartphone App

The clinical decision support smartphone app for CINV was developed using the output of the decision tree models obtained from the above analyses. The app was built on iOS and developed considering space usage and the possible variation of its users' technological skills. We created active tasks, depending on the flowcharts. In addition, the survey engine helped us to easily implement the questionnaire survey.

The app was created using two different approaches: (1) the bulk questionnaire approach and (2) the adaptive questionnaire approach. In the bulk questionnaire approach, all 14 questions regarding CINV risk factors were asked one by one. After receiving the responses of the patients on all the questions, the predictive analyzer predicted the recommendations on both phases, depending on the six flowcharts obtained by applying the decision tree algorithm. In Figure 8, the flow for the bulk approach is shown. For a better experience, the clinician has the freedom to go back and change the input and recalculate the answer. An example of a set of answers is given in Figure 9. Depending on all the answers and using the six flowcharts' logic, the system selects the result for both the acute and delayed phases and displays it.

Figure 8. Flow diagram of CINV risk prediction smartphone app using the bulk questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface.

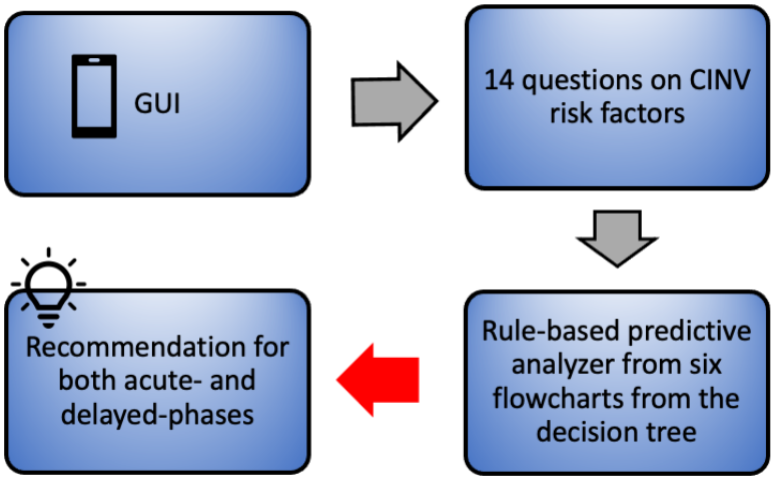
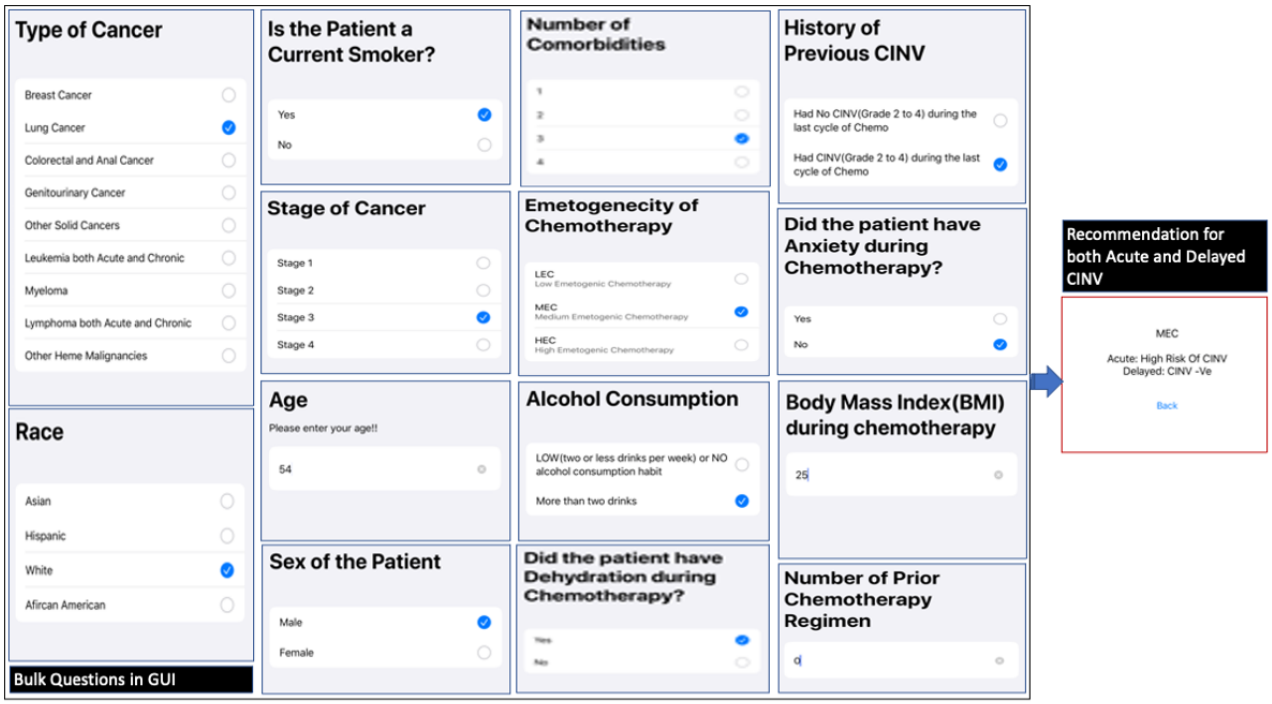


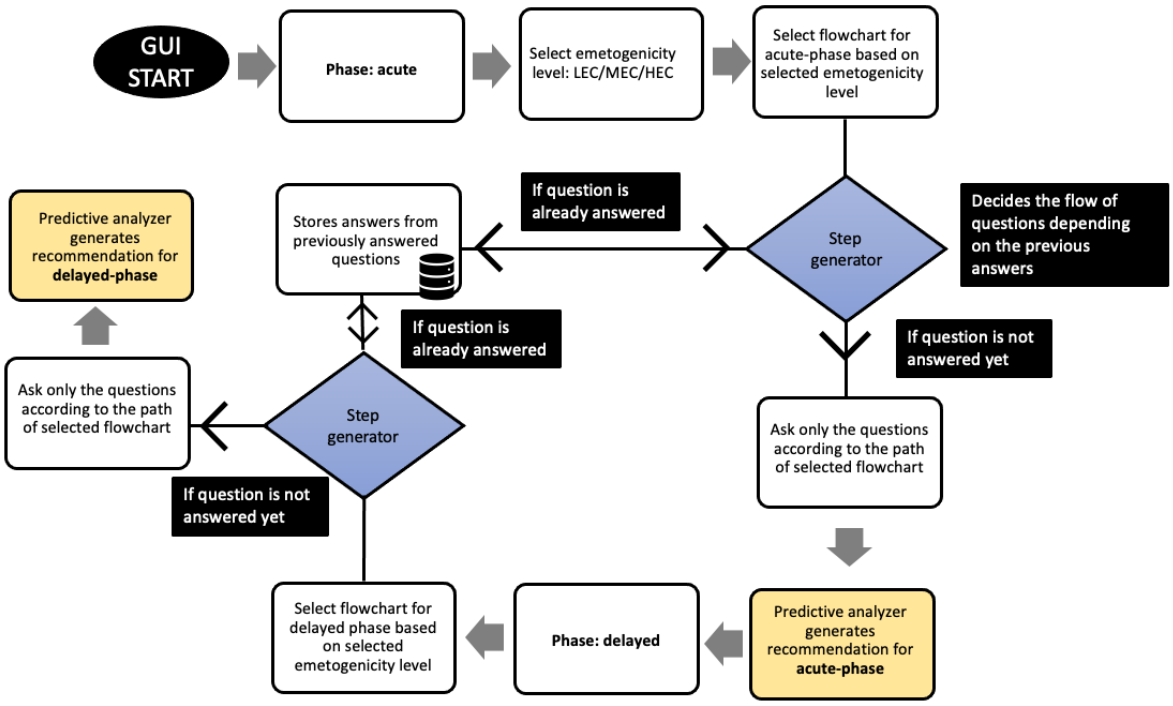
Figure 9. Application GUI for the bulk questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface.



The main limitation of the bulk questionnaire approach is that the physician at the point of care has to answer all the 14 questions to get to the final recommendation, even though not all questions are required for decision making for that patient. ResearchKit allows us to customize the questionnaire by adding features such as skipping questions or creating multiple paths, depending on the answer of the parent node of the decision trees. However, in this study, risk factors did not form a consistent hierarchy across the flowcharts, and thus skipping questions from a fixed questionnaire did not help. Moreover, some of the flowcharts had the same child under the parent node regardless of the answer, following different paths afterward. For instance, in Figure 5, the parent node is dehydration but the child node is anxiety regardless of whether dehydration is true or false. This motivated us to build a more time-energy-efficient approach called the adaptive questionnaire approach.

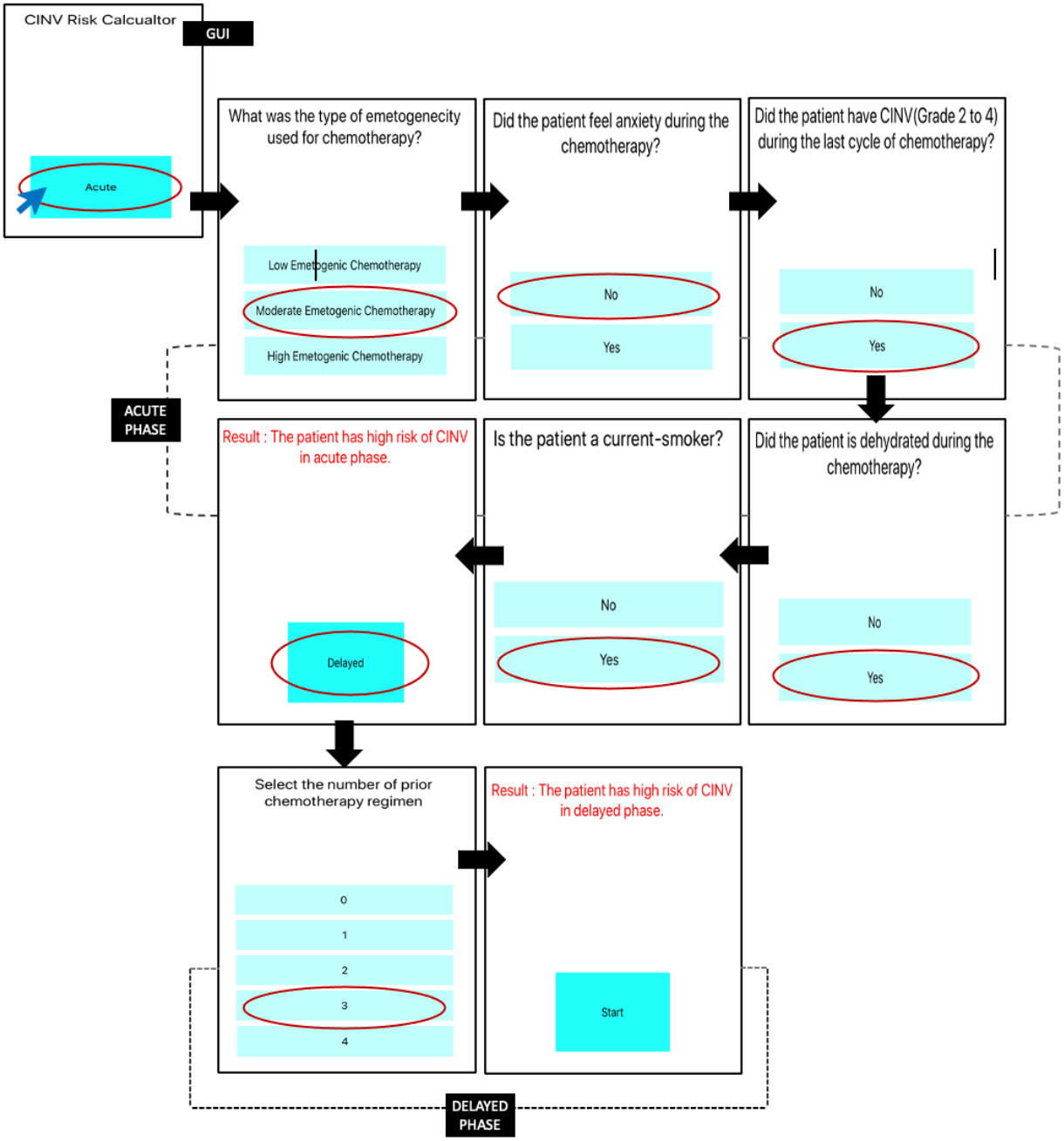
In an adaptive approach, the rule-based system first chooses a flowchart for the acute phase, depending on the emetogenicity level. A flowchart can have different paths, depending on the answers to the question as they come in the hierarchy of the decision tree. This approach follows a single path from the flowchart to generate a questionnaire for the clinician and saves all the answers in a database. Upon recommending the acute phase, the rule-based system chooses another flowchart for the delayed phase. This time, not all the questions in that flowchart are asked; instead, the app asks only the unanswered questions. There is a step generator feature at play for both acute and delayed phase prediction. The step generator determines the question paths for the patient, generates a new step if the question is unanswered, and use the answer from the saved answers for the already answered questions to generate the recommendation. In this approach, only the minimum questions needed to give a recommendation are included in the questionnaire, making the app more effective, faster, and user friendly. In Figure 10, the flow for the adaptive survey approach is shown.

Figure 10. Flow diagram of CINV risk prediction smartphone app using the adaptive questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface; HEC: high-emetogenic chemotherapy; LEC: low-emetogenic chemotherapy; MEC: moderate-emetogenic chemotherapy.



For the adaptive approach, the app's questionnaire comes in dynamic format. The flow of the adaptive approach for a single path is illustrated in Figure 11. In this scenario, the user selected MEC as the emetogenicity of chemotherapy for the acute phase, and the model chose the decision tree for acute MEC shown in Figure 4. According to this flowchart, the first question was “whether the patient had anxiety during the chemotherapy,” for which the user selected ”no“ as an answer. Following this answer, the next question was ”the history of previous chemotherapy.” The user selected “yes,” which led to the next question about ”dehydration.” Since the answer was “yes” for dehydration, the next question was about “smoking status.” Only by asking these four questions, the system identified that the patient is at high risk of CINV. Although there are 14 risk factors, our dynamic approach only asks the questions that are necessary, choosing one pathway from the flowchart, which depends on the answers to the previous questions.

Figure 11. Application GUI for the adaptive questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface.



The user started back again and now selected the delayed phase in the app for determining CINV risk. The system selected the flowchart from Figure 5 this time. One advantage of this adaptive approach is that it will not ask questions that have already been answered. For example, although in the delayed-phase flowchart, the first question was about dehydration, this was not asked, since this was already answered in the acute-phase mode. The question of anxiety was also skipped for the same reason. The third question in the delayed-phase flowchart was about "the number of prior chemotherapy regimens." Since this question was never asked, the system picked this question next and the user selected 3 as an answer. Thus, using the answers to these questions, the app generated the recommendation that the patient has a high risk of CINV in the delayed phase.

Discussion

Principal Findings

CINV is a major side effect of chemotherapy among cancer patients. Appropriate examination of patient-specific risk factors before selecting premedications for CINV is critical in cancer care [56]. Better control of CINV has both short- and long-term effects in cancer care, leading to improved therapy tolerability, less anxiety, higher patient satisfaction, and avoidance of immediate discontinuation of the treatment [28,57-59]. Our previous study on finding risk factors through a systematic literature review shed light on the prevalent risk factors of CINV, as seen in the existing literature [46]. Patient-specific factors, such as smoking and alcohol status, sex, age, and the body mass index (BMI), can play a vital role in determining

their effect on CINV. This study used data mining to discover significant relationships among the patient-related risk factors that influence the occurrence of CINV. Six independent data sets (three chemotherapy groups and two phases of CINV in each chemotherapy group) were individually analyzed to build the best-possible prediction models for CINV prediction. The risk factors used for building the models can be easily collected at the point of care or are available in the hospital EMRs. Among the popular data mining algorithms used for our study, the decision tree model performed consistently across the measures for both CINV phases.

A rule-based app can be considered an appropriate choice for its simplicity in explaining the model to a clinician and implementing it in a software application. Thus, we developed a CINV smartphone app using the results from the decision tree model because of its consistent performance and simplicity. We implemented two approaches, bulk and adaptive, to develop the CINV risk prediction app using ResearchKit. If the questions could be generated from multiple flowcharts, designing a fixed-order questionnaire might not help build an efficient app. The question hierarchy was not consistent across different tree models. Instead of asking input to all variables, we developed an adaptive approach to present a minimal number of questions for computing the prediction. The fixed (bulk)-order approach will ask 14 questions for any of those 115 decision paths, but for the adaptive approach, the maximum questions asked will be equal to the depth of that flowchart (up to 9 questions). This makes the app both time and energy efficient for the user and can reduce the physicians' time at the point of care.

The developed smartphone app for recommending patients at risk of CINV can help improve the prevention of CINV among cancer patients. The target users (ie, clinicians) can use this app at the point of care during the prescription of antiemetics. This app will help identify patients at risk of CINV based on patient-related risk factors. Having this knowledge of the patients before the prescription of antiemetics can help design a better treatment plan, leading to better CINV management. Furthermore, the app took significantly less space and was developed considering the possible variation in users' technological skills. It does not require any permission, which will help users use it more effortlessly. The oncologist will have complete access to the risk calculation algorithm in their smartphone, which will drastically reduce the amount of time required to help a large group of people and will have the flexibility to provide personalized care to every patient, improving their quality of life.

Limitations

In this study, the data were collected by retrospective record review. Prospective validation is needed to confirm the usefulness of the model in a real clinical setting. The research also shows that female patients with pregnancy-related nausea and vomiting have a higher risk of CINV. However, this information was missing from our data set. This information could considerably enhance the prediction results. The data have a lower representation of Asians and Hispanics. A multicenter or multinational study, including various populations, is needed to overcome this shortcoming. In addition, if we use EMR data to integrate with the app, there is no difference between the bulk and adaptive approaches. However, if the app is used as a prediction tool at the point of care, the adaptive approach is more time and energy efficient, thus decreasing the chances of wrong input answers. In addition, for hospitals without any EMR system, this app can be extremely beneficial for cancer patients.

Future Work

In the future, our plan is to deploy this app in point-of-care settings by integrating it into EMRs to predict the risk of CINV. We can also perform a clinical study for estimating outcomes and improvement. Currently, this app is developed only for the iOS platform, which can be expanded to Android in the future.

Conclusions

This study aimed to solve a real clinical problem, and the solution can reduce the gap between clinical practice and evidence-based guidelines for CINV management. Our study will promote the notion of precision medicine by integrating patient-related risk factors and antiemetic treatment recommendations. Hence, our efforts can lead to increased quality of the patients' life and reduced health care cost. An effort to reduce the care provider's time has high importance at the point of care. A less time-consuming decision support tool to predict patients at risk will help care providers provide better care in general.

Conflicts of Interest

None declared.

References

1. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. CA Cancer J Clin 2014;64(4):252-271 [FREE Full text] [doi: 10.3322/caac.21235] [Medline: 24890451]
2. Hawkins R, Grunberg S. Chemotherapy-induced nausea and vomiting: challenges and opportunities for improved patient outcomes. Clin J Oncol Nurs 2009 Feb;13(1):54-64. [doi: 10.1188/09.CJON.54-64] [Medline: 19193549]
3. Sun CC, Bodurka DC, Weaver CB, Rasu R, Wolf JK, Bevers MW, et al. Rankings and symptom assessments of side effects from chemotherapy: insights from experienced patients with ovarian cancer. Support Care Cancer 2005 Apr;13(4):219-227. [doi: 10.1007/s00520-004-0710-6] [Medline: 15538640]
4. Feyer P, Jordan K. Update and new trends in antiemetic therapy: the continuing need for novel therapies. Ann Oncol 2011 Jan;22(1):30-38 [FREE Full text] [doi: 10.1093/annonc/mdq600] [Medline: 20947707]

5. Heidary N, Naik H, Burgin S. Chemotherapeutic agents and the skin: an update. J Am Acad Dermatol 2008 Apr;58(4):545-570. [doi: 10.1016/j.jaad.2008.01.001] [Medline: 18342708]
6. Davila M, Bresalier RS. Gastrointestinal complications of oncologic therapy. Nat Clin Pract Gastroenterol Hepatol 2008 Dec;5(12):682-696. [doi: 10.1038/ncpgasthep1277] [Medline: 18941434]
7. Stein A, Voigt W, Jordan K. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther Adv Med Oncol 2010 Jan;2(1):51-63 [FREE Full text] [doi: 10.1177/1758834009355164] [Medline: 21789126]
8. Haslam IS, Pitre A, Schuetz JD, Paus R. Protection against chemotherapy-induced alopecia: targeting ATP-binding cassette transporters in the hair follicle? Trends Pharmacol Sci 2013 Nov;34(11):599-604. [doi: 10.1016/j.tips.2013.09.001] [Medline: 24100054]
9. Paus R, Haslam IS, Sharov AA, Botchkarev VA. Pathobiology of chemotherapy-induced hair loss. Lancet Oncol 2013 Feb;14(2):e50-e59. [doi: 10.1016/S1470-2045(12)70553-3] [Medline: 23369683]
10. Balagula Y, Lacouture ME, Cotliar JA. Dermatologic toxicities of targeted anticancer therapies. J Support Oncol 2010;8(4):149-161. [Medline: 20822032]
11. Hu C, Chan W, Delclos GP, Du XL. Adjuvant chemotherapy and risk of gastrointestinal, hematologic, and cardiac toxicities in elderly patients with stage III colon cancer. Am J Clin Oncol 2012 Jun;35(3):228-236. [doi: 10.1097/COC.0b013e318210f812] [Medline: 21383605]
12. Coates A, Abraham S, Kaye SB, Sowerbutts T, Frewin C, Fox RM, et al. On the receiving end: patient perception of the side-effects of cancer chemotherapy. Eur J Cancer Clin Oncol 1983 Feb;19(2):203-208. [doi: 10.1016/0277-5379(83)90418-2] [Medline: 6681766]
13. Cooper S, Georgiou V. The impact of cytotoxic chemotherapy: perspectives from patients, specialists and nurses. Eur J Cancer 1992 Jan;28:S36-S38. [doi: 10.1016/0959-8049(92)90635-f] [Medline: 1627406]
14. Durant JR. The problem of nausea and vomiting in modern cancer chemotherapy. CA Cancer J Clin 1984;34(1):2-6 [FREE Full text] [doi: 10.3322/canjclin.34.1.2] [Medline: 6420016]
15. Morrow GR. Chemotherapy-related nausea and vomiting: etiology and management. CA Cancer J Clin 1989;39(2):89-104 [FREE Full text] [doi: 10.3322/canjclin.39.2.89] [Medline: 2495160]
16. Bloechl-Daum B, Deuson RR, Mavros P, Hansen M, Herrstedt J. Delayed nausea and vomiting continue to reduce patients' quality of life after highly and moderately emetogenic chemotherapy despite antiemetic treatment. J Clin Oncol 2006 Sep 20;24(27):4472-4478. [doi: 10.1200/JCO.2006.05.6382] [Medline: 16983116]
17. Hilarius DL, Kloeg PH, van der Wall E, van den Heuvel JJG, Gundy CM, Aaronson NK. Chemotherapy-induced nausea and vomiting in daily clinical practice: a community hospital-based study. Support Care Cancer 2012 Jan;20(1):107-117 [FREE Full text] [doi: 10.1007/s00520-010-1073-9] [Medline: 21258948]
18. Lachaine J, Yelle L, Kaizer L, Dufour A, Hopkins S, Deuson R. Chemotherapy-induced emesis: quality of life and economic impact in the context of current practice in Canada. Support Cancer Ther 2005 Apr 01;2(3):181-187. [doi: 10.3816/SCT.2005.n.011] [Medline: 18628170]
19. Haiderali A, Menditto L, Good M, Teitelbaum A, Wegner J. Impact on daily functioning and indirect/direct costs associated with chemotherapy-induced nausea and vomiting (CINV) in a U.S. population. Support Care Cancer 2011 Jun;19(6):843-851. [doi: 10.1007/s00520-010-0915-9] [Medline: 20532923]
20. Cohen L, de Moor CA, Eisenberg P, Ming EE, Hu H. Chemotherapy-induced nausea and vomiting: incidence and impact on patient quality of life at community oncology settings. Support Care Cancer 2007 May;15(5):497-503. [doi: 10.1007/s00520-006-0173-z] [Medline: 17103197]
21. Pirri C, Katris P, Trotter J, Bayliss E, Bennett R, Drummond P. Risk factors at pretreatment predicting treatment-induced nausea and vomiting in Australian cancer patients: a prospective, longitudinal, observational study. Support Care Cancer 2011 Oct;19(10):1549-1563. [doi: 10.1007/s00520-010-0982-y] [Medline: 20811914]
22. Shih YT, Xu Y, Elting LS. Costs of uncontrolled chemotherapy-induced nausea and vomiting among working-age cancer patients receiving highly or moderately emetogenic chemotherapy. Cancer 2007 Aug 01;110(3):678-685 [FREE Full text] [doi: 10.1002/cncr.22823] [Medline: 17567835]
23. Ballatori E, Roila F, Ruggeri B, Porrozzi S, Iannopollo M, Soru G, et al. The cost of chemotherapy-induced nausea and vomiting in Italy. Support Care Cancer 2007 Jan;15(1):31-38. [doi: 10.1007/s00520-006-0094-x] [Medline: 16788840]
24. Ihbe-Heffinger A, Ehlken B, Bernard R, Berger K, Peschel C, Eichler H, et al. The impact of delayed chemotherapy-induced nausea and vomiting on patients, health resource utilization and costs in German cancer centers. Ann Oncol 2004 Mar;15(3):526-536 [FREE Full text] [doi: 10.1093/annonc/mdh110] [Medline: 14998860]
25. Viale PH, Grande C, Moore S. Efficacy and cost: avoiding undertreatment of chemotherapy-induced nausea and vomiting. Clin J Oncol Nurs 2012 Aug;16(4):E133-E141. [doi: 10.1188/12.CJON.E133-E141] [Medline: 22842698]
26. Yeh Y, McDonnell A, Klinger E, Fowler B, Matta L, Voit D, et al. Comparison of healthcare resource use between patients receiving ondansetron or palonosetron as prophylaxis for chemotherapy-induced nausea and vomiting. J Oncol Pharm Pract 2011 Sep;17(3):179-185. [doi: 10.1177/1078155210366491] [Medline: 20452991]
27. Craver C, Gayle J, Balu S, Buchner D. Clinical and economic burden of chemotherapy-induced nausea and vomiting among patients with cancer in a hospital outpatient setting in the United States. J Med Econ 2011;14(1):87-98. [doi: 10.3111/13696998.2010.547237] [Medline: 21241160]

28. Nevidjon B, Chaudhary R. Controlling emesis: evolving challenges, novel strategies. J Support Oncol 2010;8(4 Suppl 2):1-10. [Medline: 20865953]
29. Schwartzberg LS. Chemotherapy-induced nausea and vomiting: clinician and patient perspectives. J Support Oncol 2007 Feb;5(2 Suppl 1):5-12. [Medline: 17366928]
30. Richardson JL, Marks G, Levine A. The influence of symptoms of disease and side effects of treatment on compliance with cancer therapy. J Clin Oncol 1988 Nov;6(11):1746-1752. [doi: 10.1200/JCO.1988.6.11.1746] [Medline: 3183704]
31. Hesketh PJ, Van Belle S, Aapro M, Tattersall FD, Naylor RJ, Hargreaves R, et al. Differential involvement of neurotransmitters through the time course of cisplatin-induced emesis as revealed by therapy with specific receptor antagonists. Eur J Cancer 2003 May;39(8):1074-1080. [doi: 10.1016/s0959-8049(02)00674-3] [Medline: 12736106]
32. Thompson N. Optimizing treatment outcomes in patients at risk for chemotherapy-induced nausea and vomiting. Clin J Oncol Nurs 2012 Jun 01;16(3):309-313. [doi: 10.1188/12.CJON.309-313] [Medline: 22641323]
33. Basch E, Prestrud AA, Hesketh PJ, Kris MG, Feyer PC, Somerfield MR, American Society of Clinical Oncology. Antiemetics: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol 2011 Nov 01;29(31):4189-4198 [FREE Full text] [doi: 10.1200/JCO.2010.34.4614] [Medline: 21947834]
34. Basch E, Hesketh PJ, Kris MG, Prestrud AA, Temin S, Lyman GH. Antiemetics: American Society of Clinical Oncology clinical practice guideline update. J Oncol Pract 2011 Nov;7(6):395-398 [FREE Full text] [doi: 10.1200/JOP.2011.000397] [Medline: 22379425]
35. Ettinger DS, Armstrong DK, Barbour S, Berger MJ, Bierman PJ, Bradbury B, National Comprehensive Cancer Network. Antiemesis. J Natl Compr Canc Netw 2012 Apr;10(4):456-485. [doi: 10.6004/jnccn.2012.0047] [Medline: 22491046]
36. Roila F, Herrstedt J, Aapro M, Gralla RJ, Einhorn LH, Ballatori E, ESMO/MASCC Guidelines Working Group. Guideline update for MASCC and ESMO in the prevention of chemotherapy- and radiotherapy-induced nausea and vomiting: results of the Perugia consensus conference. Ann Oncol 2010 May;21 Suppl 5:v232-v243 [FREE Full text] [doi: 10.1093/annonc/mdq194] [Medline: 20555089]
37. Jones JM, Qin R, Bardia A, Linquist B, Wolf S, Loprinzi CL. Antiemetics for chemotherapy-induced nausea and vomiting occurring despite prophylactic antiemetic therapy. J Palliat Med 2011 Jul;14(7):810-814 [FREE Full text] [doi: 10.1089/jpm.2011.0058] [Medline: 21554125]
38. Glaus A, Knipping C, Morant R, Böhme C, Lebert B, Beldermann F, et al. Chemotherapy-induced nausea and vomiting in routine practice: a European perspective. Support Care Cancer 2004 Oct;12(10):708-715. [doi: 10.1007/s00520-004-0662-x] [Medline: 15278682]
39. Molassiotis A, Saunders MP, Valle J, Wilson G, Lorigan P, Wardley A, et al. A prospective observational study of chemotherapy-related nausea and vomiting in routine practice in a UK cancer centre. Support Care Cancer 2008 Feb;16(2):201-208. [doi: 10.1007/s00520-007-0343-7] [Medline: 17926070]
40. Jordan K, Gralla R, Jahn F, Molassiotis A. International antiemetic guidelines on chemotherapy induced nausea and vomiting (CINV): content and implementation in daily routine practice. Eur J Pharmacol 2014 Jan 05;722:197-202. [doi: 10.1016/j.ejphar.2013.09.073] [Medline: 24157984]
41. Molassiotis A, Brearley SG, Stamataki Z. Use of antiemetics in the management of chemotherapy-related nausea and vomiting in current UK practice. Support Care Cancer 2011 Jul;19(7):949-956. [doi: 10.1007/s00520-010-0909-7] [Medline: 20574666]
42. Lee VC. Mobile Devices and Apps for Health Care Professionals: Uses and Benefits. USA: P T Internet MediMedia USA; 2014.
43. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12(1):67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
44. Marinov M, Mosa ASM, Yoo I, Boren SA. Data-mining technologies for diabetes: a systematic review. J Diabetes Sci Technol 2011 Nov 01;5(6):1549-1556 [FREE Full text] [doi: 10.1177/193229681100500631] [Medline: 22226277]
45. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
46. Mosa A, Hossain AM, Lavoie B, Yoo I. Patient-related risk factors for chemotherapy-induced nausea and vomiting: a systematic review. Front Pharmacol 2020;11:329 [FREE Full text] [doi: 10.3389/fphar.2020.00329] [Medline: 32296333]
47. John G, Langley P. Estimating continuous distributions in Bayesian classifiers. arXiv. Preprint posted online February 20, 2013. [FREE Full text]
48. Cessie SL, Houwelingen JCV. Ridge estimators in logistic regression. Appl Stat 1992;41(1):191. [doi: 10.2307/2347628]
49. Freund Y, Schapire R. Large margin classification using the perceptron algorithm. Mach Learn Springer 1999;37(3):296. [doi: 10.1145/279943.279985]
50. Platt J. Fast training of support vector machines using sequential minimal optimization. In: Advances in Kernel Methods: Support Vector Machine. New York, NY: ACM; 1998:41-65.
51. Wang XZ, Buontempo FV, Young A, Osborn D. Induction of decision trees using genetic programming for modelling ecotoxicity data: adaptive discretization of real-valued endpoints. SAR QSAR Environ Res 2006 Oct;17(5):451-471. [doi: 10.1080/10629360600933723] [Medline: 17050186]

52. Quinlan JR. C4. 5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann; 1993.
53. Quinlan JR. Data Mining Tools See5 and C5.0. 2004. URL: http://www.rulequest. com/see5-info.html [accessed 2021-11-11]
54. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software. SIGKDD Explor Newsl 2009 Nov 16;11(1):10-18. [doi: 10.1145/1656274.1656278]
55. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine learning methods to predict diabetes complications. J Diabetes Sci Technol 2018 Mar 12;12(2):295-302 [FREE Full text] [doi: 10.1177/1932296817706375] [Medline: 28494618]
56. Berger MJ, Ettinger DS, Aston J, Barbour S, Bergsbaken J, Bierman PJ, et al. NCCN guidelines insights: antiemesis, version 2.2017. J Natl Compr Canc Netw 2017 Jul;15(7):883-893. [doi: 10.6004/jnccn.2017.0117] [Medline: 28687576]
57. Roscoe JA, Morrow GR, Colagiuri B, Heckler CE, Pudlo BD, Colman L, et al. Insight in the prediction of chemotherapy-induced nausea. Support Care Cancer 2010 Jul;18(7):869-876 [FREE Full text] [doi: 10.1007/s00520-009-0723-2] [Medline: 19701781]
58. Osoba D, Zee B, Pater J, Warr D, Latreille J, Kaizer L. Determinants of postchemotherapy nausea and vomiting in patients with cancer. Quality of Life and Symptom Control Committees of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 1997 Jan;15(1):116-123. [doi: 10.1200/JCO.1997.15.1.116] [Medline: 8996132]
59. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]

Abbreviations

AI: artificial intelligence
ASCO: American Society of Clinical Oncology
BMI: body mass index
CINV: chemotherapy-induced nausea and vomiting
EMR: electronic medical record
ESMO: European Society of Medical Oncology
GUI: graphical user interface
HEC: high-emetogenic chemotherapy
KDD: knowledge discovery in databases
LEC: low-emetogenic chemotherapy
MASCC: Multinational Association of Supportive Care in Cancer
MEC: moderate-emetogenic chemotherapy
ML: machine learning
NCCN: National Comprehensive Cancer Network
NK-1: neurokinin-1

Edited by R Kukafka, G Eysenbach; submitted 20.01.21; peer-reviewed by YC Chen, A Khayat; comments to author 25.02.21; revised version received 11.03.21; accepted 04.08.21; published 02.12.21.

Please cite as:
Mosa ASM, Rana MKZ, Islam H, Hossain AKMM, Yoo I
A Smartphone-Based Decision Support Tool for Predicting Patients at Risk of Chemotherapy-Induced Nausea and Vomiting: Retrospective Study on App Development Using Decision Tree Induction
JMIR Mhealth Uhealth 2021;9(12):e27024
URL: https://mhealth.jmir.org/2021/12/e27024
doi:10.2196/27024
PMID:34860677

Original Paper

The Effects of the ManageHF4Life Mobile App on Patients With Chronic Heart Failure: Randomized Controlled Trial

Michael P Dorsch[1,2], PharmD, MS; Karen B Farris[1], PhD; Brigid E Rowell[1], MA; Scott L Hummel[2,3,4], MD; Todd M Koelling[2,3], MD

[1]Department of Clinical Pharmacy, College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

[2]Frankel Cardiovascular Center, University of Michigan, Ann Arbor, MI, United States

[3]Division of Cardiovascular Medicine, Medical School, University of Michigan, Ann Arbor, MI, United States

[4]Ann Arbor Veterans Affairs Health System, Ann Arbor, MI, United States

Corresponding Author:
Michael P Dorsch, PharmD, MS
Department of Clinical Pharmacy
College of Pharmacy
University of Michigan
428 Church St
Ann Arbor, MI, 48109
United States
Phone: 1 734 647 1452
Email: mdorsch@med.umich.edu

Abstract

Background: The successful management of heart failure (HF) involves guideline-based medical therapy as well as self-management behavior. As a result, the management of HF is moving toward a proactive real-time technological model of assisting patients with monitoring and self-management.

Objective: The aim of this paper was to evaluate the efficacy of enhanced self-management via a mobile app intervention on health-related quality of life, self-management, and HF readmissions.

Methods: A single-center randomized controlled trial was performed. Participants older than 45 years and admitted for acute decompensated HF or recently discharged in the past 4 weeks were included. The intervention group (“app group”) used a mobile app, and the intervention prompted daily self-monitoring and promoted self-management. The control group (“no-app group”) received usual care. The primary outcome was the change in Minnesota Living with Heart Failure Questionnaire (MLHFQ) score from baseline to 6 and 12 weeks. Secondary outcomes were the Self-Care Heart Failure Index (SCHFI) questionnaire score and recurrent HF admissions.

Results: A total of 83 participants were enrolled and completed all baseline assessments. Baseline characteristics were similar between the groups except for the prevalence of ischemic HF. The app group had a reduced MLHFQ at 6 weeks (mean 37.5, SD 3.5 vs mean 48.2, SD 3.7; *P*=.04) but not at 12 weeks (mean 44.2, SD 4 vs mean 45.9, SD 4; *P*=.78), compared to the no-app group. There was no effect of the app on the SCHFI at 6 or 12 weeks. The time to first HF readmission was not statistically different between the app group and the no-app group (app group 11/42, 26% vs no-app group 12/41, 29%; hazard ratio 0.89, 95% CI 0.39-2.02; *P*=.78) over 12 weeks.

Conclusions: The adaptive mobile app intervention, which focused on promoting self-monitoring and self-management, improved the MLHFQ at 6 weeks but did not sustain its effects at 12 weeks. No effect was seen on HF self-management measured by self-report. Further research is needed to enhance engagement in the app for a longer period and to determine if the app can reduce HF readmissions in a larger study.

Trial Registration: ClinicalTrials.gov NCT03149510; https://clinicaltrials.gov/ct2/show/NCT03149510

(JMIR Mhealth Uhealth 2021;9(12):e26185) doi:10.2196/26185

KEYWORDS

mHealth; remote monitoring; self-management; self-care; heart failure; medical therapy; mobile app

Introduction

Despite major scientific advances, heart failure (HF) continues to be a common and costly condition; each year, over 1 million people are admitted to an inpatient setting for acute heart failure [1]. HF is the most common hospital discharge diagnosis among older adults in the United States, and one-fifth of HF patients are readmitted within 30 days of discharge [2].

Hospital readmissions are a substantial concern in HF and are directly linked to poor health-related quality of life (HRQOL) [3]. HF readmissions also result in significant, potentially avoidable costs to our already-strained health care system because hospitalizations account for nearly 70% of annual HF costs [1]. National attention has turned toward reducing 30-day readmissions for acute heart failure, partially because, in October 2012, the Centers for Medicare & Medicaid Services started to receive financial penalties for higher-than-expected rates of readmissions.

One of the most common causes of HF readmission—failure to recognize clinical worsening—is related to poor self-management [4,5]. HF care includes daily monitoring of weight and symptoms, taking medications as prescribed, adhering to a low-sodium diet, and assessing changes in symptoms related to self-monitoring [6]. Self-management is when a patient understands how to interpret self-monitoring to ultimately change their behaviors and improve symptoms. Increasing patients' understanding of the link between self-monitoring and self-management is key to successful HF disease management interventions [7]. Several studies have shown that self-monitoring can enhance self-management and improve HRQOL in HF [8-10]. However, currently, there are few clinically effective HF self-management tools to support HF patients in managing their condition after they transition from the hospital back into the community. Thus, there is an urgent need for low-cost solutions to help patients recognize clinical worsening and reduce HF readmissions. This study's objective was to evaluate the effectiveness of a mobile app intervention that enhances self-monitoring of HRQOL, self-management, and HF readmissions.

Methods

Study Design

This was a 12-week, prospective, single-center, open-label randomized controlled trial conducted at Michigan Medicine, the University of Michigan's academic medical center. The trial was registered on ClinicalTrials.gov (NCT03149510) and approved by the University of Michigan's Institutional Review Board. The participants were recruited from March 2017 to April 2019 by in-person recruitment from the inpatient adult hospital. The participants were randomized to the intervention ("app") or control ("no app") group in a 1:1 fashion using the Trial Randomize application created by the University of Michigan's Consulting for Statistics, Computing and Analytics Research center. The randomization methodology uses a minimization approach to reduce covariate imbalances by using nonuniform assignment probabilities for the 2 groups [11]. All of the participants provided a written consent before being fully enrolled in the clinical trial.

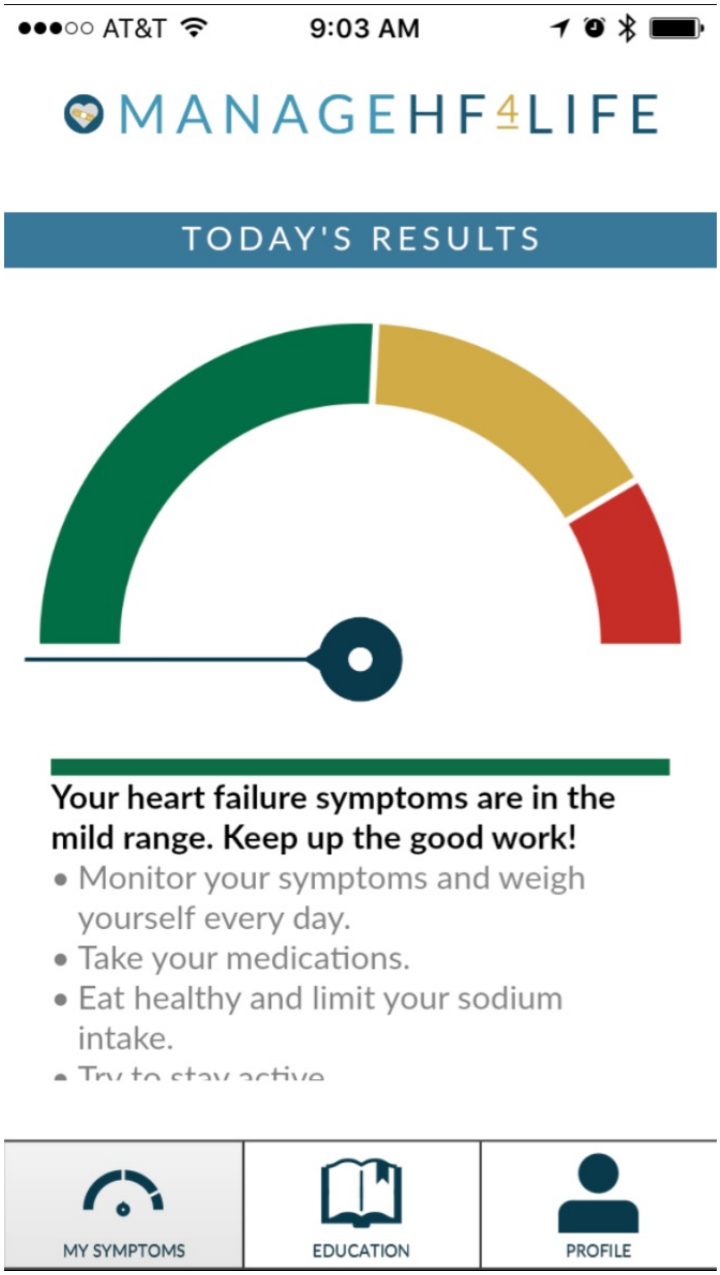
Study Participants

The participants were included if they were older than 45 years, had a left ventricular ejection fraction (EF) of ≤40% or >40% (with a left atrial size of >40 mm, brain natriuretic peptide of >200 pg/mL, or N-terminal pro–B-type natriuretic peptide of >800 pg/mL), and were currently admitted or recently discharged for acute on chronic decompensated HF. participants were excluded if they had unstable coronary syndromes within 8 weeks, primary valvular heart disease, constrictive pericardial disease, uncorrected thyroid disease, dialysis or creatinine of >4.0 mg/dL, active cancer, and pulmonary fibrosis. They were also excluded if they were a hospice candidate, if they were discharged to a setting other than home, or if they were requiring a chronic inotrope. The participants were not blinded due to the nature of the intervention. In May 2018, inclusion criteria were expanded to include HF with preserved ejection fraction, in addition to HF with reduced EF and those recently discharged to increase recruitment. Of the total 83 participants, 80 were enrolled during index hospitalization. The remaining 3 were enrolled within 4 weeks of discharge, at days 2, 4, and 28, respectively.

Intervention

The app group used a mobile app, ManageHF4Life, version 1 (The University of Michigan), along with a Fitbit (Fitbit Inc) physical activity monitor (Fitbit Charge 2) and scale (Fitbit Aria and Aria 2). Accurate self-monitoring, feedback, and self-efficacy are essential components for managing HF. The app prompted active daily self-monitoring, provided a health status indicator to promote self-management, and included standard education on HF. The daily prompt for active self-monitoring was carried out with a 9-AM push notification to complete an 8-question survey within the app. If the participants did not complete the survey by 12 PM, a reminder push notification was sent to them. The health status indicator was a stoplight (green, yellow, and red) and was generated from a rule-based model created by the investigators. The rule-based model was calculated from an equation based on the 8 survey questions and the difference between the daily weight and dry weight that was recorded in the app. The stoplight colors represented the participants' health status: the green color represented stable status, while yellow and red represented a clinical worsening state. The text below the health status indicator changed based on the color, with recommendations on self-management. An example of a health status indicator screen is shown in Figure 1, and the full mobile app layout is presented in the supplement. All intervention participants were provided with a 30-minute educational session on how to use the app. The control group received usual care upon discharge from the hospital. At Michigan Medicine, all patients receive discharge education about heart failure, which includes self-monitoring, a 2-week follow-up appointment with an advanced practice provider, and periodic phone calls from a telehealth HF nurse. The Fitbit scale was used to record the daily weight, but the Fitbit physical activity monitor was not intentionally used as part of the self-monitoring intervention.

Figure 1. Example of a health status indicator in the ManageHF4Life mobile app.



Outcome Measures

The primary outcome was the change in Minnesota Living with Heart Failure Questionnaire (MLHFQ) from baseline to 6 and 12 weeks [12]. This tool consists of 21 questions regarding the patients' perception of the effects of HF on their daily lives. Secondary outcomes were the change in self-management and HF readmission over time. Self-management was measured using the Self-Care Heart Failure Index (SCHFI), version 6.2, which was the most current version available at trial initiation [13]. The SCHFI 6.2 contains 22 questions and has 3 subscales that determine the patient's physiologic stability, response to symptoms, and ability to perform self-management. The questions in each subscale are standardized to a score of 0 to 100. Each subscale is added together to give the total SCHFI score. The SCHFI was collected at baseline, 6 weeks, and 12 weeks. Both the MLHFQ and the SCHFI were completed by participants using an automated online survey. All readmissions were reviewed in a blinded fashion for the potential to be an HF readmission. An unscheduled hospitalization was defined as an HF readmission if the primary diagnosis was HF and the length of stay either exceeded 24 hours or crossed a calendar day [14]. Outcome assessment was blinded to the randomization group. The study team contacted participants at 6 and 12 weeks to confirm the clinical outcomes and prompted the participants to complete any survey tasks. At the completion of the clinical trial, each participant in the app group received an online survey about the mobile app, which focused on its perceived usefulness and ease of use.

Statistical Analysis

The primary outcome of the trial was the change in the MLHFQ between the app and no-app groups from baseline to 6 and 12 weeks, using a modified intention-to-treat approach. Repeated measures mixed models (SAS PROC MIXED, SAS Institute) were used to determine the change in MLHFQ score over 12

weeks between the 2 groups. The group indicator (app vs no app) served as the primary covariate, and least squares mean and standard error are reported for the continuous variables over time. Based on preliminary data [9], the MLHFQ score was expected to improve from 56 to 42 on average in the app group, with no change in the no-app group (SD 11.5). Based on these assumptions, 40 participants per group (N=80) with 20% dropout will have the power of more than 83% to detect the difference at the significance level of 0.05.

For baseline characteristics, continuous variables were compared using a *t* test, and categorical variables were compared using the chi-square or Fisher exact tests, where appropriate. Repeated measures mixed models were used to compare the change in the SCHFI over time between the app and no-app groups, and data are presented in least squares mean and standard error. Cox proportional hazards survival model was used to analyze time to HF readmission between the app and no-app groups.

Results

Baseline Characteristics

A total of 83 participants were enrolled and completed all baseline assessments. Baseline characteristics were similar between the groups except for the prevalence of ischemic HF. The participants were 60.2 (SD 9.2) years old in the app group and 62 (SD 9.2) years old in the no-app group (*P*=.38). The average EF was 37.2% in the app group and 38.2% in the no-app group (*P*=.73). Most of the participants were Caucasian: 81% (34/42) app vs 83% (34/41) no app (*P*=.56); most of the participants were also New York Heart Association (NYHA) class III: 55% (23/42) app vs 66% (27/41) no app (*P*=.41) at study enrollment. The median number of days during which the app group performed self-monitoring within the app was 63 (IQR 28-84) of the 84 days (75%). Figure 2 represents the CONSORT (Consolidated Standards of Reporting Trials) diagram for this clinical trial, and Table 1 demonstrates the baseline characteristics for the participants in both groups.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

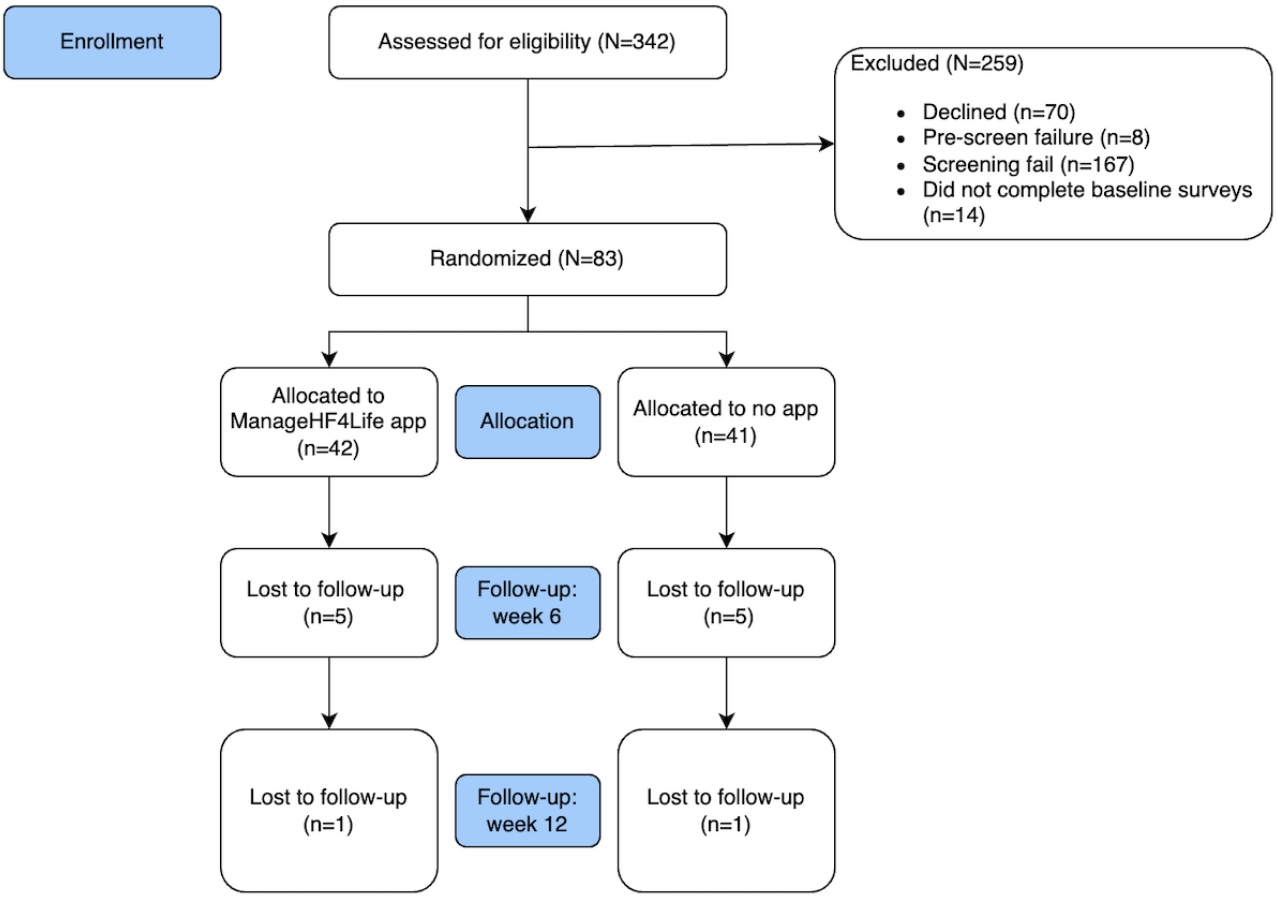


Table 1. Baseline demographics.

Variable	App (n=42)	No app (n=41)	*P* value
Age (years), mean (SD)	60.2 (9)	62 (9)	.38
Female, n (%)	14 (33)	15 (37)	.76
Race			.56
Caucasian, n (%)	34 (81)	34 (83)	
African American, n (%)	7 (17)	6 (15)	
Other, n (%)	1 (2)	1 (2)	
EF[a] (%), mean (SD)	37.2 (20)	38.8 (19)	.73
HFpEF[b], n (%)	16 (38)	19 (46)	.45
Ischemic HF[c], n (%)	19 (45)	29 (71)	.02
NYHA[d] class			
Class I, n (%)	1 (2)	0 (0)	.41
Class II, n (%)	10 (24)	5 (12)	
Class III, n (%)	23 (55)	27 (66)	
Class IV, n (%)	8 (19)	9 (22)	
Atrial fibrillation, n (%)	22 (52)	25 (61)	.43
MI[e], n (%)	10 (24)	18 (44)	.05
DM[f], n (%)	14 (33)	13 (32)	.87
Moderate or severe renal disease, n (%)	1 (2)	2 (5)	.62
Systolic BP[g] (mm Hg), mean (SD)	121.1 (23)	119.1 (21)	.68
Sodium (mmol/L), mean (SD)	138.3 (3)	137.8 (3)	.54
Hemoglobin (g/dL), mean (SD)	12.4 (2)	11.9 (2)	.32
ACEI[h], ARB[i], ARNI[j], n (%)	28 (67)	20 (49)	.10
Beta blocker, n (%)	37 (88)	35 (85)	.76
MRA[k], n (%)	18 (43)	16 (39)	.72

[a]EF: ejection fraction.

[b]HFpEF: heart failure with preserved ejection fraction.

[c]HF: heart failure.

[d]NYHA: New York Heart Association.

[e]MI: myocardial infarction.

[f]DM: diabetes mellitus.

[g]BP: blood pressure.

[h]ACEI: angiotensin converting enzyme inhibitor.

[i]ARB: angiotensin receptor blocker.

[j]ARNI: angiotensin receptor and neprilysin inhibitor.

[k]MRA: mineralocorticoid receptor antagonist.

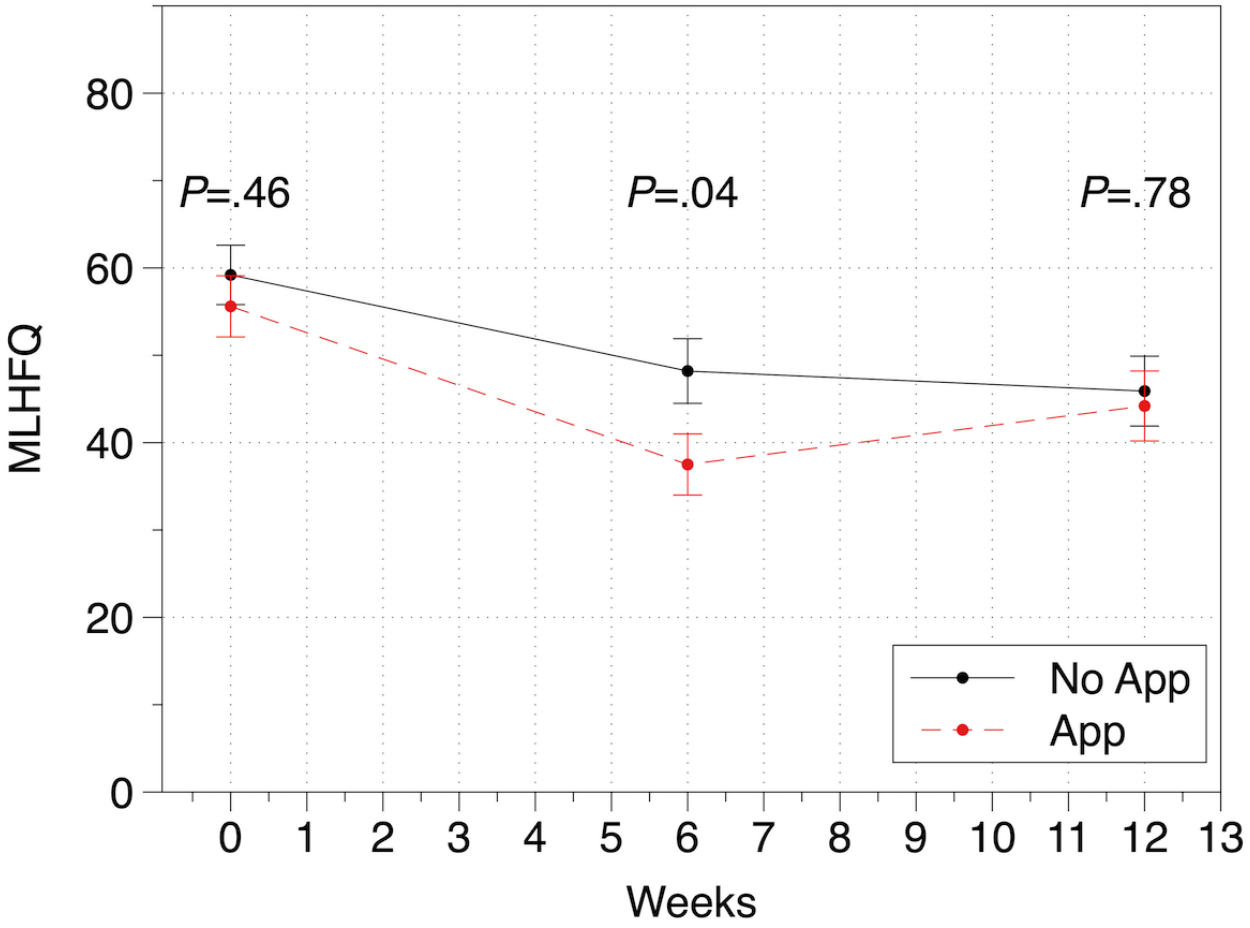
MLHFQ Scores

In the app group, the MLHFQ score changed from a baseline of 55.6 (SD 3.5) to 37.5 (SD 3.5) at 6 weeks and 44.2 (SD 4) at 12 weeks. The MLHFQ score in the no-app group changed from a baseline of 59.2 (SD 3.4) to 48.2 (SD 3.7) at 6 weeks and 45.9 (SD 4) at 12 weeks. The app group had a greater improvement in MLHFQ score at 6 weeks compared with the no-app group (*P*=.04), but not at 12 weeks (*P*=.78). Figure 3 demonstrates the change in MLHFQ total score over the course of the study between groups.

Among the emotional and physical subscales of the MLHFQ, the physical subscale showed similar results as the overall MLHFQ scale. MLHFQ physical scores changed from a baseline of 23.3 (SD 1.5) to 14.4 (SD 1.6) at 6 weeks and 17.8 (SD 1.9) at 12 weeks in the app group and a baseline of 24.4 (SD 1.5) to 20.4 (SD 1.7) at 6 weeks and 18.6 (SD 1.7) at 12 weeks in the no-app group. The app group had a greater improvement in the

MLHFQ physical subscale at 6 weeks compared with the no-app group (*P*=.01), but not at 12 weeks (*P*=.78). MLHFQ emotional scores changed from a baseline of 12 (SD 1) to 8.7 (SD 1) at 6 weeks and 9.3 (SD 1.1) at 12 weeks in the app group and a baseline of 11.9 (SD 1.2) to 10 (SD 1.1) at 6 weeks and 10 (SD 1.2) at 12 weeks in the no-app group. The app group had similar changes in the MLHFQ emotional subscale at 6 weeks compared to the no-app group (*P*=.38) and at 12 weeks (*P*=.64).

Figure 3. The change in Minnesota Living with Heart Failure Questionnaire (MLHFQ) score over time by group.



SCHFI Scores

The SCHFI total score changed from a baseline of 186.1 in the app group and 187.8 in the no-app group to 198.1 and 204.6, respectively, at 6 weeks (*P*=.40), and 196.9 and 206.1, at 12 weeks (*P*=.24). The maintenance, management, and confidence subscales of the SCHFI showed similar results. Table 2 demonstrates the change over time of the total SCHFI and 3 subscales in the app and no-app groups.

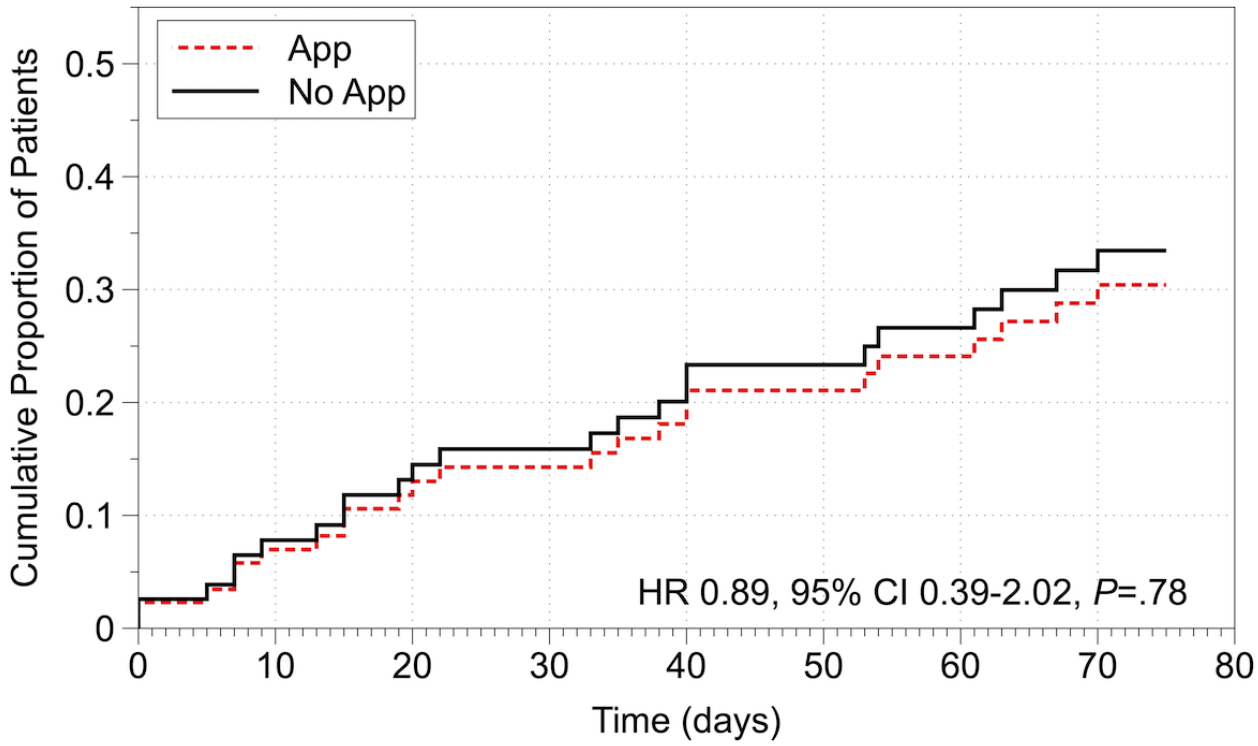
Table 2. The change in SCHFI and subscales over time by group.

Scores	Baseline			6 weeks			12 weeks		
	App	No app	*P* value	App	No app	*P* value	App	No app	*P* value
Total SCHFI, mean (SD)	186.1 (5)	187.8 (5)	.82	198.1 (5)	204.6 (5)	.40	196.9 (6)	206.1 (5)	.24
Maintenance, mean (SD)	66.6 (2)	70.4 (2)	.23	70.5 (2)	73.5 (2)	.37	69.9 (2)	74.6 (2)	.15
Management, mean (SD)	54.3 (2)	54.6 (2)	.94	55.7 (2)	60.4 (2)	.13	59.9 (2)	59 (2)	.78
Confidence, mean (SD)	64.5 (3)	62.9 (3)	.68	72.2 (3)	71.1 (3)	.79	67.7 (3)	72.6 (3)	.22

Readmissions

Over the 12-week study, 26% (11/42) of the participants had an HF readmission in the app group compared with 29% (12/41) of the participants in the no-app group (hazard ratio 0.89, 95% CI 0.39-2.02; *P*=.78). There was no significant difference in HF readmission rates between the participants in the 2 groups. Figure 4 shows a plot of the time-to-event curves for the app and no-app groups. A total of 13 non–HF-related readmissions, 8 out of 42 (19%) in the app group and 5 out of 41 (12%) in the no-app group, occurred during the 12-week follow-up. One participant HF readmission event was followed by a death.

Figure 4. Time to first heart failure readmission by group; HR: heart rate.



Mobile App Survey

In total, 86% (36/42) of the app group participants completed a survey about the app, at the end of the trial. Of the 36 participants who completed the survey, 92% (n=33) agreed or strongly agreed that they found the app useful, 94% (n=34) agreed or strongly agreed that they used the information in the app in their daily life, 89% (n=32) agreed or strongly agreed that the information they received in the app was important to them, and 97% (n=35) agreed or strongly agreed that the app was easy to use. Only 3% (n=1) agreed that the app was confusing, and no one stated that the app was difficult to understand. Moreover, 92% (n=33) thought that most people would learn to use the mobile app quickly; 75% (n=27) agreed or strongly agreed that they learned a lot from the mobile app; and 58% (n=21) said the mobile app had new information that they were not aware of before.

Discussion

Principal Findings

In recent years, smartphones have changed the landscape of the US society with 81% of the population now owning a smartphone [15]. The widespread access to smartphones can be harnessed to dramatically change health care delivery. In this study, a mobile app that used a health status indicator to communicate a clinical worsening state showed a greater improvement in HRQOL at 6 weeks, but did not sustain effects at 12 weeks when compared to a control group. From the ESCAPE (Endovascular Treatment for Small Core and Anterior Circulation Proximal Occlusion with Emphasis on Minimizing CT to Recanalization Times) trial, a decrease in the MLHFQ total score of 20 points at 1 month after an HF discharge had a lower rate of death or hospitalization compared with a 10-point decrease at 1 month [16]. In our study, the ManageHF4Life intervention demonstrated an 18-point decrease from baseline to 6 weeks compared with an 11-point decrease in the control group. This shows that the 6-week findings are clinically meaningful and deserve future investigation. This effect was also primarily driven by improvements in the physical subscale of the MLHFQ as opposed to the emotional subscale. The physical and emotional subscales of the MLHFQ have been shown to characterize how HF is affecting a patient's life. The physical subscale questions deal with the effects on the body, and the emotional subscale questions deal with the effects on the mind. The ManageHF4Life intervention is primarily targeted at the physical components of HF, so this finding aligns with the intended effects of the intervention.

Our study did not demonstrate an effect of the ManageHF4Life intervention on the secondary outcome of self-management, using the SCHFI score, compared to control. Self-management is affected directly or indirectly by depression, social support, eHealth literacy, and HF knowledge [16]. At baseline, the SCHFI total and subscale scores were higher in our study compared with those reported in the literature [13,17,18], which could have made it more difficult to demonstrate a change in self-management over time. SCHFI scores may have been higher at baseline and throughout our study, as many of the patients were followed in an advanced HF telemanagement program. This program is designed to provide clinical support and education to patients. The survey at the end of the study showed that 58% (21/36) of the participants said the ManageHF4Life intervention had new information that they were not aware of before. Increasing HF knowledge should increase self-management, but that was not the case in this study. In addition, the ManageHF4Life intervention did not improve the

emotional subscale of the MLHFQ, which aligns with the depression and social support aspects of self-management. Future interventions should target a broader support for self-management, including depression, social support, and knowledge.

A recent integrative review found 18 publications that studied the effects of mobile apps for heart failure [19]. In those studies, the total sample size ranged from 7 to 165 participants, and 7 of them were randomized controlled trials. Similar to our app, 14 studies included apps that monitored self-management components (weight, blood pressure, and HF symptoms). One mobile app, HeartMapp (University of South Florida), was most similar to our study app [20]. HeartMapp used a built-in algorithm based on the NYHA classification presenting green, yellow, orange, and red zones. The app was studied in an 18-patient, 30-day randomized controlled pilot study of patients being enrolled at hospital discharge. The study aimed to determine if the mobile app, compared to control, improved HRQOL using the Kansas City Cardiomyopathy Questionnaire and self-management behaviors using the SCHFI. It is unclear which version of the SCHFI was used because the methods do not state the version, and the subscale numbers do not match those in the SCHFI, version 6.2. Although underpowered, this study demonstrated a significant improvement in the SCHFI self-management and self-confidence subscales in the mobile app group, compared with control. Kansas City Cardiomyopathy Questionnaire measurements did not change over time in this 30-day study.

Moreover, there are some main differences between our study and the HeartMapp study. Our app used a clinician-derived rule-based model that includes patient symptoms and the change in body weight, while HeartMapp used a NYHA-based algorithm. The HeartMapp study included an active control group that was given access to some of the features of the app, while, in our study, the control group received usual care and did not have access to the mobile app. It is not possible to compare the difference in the SCHFI results in our study and HeartMapp because the methods do not state the version, and subscale numbers do not match those in SCHFI, version 6.2.

Limitations

While this is a randomized controlled trial, there are some limitations in the study. The study was open-label, so participants knew the group in which they were randomized. This could have led to a bias by participants in either group or provided undue influence on our results. The control group was "usual care" with no mobile app and did not include an attention control. Although it is common to use usual-care groups when studying mobile apps, attention control groups strengthen behavioral interventions [21]. Furthermore, the usual care in our center may be a more intensive care than some other centers in the country. Future studies of our mobile app should include a control group that receives the app, but not intervention components of interest. We gave all participants in the app group a wearable device and scale at the beginning of the study. This could have led to an intervention above and beyond the mobile app health status indicator. There are studies, however, that refute the idea that adding a wearable to an intervention improves outcomes more than the intervention alone [22]. In addition to these limitations, version 1 of our mobile app, ManageHF4Life, was very basic. It did not include contextual push notifications about self-management, adaptive content in the mobile app, or just-in-time dietary information when selecting foods [23]. Future research of the app will focus on these enhancements and other study designs to optimize the intervention and determine the effects on HF outcomes.

Conclusions

The mobile app intervention improved MLHFQ at 6 weeks but did not sustain its effects at 12 weeks, compared to control. No effect was seen on self-management measured by self-report with the SCHFI. Further versions of the app should focus on technological enhancements, and future research is needed to determine if those future versions can reduce HF readmissions in a larger study.

Acknowledgments

The study was funded by the National Institute on Aging at the National Institutes of Health, United States.

Conflicts of Interest

KBF serves as the site principal investigator for an investigator-initiated grant from AstraZeneca that is focused on implementation of a collaborative model between oncology and primary care pharmacists to improve medication adherence to oral anti-cancer agents and chronic disease medications.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 109 KB - mhealth_v9i12e26185_app1.pdf]

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]

2. Dharmarajan K, Wang Y, Lin Z, Normand ST, Ross JS, Horwitz LI, et al. Association of Changing Hospital Readmission Rates With Mortality Rates After Hospital Discharge. JAMA 2017 Jul 18;318(3):270-278 [FREE Full text] [doi: 10.1001/jama.2017.8444] [Medline: 28719692]
3. Rodríguez-Artalejo F, Guallar-Castillón P, Pascual CR, Otero CM, Montes AO, García AN, et al. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. Arch Intern Med 2005 Jun 13;165(11):1274-1279. [doi: 10.1001/archinte.165.11.1274] [Medline: 15956007]
4. Retrum JH, Boggs J, Hersh A, Wright L, Main DS, Magid DJ, et al. Patient-identified factors related to heart failure readmissions. Circ Cardiovasc Qual Outcomes 2013 Mar 01;6(2):171-177 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.112.967356] [Medline: 23386663]
5. Gilotra NA, Shpigel A, Okwuosa IS, Tamrat R, Flowers D, Russell SD. Patients Commonly Believe Their Heart Failure Hospitalizations Are Preventable and Identify Worsening Heart Failure, Nonadherence, and a Knowledge Gap as Reasons for Admission. J Card Fail 2017 Mar;23(3):252-256. [doi: 10.1016/j.cardfail.2016.09.024] [Medline: 27742454]
6. Jovicic A, Holroyd-Leduc JM, Straus SE. Effects of self-management intervention on health outcomes of patients with heart failure: a systematic review of randomized controlled trials. BMC Cardiovasc Disord 2006 Nov 02;6:43 [FREE Full text] [doi: 10.1186/1471-2261-6-43] [Medline: 17081306]
7. Clark AM, Wiens KS, Banner D, Kryworuchko J, Thirsk L, McLean L, et al. A systematic review of the main mechanisms of heart failure disease management interventions. Heart 2016 May 23;102(9):707-711. [doi: 10.1136/heartjnl-2015-308551] [Medline: 26908100]
8. Shively M, Kodiath M, Smith TL, Kelly A, Bone P, Fetterly L, et al. Effect of behavioral management on quality of life in mild heart failure: a randomized controlled trial. Patient Educ Couns 2005 Jul;58(1):27-34. [doi: 10.1016/j.pec.2004.06.007] [Medline: 15950833]
9. Varma S, McElnay JC, Hughes CM, Passmore AP, Varma M. Pharmaceutical care of patients with congestive heart failure: interventions and outcomes. Pharmacotherapy 1999 Jul;19(7):860-869. [doi: 10.1592/phco.19.10.860.31565] [Medline: 10417035]
10. Dorsch MP, Farris KB, Bleske BE, Koelling TM. A web application for self-monitoring improves symptoms in chronic systolic heart failure. Telemed J E Health 2015 Apr;21(4):267-270. [doi: 10.1089/tmj.2014.0095] [Medline: 25654615]
11. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975 Mar;31(1):103-115. [Medline: 1100130]
12. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: Reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. American Heart Journal 1992 Oct;124(4):1017-1025. [doi: 10.1016/0002-8703(92)90986-6]
13. Riegel B, Carlson B, Moser DK, Sebern M, Hicks FD, Roland V. Psychometric testing of the self-care of heart failure index. J Card Fail 2004 Aug;10(4):350-360. [doi: 10.1016/j.cardfail.2003.12.001] [Medline: 15309704]
14. Hicks KA, Mahaffey KW, Mehran R, Nissen SE, Wiviott SD, Dunn B, Standardized Data Collection for Cardiovascular Trials Initiative (SCTI). 2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. Circulation 2018 Feb 27;137(9):961-972. [doi: 10.1161/CIRCULATIONAHA.117.033502] [Medline: 29483172]
15. Anderson M. Mobile Technology and Home Broadband 2019. Pew Research Center. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2020-02-04]
16. Moser DK, Yamokoski L, Sun JL, Conway GA, Hartman KA, Graziano JA, Escape Investigators. Improvement in health-related quality of life after hospitalization predicts event-free survival in patients with advanced heart failure. J Card Fail 2009 Nov;15(9):763-769 [FREE Full text] [doi: 10.1016/j.cardfail.2009.05.003] [Medline: 19879462]
17. Chuang H, Kao C, Lin W, Chang Y. Factors Affecting Self-care Maintenance and Management in Patients With Heart Failure. J Cardiovasc Nurs 2019 Apr 17;34(4):297-305. [doi: 10.1097/jcn.0000000000000575]
18. Riegel B, Lee CS, Dickson VV, Carlson B. An update on the self-care of heart failure index. J Cardiovasc Nurs 2009;24(6):485-497 [FREE Full text] [doi: 10.1097/JCN.0b013e3181b4baa0] [Medline: 19786884]
19. Athilingam P, Jenkins B. Mobile Phone Apps to Support Heart Failure Self-Care Management: Integrative Review. JMIR Cardio 2018 May 02;2(1):e10057 [FREE Full text] [doi: 10.2196/10057] [Medline: 31758762]
20. Athilingam P, Jenkins B, Johansson M, Labrador M. A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]
21. Aycock DM, Hayat MJ, Helvig A, Dunbar SB, Clark PC. Essential considerations in developing attention control groups in behavioral research. Res Nurs Health 2018 Jun 14;41(3):320-328. [doi: 10.1002/nur.21870] [Medline: 29906317]
22. 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]
23. Dorsch MP, Cornellier ML, Poggi AD, Bilgen F, Chen P, Wu C, et al. Effects of a Novel Contextual Just-In-Time Mobile App Intervention (LowSalt4Life) on Sodium Intake in Adults With Hypertension: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e16696 [FREE Full text] [doi: 10.2196/16696] [Medline: 32663139]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
EF: ejection fraction
ESCAPE: Endovascular Treatment for Small Core and Anterior Circulation Proximal Occlusion With Emphasis on Minimizing CT to Recanalization Times
HF: heart failure
HRQOL: health-related quality of life
MLHFQ: Minnesota Living with Heart Failure Questionnaire
NYHA: New York Heart Association
SCHFI: Self-Care Heart Failure Index

Edited by L Buis; submitted 01.12.20; peer-reviewed by K Trinkley, Z Xie, R Cruz Martínez, O El-Gayar; comments to author 17.02.21; revised version received 28.04.21; accepted 29.10.21; published 07.12.21.

Please cite as:
Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM
The Effects of the ManageHF4Life Mobile App on Patients With Chronic Heart Failure: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(12):e26185
URL: https://mhealth.jmir.org/2021/12/e26185
doi:10.2196/26185
PMID:34878990

Original Paper

An Intelligent Individualized Cardiovascular App for Risk Elimination (iCARE) for Individuals With Coronary Heart Disease: Development and Usability Testing Analysis

Yuling Chen[1*], BSN; Meihua Ji[1,2*], PhD; Ying Wu[1], PhD; Qingyu Wang[1], MSN; Ying Deng[1], MSN; Yong Liu[3], ME; Fangqin Wu[1], MSN; Mingxuan Liu[1], BSN; Yiqiang Guo[1], PhD; Ziyuan Fu[1], BSN; Xiaoying Zheng[4,5], PhD

[1]School of Nursing, Capital Medical University, Beijing, China

[2]Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China

[3]Along Technology Inc, Beijing, China

[4]The Asia-Pacific Economic Cooperation Health Science Academy, Peking University, Beijing, China

[5]Institute of Population Research, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Ying Wu, PhD
School of Nursing, Capital Medical University
10 You-an-men Wai Xi-tou-tiao, Feng-tai District
Beijing, 100069
China
Phone: 86 83911766
Email: helenywu@vip.163.com

Abstract

Background: Death and disability from coronary heart disease (CHD) can be largely reduced by improving risk factor management. However, adhering to evidence-based recommendations is challenging and requires interventions at the level of the patient, provider, and health system.

Objective: The aim of this study was to develop an Intelligent Individualized Cardiovascular App for Risk Elimination (iCARE) to facilitate adherence to health behaviors and preventive medications, and to test the usability of iCARE.

Methods: We developed iCARE based on a user-centered design approach, which included 4 phases: (1) function design, (2) iterative design, (3) expert inspections and walkthroughs of the prototypes, and (4) usability testing with end users. The usability testing of iCARE included 2 stages: stage I, which included a task analysis and a usability evaluation (January to March 2019) of the iCARE patient app using the modified Health Information Technology Usability Survey (Health-ITUES); and stage II (June 2020), which used the Health-ITUES among end users who used the app for 6 months. The end users were individuals with a confirmed diagnosis of CHD from 2 university-affiliated hospitals in Beijing, China.

Results: iCARE consists of a patient app, a care provider app, and a cloud platform. It has a set of algorithms that trigger tailored feedback and can send individualized interventions based on data from initial assessment and health monitoring via manual entry or wearable devices. For stage I usability testing, 88 hospitalized patients (72% [63/88] male; mean age 60 [SD 9.9] years) with CHD were included in the study. The mean score of the usability testing was 90.1 (interquartile range 83.3-99.0). Among enrolled participants, 90% (79/88) were satisfied with iCARE; 94% (83/88) and 82% (72/88) reported that iCARE was useful and easy to use, respectively. For stage II usability testing, 61 individuals with CHD (85% [52/61] male; mean age 53 [SD 8.2] years) who were from an intervention arm and used iCARE for at least six months were included. The mean total score on usability testing based on the questionnaire was 89.0 (interquartile distance: 77.0-99.5). Among enrolled participants, 89% (54/61) were satisfied with the use of iCARE, 93% (57/61) perceived it as useful, and 70% (43/61) as easy to use.

Conclusions: This study developed an intelligent, individualized, evidence-based, and theory-driven app (iCARE) to improve patients' adherence to health behaviors and medication management. iCARE was identified to be highly acceptable, useful, and easy to use among individuals with a diagnosis of CHD.

Trial Registration: Chinese Clinical Trial Registry ChiCTR-INR-16010242; https://tinyurl.com/2p8bkrew

(JMIR Mhealth Uhealth 2021;9(12):e26439) doi:10.2196/26439

KEYWORDS

mobile health; health behavior; system; development; usability; coronary heart disease

Introduction

Coronary heart disease (CHD) is the leading cause of cardiovascular death, accounting for 7.3 million annual deaths worldwide [1,2], with about 130,000 being reported from China alone [3]. Preventive interventions focusing on unhealthy behavior (diet, physical activity, smoking) modification and adherence to secondary prevention medications can reduce at least 47% of CHD mortality and decrease 68% of CHD-related major risk factors [3-7]. Although interventions facilitating behavior change and medication adherence have been developed and widely implemented, the prevalence of unhealthy behaviors and medication nonadherence remained high among individuals with CHD, including those who have already experienced life-threatening cardiac events or underwent percutaneous coronary interventions [6,8-11]. In large international studies from Europe and China, such as the EUROASPIRE IV and the Improving Care for Cardiovascular Disease in China-Acute Coronary Syndrome (CCC-ACS) Project, over half of patients with CHD had substantial unhealthy behaviors (50%-77%), with a large proportion of patients not adhering to prescribed preventive medications (45%-83%) [8,10,12-16].

Effective intervention targeting adherence to healthy behaviors and medications requires strategies to be implemented in addressing multiple CHD-related risk factors [17]. In addition, as reported by previous studies, evidence-based interventions with incorporation of real-time monitoring, person-centered care, and tailored feedback are essential to ensure affordable and sustainable long-term benefits [17-21]. However, it is difficult for conventional interventions to provide real-time monitoring; therefore, these interventions are unable to deliver tailored feedback and person-centered care [13-15].

With the advancement in information and communication technologies, mobile health (mHealth)–based health management systems and apps have emerged [18-20]. These have shown high potential in providing individualized intervention and sending instant automatic feedback based on real-time monitoring, therefore it might have promising effects in impacting health behavior change and promoting medication adherence [18-20]. Currently, many mHealth apps are designed to promote physical activity [21], track diet intake [22], assist with smoking cessation [23], remind patients to take medications [24,25], and facilitate self-management of CHD risk factors [26]. However, many of the existing mHealth apps are not individualized; additionally, they are not inclusive of suggested strategies and often have limited functions on tailored feedback, especially automatic feedback, and inadequate information updates [27]. None of the top 5 downloaded mHealth apps in China incorporated multiple key health behaviors identified by clinical guidelines [27]. A recent systematic review on health behavior changes related to physical activity, diet, drug and alcohol use, and mental health revealed that the majority of the mHealth apps (40/52, 80%) for health behavior change only focused on a single behavior, with the other 20% focusing on 2 health behaviors [18]. Meanwhile, studies suggested that currently available mHealth apps, including those in the management of CHD, failed to address users' needs and preferences, or consider their unique characteristics during the app development phase [23-25]. Furthermore, health care providers, such as nurses, are often not involved in the design and development of the mHealth apps, despite them being recognized as important for designing interventions that are reflective of patients care needs [27,28]. In addition, currently available mHealth apps were often not guided by a behavior change theory in developing their interventions [24,29], and they often do not sufficiently emphasize evidence-based interventions [28].

In light of the imperative needs for health behavior modification and medication management among individuals with CHD, we developed an Intelligent Individualized Cardiovascular App for Risk Elimination (iCARE) through facilitating healthy behavior and medication adherence. iCARE was designed to address the gaps in conventional interventions, in which multiple CHD-related risk factors were managed through real-time monitoring, person-centered care, and automatic tailored feedback. Following clinical guidelines on CHD secondary prevention [3,30,31], the interventions of iCARE were developed based on the Intervention Mapping framework [32], and the Contemplation–Action–Maintenance (CAM) model [33,34], which was an integrated behavior change model describing the roles of multiple moderators and mediators during the motivation and volition stages that are essential for healthy behavior change and behavior maintenance in individuals with CHD (Multimedia Appendix 1). In our preliminary study, we generated a set of iCARE interventions along with a set of "IF–THEN" algorithms to improve patients' adherence to health behaviors and medications, and conducted a needs assessment on the use of an mHealth-based system among patients with CHD [33]. The interventions developed in that preliminary work were incorporated as a built-in intervention bank into the cloud platform of the iCARE system during the development stage. The findings from the needs assessment were used to guide the design of the user interface of the iCARE system.

The purposes of this study were to describe the development of iCARE and to evaluate its usability among individuals with CHD. The effect of iCARE on facilitating adherence to actual behavior change is out of scope of this study, and it will be reported separately.

Methods

Development Process of iCARE

Overview

End users' needs analysis and the development of interventions within iCARE have been published elsewhere and are

summarized as preliminary work (Multimedia Appendix 1) [33]. Based on our preliminary work, we followed the user-centered design principle [20] in developing iCARE, and considered this principle as the main methodology for designing person-centered care–delivering systems [35]. The development process of iCARE consisted of 4 phases: (1) function design, (2) iterative design, (3) expert inspections and walkthroughs of prototypes, and (4) usability testing with end users. We established a multidisciplinary team including health care professionals (nurses specialized in cardiovascular care, nursing informaticians, nursing researchers), project managers, software engineers, software architects, and interface designers to develop iCARE. Weekly meetings were held to discuss the issues raised and advance the progress of system development among multidisciplinary team members throughout all phases.

Phase 1: Functional Design of iCARE

iCARE consisted of 3 functional components: a patient app for individuals with CHD and their families, a care provider app for health care providers such as nurses and physicians, and a cloud platform. The key functions and modules of each component were first drafted in a mind mapping software (XMind version 8; XMind, Ltd.) based on the results of the needs assessment and the overall aims of the system. Through brainstorming activities within the professional team, a detailed contextual document that described the framework, modules of each app and the cloud platform, functions, and design principles were formulated to guide the development of iCARE. To ensure the interventions are more effective and reflective of the care management of individuals with CHD, the components and schematic diagram of iCARE were designed following the nursing process [35].

Phase 2: Iterative Design of iCARE

Overview of Stages

In this phase, the system architecture, database, protective measures to secure individuals' personal identification information and health data, and user interface were determined. The patient app and care provider app were developed to be compatible with Android-based smartphones, as they are more popular and affordable in China [27]. We will also develop iCARE to be compatible on iOS-based smartphones once the app is identified as being stable in Android-based smartphones.

Design Architecture of iCARE

To increase the scalability and ensure the reliability of the software, developers need to respond to user's needs and consider continuous delivery to establish a culturally and environmentally adaptive software [36]. Therefore, the MicroService architecture [36-38], an emerging architectural design, was applied in the development of iCARE. MicroService is used for handling complex systems that require highly repetitive and frequent changes and allows continuous delivery of software in short circles; it was identified with increased deployability and modifiability among researchers. Compared with the monolithic approach, the MicroService architecture allows us to organize iCARE into a set of small, single-responsibility units, and self-contained services that can be developed, operated, tested, and deployed independently [36]. The MicroService architecture uses a continuous delivery approach to handle data queries, which makes it easier to achieve zero-downtime releases in responding to the rapid evolving needs [35]. After the MicroService architecture is applied, a majority of the services were allowed to use NoSQL databases to capture, query, and administer data in iCARE. This greatly reduced the time and effort in database migration, thereby substantially improving the efficiency in system development.

iCARE involves transmission of sensitive data, such as personal identification information, which makes the system subject to external and internal threats. Therefore, to ensure system security, protective measures were applied within iCARE. Based on the MicroService architecture, access control was performed through the User Account and Authentication service. The 2-way authentication HTTPS was applied to ensure the security of data transmission between the app and the back end services. To ensure secured data transmission and storage, all sensitive data in the apps were encrypted, and access to the data was password protected with sophisticated protective mechanisms, including verifying password strength, limiting number of login errors, periodically changing password, etc. Meanwhile, the system has built-in functions in providing database backup and recovery. In this study, we utilized MongoDB, a document-based database, and one of the leading NoSQL databases. MongoDB is used in the MicroService architecture owing to its ability to provide flexible schema, redundancy, automation, and scalability. The security of the MongoDB database is mainly achieved by strengthening the security of the operating system, authentication, and database. The development of iCARE was based on international and national data standards (Table S1 in Multimedia Appendix 1).

User Interface Design and Visualization Design

To fully reflect end users' needs and achieve the overall objectives of iCARE, the health professional team initially drafted the user interfaces of the patient app using Axure RP 8 software (Axure Software Solutions, Inc.). First, the user interface of the patient app was designed with diet, physical activity, smoking, and medication adherence as the main functional structure. Second, as the interventions of iCARE were developed based on the CAM behavior change model, the user interface of the patient app was designed to consider the characteristics of the interventions that addressed the major moderators and mediators identified within the CAM model, with patients' risk perception, outcome expectation, action planning, self-efficacy, social support, perceived enjoyment, perceived effectiveness, and coping planning addressed within the app designing process.

Based on the initial draft, the interface designer prepared a mock-up (draft) of the planned user interfaces and their interactions and workflow using Flinto version 26.0.5 [39] following the 6 user-friendly design principles: structure, simplicity, visibility, feedback, tolerance, and reuse [40]. To ensure workflow efficiency, prevent information-entry errors, and increase end users' positive experience in using the app, contrasting color, large font, distinctive graphical shapes, etc. were utilized to indicate different functions and status, as well as to increase the accessibility and effectiveness of the app.

Meanwhile, visualization (infographics, figures, and chart) was used to enhance individuals' perception on the risks of nonadherence to health behavior, prescribed medications, etc.

Phase 3: Expert Inspections and Walkthroughs of the iCARE Prototypes

iCARE was programed using Java and Nodejs, with Linux and Docker as the operating environment. Following user-centered design principles, and based on the functions, content, and architecture of iCARE, the software engineers developed the fully functional prototypes (alpha and beta versions) of the patient app and care-provider app [41].

Following the agile approach [42], software engineers iteratively identified and solved technical/implementational issues during the development process. The alpha version was initially developed, released, tested, and retested among technology professionals. The Gitlab and Docker tools were implemented for the release management throughout different stages and environments. Meanwhile, to ensure testability of iCARE, we used the test-first mindset and practices to define the acceptance criteria for the system [36].

After resolving technical issues or bugs in the alpha version, a full version of the app, the beta version, was formulated. Multiple testing cycles of the beta version of the app were applied among health care professionals in our team until a final consensus was reached. The testing of the beta version lasted for 3 months before the fully functional beta version was found to be stable and ready for usability testing.

Phase 4: Usability Testing of iCARE With End Users

Overview of Stages

The usability testing of iCARE included 2 stages. Stage I included 2 steps: a task analysis and a usability evaluation of the patient app using the modified Health Information Technology Usability Survey (Health-ITUES) [43]. Stage II included a usability evaluation using the Health-ITUES among end users who used the app for 6 months. The usability testing was designed following the International Organization for Standardization (ISO) standard 9241-11 [44]. In this study, we only tested the usability of the patient app among individuals with CHD. As iCARE was in the testing stage, and the main focus of this study was to test the experience of patients and the usability of the patient app, and therefore, we did not include health care providers (such as nurses) to complete the Health-ITUES.

Eligible patients who were hospitalized in 3 cardiac units of 2 university-affiliated hospitals in Beijing, China, participated in stage I usability testing. Patients who had a documented diagnosis of CHD and reported at least one unhealthy behavior were included in this study. A more detailed description of study participants and procedures is provided in Multimedia Appendix 1. The study was approved by the Institutional Review Committee of the Capital Medical University (Approval No. 2015SY45) and the study hospitals (Beijing An-Zhen Hospital, Approval No. 2015030; Beijing Chao-Yang Hospital, Approval No. 20211224). All enrolled participants provided written informed consent. Fully functional beta version 1 was used for task analysis and first-step usability testing with the Health-ITUES; beta version 2 was used for stage II usability testing with the same questionnaire. As the usability study is the preliminary work of our randomized controlled trial (RCT), which tests the effects of iCARE, the entire study has been registered with the Chinese Clinical Trial Registry (No. ChiCTR-INR-16010242).

Task Analysis

To identify usability problems as suggested by Maramba et al [45], individuals with CHD were asked to complete 8 tasks regarding the main functions of the app (Table S2 in Multimedia Appendix 1). Eight videos on how to complete the tasks were made available to all participants (Multimedia Appendices 2-9). All participants watched each video and practiced the exercises on the patient app until they were fully confident in completing each task. Based on the ISO standard 9241-11 [44], using standardized evaluation forms, we described the usability of the patient app in terms of the duration in completing each task, the level of task completeness, usability error, and usability problems. The duration in completing each task was recorded using a timer. The level of task completeness (Table S3 in Multimedia Appendix 1) was ranked on a scale of 1 (no problem) to 4 (significant problem) [44]. The completeness rates for each task were calculated by the proportion of participants who successfully completed the tasks (scored as "1"). Any usability-related problems were identified when patients were not able to complete the task. The severity of usability problems (Table S4 in Multimedia Appendix 1) was ranked from 0 "I don't agree that this is a usability problem at all" to 4 "usability catastrophes: imperative to fix this before product can be released," to reflect the level of severity of the identified problems [46,47]. The usability problem rates for each task were reflected by the proportion of participants who reported any problems when completing the tasks. Based on the ISO standard 9241-11 [44], the effectiveness of the app refers to the level of task completeness by the users; the efficacy of the app is expressed by the time (in seconds) the end user required in completing the tasks.

Usability Evaluation Using the Questionnaire

The usability evaluation was completed using the modified Health-ITUES [43]. Stage I usability evaluation was conducted among participants who completed the task analysis on beta version 1 (January 2019 to March 2019). All participants were given a small gift (around US $4) after they completed the questionnaire to compensate for their time. The iCARE system was then refined according to the results of stage I usability testing, and the updated version was used in the RCT. Stage II usability evaluation was performed in June 2020 among individuals who enrolled in the RCT after they used the patient app for 6 months. The RCT was a multicenter open-labeled study that tested the effects of iCARE interventions on major cardiovascular risk reduction and facilitation of adherence to health behaviors and medication among individuals with CHD. The RCT has 3 groups, including the intervention group (received fully functional iCARE to provide person-centered interventions that use multiple formats such as comics, videos, pictures, words to address all the factors in the CAM model plus routine care), control group 1 (received a person-centered

intervention presented in text format only but did not address all factors in the CAM model plus routine care), and control group 2 (received routine care). Participants who were randomized into the intervention arm in the RCT and used the app for at least six months were invited to complete the modified Health-ITUES. Details of the RCT protocol and the results of the trial will be published in the future. We also retrieved the number of times patients accessed the app and analyzed the patients' preference by identifying the most commonly used functions of the app over the past 6 months.

Measures

The English version of the Health-ITUES has a Cronbach α of .85-.92 and a criterion validity of 0.46-0.70 [43]. To evaluate the usability of the patient app, the questionnaire was customized to address the type of tool (patient app), the user (individuals with CHD), and tasks (for reducing cardiovascular risk factors) of the target app. The Chinese version of the Health-ITUES in evaluating the patient app was shown to be reliable and valid among individuals with CHD in this study, with the Cronbach α at .74-.90 and expert validity at 0.87-0.95. The Health-ITUES comprises 4 dimensions with 20 items: impact (3 items), perceived usefulness (9 items), perceived ease of use (5 items), and user control (3 items). Each item was rated from "1" (strongly disagree) to "5" (strongly agree) on a 5-point Likert scale. Total scores ranged from 20 to 100, with higher scores indicating better usability. In addition, iCARE was considered satisfactory, useful, and easy to use if the score of patients' responses was 4 or more on item 7 (using iCARE is useful for self-management of CHD-related risk factors), item 9 (I am satisfied with iCARE for self-management of CHD-related risk factors), and item 14 (learning to operate iCARE is easy for me), respectively.

Data Analysis

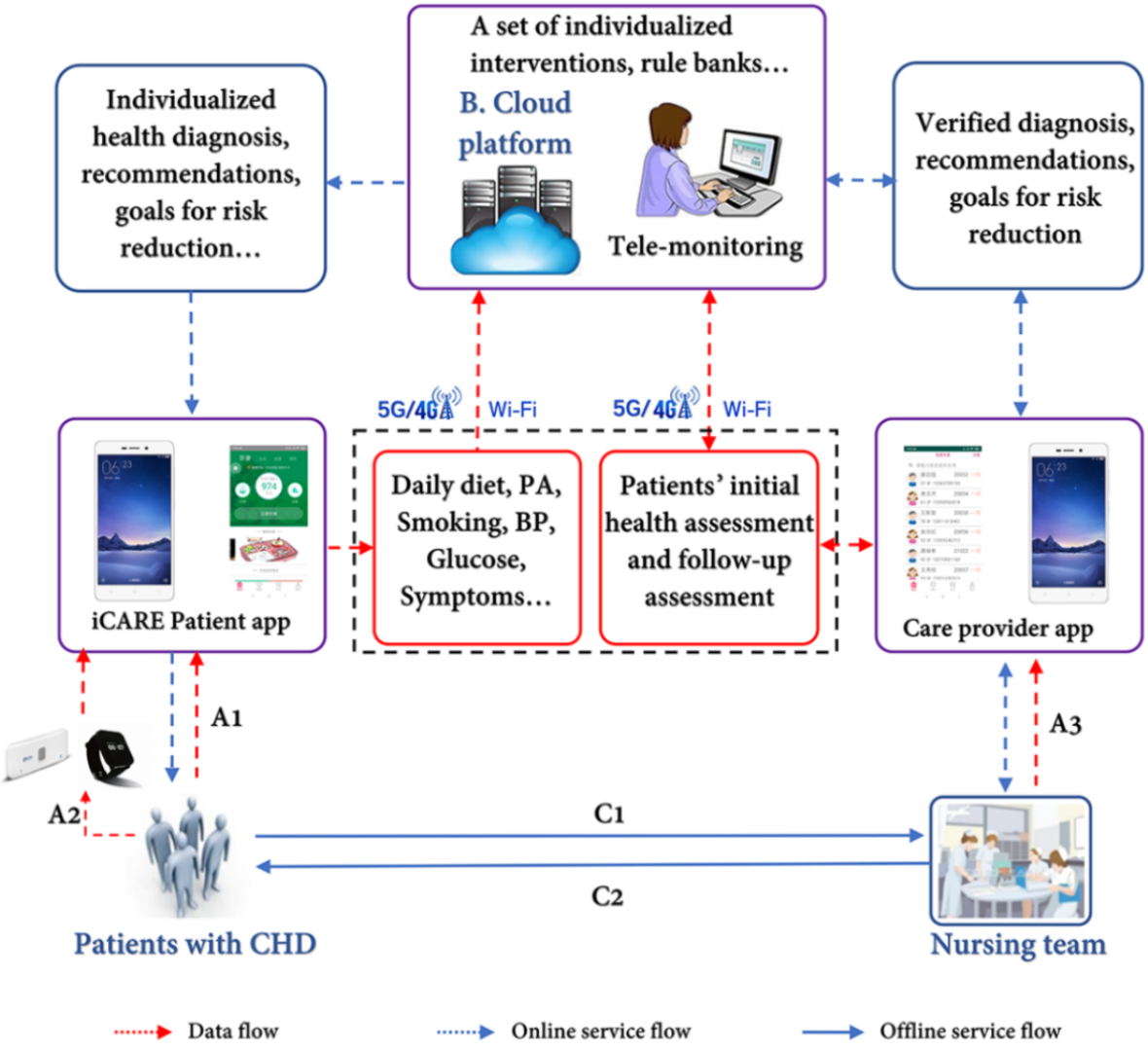
Statistical analysis was conducted using SPSS version 24.0 software (IBM, Corp). Categorical data were described as frequencies and proportions. Continuous data were tested for normality using the one-sample Kolmogorov–Smirnov test, and was described as means and SDs, or medians and interquartile ranges, as appropriate. Comparisons of participants' responses to the Health-ITUES between the first and second usability testing groups, and comparisons between the usage in terms of different groups (working status, educational levels, gender, used the device or not) were performed with the Mann–Whitney U test (scores of the Health-ITUES) and chi-square test (rates of satisfaction, usefulness, and easy to use), as appropriate. A P value <.05 was considered statistically significant. Regarding the sample size for usability testing, according to a previous study [43], the average score for each item in the Health-ITUES is 4 points and the maximum SD for each item is 0.8. Based on this information, considering 75% of the enrolled participants could complete the questionnaire and following a 2-sided 1-sample t test (unpaired), we estimated that we would need a total of 77 individuals in the first usability evaluation to achieve 80% power with α=.05, with an expected average score of 4.3 on each item of the Health-ITUES. As the second usability testing was conducted among individuals who were enrolled in the ongoing RCT, we did not calculate the sample size for that test.

Results

Phase 1: Functional Design of iCARE

The functional components of iCARE illustrating the system workflow are shown in Figure 1.

Figure 1. The components and schematic diagram of the iCARE. (A1) Initial and periodic assessment through daily monitoring via manual entry by patients. (A2) Daily monitoring by wearable devices. (A3) Patients' initial assessment data, follow-up data. (C1) Initial and periodic assessment by cardiovascular nurses. (C2) Further evaluation, interventions, follow-up. iCARE: Intelligent Individualized Cardiovascular Application for Risk Elimination; PA: physical activity; BP: blood pressure; CHD: coronary heart disease.



An initial health assessment was conducted by a nurse in our research team before patients were discharged from the hospital. The assessment included patients' demographics (age, gender, education, etc.), current health behaviors (daily diet, physical activity, smoking status), preferences regarding health behaviors (such as preferred fruits, vegetables, types of physical activities), and medication adherence in the past 30 days before admission if patients were on medication. Patients' health information including blood pressure, glucose level, lipid level, etc. was also collected. After the initial assessment, a summarized health report that described patient's medical and nursing diagnosis, unhealthy behaviors, and modifiable risk factors, along with established goals for risk reduction, was generated for patient's easy access through the patient app. The tailored goals for risk reduction were created based on the recommendations in the cardiovascular secondary prevention guidelines, with consideration of patients' age, gender, preferences (such as preferred time and types of physical activity), physical activity levels, left ventricular ejection fraction, comorbidity, metabolic status, musculoskeletal condition or disease, cardiac risk profile, and the current level of habitual physical activity. When the health care providers completed the baseline assessments on the care provider app and uploaded the data to the cloud platform, the recommended final goals for cardiovascular risk reduction were generated based on the cardiovascular secondary prevention guidelines and patients' specific health issues. They were then displayed in patients' health report and on the home page of the patient app. Patients can personally modify their goals to be achieved at different stages if they felt the goals sent to the patient app were not achievable at the moment. When we designed the algorithm for the goals of healthy diet, we considered patients' gender, age, and physical activity levels to determine the requirement of daily energy intake. Health-related data on physical activity, diet, blood pressure, lipid profile, blood sugar, heart rate, and weight were entered by patients either manually or through wearable devices. A smart watch (Ustone) was used to count steps and patients' heart rate, and a sensor (*Youyi Tang*; only for patients with diabetes) was connected to iCARE to record blood glucose, and the information was uploaded to the cloud platform for analysis

after patients logged into the app and selected the corresponding wearable device, which is to be synchronized, by clicking the upload button. Patients can also upload a photocopy of their laboratory results via the patient app if they prefer. Researchers will check the uploaded information in the care provider app or access it via cloud platform, and manually edit and send tailored feedback based on the laboratory results. A reminder on daily health data entry was sent to patients if they failed to enter data for 3 days, and a call from nurses was made if no further action was detected within 4 days after the reminder was sent. Instant and individualized feedback and tailored recommendations for changing behavior were automatically sent to the care provider app for verification by cardiovascular nurses based on the built-in algorithms (examples are shown in Multimedia Appendix 1). The verified recommendations of interventions were sent to the patient app for implementation by patients. Patient's health behaviors and medication adherence were evaluated through daily monitoring and periodic assessment (every 3 months).

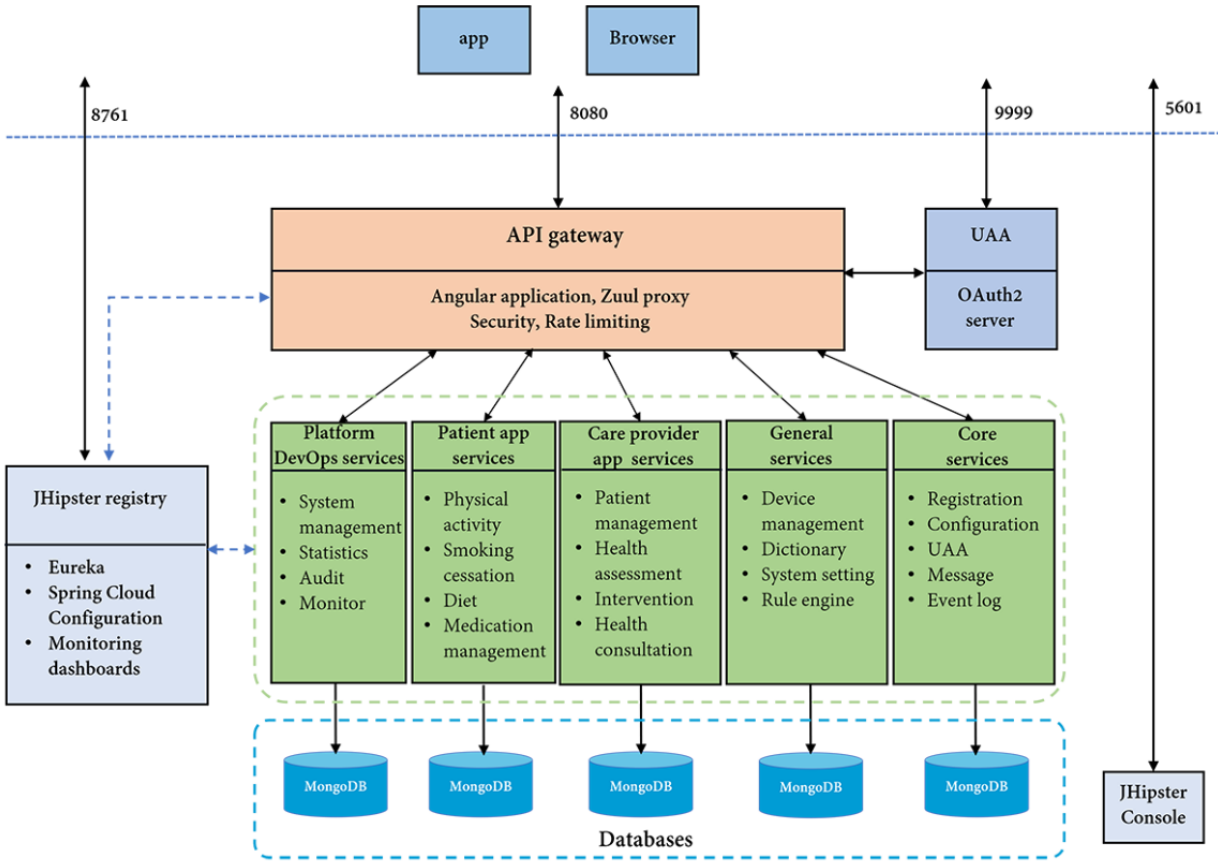
The overall functional modules of iCARE are displayed in Figure S1 in Multimedia Appendix 1. The patient app was developed with functions allowing individual end users to input personal data related to health behaviors, prescribed medications, and physiological indicators, and to review individual health data and health report, as well as to receive recommended interventions. The care provider app was developed with the functions allowing health care providers to view patients' health data and health report, make health assessment, verify recommended interventions, provide health consultation as necessary, and manage follow-up visits. The cloud platform was developed for health care providers and system manager with authorized access rights to view patients' overall health behaviors, prescribed medications, and physiological indicators. It was also structured to create and edit the intervention pool, knowledge, rule, and algorism; assign roles of users; and conduct data analysis. However, having access to the cloud platform and care provider app does not allow nurses or physicians to modify the data entered by patients.

Phase 2: Iterative Design of iCARE

Architecture Design

The MicroService architecture of iCARE is displayed in Figure 2. It included a set of microservices, such as services for the platform DevOps, patient app, and care provider app, and general and core services. Each individual microservice has its separate database (MongoDB) with its own domain data. Instead of calling services directly, the app and browser get access to the different services through the application programming interface gateway which will forward the request to the appropriate services on the back end. All services were connected with the JHipster registry.

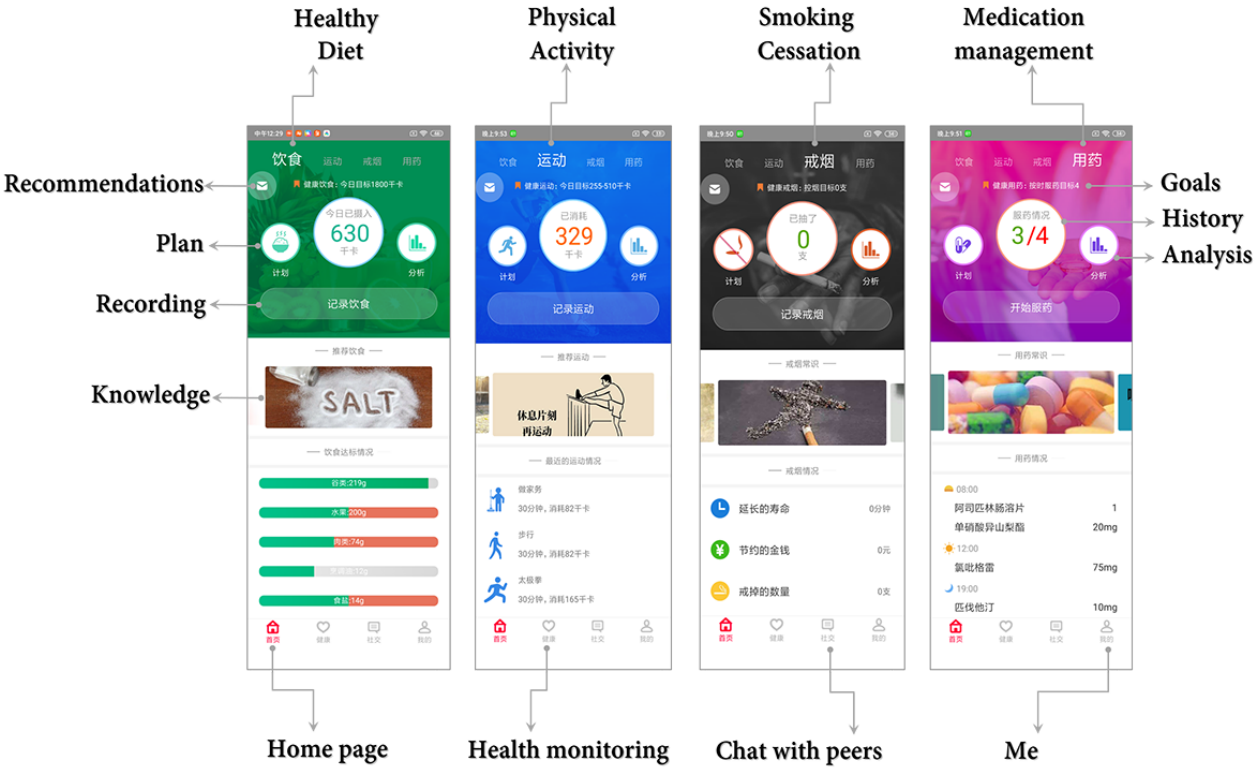
Figure 2. The system architecture of the iCARE (Intelligent Individualized Cardiovascular Application for Risk Elimination). API: Application Programming Interface; UAA: User Account and Authentication.



User Interface Design and Visualization Design

Following the user-centered design principle, the user interfaces for all functional modules were generated by health care professionals. The user interfaces of the patient app's home page are displayed in Figure 3, which included diet, physical activity, smoking, and medication management with different colors for easy identification. This arrangement allows easy access of the selected modules without complicated manipulation of the app.

Figure 3. The user interface of the home page of the patient app.



To address the major moderators and mediators identified within the CAM model, techniques were applied to augment the functions of the related features. For example, to increase patient's risk perception, we applied visual comics to increase individuals' awareness on potential cardiac events and the severity of a risk (Figure 4A). Visualization was also used to increase patients' perception on the effectiveness of adhering to healthy behaviors and medication treatment. As shown in Figure 4E, if good adherence was identified on health behaviors, medication adherence, blood pressure, blood glucose, and lipid levels, the Cardiac Health Score will increase. To promote positive outcome expectation, we also used videos, such as visualization of their future life with or without modifying their unhealthy lifestyles, to remind users about the positive effect of following suggested interventions (Figure 4B). Individualized action plans such as plans for physical activity were generated based on patients' assessment and their preferences, and they were available on the home page for easy reference (Figure 4C). In addition, as shown in Figure 4D, to increase patients' self-efficacy, a peer rank was designed in iCARE to increase patients' confidence in maintaining healthy behaviors. We also sent health promotion messages regarding the positive expectations from changing unhealthy behaviors to the patient app. When the patient was at the action stage of behavior change, a predesigned questionnaire was sent to the patient app and patients were asked to respond to the questionnaire to understand their possible barriers. According to their response, iCARE would automatically match the predesigned countermeasures with identified barriers, and push a coping plan to the patient. Patients can revise the coping plan according to their personal features and preferences and form a final coping plan.

Figure 4. Examples of user interfaces to reflect major moderators and mediators identified within the CAM (Contemplation-Action-Maintenance) model. (A) Risk Perception; (B) Outcome Expectation; (C) Action Planning; (D) Self-efficacy; (E) Perceived Effectiveness; (F) Behavioral Enjoyment; (G) Social Support; (H) Coping Planning. (B1) Video format; (E1) Visualization of iCARE Cardiac Health Score; (E2) Visualization of 10 years of cardiovascular risk; (E3) Visualization of atherosclerotic plaque; (F1) Trend chart; (H1) Read-out mode. iCARE: Intelligent Individualized Cardiovascular App for Risk Elimination; PA: physical activity.

Phase 3: Expert Inspections and Walkthroughs of the iCARE Prototypes

The software engineers translated the user interfaces into prototypes. The finalized alpha version of the iCARE prototype was developed, tested, and discussed among technology professionals to address technical issues throughout the development process. The steps involved in testing the iCARE alpha version are presented in Figure S2 in Multimedia Appendix 1. To verify the basic functionality of the developed modules, engineers tested each module independently with simulated data 10-20 times. To verify the stability of each module after they were finalized, a dedicated software evaluator tested the module 5-10 times with simulated patients' data.

The beta version of the iCARE prototype was released after the technical issues or bugs identified in the alpha version were resolved. The beta version of iCARE was released in March 10, 2019. An example regarding the logic of the interfaces on medication management is displayed in Figure S3 in Multimedia Appendix 1.

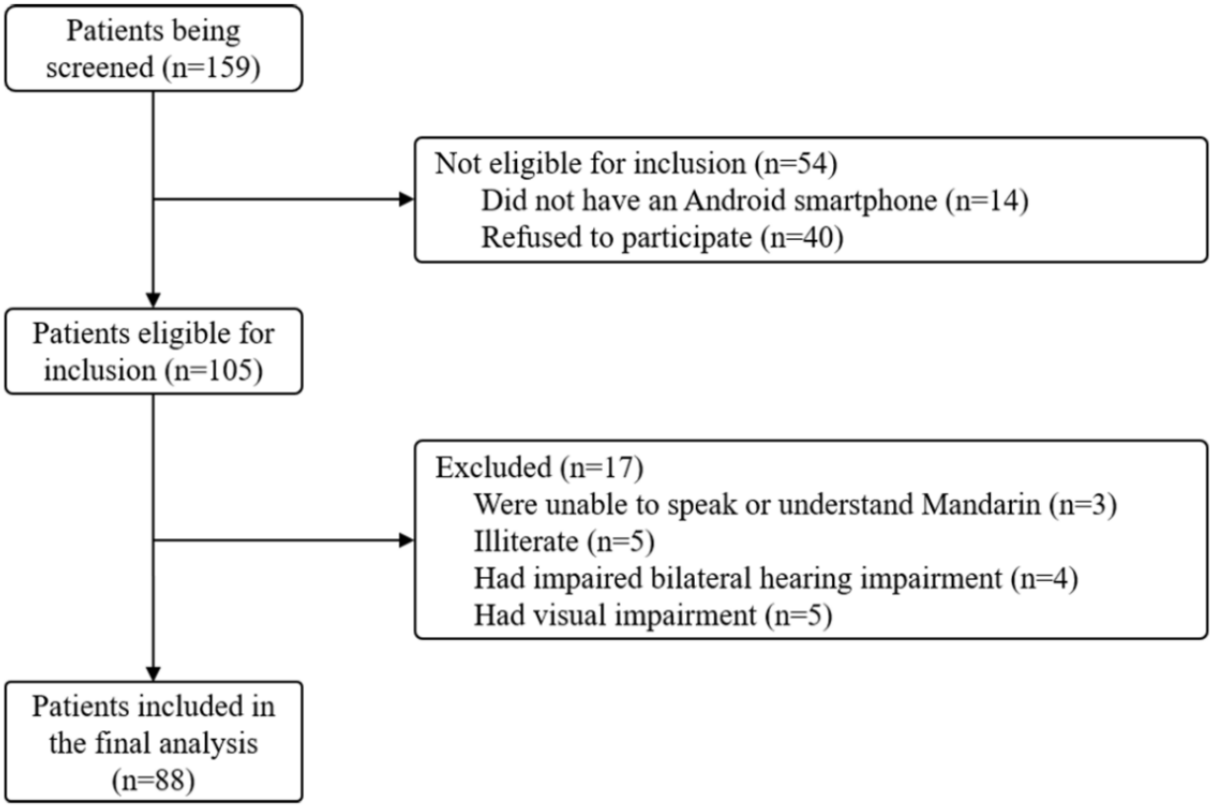
Phase 4: Usability Testing of iCARE With End Users

Overview of Stages

After the fully functional iCARE prototype was formulated, we carried out a 2-stage usability testing of the iCARE patient app. Stage I was conducted among 88 patients with CHD after iCARE was released. Figure 5 shows the flowchart of participant recruitment for the first-stage usability testing. A total of 159 patients were diagnosed with CHD during the study period, among which 88 eligible patients participated in the study and were included in the final analysis. There were 40 patients eligible for the study but refused to participate due to various reasons, such as lack of time, conflict of schedule, or simply being uninterested. The basic characteristics of enrolled participants for the first-step usability testing are presented in

Table S5 in Multimedia Appendix 1, and they were largely male (72%, 63/88), with a mean age of 60 (SD 9.9) years.

Figure 5. The flow of patient recruitment for the first step usability evaluation study.



Task Analysis

Figure 6 shows the effectiveness and efficacy of the patient app. Regarding the effectiveness of the app, the results showed that a majority of participants (80%, 70/88) perceived the tasks as easy to complete. The usability problems in completing the 8 tasks were rated from 0% (0/88) to 12.5% (11/88). In terms of the efficacy of the patient app, the average time used to complete tasks 1 through 8 ranged from 8 to 39 seconds.

Figure 6. The effectiveness and efficacy evaluation of the patient App based on task analysis.

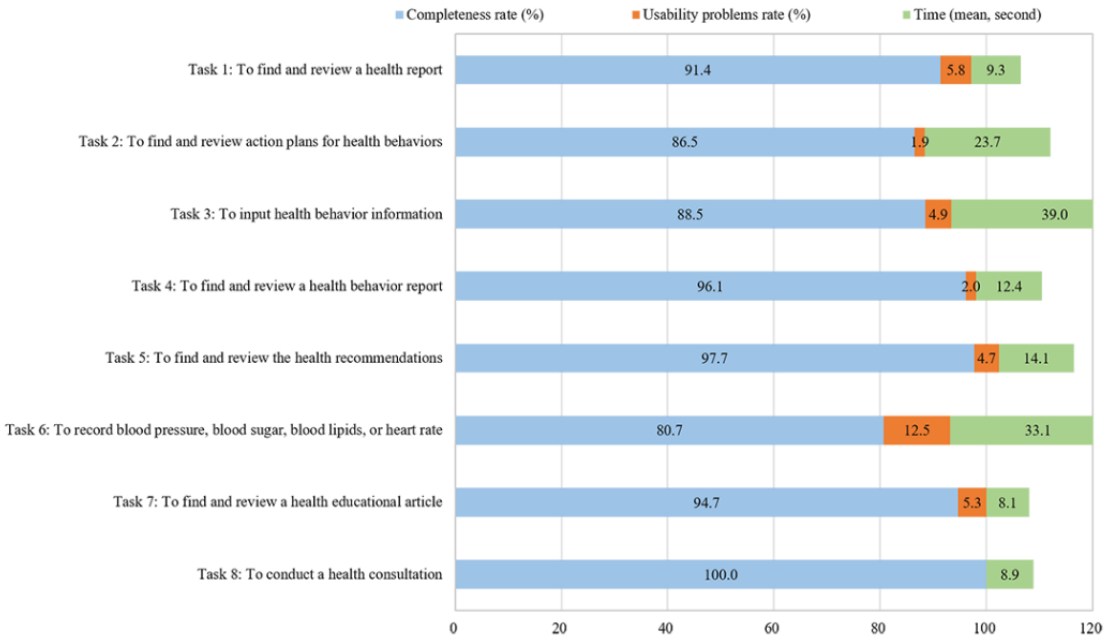


Table S6 in Multimedia Appendix 1 summarizes the usability problems that emerged from task analysis. A majority (6/17) of the usability problems were related to “ease of input, screen readability, and glanceability” (category 5), which most frequently occurred while entering data for blood pressure, blood sugar, blood lipids, or heart rate (task 6). Regarding the

severity of the identified usability problems, 4 were identified as usability catastrophes. For example, regarding task 6 "to record blood pressure, blood sugar, blood lipids, or heart rate," usability catastrophe was identified when patients reported that the icon was not easy to manipulate in recording blood pressure, as the distance between the 2 circles was too close and it was hard to manipulate.

Usability Evaluation With the Health-ITUES

Table 1 illustrates participants' perceptions on using iCARE in the first-stage usability testing. The mean total score of the Health-ITUES among participants was 90.1 (IQR 83.3-99.0), with a mean score on each item at 4.5 (SD 0.50). The mean scores of items related to the impact, perceived usefulness, perceived ease of use, and user control of the patient app ranged from 4.3 to 4.6. Meanwhile, based on patients' response to items 7, 9, and 14 of the Health-ITUES, 90% (79/88), 94% (83/88), and 82% (72/88) of enrolled patients perceived iCARE as satisfactory, useful, and easy to use, respectively.

After completion of stage I usability testing, the iCARE beta version was updated and revised based on the results of initial usability evaluation and task analysis. Such revisions included, but not limited to, increase of font sizes, modification of icons (ie, increasing the distance between the 2 circles of the icon for blood pressure), and utilization of more colors for easy identification in related user interfaces. The finalized iCARE version was released and used in the RCT study. Stage II usability testing was conducted among 61 (85% [52/61] male, mean age 53 [SD 8.2] years) individuals in the RCT intervention arm after they used the app for 6 months. The mean total score of the Health-ITUES was 89.0 (IQR 77.0-99.5), with the mean score on each item being 4.3 (SD 0.64). The mean scores of items related to the subdimensions (impact, perceived usefulness, perceived ease of use, and user control) ranged from 4.2 to 4.5. Among enrolled participants, 89% (54/61) were satisfied with the use of the app, 93% (57/61) perceived the app as useful, and 70% (43/61) as easy to use, based on patients' response to related items (7, 9, and 14). In addition, 52% (32/61) and 11% (7/61) reported using the patient app at least once a week or every day, respectively.

As shown in Figure S4 in Multimedia Appendix 1, no significant changes were found in terms of the total score of the Health-ITUES (P=.25) as well as the 4 dimensions of the Health-ITUES (P=.52, .68, .14, and .10 for impact, perceived usefulness, perceived ease of use, and user control, respectively) between participants in the first and second stages of usability testing. There was also no significant difference identified in terms of the rates of satisfaction (P=.85), usefulness (P=.86), and ease of use (P=.12) between first and second usability evaluations.

The number of times individuals accessed the patient app showed that they used it for a mean of 33.5 (SD 75.3) times over 6 months. The most frequently visited screens were related to health recommendation, followed by daily monitoring of health indicators, health behaviors, and medication adherence (26179, 1649, and 942 times, respectively). There were no significant differences in the usage of the patient app among patients in terms of their working status (retired verses not retired; P=.13), educational levels (P=.34), and gender (P=.97). In stage II usability study, 38/61 patients used the smart watch, and 23 patients did not use the smart watch. The mean usage of the app among patients who used the smart watch was higher than those who did not (97.7 vs. 12.5, P<.001) over 6 months. As the blood glucose monitor was not distributed to patients with diabetes (10/61 patients) in stage II testing, it was not available for testing the differences in this regard.

Table 1. Summary of the Health-ITUES[a] questionnaire.

Items	First testing, mean (SD)	Second testing, mean (SD)
Impact (Cronbach α: .863)	4.6 (0.62)	4.5 (0.76)
1. I think iCARE[b] can be a positive addition for heart health for patients living with CHD[c].	4.6 (0.71)	4.4 (0.74)
2. I think iCARE can improve the quality of life of persons living with CHD.	4.6 (0.61)	4.6 (0.61)
3. iCARE is an important part of meeting my information needs related to self-management of CHD-related risk factors.	4.5 (0.76)	4.4 (0.96)
Perceived usefulness (Cronbach α: .901)	4.6 (0.51)	4.4 (0.63)
4. Using iCARE makes it easier to self-manage my CHD-related risk factors.	4.5 (0.70)	4.5 (0.67)
5. Using iCARE enables me to self-manage my CHD-related risk factors more quickly.	4.6 (0.60)	4.5 (0.67)
6. Using iCARE makes it more likely that I can self-manage my iCARE-related risk factors.	4.6 (0.60)	4.4 (0.88)
7. Using iCARE is useful for self-management of CHD-related risk factors.	4.6 (0.70)	4.6 (0.62)
8. I think iCARE presents a more equitable process for self-management of CHD-related risk factors.	4.6 (0.70)	4.5 (0.72)
9. I am satisfied with iCARE for self-management of CHD-related risk factors.	4.6 (0.68)	4.5 (0.74)
10. I self-manage my CHD-related risk factors in a timely manner because of iCARE.	4.5 (0.77)	4.3 (0.83)
11. Using iCARE increases my ability to self-manage my CHD-related risk factors.	4.6 (0.63)	4.4 (0.74)
12. I am able to self-manage my CHD-related risk factors whenever I use iCARE.	4.4 (0.78)	4.3 (0.83)
Perceived ease of use (Cronbach α: .899)	4.3 (0.76)	4.2 (0.76)
13. I am comfortable with my ability to use iCARE.	4.3 (0.88)	4.3 (0.73)
14. Learning to operate iCARE is easy for me.	4.3 (0.90)	4.0 (1.05)
15. It is easy for me to become skillful at using iCARE.	4.3 (0.95)	4.0 (1.06)
16. I find iCARE easy to use.	4.5 (0.77)	4.2 (0.96)
17. I can always remember how to log on to and use iCARE.	4.3 (0.97)	4.3 (0.85)
User control (Cronbach α: .743)	4.4 (0.67)	4.2 (0.81)
18. iCARE gives error messages that clearly tell me how to fix problems.	4.3 (0.91)	4.2 (0.90)
19. Whenever I make a mistake using iCARE, I recover easily and quickly.	4.4 (0.89)	4.2 (0.88)
20. The information (such as online help, on-screen messages, and other documentation) provided with iCARE is clear.	4.7 (0.60)	4.2 (0.96)

[a]The Modified Health-Information Technology Usability Survey was adopted from Schnall et al [43]. ©[2021] Capital Medical University, Beijing, China. All rights reserved. Adapted with permission from the "Health IT Usability Evaluation Scale."

[b]iCARE: Individualized, Intelligent and Integrated Cardiovascular App for Risk Elimination.

[c]CHD: coronary heart disease.

Discussion

Principal Findings

In this study, we developed an intelligent and individualized health care management system for individuals with CHD, iCARE, following a 4-phase user-centered approach. iCARE was designed to facilitate patients' adherence to multiple health behaviors (diet, physical activity, and smoking) and preventive medications. The design and development of iCARE were realized through incorporation of individualized interventions, which were developed through a theory-driven and evidence-based approach and following the principles of the nursing process. iCARE was developed to reflect designated interventional strategies that target the mediators and moderators of the CAM model by using an individualized approach and visualization techniques. iCARE is an intelligent health management system that has a set of "IF–THEN" algorithms that trigger real-time monitoring, person-centered care, and automatic tailored feedback, based on data from initial assessment and health monitoring via manual entry or wearable devices. iCARE includes functions in facilitating health assessment and monitoring, health behavior and medication management, intervention implementation, and health counseling. The results of the 2-stage usability testing revealed that iCARE has acceptable usability. The patient app was perceived as highly acceptable among individuals with CHD, with a majority perceiving that the app was satisfactory, useful, and easy to use.

Comparison With Prior Work

Evidence has shown that lack of scalability in health care management system can hamper system growth and prevent it from providing person-centered care [48]. A health care system adopting multiple services, such as iCARE, requires software developers to build and deploy the system in a reliable and timely fashion. Such system should be scalable to adapt to the evolving needs of patients and health care professionals [48]. The key barrier to achieve this is the selection of a suitable software architecture [36]. In this study, the MicroService architecture [49], an emerging architectural style for developing distributed apps, was applied in the development of iCARE. With the application of this architecture, our system may have the advantages in realizing availability, flexibility, scalability, and allowing multiple services to be scaled up independently [49,50]. Implementation of the MicroService architecture in this study allowed us to develop, modify, and deploy iCARE promptly based on the needs identified in the task analysis and issues revealed during the first and second usability evaluations.

Overall, the iCARE patient app was perceived as useful and satisfactory despite several commonly identified usability problems that were determined to be similar to those of other health-related self-management apps [51]. The 2-stage usability evaluation revealed that the mean scores on each item of the Health-ITUES were all above 4.0, and the end users commented that iCARE had much better usability in terms of perceived usefulness. There are multiple reasons for this: First, the development of the iCARE interventions was evidence based and followed a theory-driven approach. Second, iCARE was designed based on the user-centered design principle, which was shown to be effective in increasing user's engagement, thereby making the interactive system useful for targeted end users [35]. In our preliminary work, we carefully assessed the needs and preferences of individuals with CHD not only during the user interface design phase, but also during the development and modification phases. In addition, as suggested by Dawson et al [28], to promote sustained and significant behavior change among users, we encouraged nurses and other health professionals to be actively involved in the inspections and walkthroughs of the iCARE prototypes. Second, as outlined in our previous work [33], to ensure the interventions of iCARE are evidence based and compliant with clinical guidelines, we structured them based on available cardiovascular secondary prevention guidelines [3,30,31]. Meanwhile, an integrated behavior change theory, the CAM model, was used to guide the design of iCARE, to ensure interventions are empirically based [33]. Third, based on a previous study [52], a systematic and problem-solving approach along with a patient-centered and goal-oriented method was used in designing our system. To ensure the functions of iCARE are acceptable and useful, the components and schematic diagram of iCARE were designed following the principles of the nursing process, which are identified as essential to ensure nurses deliver holistic and patient-centered care [35]. These may have contributed to the good functionality of iCARE in this study.

Meanwhile, the perceived ease of use of a system by users is another essential element in developing useful tools [53]. However, achieving the balance between usefulness and ease of use is challenging as these 2 components are often contradictory during app development [54]. When designing a system, some designers sacrifice usefulness to provide ease of use, or vice versa [54]. For the health management of individuals with CHD, comprehensive CHD prevention strategies were recommended by the cardiovascular secondary prevention guidelines [3,30,31]. Therefore, the system should be designed with multiple functionalities to incorporate comprehensive CHD prevention strategies to achieve its optimal usefulness. However, increasing the amount of functionality in a system may also increase the treatment-related burden on patients and affect effective care delivery [55]. This is reflected in our case as well: the multiple functionalities included in our system increased the complexity of the system, which might have contributed to the lower scores on the ease of use functionality of the patient app during the first and second usability testing phases.

Strategies to address the ease-of-use functionality of the system are vital to reduce unnecessary burden on individuals. As identified by Tsopra et al [56], interfaces designed following usability principles suggested by the Healthcare Information and Management Systems Society [57], including effective information presentation, consistency, efficient interactions, effective use of language, and minimizing cognitive load, were perceived as ease of use. In our system, we applied several techniques to improve its ease of use, including the following: (1) the use of larger font sizes and different colors to improve readability; (2) the use of a read-out mode to minimize individuals' cognitive load; and (3) limiting the amount of text message and replacing it with figures, infographics, and chart

to ensure effective and visual information presentation. These techniques are also in agreement with a previous finding [58], in which the use of color-coded visual layout improved end user's perceived ease of use for a system. The findings of the usability testing in our study indicated that most participants perceived the patient app as easy to use.

However, we are fully aware that some design and functional elements of the app still need to be addressed to improve its ease of use. For example, the number of interfaces on the patient app should be reduced to promote efficient interactions; some icons should be designed as simpler and meaningful as possible to be easily recognized; and the system should provide more guidance and assistance to end users to avoid human errors and increase their perceived user control. In addition to the lessons learned during our system designing and developing process, future studies should pay more attention to maximize the learnability component through minimizing cognitive loads during the designing process, and improve the memorability of the system that can assist users to reestablish proficiency after a long period of inactivity [59].

Limitations

This study has several limitations. First, given the nature of the study, participants who were willing to participate in this study may be more motivated and more proficient in using the mHealth apps than the general population. Second, this study was conducted among patients who were admitted to hospitals in Beijing, which is a technology-advanced city in China. However, the 2 study hospitals are the most recognized hospitals in cardiovascular disease care in China, and patients admitted to these hospitals are from all over the country, including those from rural and underserved areas. Third, in the second stage of usability testing, we only subjectively evaluated the usability of the patient app; therefore, some usability problems within our system may not have been discovered adequately. Fourth, although the needs assessment was analyzed among patients with CHD, they were not actively involved in the interface designing process, which might influence the usability of the patient app. However, their response to stage I usability testing was applied to modify the system, and the updated version was used in the RCT and second usability evaluation. Fifth, the care provider app only has a few basic functions in its current stage; therefore, its usability was not considered in this study. However, we will expand the care provider app to include more fully functional components and evaluate its usability in the future. In addition, in its current stage, the data from external devices cannot be automatically uploaded to the patient app which might influence its usability. Finally, the iCARE system was developed to be only compatible with Android-based smartphones, and thus patients who used an iOS-based smartphones were excluded. Although iPhones are popular in China, they are very expensive, and are considered as high-end mobile phones that are similar to the high-end Huawei mobile phones; besides, iPhone users are usually mid- or high-income population that live in tech-developed cities, such as Shanghai or Beijing. Moreover, after completing our RCT, we will develop an updated iCARE system that will be compatible with both Android and iOS versions, as many patients who used iPhones expressed their interest and intention in using iCARE to manage their diseases and health. Therefore, we will perform additional usability testing in a more representative population.

Conclusions

This study developed the iCARE system with the aim to facilitate patients' adherence to multiple health behaviors and preventive medications through incorporation of theory-driven, evidence-based, individualized interventions and tailored automatic feedback, following the CAM model and principles of the nursing process. iCARE was demonstrated to be highly satisfactory, useful, and easy to use among individuals with CHD and had acceptable usability. We are currently evaluating the effectiveness of iCARE in patients with CHD via a randomized clinical trial to better understand the real-time usability of the app, as well as patients' experience in using the app. In future studies, we will revise iCARE based on the findings of this study and develop a version that is compatible with iOS-based smartphones.

Acknowledgments

This study was supported by the National Natural Science Funding of China (Grant/Award Number 71573181). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We extend our thanks to all study participants. Special thanks go to Karen V. Lamb, who is a Visiting Professor at Capital Medical University, a Visiting Associate Clinical Professor from University of Illinois Chicago College of Nursing Associate, and a Professor Emerita at Rush, for her review of the manuscript. In addition, the authors also thank the following team members who were involved in and made significant contribution to the development of iCARE (Intelligent Individualized Cardiovascular App for Risk Elimination): Meng Fan and Chaoqun Li (software engineers); and Lei Gu (a user interface designer at Along Technology Inc., Beijing, China); Yuzhi Shen, Lingling Qin, and Junqiu Yang (clinical nurses specialized in cardiovascular care at Beijing Chao-Yang Hospital); and Xinchun Yang (a cardiologist at Beijing Chao-Yang Hospital); Fang Li and Lixin Zhang (clinical nurses specialized in cardiovascular care at Beijing An-Zhen Hospital); and Shaoping Nie and Jinghua Liu (cardiologist at Beijing An-Zhen Hospital).

Authors' Contributions

YC, QW, YG, ML, and ZF managed the data collection. YC and MJ contributed to the drafting, interpretation, and critical revision of the manuscript. YC took part in the conception, study design, and led the data analysis and interpretation. YL, YC, FW, XZ, and YD participated in the design and development of the iCARE (Intelligent Individualized Cardiovascular App for Risk

Elimination) system. YW led the whole conduction process of the project, including the study design, designing and developing iCARE, quality control of study conduction, data analysis and interpretation, and writing of the report. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preliminary work, study participants and procedures, supplementary Tables and Figures.
[DOCX File , 920 KB - mhealth_v9i12e26439_app1.docx]

Multimedia Appendix 2
Task 1.
[MP4 File (MP4 Video), 4261 KB - mhealth_v9i12e26439_app2.mp4]

Multimedia Appendix 3
Task 2.
[MP4 File (MP4 Video), 4900 KB - mhealth_v9i12e26439_app3.mp4]

Multimedia Appendix 4
Task 3.
[MP4 File (MP4 Video), 4312 KB - mhealth_v9i12e26439_app4.mp4]

Multimedia Appendix 5
Task 4.
[MP4 File (MP4 Video), 3246 KB - mhealth_v9i12e26439_app5.mp4]

Multimedia Appendix 6
Task 5.
[MP4 File (MP4 Video), 1447 KB - mhealth_v9i12e26439_app6.mp4]

Multimedia Appendix 7
Task 6.
[MP4 File (MP4 Video), 4300 KB - mhealth_v9i12e26439_app7.mp4]

Multimedia Appendix 8
Task 7.
[MP4 File (MP4 Video), 3515 KB - mhealth_v9i12e26439_app8.mp4]

Multimedia Appendix 9
Task 8.
[MP4 File (MP4 Video), 2132 KB - mhealth_v9i12e26439_app9.mp4]

References

1. Mendis S, Puska PB, Norrving B. Global Atlas on Cardiovascular Disease Prevention and Control. Geneva, Switzerland: World Health Organization; 2011.
2. Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. Curr Probl Cardiol 2010 Feb;35(2):72-115 [FREE Full text] [doi: 10.1016/j.cpcardiol.2009.10.002] [Medline: 20109979]
3. National Center for Cardiovascular Disease of China. Report on Cardiovascular Disease in China 2019 (in Chinese). Beijing, China: China Publishing House; Sep 2020:833-854.
4. Devaraj S, Kriska A, Orchard T, Miller R, Costacou T. Cardiovascular health in early adulthood predicts the development of coronary heart disease in individuals with type 1 diabetes: 25 year follow-up from the Pittsburgh Epidemiology of Diabetes Complications study. Diabetologia 2021 Mar;64(3):571-580. [doi: 10.1007/s00125-020-05328-9] [Medline: 33190165]

5. Lv J, Yu C, Guo Y, Bian Z, Yang L, Chen Y, China Kadoorie Biobank Collaborative Group. Adherence to healthy lifestyle and cardiovascular diseases in the Chinese population. J Am Coll Cardiol 2017 Mar 07;69(9):1116-1125 [FREE Full text] [doi: 10.1016/j.jacc.2016.11.076] [Medline: 28254173]
6. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med 2007 Jun 07;356(23):2388-2398. [doi: 10.1056/NEJMsa053935] [Medline: 17554120]
7. Han C, Liu F, Yang X, Chen J, Li J, Cao J, et al. Ideal cardiovascular health and incidence of atherosclerotic cardiovascular disease among Chinese adults: the China-PAR project. Sci China Life Sci 2018 May;61(5):504-514. [doi: 10.1007/s11427-018-9281-6] [Medline: 29721777]
8. Hu G, Zhou M, Liu J, Smith SC, Ma C, Ge J, CCC-ACS Investigators. Smoking and provision of smoking cessation interventions among inpatients with acute coronary syndrome in China: Findings from the Improving Care for Cardiovascular Disease in China-Acute Coronary Syndrome Project. Glob Heart 2020 Oct 23;15(1):72 [FREE Full text] [doi: 10.5334/gh.784] [Medline: 33150137]
9. Du H, Li L, Bennett D, Guo Y, Key TJ, Bian Z, China Kadoorie Biobank Study. Fresh fruit consumption and major cardiovascular disease in China. N Engl J Med 2016 Apr 07;374(14):1332-1343 [FREE Full text] [doi: 10.1056/NEJMoa1501451] [Medline: 27050205]
10. Kotseva K, Wood D, De Bacquer D, De Backer G, Rydén L, Jennings C, EUROASPIRE Investigators. EUROASPIRE IV: A European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. Eur J Prev Cardiol 2016 Apr;23(6):636-648. [doi: 10.1177/2047487315569401] [Medline: 25687109]
11. Tang L, Patao C, Chuang J, Wong ND. Cardiovascular risk factor control and adherence to recommended lifestyle and medical therapies in persons with coronary heart disease (from the National Health and Nutrition Examination Survey 2007-2010). Am J Cardiol 2013 Oct 15;112(8):1126-1132. [doi: 10.1016/j.amjcard.2013.05.064] [Medline: 23827404]
12. Teo K, Lear S, Islam S, Mony P, Dehghan M, Li W, PURE Investigators. Prevalence of a healthy lifestyle among individuals with cardiovascular disease in high-, middle- and low-income countries: The Prospective Urban Rural Epidemiology (PURE) study. JAMA 2013 Apr 17;309(15):1613-1621. [doi: 10.1001/jama.2013.3519] [Medline: 23592106]
13. Wood DA, Kotseva K, Connolly S, Jennings C, Mead A, Jones J, EUROACTION Study Group. Nurse-coordinated multidisciplinary, family-based cardiovascular disease prevention programme (EUROACTION) for patients with coronary heart disease and asymptomatic individuals at high risk of cardiovascular disease: a paired, cluster-randomised controlled trial. Lancet 2008 Jun 14;371(9629):1999-2012. [doi: 10.1016/S0140-6736(08)60868-5] [Medline: 18555911]
14. Rautio N, Jokelainen J, Pölönen A, Oksa H, Peltonen M, Vanhala M, et al. Changes in lifestyle modestly reduce the estimated cardiovascular disease risk in one-year follow-up of the Finnish diabetes prevention program (FIN-D2D). Eur J Cardiovasc Nurs 2015 Apr;14(2):145-152. [doi: 10.1177/1474515114521713] [Medline: 24452452]
15. Lv J, Liu Q, Ren Y, He P, Wang S, Gao F, Community Interventions for Health (CIH) collaboration. A community-based multilevel intervention for smoking, physical activity and diet: short-term findings from the Community Interventions for Health programme in Hangzhou, China. J Epidemiol Community Health 2014 Apr;68(4):333-339 [FREE Full text] [doi: 10.1136/jech-2013-203356] [Medline: 24297972]
16. Du H, Venkatakrishnan A, Youngblood GM, Ram A, Pirolli P. A group-based mobile application to increase adherence in exercise and nutrition programs: a factorial design feasibility study. JMIR Mhealth Uhealth 2016;4(1):e4 [FREE Full text] [doi: 10.2196/mhealth.4900] [Medline: 26772910]
17. Maron DJ, Mancini GBJ, Hartigan PM, Spertus JA, Sedlis SP, Kostuk WJ, COURAGE Trial Group. Healthy behavior, risk factor control, and survival in the COURAGE trial. J Am Coll Cardiol 2018 Nov 06;72(19):2297-2305 [FREE Full text] [doi: 10.1016/j.jacc.2018.08.2163] [Medline: 30384885]
18. 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]
19. Hong YA, Zhou Z, Fang Y. Digital divide and health disparities in China: data from a national longitudinal survey of CHARLS. Stud Health Technol Inform 2017;245:1301. [Medline: 29295384]
20. Sobrinho A, da Silva LD, Perkusich A, Pinheiro ME, Cunha P. Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries. BMC Med Inform Decis Mak 2018 Jan 12;18(1):7 [FREE Full text] [doi: 10.1186/s12911-018-0587-9] [Medline: 29329530]
21. 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]
22. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]

23. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
24. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
25. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
26. 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]
27. Xiao Q, Lu S, Wang Y, Sun L, Wu Y. Current status of cardiovascular disease-related smartphone apps downloadable in China. Telemed J E Health 2017 Mar;23(3):219-225. [doi: 10.1089/tmj.2016.0083] [Medline: 27356156]
28. Xie B, Su Z, Zhang W, Cai R. Chinese cardiovascular disease mobile apps' information types, information quality, and interactive functions for self-management: systematic review. JMIR Mhealth Uhealth 2017 Dec 14;5(12):e195 [FREE Full text] [doi: 10.2196/mhealth.8549] [Medline: 29242176]
29. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematicrReview of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
30. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016 Dec 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
31. 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 endorsed by the World Heart Federation and the Preventive Cardiovascular Nurses Association. J Am Coll Cardiol 2011 Nov 29;58(23):2432-2446 [FREE Full text] [doi: 10.1016/j.jacc.2011.10.824] [Medline: 22055990]
32. Bartholomew LC, Markham CM, Ruiter RAC, Fernandez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. Hoboken, NJ: John Wiley & Sons; Feb 1, 2016:57-854.
33. Chen Y, Wu F, Wu Y, Li J, Yue P, Deng Y, et al. Development of interventions for an intelligent and individualized mobile health care system to promote healthy diet and physical activity: using an intervention mapping framework. BMC Public Health 2019 Oct 17;19(1):1311 [FREE Full text] [doi: 10.1186/s12889-019-7639-7] [Medline: 31623589]
34. Yue P, Wu Y, Zhang Y, Chen Y, Li J, Xu Y, et al. Contemplation-action-maintenance model of behaviour change for persons with coronary heart disease: A qualitative study. J Clin Nurs 2021 May;30(9-10):1464-1478. [doi: 10.1111/jocn.15699] [Medline: 33555622]
35. Dawson RM, Felder TM, Donevant SB, McDonnell KK, Card EB, King CC, et al. What makes a good health 'app'? Identifying the strengths and limitations of existing mobile application evaluation tools. Nurs Inq 2020 Apr;27(2):e12333. [doi: 10.1111/nin.12333] [Medline: 31854055]
36. Chen L. Microservices: Architecting for continuous delivery and DevOps. New York, NY: IEEE; 2018 Presented at: IEEE International Conference on Software Architecture (ICSA); April 30-May 4, 2018; Seattle, WA p. 39-397. [doi: 10.1109/icsa.2018.00013]
37. Yang Y, Zu Q, Liu P, Ouyang D, Li X. MicroShare: Privacy-preserved medical resource sharing through MicroService architecture. Int J Biol Sci 2018;14(8):907-919 [FREE Full text] [doi: 10.7150/ijbs.24617] [Medline: 29989095]
38. Taherizadeh S, Stankovski V, Grobelnik M. A capillary computing architecture for dynamic internet of things: orchestration of microservices from Edge devices to Fog and Cloud providers. Sensors (Basel) 2018 Sep 04;18(9):2938 [FREE Full text] [doi: 10.3390/s18092938] [Medline: 30181454]
39. The App Design App. URL: https://www.flinto.com/ [accessed 2021-11-23]
40. Lumsden J. Handbook of Research on User Interface Design and Evaluation for Mobile Technology. 2008 Feb. URL: https://hci.stanford.edu/publications/2007/MobilePrototyping.pdf [accessed 2021-11-11]
41. MOBGEN. High-Fidelity Prototyping: What, When, Why and How?. URL: https://blog.prototypr.io/high-fidelity-prototyping-what-when-why-and-how-f5bbde6a7fd4/[accessed [accessed 2021-11-19]
42. Kushniruk AW, Borycki EM. Integrating low-cost rapid usability testing into agile system development of healthcare IT: a methodological perspective. Stud Health Technol Inform 2015;210:200-204. [Medline: 25991130]
43. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]

44. Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs)-Part 11: Guidance on Usability. 1998. URL: https://www.iso.org/standard/16883.html [accessed 2021-11-19]
45. 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]
46. Nielsen J. Severity Ratings for Usability Problems. 1994 Nov 1. URL: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [accessed 2021-11-11]
47. Vélez O, Okyere PB, Kanter AS, Bakken S. A usability study of a mobile health application for rural Ghanaian midwives. J Midwifery Womens Health 2014;59(2):184-191 [FREE Full text] [doi: 10.1111/jmwh.12071] [Medline: 24400748]
48. Roca S, Sancho J, García J, Alesanco. Microservice chatbot architecture for chronic patient support. J Biomed Inform 2020 Feb;102:103305. [doi: 10.1016/j.jbi.2019.103305] [Medline: 31622802]
49. Newman S. Building Microservices: Designing Fine-Grained Systems. Sebastopol, CA: O'Reilly Media, Inc; Jul 23, 2021.
50. Garcia-Moreno FM, Bermudez-Edo M, Garrido JL, Rodríguez-García E, Pérez-Mármol JM, Rodríguez-Fórtiz MJ. A microservices e-health system for ecological frailty assessment using wearables. Sensors (Basel) 2020 Jun 17;20(12):3427 [FREE Full text] [doi: 10.3390/s20123427] [Medline: 32560529]
51. 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]
52. Kim H, Park H, Min YH, Jeon E. Development of an obesity management ontology based on the nursing process for the mobile-device domain. J Med Internet Res 2013 Jun 28;15(6):e130 [FREE Full text] [doi: 10.2196/jmir.2512] [Medline: 23811542]
53. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
54. Ease of Use. URL: https://www.interaction-design.org/literature/topics/ease-of-use [accessed 2020-11-19]
55. Tran V, Barnes C, Montori VM, Falissard B, Ravaud P. Taxonomy of the burden of treatment: a multi-country web-based qualitative study of patients with chronic conditions. BMC Med 2015 May 14;13:115 [FREE Full text] [doi: 10.1186/s12916-015-0356-x] [Medline: 25971838]
56. Tsopra R, Jais J, Venot A, Duclos C. Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. J Am Med Inform Assoc 2014 Feb;21(e1):e107-e116 [FREE Full text] [doi: 10.1136/amiajnl-2013-002042] [Medline: 24008427]
57. Belden J, Grayson R, Barnes J. Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating. 2009 Jun 30. URL: https://www.researchgate.net/publication/277829258_Defining_and_Testing_EMR_Usability_Principles_and_Proposed_Methods_of_EMR_Usability_Evaluation_and_Rating [accessed 2021-11-11]
58. Lazard AJ. Design cues for tobacco communication: Heuristic interpretations and usability of online health information about harmful chemicals. Int J Med Inform 2020 Sep;141:104177 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104177] [Medline: 32521448]
59. Shukairy A. 5 usability design tips for a better user experience design. URL: https://www.invespcro.com/blog/usability-design-for-a-better-user-experience/ [accessed 2020-11-19]

Abbreviations

CAM: Contemplation–Action–Maintenance model
CHD: coronary heart disease
Health-ITUES: Health Information Technology Usability Survey
iCARE: Intelligent Individualized Cardiovascular App for Risk Elimination
ISO: International Organization for Standardization
mHealth: mobile health
RCT: randomized controlled trial

Edited by R Kukafka, G Eysenbach; submitted 12.12.20; peer-reviewed by J Bruthans, S Donevant; comments to author 08.02.21; revised version received 18.03.21; accepted 05.10.21; published 13.12.21.

Please cite as:
Chen Y, Ji M, Wu Y, Wang Q, Deng Y, Liu Y, Wu F, Liu M, Guo Y, Fu Z, Zheng X
An Intelligent Individualized Cardiovascular App for Risk Elimination (iCARE) for Individuals With Coronary Heart Disease: Development and Usability Testing Analysis
JMIR Mhealth Uhealth 2021;9(12):e26439
URL: https://mhealth.jmir.org/2021/12/e26439
doi:10.2196/26439
PMID:34898449

Original Paper

Walking and Daily Affect Among Sedentary Older Adults Measured Using the StepMATE App: Pilot Randomized Controlled Trial

Alycia N Bisson[1,2], BSc, MA, PhD; Victoria Sorrentino[1], BSc, MA; Margie E Lachman[1], PhD

[1]Psychology Department, Brandeis University, Waltham, MA, United States

[2]Psychiatry Department, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Alycia N Bisson, BSc, MA, PhD
Psychiatry Department
Brigham and Women's Hospital
Room 394
221 Longwood Ave
Boston, MA, 02115
United States
Phone: 1 5082598151
Email: alyciansullivan@brandeis.edu

Abstract

Background: Although fitness technology can track and encourage increases in physical activity, few smartphone apps are based on behavior change theories. Apps that do include behavioral components tend to be costly and often do not include strategies to help those who are unsure of how to increase their physical activity.

Objective: The aim of this pilot study is to test the efficacy of a new app, StepMATE, for increasing daily walking in a sample of inactive adults and to examine daily relationships between walking and self-reported mood and energy.

Methods: The participants were middle-aged and older adults aged ≥50 years (mean 61.64, SD 7.67 years). They were randomly assigned to receive either a basic, pedometer-like version of the app or a version with supports to help them determine where, when, and with whom to walk. Of the 96 participants randomized to 1 of 2 conditions, 87 (91%) completed pretest assessments and 81 (84%) successfully downloaded the app. Upon downloading the app, step data from the week prior were automatically recorded. The participants in both groups were asked to set a daily walking goal, which they could change at any point during the intervention. They were asked to use the app as much as possible over the next 4 weeks. Twice per day, pop-up notifications assessed mood and energy levels.

Results: Although one group had access to additional app features, both groups used the app in a similar way, mainly using just the walk-tracking feature. Multilevel models revealed that both groups took significantly more steps during the 4-week study than during the week before downloading the app (γ=0.24; P<.001). During the study, the participants in both groups averaged 5248 steps per day compared with an average of 3753 steps per day during the baseline week. Contrary to predictions, there were no differences in step increases between the two conditions. Cognition significantly improved from pre- to posttest (γ=0.17; P=.02). Across conditions, on days in which the participants took more steps than average, they reported better mood and higher energy levels on the same day and better mood on the subsequent day. Daily associations among walking, mood, and energy were significant for women but not for men and were stronger for older participants (those aged ≥62 years) than for the younger participants.

Conclusions: Both groups increased their steps to a similar extent, suggesting that setting and monitoring daily walking goals was sufficient for an initial increase and maintenance of steps. Across conditions, walking had benefits for positive mood and energy levels, particularly for women and older participants. Further investigations should identify other motivating factors that could lead to greater and more sustained increases in physical activity.

Trial Registration: ClinicalTrials.gov NCT03124537; https://clinicaltrials.gov/ct2/show/NCT03124537

(JMIR Mhealth Uhealth 2021;9(12):e27208) doi:10.2196/27208

KEYWORDS

physical activity; fitness technology; intervention; behavioral science; aging; mobile phone

Introduction

Background

The benefits of physical activity for lifelong health, well-being, and cognition are well-documented; yet, most American adults lead an inactive lifestyle [1-3]. According to the Centers for Disease Control and Prevention, only 53% of adults meet the guidelines for aerobic activity, and even fewer older adults meet these guidelines [4]. Fitness technologies such as Fitbit, Apple Watch, or smartphone apps can track and encourage physical activity without the need for additional equipment or a gym membership. Indeed, the *health and fitness* category is one of the most popular categories in the iTunes and Google Play app stores, with almost 230,000 apps available in 2017.

A recent review of 37 Fitbit-based interventions reported that studies were associated with increases in daily steps, moderate to vigorous physical activity, and decreases in body weight [5]. Participants took approximately 950 steps per day more than the controls who were not given a Fitbit. The behavioral components included in the given interventions were related to the success of the interventions [5]. Goal-setting was described as the most promising component; however, a combination of intervention tools may be necessary to encourage changes in physical activity [5]. Another recent review suggested that there is little to moderate evidence that mobile health or eHealth interventions are successful for increasing physical activity in older adults [6].

Although many devices and smartphone apps currently track physical activity, encourage users to meet step goals, and link with other personal data, few stand-alone smartphone apps include additional features that address barriers unique to inactive adults [7]. Furthermore, apps that do include behavior change strategies typically cost more money and do not provide features such as action planning and environmental supports. Focus groups have identified a need for physical activity apps to promote autonomy and self-regulation, while also providing adaptability and flexibility to accommodate individual needs [8].

Implementation intentions involve behavioral strategies such as creating a specific plan to reach a goal [9]. Using walking as an example, implementation intentions could include action planning, which involves creating a plan that includes the time and place that walking would occur [10]. Meta-analyses have shown that action planning is associated with increases in physical activity [11,12]. A recent study tested whether an implementation intentions intervention was more successful in increasing physical activity than just using a Fitbit [13]. The intervention group participants, who were given step goals, personalized walking routes, and a daily schedule to fill out, significantly increased their daily steps over 1 month compared with the control group participants who only wore a Fitbit [13]. Although action planning and environmental supports are rarely incorporated into fitness technology, such strategies may directly address common barriers that prevent adults from engaging in physical activity [7].

Physical Activity and Affect

Along with the benefits of exercise to physical and cognitive health, many have shown the importance of physical activity for mood and affective states [14,15]. In fact, a recent meta-analysis reported that improved executive functioning and mood, along with decreases in stress, are among the most consistently reported outcomes after exercise [14]. These effects have been echoed in multiple populations and various activity domains, including vigorous activities such as cycling and lower-intensity activities such as yoga or walking. A study showed that patients with multiple sclerosis were more likely to report improved mood after a single 20-minute bout of walking or yoga than after an equivalent period of rest [16]. Others have also found that positive exercise experiences are linked to increases in motivational self-efficacy and exercise intentions, which then predict future exercise behavior [17].

New technologies have made it possible to examine the relationships among physical activity, mood, and affect in real time using accelerometry along with methods such as experience sampling or ecological momentary assessments (EMAs). A study used a newly developed smartphone app and EMAs to test whether self-reported happiness and physical activity are linked [18]. The results from >10,000 app users showed that more active individuals reported being happier than those who were inactive. Daily relationships also emerged; people were happier on more physically active days than on less active ones [18].

A review paper by Liao et al [19] summarized 14 studies that used EMAs to examine short-term relationships between physical activity and affect. The authors found evidence for reciprocal relationships; current positive affect predicted increased physical activity within the next few hours, and physical activity engagement predicted greater positive affect within the next few hours [19]. Thus, it seems that positive affect predicts subsequent physical activity, which also predicts future positive affect.

It is possible that men and women experience differential effects of exercise on mood; however, very little work has examined sex differences. A study found that in young adults, women were more likely to report improvements in mood after exercise than men [20]. The same study found that women were more likely than men to report reduced fatigue after a 30-minute bout of exercise [20]. It is possible that women are more sensitive to mood changes after exercise.

In sum, physical activity and affect have been linked at both the within-person and between-person levels. The effects are similar across various domains of physical activity and in healthy and nonhealthy adult populations. Affective changes can be seen from acute (20-minute) bouts of activity to regular activity over the course of months. Although prior studies have examined affective improvements in the context of structured exercise, no studies to our knowledge have tested whether the number of steps one takes per day is predictive of contemporaneous changes in affect. Furthermore, few studies have closely examined whether the effects differ between men and women.

Physical Activity, Sleep, and Energy

Another consistent finding in the literature is the relationship between physical activity and sleep. When examining average levels of physical activity, people who are more active tend to sleep better than those who are less active [21]. Most of this work has focused on high-impact physical activity or on populations with sleep disorders or other health problems. Daily studies suggest that on the days in which people are more active, they sleep better and longer than on less active days [22-24]. Recent work found that women who average more steps per day over the course of a month reported better sleep quality than inactive women, not men [22].

A study examined the relationship among physical activity, affect, and insomnia symptoms in a sample of inactive adults with insomnia [25]. Those who were asked to engage in consistent walking reported significant decreases in insomnia symptoms, along with improved affect, over the 6-month intervention [25]. Taken together, the results suggest that even low-impact physical activity such as walking or yoga can improve sleep in adults. Those who sleep well will likely report higher energy levels during the day; however, self-reports of energy are also affected by other things that happen on any given day. Although the link between physical activity and sleep has been studied, less is known about how daily physical activity is related to self-reported daily energy levels.

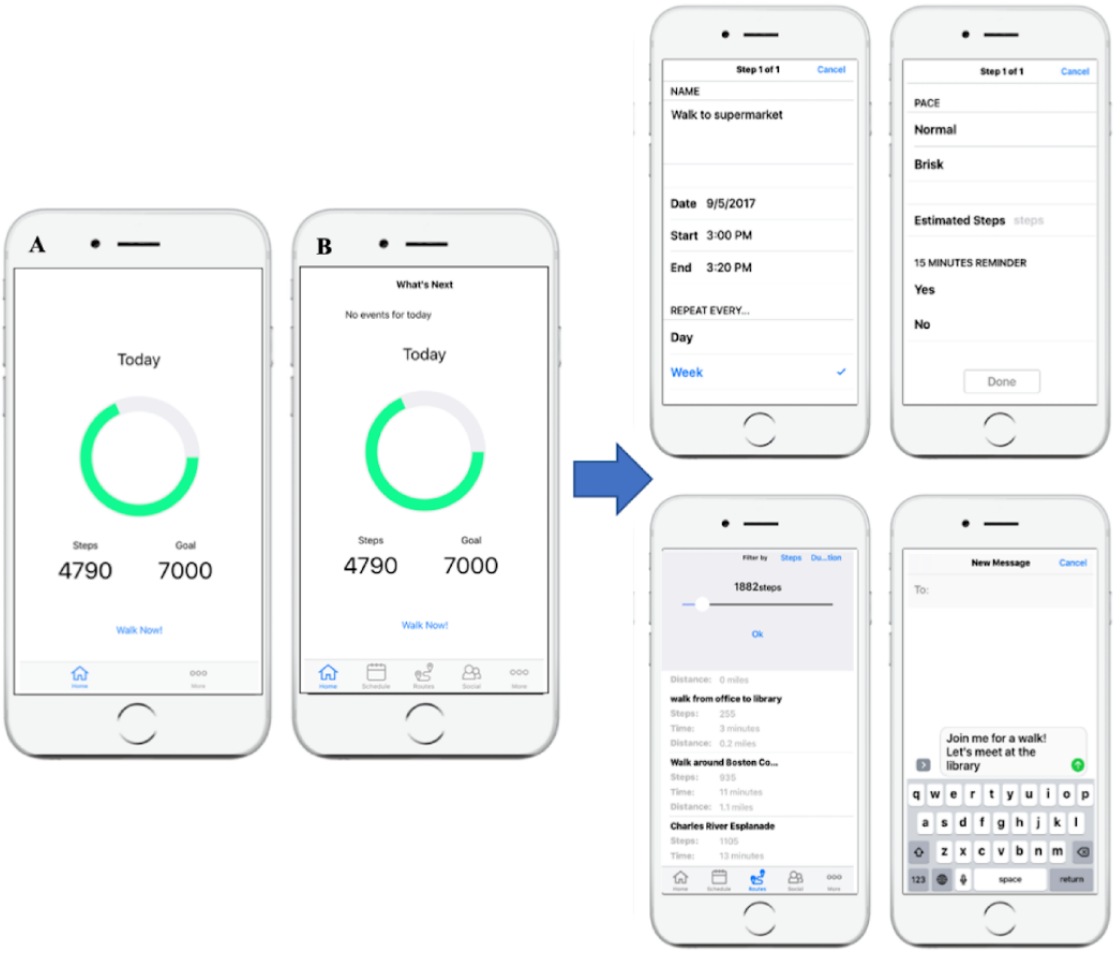
This study aims to test whether an iPhone (Apple Inc) app—StepMATE—with behavioral supports was associated with increases in daily walking among a sample of inactive but otherwise healthy adults. We also aim to examine whether a version of the app with additional action planning strategies is more successful than a version with only step-tracking and daily step goals. Finally, we aim to assess within-person fluctuations in steps, mood, and energy and whether there are differences in these relationships based on demographic characteristics, including age and sex.

Methods

Study Details

On the basis of our previous study on midlife adults that assessed barriers to being physically active [13], along with the findings from pilot interviews of 9 older adults, we found that perceived lack of time was a common barrier preventing people from getting enough exercise. Other barriers reported in these studies were not knowing where to exercise and not wanting to exercise alone. The StepMATE app (Figure 1) was developed by Beneufit using Apple ResearchKit, with feedback from university researchers. StepMATE is a fully automated app that includes behavioral supports to help people plan where and when to walk and social supports to help find others who might want to walk with them.

Figure 1. Screenshots of the StepMATE app. (A) The home screen for the control group. (B) The home screen for the treatment group members, who had access to the additional features shown in the screenshots on the right.

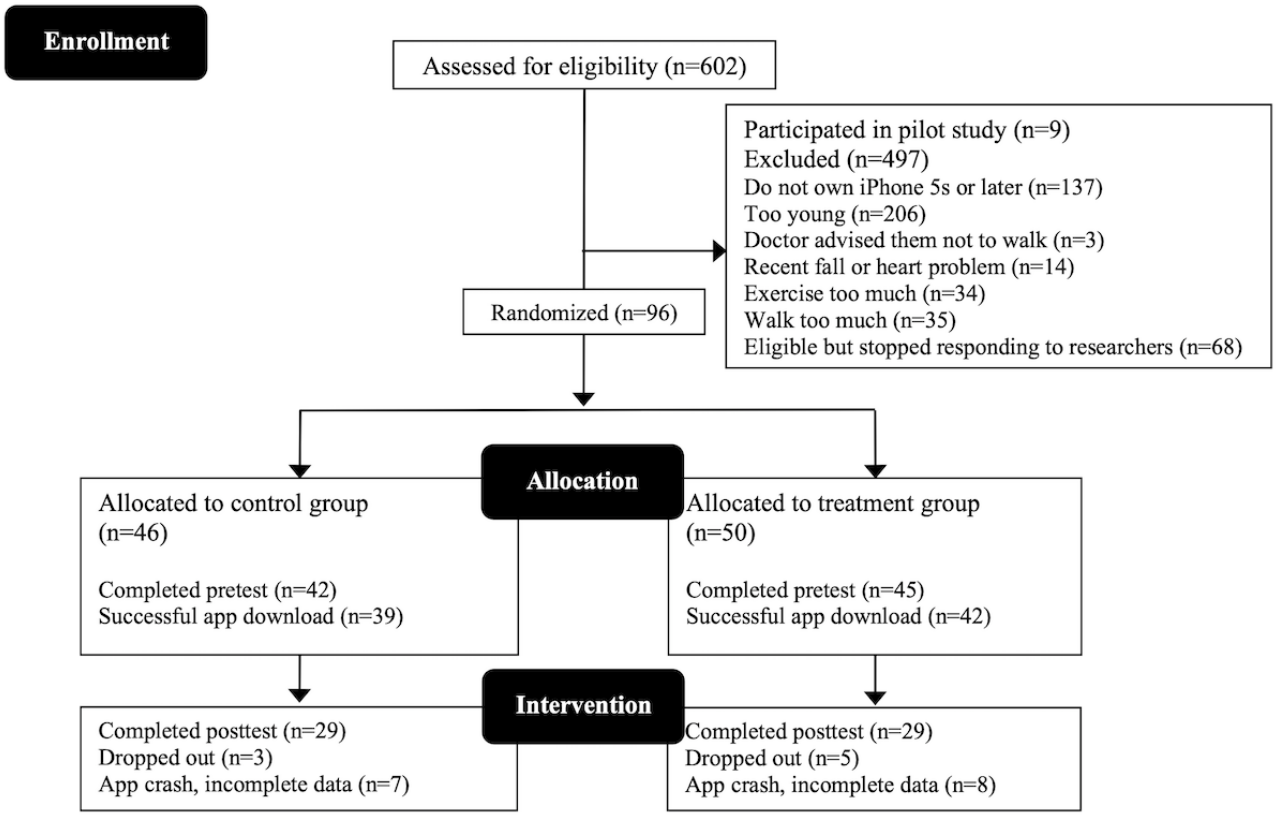


The participants in the StepMATE pilot randomized controlled trial were randomly assigned to receive 1 of 2 versions of the app. The control group was only given step-tracking and goal-setting functions, similar to those of a Fitbit, and the treatment group was given a version of the app with additional social and environmental supports. First, we tested whether the app was associated with increases in average daily steps over the 4 weeks and whether there were differences between the 2 conditions. It was hypothesized that the additional supports would result in greater increases in walking for the treatment group compared with the control group. Next, we examined whether there were changes over time or between-group differences in other outcomes, including sleep, exercise control, exercise self-efficacy, social engagement, and memory. We hypothesized that the participants would report improvements in these outcomes from pre- to posttest, with greater improvements in the treatment group. Finally, within-person relationships between daily steps and self-reported mood and energy were modeled. We hypothesized that on the days when the participants took more steps than average, they would report higher energy levels and better mood than on less active days. Drawing from prior findings on sex differences in daily relationships between exercise and other outcomes, the interactions between daily steps and sex on mood and energy were examined. It was hypothesized that daily steps would be more closely related to mood and energy in women than in men. Exploratory analyses examined whether there were interactions between daily steps and age in predicting mood and energy.

Participants

The participants were recruited on the web on a rolling basis between January 2018 and March 2019 using Facebook, Craigslist, and FindParticipants. Participants were also recruited locally in eastern and central Massachusetts through flyers at senior centers, libraries, cafes, and community events. As the study did not require an in-person meeting, participants were recruited from locations across the continental United States. The CONSORT table is presented in Figure 2, detailing recruitment, enrollment, and exclusion criteria for this study.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram.



The participants were required to own an iPhone with a built-in accelerometer to measure steps (iPhone 5s or newer). Only those who reported exercising less than the Centers for Disease Control and Prevention guidelines of 150 minutes of moderate to vigorous exercise per week were eligible. The participants also needed to report walking for exercise no more than 30 minutes per day [26]. Participants were ineligible if a physician advised them not to walk because of health conditions or if they had had a cardiac event or fall within the last 6 months. Screening for cognitive impairment was conducted over the phone using a shortened version of the Short Portable Mental Status Questionnaire [27]. Participants were ineligible if they made ≥3 errors on this questionnaire.

All procedures were approved by the university institutional review board. An a priori power analysis for the primary outcome variable, number of steps, was conducted using G*Power (version 3.1; Heinrich Heine University) [28], which indicated that 31 participants per condition were required with an estimated effect size of d=0.10, with 95% power at P=.05.

After an iPhone software update, the StepMATE app crashed and did not work properly for approximately 2 weeks. Of the

81 participants who downloaded the app, 18 (22%) were affected, and daily step data were lost for these participants. Analyses were conducted on an intent-to-treat basis. For the questionnaire data, we analyzed all participants who completed the pretest measures (87/96, 91%), and for the daily step analyses, all participants with sufficient step data were included in the analyses (80/81, 99%). Sample sizes are included in all results tables for clarity.

Pre- and Posttest Measures

Social Engagement

Social engagement was measured using the Lubben Social Network Scale [29]. This scale comprises 12 items (6 related to family and 6 to friends) that ask about the size of one's social network (eg, How many friends or family members do you feel at ease with that you can talk about private matters?) and the closeness of the relationships (eg, How often do you see or hear from the friend or family member with whom you have the most contact?). A composite score was calculated by summing the responses of the 12 items, with a final score ranging from 0 to 60, where a higher score indicates more social engagement. The Cronbach α value for the internal consistency in this sample was .862.

Exercise Control

Control over exercise was measured using the 6-item Exercise Control Beliefs Scale [30]. The items assess one's beliefs about one's control over exercise (eg, I am confident in my ability to do an exercise routine), with answer choices ranging from strongly disagree (score=1) to strongly agree (score=5). The 6 items were averaged to create a mean exercise control score, with a higher score indicating greater control over exercise. The Cronbach α value for the internal consistency in this sample was .604.

Exercise Self-efficacy

A modified version of the Bandura Exercise Self-Efficacy Scale [31] was used in this study. This 9-item scale assesses how sure an individual is that they would exercise under different conditions or constraints (eg, How sure are you that you will exercise when you are feeling tired or under pressure to get things done?), with answer choices ranging from not sure at all (score=1) to very sure (score=4). The 9 items were averaged to create a composite score, where a higher score indicates greater exercise self-efficacy. The Cronbach α value for the reliability of this scale was .935.

Cognitive Performance

Cognition was assessed using a shortened version of the Brief Test of Adult Cognition by Telephone (BTACT) [32]. This version of the BTACT assesses 5 cognitive dimensions, including 2 measures of episodic verbal memory (immediate and delayed free recall of 15 words), working memory (backward digit span), verbal fluency (the number of words produced from a given category within 60 seconds), and processing speed (counting backward from 100 in 30 seconds). The primary outcome measure was a composite of all cognitive tests. The scores on both occasions (pre- and posttest) were standardized based on the scores at the pretest to create cognitive composites. The BTACT is a reliable assessment of cognitive functioning; its psychometric properties have been reported in another manuscript [32].

Sleep

Sleep duration and quality were measured using the Pittsburgh Sleep Quality Index (PSQI) [33]. The PSQI global score could range from 0 to 21, with a higher score indicating *worse* sleep. The Cronbach α value for the reliability of the 7 subscales of the PSQI was .77. In this study, we examined the PSQI global score, along with raw scores for duration (average number of hours slept during the past month) and latency (average number of minutes taken to fall asleep during the past month).

Daily Measures

App Engagement

To assess the use of various app features, for each participant, the total number of SMS text messages sent to contacts, number of routes saved, number of scheduled events, and number of times the *Walk Now* feature was used were computed.

Physical Activity

Every day, over the course of a month, physical activity was assessed using the total number of steps taken each day. Daily steps were quantified using the iPhone's built-in accelerometer and recorded through the StepMATE app. When the participants downloaded StepMATE at the beginning of the study, the app automatically and retroactively recorded daily steps from the week before the start of the study. During the 4-week intervention, the participants were asked to carry their phone with them during the day; however, they were not specifically instructed to do so during the baseline week before the intervention began.

Although data indicate that older adults typically average between 2000 and 9000 steps per day [34], there are likely times when the participants walked without carrying their iPhone. For the days when the iPhone recorded fewer than 500 steps, that day of steps was coded as missing. Weekly step averages were calculated for weeks with 4 or more days with 500 or more daily steps. Of the 81 participants included in this intent-to-treat analysis, 11 (14%) had missing or incomplete baseline data.

Daily Affect: Mood and Energy

Twice, at random times each day, mood and energy levels were assessed. A pop-up notification asked the participants to rate their current mood (unhappy, neutral, or happy) and energy (low, neutral, or high) on a slider scale. The scores were converted by using the StepMATE app to a 0-10 scale, with 0 indicating low mood or energy and 10 indicating high mood or energy. The 2 daily ratings were averaged to provide a daily average of the participants' mood and energy.

Covariates

Age, sex, education, and health were covariates in the current set of analyses because they were expected to be related to the outcomes. In models where time×condition interactions were not estimated, condition was included as a covariate. Age was continuous, sex was coded as 1=male and 2=female, and education was number of years in school. Health was measured

using the general health subscale from the Medical Outcomes Study 36-item Short-Form Health Survey [35]. Condition was coded as 0=control group and 1=treatment group.

Design and Procedures

App

The StepMATE app was designed to help participants determine when, where, and with whom they would add physical activity to their day. Daily step goals were set by the participants in both the intervention and control groups, and all participants had the ability to change their step goal at any point throughout the study. For the *when* component, those in the intervention condition had a scheduling feature in the app. The participants could schedule a block of time to go for a walk, and they had the option to create a reminder, set recurring events, and estimate the number of steps they would get in that walk. Once an event was created, it appeared in both the StepMATE app and the iPhone's built-in calendar.

For the *where* component, those in the intervention group were able to create, name, and save walking routes in the app. When the participants in the intervention group hit *Walk Now*, StepMATE began keeping track of their geographical location, distance, number of steps, and total time of the walk. This information was then saved after the walk was finished so that the time it took to walk a route could be compared if the same route was walked again. When a user created multiple routes, they could be filtered by number of steps or duration so that the user could easily find a walk that fit the amount of time they had or the number of steps they needed to achieve their daily walking goal. Those in the control group also had a *Walk Now* button; however, when the control participants hit *Walk Now*, the app would simply track the number of steps taken in that walk. These participants were not able to name or save their walks, nor could they view their walks on a map.

For the *with whom* (social feature) component, those in the intervention condition had the option to text one of their iPhone contacts through the app and invite them for a walk. Those in the control condition did not have access to this feature. Multimedia Appendix 1 includes screenshots of the app, video tutorials of the app features, and differences between the 2 versions, as well as descriptions of other app functions.

Procedures

A research assistant used Microsoft Excel for the block randomization procedures. Blocks of 10 consecutive ID numbers were randomly assigned to 1 of the 2 conditions. The app developer received the lists of ID numbers and associated treatment condition so that when an ID number was provided during the app download, the correct version of the app would install on the participant's phone. Upon meeting the inclusion criteria and consenting to participate in the study, each participant was assigned an ID number that was paired to the condition generated from the block randomization. Next, the participants were administered a shortened version of the BTACT, and they completed the prestudy questionnaires on the web through Qualtrics, including self-assessed social engagement, exercise control, exercise self-efficacy, and sleep. The participants filled out their ID number at the beginning of the Qualtrics survey so that their self-report data could easily be linked to their step data.

Subsequently, the researchers scheduled a phone call to help the participants to download the app, set up their account, including daily walking goals, and thoroughly explain the app features. Those in the control condition downloaded a version with only the daily step goals and the ability to track time, distance, and steps within a walk. Those in the treatment condition had access to these and additional features, including schedules, maps, and social features. The participants were blinded to which condition they were assigned to receive. Although the researchers were aware of the condition assignment for the purposes of helping with app downloads and troubleshooting issues, all measures—except for the cognitive assessments—were carried out on the web without researcher involvement. Randomization was checked by comparing the covariates (age, sex, education, and general health) between the conditions using independent samples *t* tests (2-tailed). No significant differences were found between the conditions for any of these variables.

The participants in both groups were asked to use the app for 1 month and do their best to answer the daily mood and energy questions. All participants were sent a pouch to wear around their waists and were encouraged to use it to carry their phone with them as much as possible until they went to bed each night. After the first and third weeks, the participants received an email letting them know how many weeks had elapsed in the study and how many weeks remained. After the second week, the researchers called the participants to ask some open-ended feedback questions and ensure that there were no problems with the app. If any problems arose during the intervention, the participants had access to a *Help* section within the app that included a phone number and email address to contact the researchers. This information was also included in the paper intervention materials that were mailed to them and attached to all email communications.

At the completion of the 1-month study, the participants in both groups were again administered the shortened version of the BTACT and asked some open-ended feedback questions, after which they were asked to complete the poststudy questionnaires on the web through Qualtrics. After completing the questionnaires, the participants were sent a US $25 Amazon gift card through email. After the posttest, the participants in the control condition were given the opportunity to download the full version of the app, and all participants were encouraged to retain and use the app for their personal use.

Data Analysis

Data analyses were conducted using RStudio (version 1.2.1335; RStudio, PBC) [36]. First, the difference in app engagement between the conditions was examined. We compared use of the *Walk Now* feature between the conditions using independent samples *t* tests. The use of the schedule and social functions was tallied for the intervention condition.

We tested the remainder of our hypotheses with multilevel mixed effects modeling with the lme4 package [37], controlling for age, sex, education, and health. Using the following model,

we tested whether weekly average steps increased from the baseline week to the 4 intervention weeks. Sensitivity analyses tested whether this effect differed if the baseline week was excluded. Interactions were specified to determine whether the change in weekly step averages differed between the conditions.

$$\text{Level 1: Step Average}_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Week}) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Age}_j) + \gamma_{02}\,(\text{Sex}_j) + \gamma_{03}\,(\text{Condition}_j) + \gamma_{04}\,(\text{Education}_j) + \gamma_{05}\,(\text{Health}_j) + u_{0}\,j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\,(\text{Condition}_j)$$

Next, we used the following model to examine changes in the other outcome measures between pre- and posttest, including social engagement, exercise control and self-efficacy, memory, and sleep. Interactions were examined to determine whether the change in outcomes differed between the conditions.

$$\text{Level 1: Outcome Measure}_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Time}) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Age}_j) + \gamma_{02}\,(\text{Sex}_j) + \gamma_{03}\,(\text{Condition}_j) + \gamma_{04}\,(\text{Education}_j) + \gamma_{05}\,(\text{Health}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\,(\text{Condition}_j)$$

Finally, within-person relationships among daily steps, mood, and energy levels were tested. The following models tested whether daily steps were associated with same-day mood and energy. Lagged analyses were used to determine whether steps predicted next-day mood and energy, controlling for previous day mood and energy. To parse out between-person and within-person effects, the models included both weekly average steps and daily deviation from average steps as predictors. Exploratory analyses examined whether sex or age moderated these effects. In instances when significant interactions with sex were found, separate models were run with men and women to probe the interaction. When significant age interactions were found, separate models were run by using a median split of age in our sample (62 years).

$$\text{Level 1: Daily Mood or Energy}_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Daily Steps}) + r_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Age}_j) + \gamma_{02}\,(\text{Sex}_j) + \gamma_{03}\,(\text{Condition}_j) + \gamma_{04}\,(\text{Education}_j) + \gamma_{05}\,(\text{Health}_j) + \gamma_{06}\,(\text{Average Steps}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Results

Participants

These analyses included adults aged ≥50 years (mean 61.87, SD 7.82 years). Of the 87 participants, 61 (70%) were women, 75 (86%) were White (86%), 2 (2%) reported being Asian, 9 (10%) reported being Black or African American, and 1 (1%) did not wish to report race. The participants were well-educated, with an average of 16.45 (SD 2.56) years of education. Health, on average, was 69.25 (SD 17.40; as reported on the 36-item Short-Form Health Survey general health subscale, with a 0-100 range). Of the 87 participants, 27 (31%) reported working full time, 17 (20%) reported working part time, 34 (39%) were retired, 6 (7%) reported that they were self-employed, and 3 (3%) reported being a homemaker.

Correlations Among Primary Outcome Variables

Zero-order correlations were computed among all outcome variables and covariates at pre- and posttest (Multimedia Appendix 1, Table S1). The average number of steps taken throughout the intervention was positively correlated with exercise self-efficacy at posttest (r=0.33; P=.01). Average steps were also significantly correlated with sleep duration (r=–0.29; P=.03) and sleep latency (r=0.27; P=.04). At pretest, average mood was significantly correlated with age (r=0.26; P=.02), health (r=0.29; P=.01), social engagement (r=0.30; P=.007), exercise control (r=0.32; P=.004), exercise self-efficacy (r=0.24; P=.03), PSQI global score (r=–0.41; P<.001), sleep latency (r=–0.36; P=.001), and average energy (r=0.67; P<.001). At posttest, average mood was significantly correlated with social engagement (r=0.34; P=.008), PSQI global score (r=–0.27; P=.04), and average energy (r=0.70; P<.001). At pretest, average energy was significantly correlated with health (r=0.40; P<.001), social engagement (r=0.30; P=.008), exercise control (r=0.36; P=.001), exercise self-efficacy (r=0.23; P=.04), PSQI global score (r=–0.42; P<.001), sleep latency (r=–0.26; P=.02), and average mood (r=0.67; P<.001). At posttest, average energy was significantly correlated with health (r=0.40; P=.002), social engagement (r=0.27; P=.04), exercise control (r=0.39; P=.003), exercise self-efficacy (r=0.42; P=.01), PSQI global score (r=–0.33; P=.01), and average mood (r=0.70; P<.001).

App Use

The participants in both groups were able to use the *Walk Now* feature; however, only the treatment group participants were able to see their walking routes on a map and name and save them. In the control condition, of the 39 participants, 28 (72%) used the *Walk Now* feature, with an average of 2363 (SD 1616) steps per walk. In the treatment condition, of the 42 participants, 24 (57%) used the *Walk Now* feature, with an average of 1939 (SD 791) steps per walk. An independent samples t test showed that the group difference in average steps per walk was not significant (t_{50}=1.17; P=.25). On the basis of the use of the *Walk Now* feature, those in the control group took an average of 9 (SD 11) walks, whereas those in the intervention group took an average of 11 (SD 24) walks over the course of the 1-month study. An independent samples t test showed that there was no significant group difference in the average number of walks taken (t_{79}=–0.56; P=.58).

Of the participants in the treatment group, only 5 used the schedule feature at least once; 1 participant used the schedule feature 3 times, whereas the other 4 used it once. Only 4 of the treatment group participants used the social feature to text friends through the app; each of these participants used it once during the 1-month intervention.

In terms of correlations between the covariates and app use, those who were older used the *Walk Now* feature more often (r=0.26; P=.02). Age was not significantly correlated with use of the schedule or social features. Neither sex nor education was significantly correlated with use of the *Walk Now* feature, schedule, or social features.

Weekly Steps

Table 1 shows the average daily steps by condition and week.

After controlling for age, sex, education, health, and condition, we found that there was a significant main effect of time (including the baseline week) in predicting weekly average steps (γ=0.24; *P*<.001; Table 2; Figure 3). Average daily steps were significantly higher during the 4-week study than during the baseline week. Analyses were rerun with the baseline week excluded to determine whether average daily steps increased over the course of the 4-week study. There was no significant change in average steps over the 4 intervention weeks (γ= –0.12; *P*=.1; Table 2). Weekly step averages did not differ between the control and intervention groups, nor were there any significant time×condition interactions.

Table 1. Descriptive statistics of baseline measures by condition (N=87)[a].

Characteristics	Control condition		Treatment condition		Combined	
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, N	Values, mean (SD)
Age (years)	42 (48)	61.64 (7.67)	45 (52)	61.51 (8.05)	87	61.57 (7.82)
Sex	42 (48)	71[b]	45 (52)	69[b]	87	70[b]
Education (years)	41 (48)	16.71 (2.21)	45 (52)	16.22 (2.85)	86	16.45 (2.56)
Health	42 (48)	66.31 (16.04)	45 (52)	72.00 (18.32)	87	69.25 (17.40)
Baseline steps	34 (49)	4043.68 (2872.70)	36 (51)	3411.67 (1631.32)	70	3718.64 (2323.34)
Week 1 steps	39 (49)	5530.37 (2286.37)	41 (51)	5046.48 (2426.88)	80	5282.37 (2357.09)
Week 2 steps	38 (48)	5036.25 (2231.26)	41 (52)	4958.95 (2913.55)	79	4996.13 (2606.02)
Week 3 steps	36 (51)	5667.82 (2408.19)	34 (49)	4897.45 (2813.94)	70	5293.64 (2622.80)
Week 4 steps	33 (41)	5175.53 (2501.43)	42 (52)	4570.65 (2425.45)	81	4800.66 (2328.22)
Average steps[c]	39 (48)	5082.83 (2242.57)	42 (52)	4570.65 (2425.45)	81	4813.94 (2339.82)
Average mood[c]	39 (48)	6.85 (1.99)	42 (52)	6.56 (2.10)	81	6.71 (2.06)
Average energy[c]	39 (48)	5.73 (2.07)	42 (52)	5.80 (2.24)	81	5.76 (2.16)

[a]There were no significant differences in age, sex, education, baseline steps, or health between the conditions at baseline.

[b]Percentage values.

[c]Averages across the 4-week intervention.

Table 2. Unstandardized coefficients from multilevel week×condition interaction on daily steps (N=80)[a].

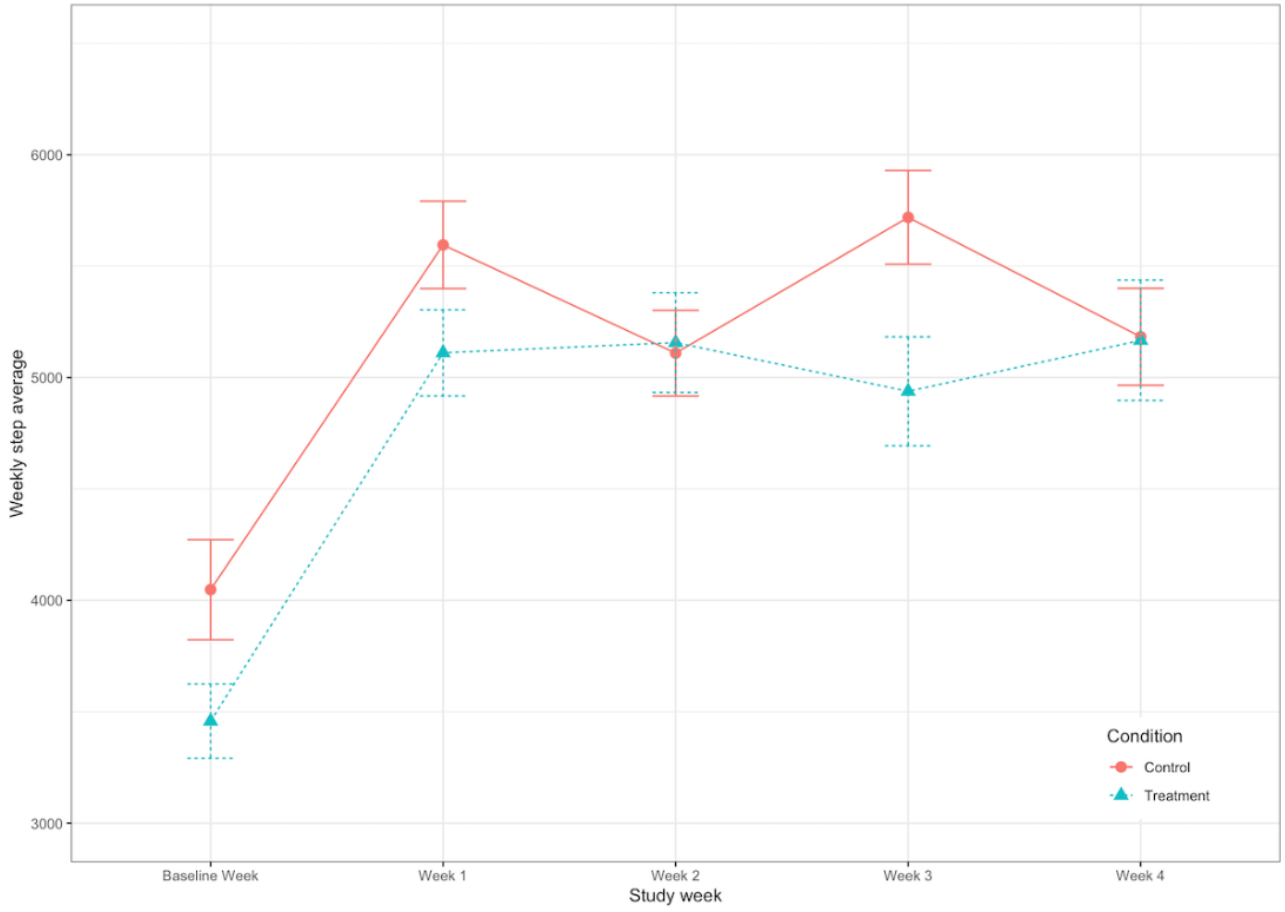
Outcome	Model 1[b]: steps with baseline			Model 2[c]: steps without baseline		
	β	SE	*P* value	β	SE	*P* value
Intercept	5.13	2.61	.05	5.47	2.78	.05
Week	.24	0.05	<.001	–.12	0.07	.1
Age (years)	–.02	0.03	.64	–.01	0.04	.77
Sex	–1.13	0.52	.03	–1.16	0.55	.04
Condition	–.51	0.55	.35	–.57	0.63	.37
Education	–.03	0.10	.72	–.03	0.10	.75
Health	.03	0.01	.07	.03	0.01	.04
Days of app use	.03	0.05	.52	.04	0.05	.4
Week×condition interaction	.02	0.07	.76	.02	0.10	.82

[a]Daily steps were rescaled by dividing the number of steps by 1000; therefore, the β value reflects change per 1000-step increase.

[b]Model 1: level 1 variance=4.44 (SD 2.11); level 2 variance=6.05 (SD 2.46). Akaike information criterion=11319.5; Bayesian information criterion=11383.1; log likelihood=–5648.8.

[c]Model 2: level 1 variance=4.98 (SD 2.23); level 2 variance=5.67 (SD 2.38). Akaike information criterion=9038.6; Bayesian information criterion=9099.8; log likelihood=–4508.3.

Figure 3. Weekly step averages by condition. The error bars refer to SE of the mean. There was a significant positive main effect of week; however, time×condition interactions were not significant.



Group Differences in Changes for Other Outcome Variables

Table 3 details the pre- and post-intervention outcomes by condition. After controlling for age, sex, education, and health, we found that there were no changes in the PSQI global score, sleep duration, or sleep latency between the pre- and posttests. There was a significant main effect of time in predicting cognitive performance (γ=0.17; *P*=.02; Multimedia Appendix 1, Table S2). Cognitive performance increased between the pre- and posttests. There were no significant main effects for time, condition, or significant time×condition interactions for other outcomes, including sleep, social engagement, exercise control, or exercise self-efficacy (Multimedia Appendix 1, Tables S2 and S3).

Table 3. Descriptive statistics for pre- and posttest variables by condition[a].

	Pretest			Posttest		
	Control (n=42), mean (SD)	Treatment (n=45), mean (SD)	Combined (N=87), mean (SD)	Control (n=29), mean (SD)	Treatment (n=29), mean (SD)	Combined (N=58), mean (SD)
PSQI[b] global score	5.71 (3.78)	5.27 (4.14)	5.48 (3.96)	5.45 (3.69)	5.59 (4.44)	5.52 (4.05)
Sleep duration	6.69 (1.13)	6.87 (1.13)	6.78 (1.13)	6.59 (1.03)	7.03 (1.21)	6.81 (1.13)
Sleep latency	18.08 (13.45)	24.47 (26.80)	21.39 (21.54)	19.35 (12.09)	26.37 (27.72)	22.86 (21.49)
Cognitive performance	0.05 (0.61)	0.02 (0.58)	0.03 (0.59)	0.18 (0.29)	0.18 (0.35)	0.18 (0.32)
Social engagement	36.14 (7.65)	33.07 (10.86)	34.55 (9.52)	35.10 (8.62)	30.55 (12.15)	32.83 (10.69)
Exercise control	4.32 (0.51)	4.30 (0.57)	4.31 (0.54)	4.22 (0.63)	3.98 (0.68)	4.10 (0.66)
Exercise self-efficacy	2.57 (0.69)	2.82 (0.81)	2.70 (0.76)	2.40 (0.85)	2.54 (0.97)	2.47 (0.90)

[a]There were no significant differences between the conditions for any of these variables: Pittsburgh Sleep Quality Index, cognitive performance (Brief Test of Adult Cognition by Telephone cognitive composite), and social engagement (Lubben Social Network Scale).

[b]PSQI: Pittsburgh Sleep Quality Index.

Daily Affect

Overview

The relationships among steps, mood, and energy were examined next. As there were no condition differences in weekly steps and no significant time×condition interactions for any of the pre- and post-intervention outcomes, time×condition interactions were not estimated for subsequent analyses. Condition was, however, included as a covariate.

Mood

After accounting for covariates, on the days when the participants took more steps than they did on average, they reported better mood (γ=0.06; *P*<.001; Table 4). There was a significant interaction between daily steps and sex (γ=0.08; *P*=.01; Table 4). The relationship between daily steps and mood was significant for women (γ=0.09; *P*<.001) but not for men (γ=0.009; *P*=.63; Figure 4). There was also a significant interaction between daily steps and age (γ=0.005; *P*=.02; Table 4). The relationship between daily steps and energy was stronger for adults aged ≥62 years (γ=0.08; *P*<.001) than it was for those aged ≤62 years (γ=0.04; *P*=.05; Figure 5).

Table 4. Unstandardized coefficients from multilevel effects of daily steps on same-day mood (N=79)[a].

Outcome	Model 1[b]: mood			Model 2[c]: mood			Model 3[d]: mood		
	β	SE	*P* value	β	SE	*P* value	β	SE	*P* value
Intercept	3.18	1.90	.01	3.22	1.90	.09	3.22	1.90	.09
Daily steps	.06	0.02	<.001	–.07	0.05	.19	–.23	0.12	.06
Age (years)	.04	0.02	.13	.04	0.02	.13	.04	0.02	.14
Sex	–.47	0.40	.25	–.50	0.40	.22	–.47	0.40	.25
Condition	–.50	0.37	.18	–.49	0.37	.19	–.49	0.37	.18
Education	.05	0.07	.49	.05	0.07	.49	.05	0.07	.49
Health	.03	0.01	.02	.03	0.01	.02	.03	0.01	.02
Average steps	–.09	0.08	.29	–.09	0.08	.28	–.09	0.08	.29
Daily steps×sex	__[e]	—	—	.08	0.03	.01	__[e]	—	—
Daily steps×age	__[e]	—	—	__[e]	—	—	.005	0.002	.02

[a]Daily steps were rescaled by dividing the number of steps by 1000; therefore, the β value reflects change in mood per 1000-step increase.

[b]Model 1: level 1 variance=2.38 (SD 1.54); level 2 variance=1.60 (SD 1.27). Akaike information criterion=4981.2; Bayesian information criterion=5033.8; log likelihood=–2480.6.

[c]Model 2: level 1 variance=2.38 (SD 1.54); level 2 variance=1.60 (SD 1.26). Akaike information criterion=4976.8; Bayesian information criterion=5034.7; log likelihood=–2477.4.

[d]Model 3: level 1 variance=2.39 (SD 1.55); level 2 variance=1.60 (SD 1.26). Akaike information criterion=4977.7; Bayesian information criterion=5035.5; log likelihood=–2477.8.

[e]Outcome was not included in the model.

Figure 4. Within-person relationships between daily steps and mood by sex. A score of 0 denotes within-person average daily steps. Shaded areas represent 95% CIs.

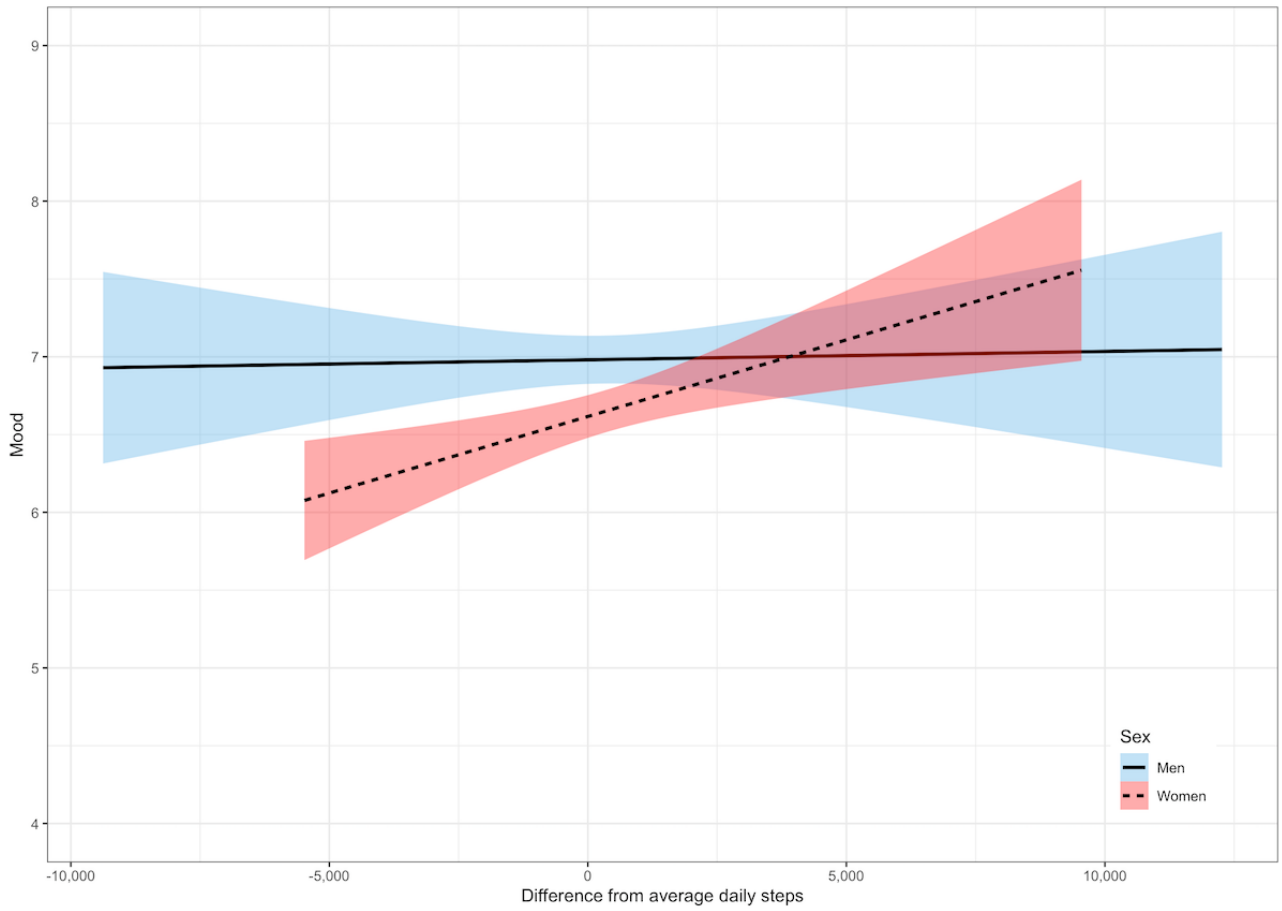
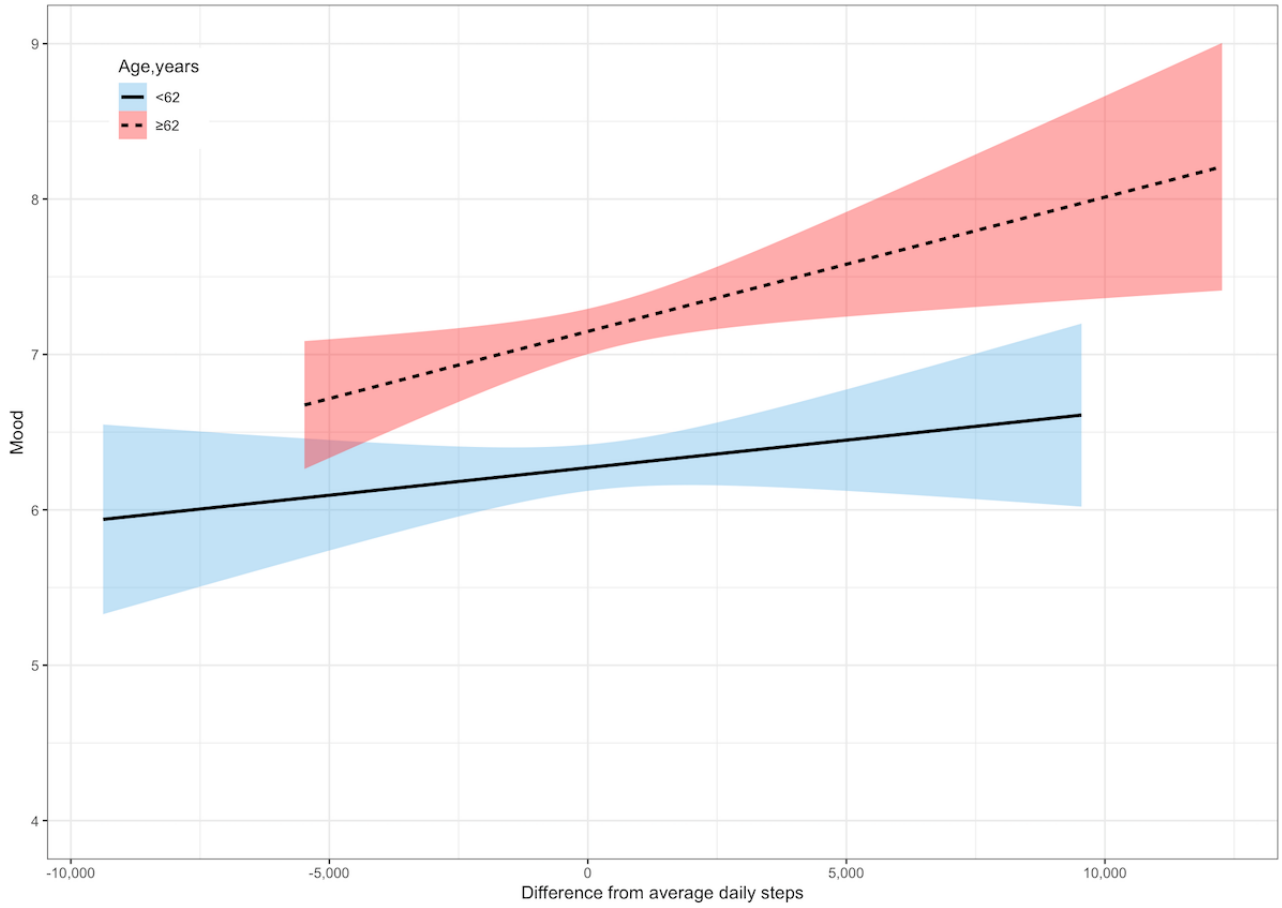


Figure 5. Within-person relationships between daily steps and mood by age group. A score of 0 denotes within-person average daily steps. Shaded areas represent 95% CIs.



Energy

After adjusting for covariates, on the days the participants took more steps than average, they reported having more energy (γ=0.11; *P*<.001; Table 5). There was a significant interaction between daily steps and sex (γ=0.15; *P*<.001). The relationship between daily steps and energy was significant for women (γ=0.16; *P*<.001) but not for men (γ=0.009; *P*=.68; Figure 6). There was also a significant interaction between daily steps and age (γ=0.005; *P*=.01). The relationship between daily steps and energy was stronger for adults aged ≥62 years (γ=0.16; *P*<.001) than it was for those aged ≤62 years (γ=0.05; *P*=.02; Figure 7).

Table 5. Unstandardized coefficients from multilevel effects of daily steps on same-day energy (N=79)[a].

Outcome	Model 1[b]: energy			Model 2[c]: energy			Model 3[d]: energy		
	β	SE	*P* value	β	SE	*P* value	β	SE	*P* value
Intercept	2.93	1.71	.1	3.02	1.71	.08	2.99	1.71	.08
Daily steps	.11	0.02	<.001	–.14	0.06	.02	–.25	0.14	.07
Age (years)	–.002	0.02	.94	–.0009	0.02	.97	–.003	0.02	.9
Sex	–.62	0.36	.09	–.67	0.36	.07	–.62	0.36	.1
Condition	–.13	0.33	.7	–.11	0.33	.73	–.12	0.33	.71
Education	.09	0.06	.15	.09	0.06	.16	.09	0.06	.15
Health	.04	0.01	<.001	.04	0.01	<.001	.04	0.01	<.001
Average steps	–.01	0.07	.88	–.01	0.07	.84	–.01	0.07	.89
Daily steps×sex	—[e]	—	—	.15	0.04	<.001	—[e]	—	—
Daily steps×age	—[e]	—	—	—[e]	—	—	.006	0.002	.01

[a]Daily steps were rescaled by dividing the number of steps by 1000; therefore, the β value reflects change in energy per 1000-step increase.

[b]Model 1: level 1 variance=1.88 (SD 1.37); level 2 variance=2.06 (SD 1.44). Akaike information criterion=5303.3; Bayesian information criterion=5355.8; log likelihood=–2641.6.

[c]Model 2: level 1 variance=1.87 (SD 1.37); level 2 variance=2.04 (SD 1.43). Akaike information criterion=5286.9; Bayesian information criterion=5344.7; log likelihood=–2638.4.

[d]Model 3: level 1 variance=1.88 (SD 1.37); level 2 variance=2.05 (SD 1.43). Akaike information criterion=5298.7; Bayesian information criterion=5356.6; log likelihood=–2638.4.

[e]Outcome was not included in the model.

Figure 6. Within-person relationships between daily steps and energy by sex. A score of 0 denotes within-person average daily steps. Shaded areas represent 95% CIs.

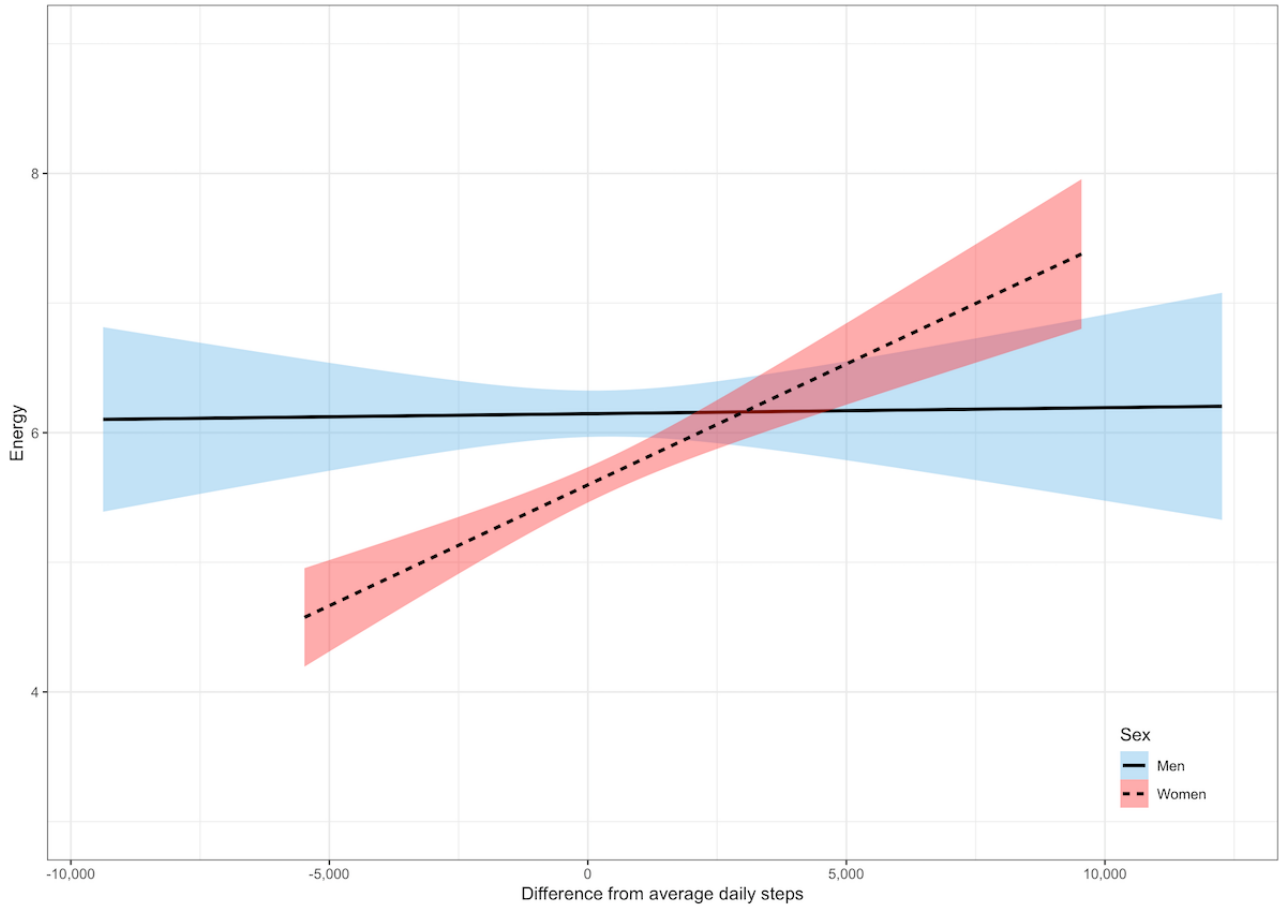
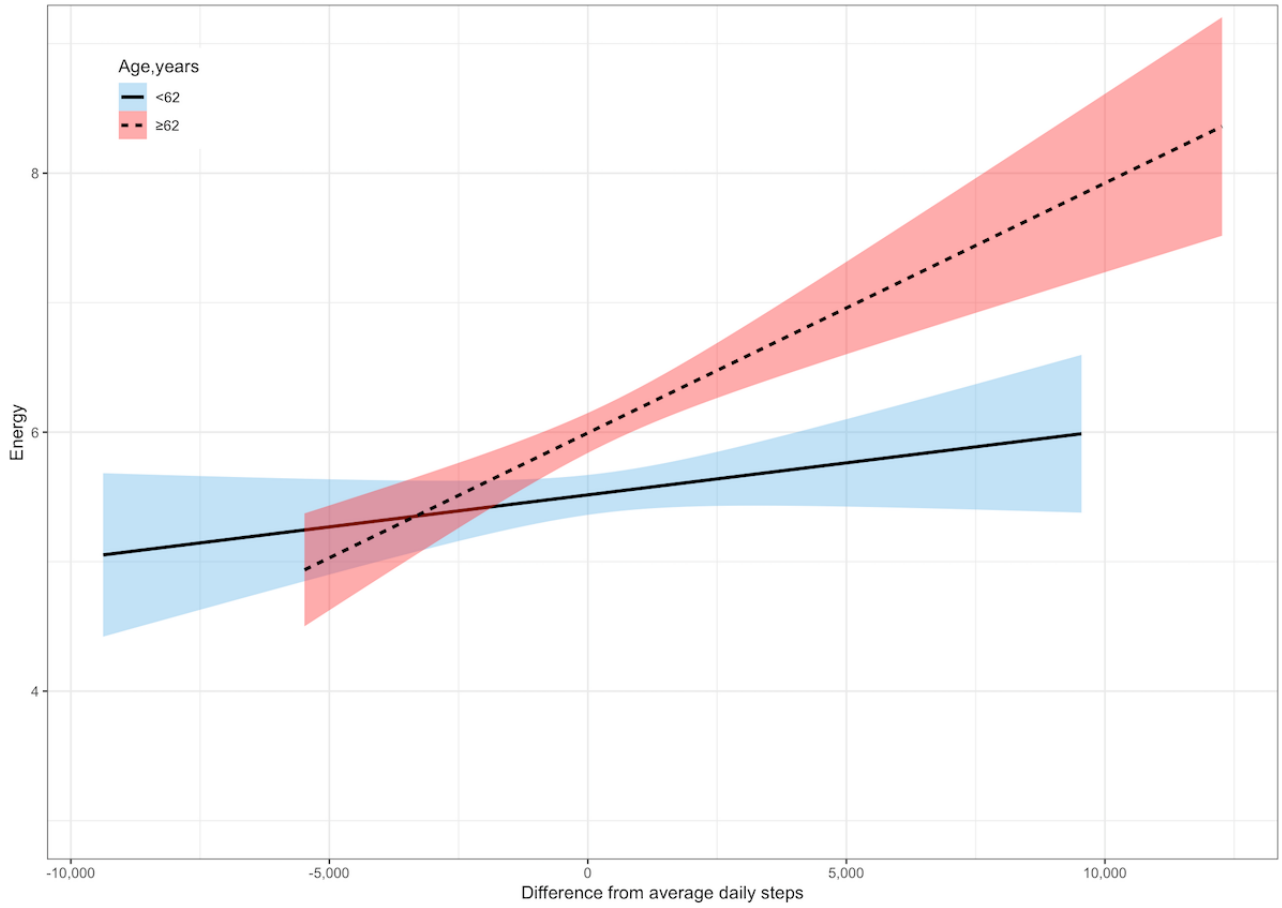


Figure 7. Within-person relationships between daily steps and energy by age. A score of 0 denotes within-person average daily steps. Shaded areas represent 95% CIs.



Lagged Analyses

The effects of steps on next-day mood and energy were also tested. In these analyses, prior-day mood or energy was controlled to determine whether prior-day steps predicted the participants' mood and energy on the following day above and beyond how they felt on the previous day. After controlling for age, sex, condition, education, health, prior-day mood, and average monthly steps, we found that prior-day steps significantly predicted next-day mood (γ=0.04; *P*=.01; Table 6). There was also a significant interaction between prior-day steps and sex (γ=0.08; *P*=.02). This relationship was significant for women (γ=0.06; *P*=.002) but not for men (γ=–0.007; *P*=.70). There was no significant interaction between prior-day steps and age in predicting next-day mood, unlike our findings in same-day relationships.

Table 6. Unstandardized coefficients from multilevel effects of daily steps on lagged (next-day) mood (N=75)[a].

Outcome	Model 1[b]: next-day mood			Model 2[c]: next-day mood			Model 3[d]: next-day mood		
	β	SE	*P* value	β	SE	*P* value	β	SE	*P* value
Intercept	1.89	1.42	.19	1.99	1.43	.17	1.90	1.42	.19
Prior-day steps	.04	0.02	.01	–.08	0.06	.14	–.005	0.01	.97
Prior-day mood	.30	0.03	<.001	.30	0.03	<.001	.30	0.03	<.001
Age (years)	.04	0.02	.05	.04	0.02	.05	.04	0.02	.06
Sex	–.42	0.30	.17	–.44	0.30	.15	–.42	0.30	.17
Condition	–.44	0.27	.11	–.44	0.28	.12	–.44	0.28	.11
Education	.03	0.05	.58	.03	0.05	.61	.03	0.05	.58
Health	.02	0.01	.04	.02	0.01	.04	.02	0.01	.04
Average steps	–.05	0.06	.38	–.06	0.06	.37	–.05	0.06	.38
Prior-day steps×sex	__[e]	—	—	.08	0.03	.02	__[e]	—	—
Prior-day steps×age	__[e]	—	—	__[e]	—	—	.0007	0.002	.71

[a]Daily steps were rescaled by dividing the number of steps by 1000; therefore, the β value reflects change in mood per 1000-step increase.

[b]Model 1: level 1 variance=1.20 (SD 1.10); level 2 variance=1.48 (SD 1.22). Akaike information criterion=4042.8; Bayesian information criterion=4098.7; log likelihood=–2010.4.

[c]Model 2: level 1 variance=1.22 (SD 1.10); level 2 variance=1.47 (SD 1.21). Akaike information criterion=4039.5; Bayesian information criterion=4100.4; log likelihood=–2007.7.

[d]Model 3: level 1 variance=1.20 (SD 1.10); level 2 variance=1.48 (SD 1.22). Akaike information criterion=4044.7; Bayesian information criterion=4105.6; log likelihood=–2010.3.

[e]Outcome was not included in the model.

Prior-day steps did not predict next-day energy, nor was there a significant prior-day steps×sex or prior-day steps×age interactions in predicting next-day energy (Table 7). The alternative directional relationships were also tested, but prior-day mood and energy did not significantly predict next-day steps.

Table 7. Unstandardized coefficients from multilevel effects of daily steps on lagged (next-day) energy (N=75)[a].

Outcome	Model 1[b]: next-day energy			Model 2[c]: next-day energy			Model 3[d]: next-day energy		
	β	SE	*P* value	β	SE	*P* value	β	SE	*P* value
Intercept	2.01	1.40	.16	1.99	1.40	.16	2.03	1.40	.15
Prior-day steps	.01	0.02	.75	.03	0.07	.71	–.11	0.15	.48
Prior-day energy	–.24	0.03	<.001	.24	0.03	<.001	.24	0.03	<.001
Age (years)	–.001	0.02	.93	–.002	0.02	.93	–.002	0.02	.92
Sex	–.47	0.30	.12	–.46	0.30	.12	–.47	0.30	.12
Condition	–.16	0.27	.55	–.17	0.27	.54	–.16	0.27	.55
Education	.07	0.05	.15	.07	0.05	.15	.07	0.05	.15
Health	.03	0.01	<.001	.03	0.01	<.001	.03	0.01	<.001
Average steps	.01	0.06	.81	.01	0.06	.80	.02	0.06	.80
Prior-day steps×sex	__[e]	—	—	–.01	0.04	.76	__[e]	—	—
Prior-day steps×age	__[e]	—	—	__[e]	—	—	.002	0.002	.45

[a]Daily steps were rescaled by dividing the number of steps by 1000; therefore, the β value reflects change in energy per 1000-step increase.

[b]Model 1: level 1 variance=1.11 (SD 1.06); level 2 variance=2.03 (SD 1.43). Akaike information criterion=4394.8; Bayesian information criterion=4450.7; log likelihood=–2186.4.

[c]Model 2: level 1 variance=1.11 (SD 1.05); level 2 variance=2.03 (SD 1.43). Akaike information criterion=4396.8; Bayesian information criterion=4457.7; log likelihood=–2186.4.

[d]Model 3: level 1 variance=1.12 (SD 1.06); level 2 variance=2.03 (SD 1.43). Akaike information criterion=4396.3; Bayesian information criterion=4457.2; log likelihood=–2186.1.

[e]Outcome was not included in the model.

Discussion

Principal Findings

This study was a pilot randomized controlled trial, which examined the effectiveness of StepMATE, a newly developed iPhone app aimed at increasing daily steps in a sample of middle-aged and older adults. The app included behavioral supports, including goal-setting and feedback, action planning, and social supports to encourage changes in behavior. Average daily steps were significantly higher during the 4-week intervention than during the baseline week, and these increases were maintained over the course of the study. The increase in steps, however, did not differ between the 2 groups.

Contrary to the hypotheses, there were no differences in physical activity outcomes between the control condition participants with the basic pedometer-like version of the StepMATE app and the treatment condition participants who had access to the app's full behavioral strategies. This is likely because the participants in both conditions used the features to a similar extent. Despite having a different version of the app, the treatment condition participants rarely used the additional features available to them. It is possible that the extra supports were not needed or that the participants may have considered the extra features difficult or too time-consuming to use.

It is also possible that self-monitoring and goal-setting are enough to encourage increases in daily walking, as other studies have shown [5,6]. The qualitative feedback from the participants echoes this notion:

> *I love when I look at my steps for the day and see that I get close or exceed my daily step goals! I am a person who needs to exercise more, and this app reminds me to keep it moving!*
>
> *Kept track of steps, spot on. Mood questions made me aware of steps, I am checking steps more often and more aware of reaching my goal.*

These results are consistent with prior findings that goal-setting is among the most successful behavior change techniques for increasing physical activity [5]. Yet, there was only an initial increase in steps, with no further incremental change throughout the intervention period.

Although there was a significant increase in cognition from pre- to posttest, there were no significant changes for any of the other outcomes, including social engagement, exercise control, or exercise self-efficacy. Although this is consistent with other findings that exercise is associated with improvements in cognition [38,39], the increase in cognitive performance between pre- and posttest could be due to retest effects. The same version of the test was administered on both occasions. We also tested whether a change in steps between baseline and the end of the intervention was correlated with cognitive performance; no significant correlations emerged.

Contrary to the hypotheses, there were no significant improvements in sleep. The PSQI global score at pretest indicated that the participants in general were good sleepers, with an average sleep duration of just under 7 hours and average sleep latencies under 20 minutes. It is possible that a ceiling effect could explain the lack of change in sleep over the 4-week

study. To determine whether sleep improved for those with poorer sleep at baseline, post hoc analyses were conducted with a median split of PSQI global scores. There were no changes from pre- to posttest for either good sleepers (PSQI global score of 4 or lower) or those with scores higher than 4. It is also possible that because our pilot study only lasted 4 weeks, it was not long enough to elicit changes in our outcome measures. Future work should assess physical activity and subsequent changes in outcomes over longer time periods, with longer baseline and follow-up periods. This would allow researchers to assess whether changes in physical activity are maintained even after the novelty of a behavior change intervention has worn off.

Although the differences between the conditions were not significant, there was within-person variability over time across the 2 conditions. The within-person hypotheses of daily steps predicting mood and energy were supported and add to prior literature on exercise and affect by suggesting that, similar to findings with more intensive or structured exercise, walking can also elicit mood and energy benefits [18,19]. Although others have shown that the effects of exercise on mood are similar across adulthood [15,40], our findings suggest that there are some differences by age. Furthermore, we provide additional support for prior findings that women seem to experience greater mood and energy benefits of exercise than men [20]. Prior work suggests that women may be more aware of internal states than men; therefore, it is possible that women are more sensitive to changes in mood or energy [41]. In our sample, variations in self-reported mood and energy were higher in women than in men. It is encouraging that these results show that even a low-impact activity such as daily walking can be associated with improvements in self-reported mood and energy, at least for women. It is also promising because those who get more enjoyment out of being active are more likely to continue being active [42]. Walking is an easily accessible form of daily activity, and daily steps are a metric that most American adults can track daily with a smartphone or pedometer.

Limitations

This study includes some limitations that are worthy of consideration. The app was only available to users of iPhones with step-tracking capabilities; therefore, there may have been selection bias in only recruiting users who have a relatively new iPhone. The generalizability of the study is also limited by a relatively small sample consisting of mostly White, well-educated adults. According to the Pew Research Center, White individuals and those with higher education and higher household income are more likely to be smartphone owners [43]. Of smartphone owners, iPhone owners in particular are more likely to be White, with higher education and income [44]. Although we do have data on whether the participants were working full time or part time or retired, future work could address whether those in certain professions are more or less likely to engage in physical activity. This could aid in the development of targeted interventions for groups that are most inactive. Another limitation is that we did not assess whether the participants were using fitness technology or apps before enrolling in our study. We specifically recruited individuals who believed that they needed to increase their physical activity; therefore, it is likely that even if the participants used these devices in the past, they were not successful in changing long-term behaviors.

As this study was conducted on a rolling basis over the course of a year at different locations, it is possible that seasonal or geographical factors may have played a role in the findings. The validity of the baseline week steps is also unclear. It is possible that the participants' steps during the week before the intervention may not be representative of their typical daily walking. During the 4-week intervention, the participants were given a pouch for their phone and were specifically asked to carry their phone with them during the day. They were not explicitly instructed to do so before the intervention began. Of the 87 participants, 11 (13%) did not have step-tracking enabled on their iPhone before the study; therefore, they did not have any baseline data. These participants were still included in all analyses because they had step data during the intervention. Post hoc sensitivity analyses revealed that the results did not change if these participants were excluded.

Future studies should aim to collect baseline data for longer periods to obtain a more accurate estimate of normative physical activity levels before an intervention. The study itself was short; 1 month may not be long enough to observe changes in physical activity. Future work could examine whether there is a threshold of intervention duration that must be met to observe physical activity increases. Follow-up assessments after the interventions are completed would also enable examination of long-term benefits and maintenance of any effects.

Measuring physical activity with a smartphone poses limitations. First, the accuracy of measurement may be a limitation. Although some studies and meta-analyses suggest that smartphones—and iPhones in particular—provide accurate and valid measures, especially in terms of differentiating walking from sedentary behaviors [45,46], others suggest that iPhones may be prone to underestimating steps [47]. There could also be accuracy differences based on the iPhone model. The participants may have forgotten to carry their phones with them at different points through the day. It is possible that the participants could have given their phones to others to increase their step counts. The qualitative feedback from the participants suggests that most of them kept their phones on their person for most of the day. As the participants kept their phones with them throughout the day, many of the steps may not have been taken with the intention of walking for exercise. The goal was to capture a full picture of daily activity in our study because walking is an exercise modality that can easily be incorporated into one's regular routine throughout the day. Thus, we did not differentiate whether the steps were taken for exercise.

The intervention was personalized by allowing the participants to use the app at their convenience and to set and change their walking goals as often as desired. This was designed to mirror what would happen if an individual independently downloaded a new walking app and used it on their own. It is possible that the participants did not use the features of the app because they were not specifically asked to do so. In contrast, in another study that used similar behavioral features [13], the participants were reminded daily to use the calendars and maps and to set goals.

In this study, the participants also may not have set their step goals high enough to challenge themselves or encourage increases in walking. Future studies should continue to examine which behavioral supports are most successful in increasing physical activity in older adult populations and find best practices for incorporating these supports into successful physical activity interventions.

Finally, the participants could have encountered some difficulties in using the app and might have preferred a lower-tech intervention for increasing steps. Future work should compare how different age groups can be motivated to increase their activity, especially by making technology more user-friendly and age appropriate. Technology has the potential to assess multiple outcomes (eg, health data and EMAs) in real time, such as through a smartphone, a device that most adults already carry around with them daily [48].

Conclusions

This study tested whether a new walking app, StepMATE, increased daily walking in a sample of inactive older adults. Weekly step averages were significantly higher during the 4-week study than during the baseline week for both intervention groups, and increases were maintained over the course of the 4-week intervention. However, the treatment condition generally did not use the app's additional behavioral strategies; thus, both conditions used similar app features. The components that were similar in both conditions, including self-monitoring of steps and daily walking goals, may be sufficient to encourage increases in walking without the need for additional supports. We also found a significant increase in cognition over the course of the study. Future studies should explore how to make apps more user-friendly and accessible to older adults. More daily steps were associated with better same-day mood and energy for women—but not for men—and were also associated with better next-day mood for women. Relationships among walking, mood, and energy were more apparent for older participants than for the younger ones. Future work could more closely examine sex and age differences in the relationship among walking, mood, and energy. Such research could uncover which features of apps are the most successful and motivating for both men and women across adulthood and could lead to the development of large-scale technology-based interventions for increasing physical activity.

Acknowledgments

We acknowledge grant support from the National Institute on Aging Grant # P30 AG048785 for a Roybal Center for Translational Research on Aging. We thank Jeff Broderick and Beneufit, Inc for developing the StepMATE App and assisting our team with any issues or questions that arose over the course of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables and figures from supplementary analyses. Descriptions of different versions of the app.
[DOCX File , 1021 KB - mhealth_v9i12e27208_app1.docx]

Multimedia Appendix 2
R codes for pre-post analyses.
[PDF File (Adobe PDF File), 273 KB - mhealth_v9i12e27208_app2.pdf]

Multimedia Appendix 3
R codes for daily analyses.
[PDF File (Adobe PDF File), 309 KB - mhealth_v9i12e27208_app3.pdf]

Multimedia Appendix 4
StepMATE demonstration: control group.
[MP4 File (MP4 Video), 11301 KB - mhealth_v9i12e27208_app4.mp4]

Multimedia Appendix 5
StepMATE demonstration: treatment group.
[MP4 File (MP4 Video), 29009 KB - mhealth_v9i12e27208_app5.mp4]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1411 KB - mhealth_v9i12e27208_app6.pdf]

References

1. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
2. Ward B, Clarke T, Nugent C, Schiller J. Early release of selected estimates based on data from the 2015 National Health Interview Survey. National Center for Health Statistics. 2016. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/earlyrelease201605.pdf [accessed 2021-10-17]
3. Hillman CH, Erickson KI, Kramer AF. Be smart, exercise your heart: exercise effects on brain and cognition. Nat Rev Neurosci 2008 Jan;9(1):58-65. [doi: 10.1038/nrn2298] [Medline: 18094706]
4. Clarke T, Norris T, Schiller J. Early release of selected estimates based on data from the 2016 National Health Interview Survey. National Center for Health Statistics. 2017. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/earlyrelease201705.pdf [accessed 2021-10-17]
5. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
6. McGarrigle L, Todd C. Promotion of physical activity in older people using mHealth and eHealth technologies: Rapid review of reviews. J Med Internet Res 2020 Dec 29;22(12):e22201 [FREE Full text] [doi: 10.2196/22201] [Medline: 33372894]
7. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2017 Jan 11;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
8. D'Addario M, Baretta D, Zanatta F, Greco A, Steca P. Engagement features in physical activity smartphone apps: focus group study with sedentary people. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e20460 [FREE Full text] [doi: 10.2196/20460] [Medline: 33196450]
9. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]
10. Gollwitzer P, Sheeran P. Implementation intentions and goal achievement: a meta-analysis of effects and processes. Adv Exp Soc Psychol 2006;38:69-119 [FREE Full text] [doi: 10.1016/s0065-2601(06)38002-1]
11. Bélanger-Gravel A, Godin G, Amireault S. A meta-analytic review of the effect of implementation intentions on physical activity. Health Psychol Rev 2013 Mar;7(1):23-54. [doi: 10.1080/17437199.2011.560095]
12. Carraro N, Gaudreau P. Spontaneous and experimentally induced action planning and coping planning for physical activity: a meta-analysis. Psychol Sport Exerc 2013 Mar;14(2):228-248. [doi: 10.1016/j.psychsport.2012.10.004]
13. Robinson SA, Bisson AN, Hughes ML, Ebert J, Lachman ME. Time for change: using implementation intentions to promote physical activity in a randomised pilot trial. Psychol Health 2019 Feb;34(2):232-254 [FREE Full text] [doi: 10.1080/08870446.2018.1539487] [Medline: 30596272]
14. Basso JC, Suzuki WA. The effects of acute exercise on mood, cognition, neurophysiology, and neurochemical pathways: a review. Brain Plast 2017 Mar 28;2(2):127-152 [FREE Full text] [doi: 10.3233/BPL-160040] [Medline: 29765853]
15. Hogan CL, Mata J, Carstensen LL. Exercise holds immediate benefits for affect and cognition in younger and older adults. Psychol Aging 2013 Jun;28(2):587-594 [FREE Full text] [doi: 10.1037/a0032634] [Medline: 23795769]
16. Ensari I, Sandroff BM, Motl RW. Effects of single bouts of walking exercise and yoga on acute mood symptoms in people with multiple sclerosis. Int J MS Care 2016;18(1):1-8 [FREE Full text] [doi: 10.7224/1537-2073.2014-104] [Medline: 26917992]
17. Parschau L, Fleig L, Warner LM, Pomp S, Barz M, Knoll N, et al. Positive exercise experience facilitates behavior change via self-efficacy. Health Educ Behav 2014 Aug;41(4):414-422. [doi: 10.1177/1090198114529132] [Medline: 24722218]
18. Lathia N, Sandstrom GM, Mascolo C, Rentfrow PJ. Happier people live more active lives: using smartphones to link happiness and physical activity. PLoS One 2017 Jan 4;12(1):e0160589 [FREE Full text] [doi: 10.1371/journal.pone.0160589] [Medline: 28052069]
19. Liao Y, Shonkoff ET, Dunton GF. The acute relationships between affect, physical feeling states, and physical activity in daily life: a review of current evidence. Front Psychol 2015 Dec 23;6:1975 [FREE Full text] [doi: 10.3389/fpsyg.2015.01975] [Medline: 26779049]
20. McDowell CP, Campbell MJ, Herring MP. Sex-related differences in mood responses to acute aerobic exercise. Med Sci Sports Exerc 2016 Sep;48(9):1798-1802. [doi: 10.1249/MSS.0000000000000969] [Medline: 27128666]
21. Kredlow MA, Capozzoli MC, Hearon BA, Calkins AW, Otto MW. The effects of physical activity on sleep: a meta-analytic review. J Behav Med 2015 Jun;38(3):427-449. [doi: 10.1007/s10865-015-9617-6] [Medline: 25596964]
22. Sullivan Bisson AN, Robinson SA, Lachman ME. Walk to a better night of sleep: testing the relationship between physical activity and sleep. Sleep Health 2019 Oct;5(5):487-494 [FREE Full text] [doi: 10.1016/j.sleh.2019.06.003] [Medline: 31358470]
23. Kishida M, Elavsky S. An intensive longitudinal examination of daily physical activity and sleep in midlife women. Sleep Health 2016 Mar;2(1):42-48. [doi: 10.1016/j.sleh.2015.12.001] [Medline: 29073452]

24. Master L, Nye RT, Lee S, Nahmod NG, Mariani S, Hale L, et al. Bidirectional, daily temporal associations between sleep and physical activity in adolescents. Sci Rep 2019 May 22;9(1):7732 [FREE Full text] [doi: 10.1038/s41598-019-44059-9] [Medline: 31118441]
25. Hartescu I, Morgan K, Stevinson CD. Increased physical activity improves sleep and mood outcomes in inactive people with insomnia: a randomized controlled trial. J Sleep Res 2015 Oct;24(5):526-534 [FREE Full text] [doi: 10.1111/jsr.12297] [Medline: 25903450]
26. Centers for Disease Control and Prevention. 2014. URL: http://www.cdc.gov/physicalactivity/data/ [accessed 2021-10-17]
27. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975 Oct;23(10):433-441. [doi: 10.1111/j.1532-5415.1975.tb00927.x] [Medline: 1159263]
28. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
29. Lubben J. Assessing social networks among elderly populations. Fam Community Health 1988 Nov;11(3):42-52 [FREE Full text] [doi: 10.1097/00003727-198811000-00008]
30. Neupert SD, Lachman ME, Whitbourne SB. Exercise self-efficacy and control beliefs: effects on exercise behavior after an exercise intervention for older adults. J Aging Phys Act 2009 Jan;17(1):1-16 [FREE Full text] [doi: 10.1123/japa.17.1.1] [Medline: 19299835]
31. Bandura A. Self-efficacy: The Exercise of Control. New York: Worth Publishers; 1997.
32. Lachman ME, Agrigoroaei S, Tun PA, Weaver SL. Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. Assessment 2014 Aug;21(4):404-417 [FREE Full text] [doi: 10.1177/1073191113508807] [Medline: 24322011]
33. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
34. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011 Jul 28;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
35. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
36. RSTUDIO CONNECT 2021.09. RStudio. URL: http://www.rstudio.com [accessed 2021-10-17]
37. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
38. Sng E, Frith E, Loprinzi PD. Temporal effects of acute walking exercise on learning and memory function. Am J Health Promot 2018 Sep;32(7):1518-1525. [doi: 10.1177/0890117117749476] [Medline: 29284283]
39. Schmidt-Kassow M, Zink N, Mock J, Thiel C, Vogt L, Abel C, et al. Treadmill walking during vocabulary encoding improves verbal long-term memory. Behav Brain Funct 2014 Jul 12;10:24 [FREE Full text] [doi: 10.1186/1744-9081-10-24] [Medline: 25015595]
40. Kilpatrick MW, Greeley SJ, Collins LH. The impact of continuous and interval cycle exercise on affect and enjoyment. Res Q Exerc Sport 2015;86(3):244-251. [doi: 10.1080/02701367.2015.1015673] [Medline: 25811234]
41. Pennebaker JW. The Psychology of Physical Symptoms. New York: Springer; 1982.
42. Lachman M, Lipsitz L, Lubben J, Castaneda-Sceppa C, Jette A. When adults don't exercise: behavioral strategies to increase physical activity in sedentary middle-aged and older adults. Innov Aging 2018 Jan;2(1):igy007 [FREE Full text] [doi: 10.1093/geroni/igy007] [Medline: 30003146]
43. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-10-17]
44. Demographics of mobile device ownership and adoption in the United States. Renalis. 2019. URL: https://www.renalis.health/research/2019/9/13/demographics-of-mobile-device-ownership-and-adoption-in-the-united-states-2019 [accessed 2021-10-17]
45. 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]
46. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. J Med Internet Res 2012 Oct 05;14(5):e130 [FREE Full text] [doi: 10.2196/jmir.2208] [Medline: 23041431]
47. Amagasa S, Kamada M, Sasai H, Fukushima N, Kikuchi H, Lee I, et al. How well iPhones measure steps in free-living conditions: cross-sectional validation study. JMIR Mhealth Uhealth 2019 Jan 09;7(1):e10418 [FREE Full text] [doi: 10.2196/10418] [Medline: 30626569]
48. Nugent NR, Pendse SR, Schatten HT, Armey MF. Innovations in technology and mechanisms of change in behavioral interventions. Behav Modif 2019 Apr 29:145445519845603 (forthcoming). [doi: 10.1177/0145445519845603] [Medline: 31030527]

Abbreviations

BTACT: Brief Test of Adult Cognition by Telephone
EMA: ecological momentary assessment
PSQI: Pittsburgh Sleep Quality Index

Edited by L Buis; submitted 18.01.21; peer-reviewed by M Elliott, L Bechard; comments to author 08.03.21; revised version received 21.04.21; accepted 24.09.21; published 01.12.21.

Please cite as:
Bisson AN, Sorrentino V, Lachman ME
Walking and Daily Affect Among Sedentary Older Adults Measured Using the StepMATE App: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(12):e27208
URL: https://mhealth.jmir.org/2021/12/e27208
doi:10.2196/27208
PMID:34855609

Original Paper

A Personalized Approach Bias Modification Smartphone App ("SWiPE") to Reduce Alcohol Use: Open-Label Feasibility, Acceptability, and Preliminary Effectiveness Study

Victoria Manning[1,2*], BSc, MSb, PhD; Hugh Piercy[1,2*], BA; Joshua Benjamin Bernard Garfield[1,2*], BSc, PhD; Stuart Gregory Clark[3*], BPsych; Mah Noor Andrabi[4*], BSc; Dan Ian Lubman[1,2*], BSc, MB ChB, PhD

[1]Monash Addiction Research Centre, Eastern Health Clinical School, Monash University, Melbourne, Australia

[2]Turning Point, Eastern Health, Melbourne, Australia

[3]School of Psychological Sciences, Monash University, Melbourne, Australia

[4]School of Medicine, Monash University, Melbourne, Australia

[*]all authors contributed equally

Corresponding Author:
Victoria Manning, BSc, MSb, PhD
Turning Point
Eastern Health
110 Church St
Richmond
Melbourne, 3121
Australia
Phone: 61 0428337961
Email: victoria.manning@monash.edu

Abstract

Background: Approach bias modification (ApBM), a computerized cognitive intervention that trains people to "avoid" alcohol-related images and "approach" nonalcohol images, reduces the likelihood of relapse when administered during residential alcohol treatment. However, most individuals experiencing alcohol problems do not require, do not seek, or have difficulty accessing residential treatment. Smartphone-delivered ApBM could offer an easily accessible intervention to reduce alcohol consumption that can be personalized (eg, allowing selection of personally relevant alcohol and positive nonalcohol training images) and gamified to optimize engagement.

Objective: We examined the feasibility, acceptability, and preliminary effectiveness of "SWiPE," a gamified, personalized alcohol ApBM smartphone app, and explored alcohol consumption and craving outcomes in people drinking at hazardous levels or above (Alcohol Use Disorders Identification Test [AUDIT] score ≥8) who wanted to reduce their alcohol use.

Methods: In this open-label trial, frequency and quantity of alcohol consumption, alcohol dependence severity, and craving were measured prior to participants downloading SWiPE. Participants (n=1309) were instructed to complete at least 2 sessions per week for 4 weeks. Recruitment and completion rates were indicators of feasibility. Functionality, aesthetics, and quality ratings were indicators of acceptability. Participants were prompted to report frequency and quantity of alcohol consumption weekly during training and 1 month after training. They completed measures of craving and dependence after 4 weeks of training.

Results: We recruited 1309 participants (mean age 47.0, SD 10.0 years; 758/1309, 57.9% female; mean AUDIT score 21.8, SD 6.5) over 6 months. Participants completed a median of 5 sessions (IQR 2-9); 31.2% (409/1309) completed ≥8 sessions; and 34.8% (455/1309) completed the posttraining survey. Mean Mobile Application Rating Scale scores indicated good acceptability for functionality and aesthetics and fair acceptability for subjective quality. Among those who completed the posttraining assessment, mean past-week drinking days reduced from 5.1 (SD 2.0) pre-training to 4.2 (SD 2.3) in week 4 (t_{454}=7.87; P<.001), and mean past-week standard drinks reduced from 32.8 (SD 22.1) to 24.7 (SD 20.1; t_{454}=8.58; P<.001). Mean Craving Experience Questionnaire frequency scores reduced from 4.5 (SD 2.0) to 2.8 (SD 1.8; t_{435}=19.39; P<.001). Severity of Dependence scores reduced from 7.7 (SD 3.0) to 6.0 (SD 3.2; t_{435}=12.44; P<.001). For the 19.4% (254/1309) of participants who completed a 1-month follow-up, mean past-week drinking days and standard drinks were 3.9 (SD 2.5) and 23.9 (SD 20.7), respectively, both significantly lower than at baseline (P<.001).

Conclusions: The findings suggest SWiPE is feasible and acceptable and may be effective at reducing alcohol consumption and craving in a predominantly nontreatment-seeking sample of adult Australians drinking at hazardous levels. SWiPE's efficacy, relative to a control condition, now needs establishing in a randomized controlled trial. Smartphone-delivered personalized ApBM could be a highly scalable, widely accessible support tool for reducing alcohol use.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12620000638932; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12620000638932p

International Registered Report Identifier (IRRID): RR2-10.2196/21278

(JMIR Mhealth Uhealth 2021;9(12):e31353) doi:10.2196/31353

KEYWORDS

alcohol; hazardous alcohol use; alcohol use disorder; approach bias modification; cognitive bias modification; smartphone app; ehealth; mobile phone app; mhealth; digital health

Introduction

Alcohol remains the most widely used drug globally [1,2] and is a leading cause of injury, chronic disease, and mortality (contributing to 3 million deaths per year), accounting for 5.1% of the global burden of disease [3]. In 2019, 16.8% of Australians aged over 14 years reported drinking above the recommended national guidelines of 14 standard drinks per week, while 25% drank more than the recommended 4 standard drinks in 1 day at least monthly [4]. Unsurprisingly, alcohol remains a drug of concern for almost 50% of all treatment contacts in Australia's publicly funded addiction treatment services [5], although recent modelling suggests this system likely only meets 27% to 56% of the potential demand for treatment [6]. Barriers to seeking alcohol treatment include limited treatment availability, limited time, poor knowledge of treatment options, fear of stigma, geographical distance, privacy concerns, or a belief that the individual can address their alcohol problems without professional help [7].

With approximately 90% of Australians now owning a smartphone [8], development of app interventions offers the potential to overcome many of these barriers. Although hundreds of apps claim to help people reduce alcohol use, very few have been evaluated. A systematic review of studies completed by the end of 2019 identified only 12 trials of alcohol reduction apps designed for adults [9]. Only 5 of these studies were randomized controlled trials (RCTs), of which only 2 demonstrated efficacy. The clarity of these findings is limited because of the disparate interventions used by different apps (including normative feedback, self-monitoring, psychoeducation, action planning, goal setting, problem-solving skills, and identifying or managing triggers and cravings) that target various mechanisms, although broadly speaking, most of them aimed to strengthen "reflective" cognitive processes used to control behavior.

According to the "incentive-sensitization" model [10], repeated use of addictive drugs sensitizes the neural reward system, strengthening the attention-grabbing and motivational properties of alcohol and its associated cues [11] (such as physical and social contexts, sights, sounds, scents), leading to "attentional bias" [12] toward these cues and cue-induced craving [13]. This also leads to the development of "approach bias" (the automatic, impulsive action tendency to approach alcohol-related cues) [12]. Craving [14,15], approach bias [16], and attention bias [17] have all been found to predict heavy alcohol use or relapse. Since alcohol-related cues are ubiquitous and nearly impossible to avoid, in Australia (like many other countries), the craving and cognitive bias that can be elicited by these cues pose a serious challenge for people seeking to reduce or cease their drinking.

Alcohol approach bias can be reduced, or even reversed, through a form of computerized "brain training" known as approach bias modification (ApBM) [18-21]. In ApBM, individuals are repeatedly presented with alcohol-related images, to which they must make an "avoidance" movement (eg, "pushing away" images using a joystick), and nonalcohol-related images, to which they must make an "approach movement" (eg, "pulling" the image toward oneself using a joystick). This trains individuals to automatically "avoid" alcohol-related cues. Several RCTs have shown that, when delivered as an adjunctive intervention during residential treatment for alcohol use disorders (AUD), 4 to 12 sessions of ApBM (typically lasting 10-15 minutes per session) can reduce likelihood of posttreatment relapse [18-20,22,23].

Although residential treatment settings are appropriate for people with severe AUD [24], there is a much larger population of people with less severe alcohol use problems that adversely impact health and quality of life [25,26] who want to reduce or cease drinking. Smartphone-delivered ApBM may be particularly advantageous for this broader population. Using a smartphone, people could complete ApBM training sessions at times and in places that are most convenient for them (eg, at times or in situations where they are vulnerable to experiencing heightened craving). Generalization of training effects may be enhanced by completing ApBM in naturalistic environments rather than in clinical settings.

Thus far, we are aware of only 2 previous studies examining ApBM smartphone apps. In the United Kingdom, Crane et al [27] tested apps containing various combinations of 5 different modules (including an ApBM module) among people drinking at hazardous levels. Despite initially reporting that combining ApBM with normative feedback reduced participants' weekly alcohol consumption [27], they later reported a lack of evidence for efficacy after re-analyzing outcomes with a larger sample [28]. In the Netherlands, Laurens et al [29] tested an ApBM app with people who were concerned about, or wished to reduce,

their drinking. Over a 3-week training period, weekly alcohol consumption declined by a mean of 7.2 standard drinks, relative to pretraining [29]. Participant feedback was generally positive, though participants noted the monotony and repetitiveness of the ApBM training, suggesting that game-like features could make it more engaging. Participants also noted the lack of personalization (participants were all trained using the same standardized set of beverage images and participants).

In our research on treatment seekers for AUD [19,22,30], we have observed that participants tend to drink a limited range of beverages. Thus, use of a standard image set of beverages for all participants reduces the relevance of the training to individuals (eg, being repeatedly trained to avoid images of beer may have little impact for someone who only drinks wine). Since approach bias is the product of associative learning [31], it is likely to be specific to stimuli resembling the drinks frequently consumed by an individual. Designing ApBM tasks where individuals can use their own "personalized" images is therefore likely to be more engaging and more "potent" at reducing approach bias. Personalization can be easily implemented in smartphones by allowing participants to incorporate their own photos of the beverages they most wish to "avoid."

It is not only the "avoidance" stimuli that could be personalized. In almost all alcohol ApBM research to date [18-20,22,23,27,29,30], participants have been systematically trained to approach nonalcoholic beverages. However, these images are likely to be monotonous and of relatively little personal relevance to patients [29]. Recently, we have begun exploring the use of images representing positive, personal goals or personally preferred healthy sources of pleasure (eg, images symbolizing friends, family, social connection, pets, exercise, financial gain) as "approach" stimuli in ApBM training for substance use disorders [32,33]. This responds to recommendations that these should align with patients' goals for behavioral change or offer alternative strategies to manage cravings [34-37]. In this way, personalized ApBM can simultaneously be used to weaken motivations to drink and reinforce positive goals, which may further increase its overall therapeutic benefit. In a smartphone app, people could use their own photographs of friends, family, or hobbies as approach stimuli, making the positive "approach" stimuli highly tailored to the individual's motives for reducing their alcohol use. Including gamified aspects in the task may also improve engagement even further, enhancing completion rates and thereby further enhancing efficacy.

Drawing on the aforementioned body of research, we recently developed "SWiPE," a novel, world-first, personalized alcohol ApBM app. We aimed to test its feasibility, acceptability, and preliminary effectiveness in an open-label, single-group pilot study in people reporting hazardous alcohol use (ie, a score of 8 or more on the Alcohol Use Disorders Identification Test [AUDIT], a commonly used AUD screening tool [38]) recruited from the general community. In addition, we collected data regarding drinking, alcohol craving, and alcohol dependence severity outcomes following training, to inform the design of a future RCT of this app. As previously stated in the published protocol [33], we hypothesized that:

1. We would recruit at least 500 participants within 6 months of launching the app and that at least 60% of participants would complete 8 sessions of ABM, supporting its feasibility.
2. Mean ratings of SWiPE would be greater than 3 on the "functionality," "aesthetics," and "app subjective quality" subscales of the user version of the Mobile Acceptability Rating Scale (uMARS) [39], demonstrating adequate acceptability.
3. There would be statistically significant decreases in number of standard drinks per week, number of days on which alcohol was used in the past 7 days, alcohol craving, and Severity of Dependence Scale (SDS) [40] scores at the end of the 4-week intervention, relative to pretraining scores, suggesting its potential effectiveness.
4. There would be "dose-response" relationships, whereby the degree of reduction between the pretraining and 4-week assessments in measures of alcohol consumption, craving, and dependence severity will be related to the number of ApBM sessions completed over this period (ie, more sessions will be associated with larger reductions), consistent with positive changes being related to engagement with ApBM training.

We also explored participants' reaction time (RT) and error rate data from their ApBM sessions to inform further refinement of the technical parameters of the app.

Methods

Design

This was a single-group, open-label, feasibility study. Analyses of drinking, craving, and dependence severity used a repeated measures design.

Participants

Participants were recruited from across Australia using advertisements on Facebook that directed them to a screening questionnaire hosted on Qualtrics. In addition, online and radio promotions referred participants to a website that contained a brief lay description of ApBM and a link to the screening questionnaire. Participants were required to be aged ≥18 years, have an AUDIT score of at least 8, own a recently updated (ie, within the past year) Android or Apple iOS smartphone with an Australian phone number, and express a desire to reduce or cease their drinking.

Measures

Demographic Information

Participants entered their age, gender, and postcode of residence in an online survey hosted on Qualtrics.

Alcohol Problem Severity

The AUDIT was included in the baseline survey to measure the severity of alcohol use and related problems during the past year [38]. The SDS was used to measure severity of psychological dependence in the past month [41], with wording slightly modified to enhance its relevance to alcohol, as recommended by Gossop et al [40].

Alcohol Craving

The frequency scale of the Craving Experience Questionnaire (CEQ-F) [42] was used to measure past-week frequency of alcohol cravings. The CEQ-F is a 10-item scale, with each item rated on a scale of 0 to 10. The scale can further be broken down into 3 subscales: "intensity," "imagery," and "intrusiveness." In addition to the CEQ-F, we also utilized a single-item visual analogue scale (VAS) to measure current intensity of alcohol craving immediately before and after each ABM session. Participants were asked "How strongly are you craving alcohol right now?" with a line displayed below the question and a slider that they could place between ends anchored with the words "not at all" on the left end and "extremely" on the right. Participants' placement of the slider was converted to a number ranging from 0 to 100.

Alcohol Consumption

At baseline, participants were asked to estimate the number of days on which they consumed alcohol in the past 28 days. In addition, they were asked to use a calendar chart to enter the number of standard drinks consumed on each of the past 7 days, to allow calculation of the total amount of alcohol consumed, and number of days on which any alcohol was consumed in the past week. To maximize accuracy of self-report, an infographic showing how much wine, beer, or spirits corresponds to 1 standard drink (which, in Australia, is defined as 10 g or 12.7 mL of pure alcohol) was displayed with the calendar chart, and this infographic contained a link to the Australian Government's Department of Health standard drinks guide [43]. This 7-day drinking assessment was repeated at weekly intervals over the course of the intervention to gather complete drinking data for each week of the 4-week intervention period. At 28 days after the end of the 4-week intervention period, participants were again asked to complete the alcohol consumption calendar chart, estimating the number of days on which they consumed alcohol in the past 28 days and the number of standard drinks consumed on each day in the past week.

App Acceptability

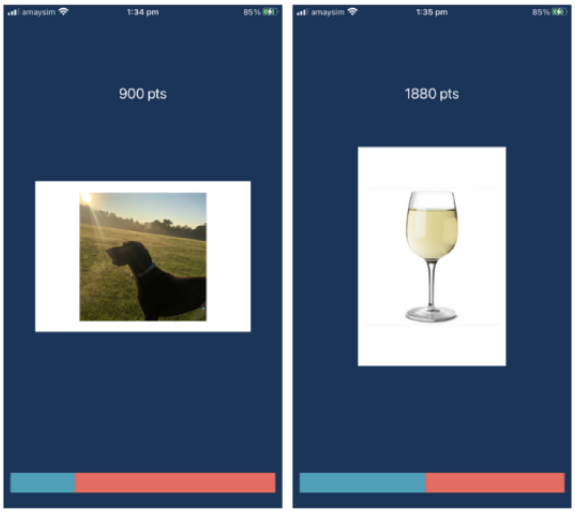
At the end of the 4-week intervention, participants completed the "functionality," "aesthetics," and "app subjective quality" subscales of the uMARS [39]. Individual items of the uMARS range from 1 to 5, with 1 corresponding to very negative, 3 corresponding to neutral/indifferent, and 5 corresponding to very positive assessments, and scores for each subscale are calculated from the mean of individual item scores.

Intervention

Prior to commencing the intervention, participants were prompted to select 6 alcohol-related images that represent the drinks they most frequently consume. Participants could either take photographs using their phone or select images from a library of 72 alcohol-related images chosen to represent a broad range of alcoholic beverages and brands commonly consumed in Australia. Participants were then prompted to "choose 6 images that represent your goals and motivations for reducing drinking. These could be images of family, friends, pets, hobbies, etc. but must not contain alcoholic beverages." Again, participants could either use photographs from their phone or select images from a library of 72 images representing a range of healthy activities, positive goals, and sources of pleasure (including family or friends enjoying time together; financial success; employment; exercise, sports, and recreational activities; healthy foods; pets; travel and holidays), which did not contain any depiction of alcohol. Images included in the alcohol and positive image libraries were selected in consultation with a focus group of people with lived experience of treatment for alcohol use problems (see the protocol [33] for further details of consumer input into the development of the app). It should be noted that if participants used their own photographs, these images were not uploaded to a SWiPE server. To maintain their privacy, images were stored only on the participant's phone, and the SWiPE app only used these files locally while the participant was completing a training session.

After selecting their 12 images, participants were presented with instructions for the ApBM task. Images were displayed with a white "frame" around them, which was in either landscape or portrait orientation. When the frame was in landscape orientation, the participant was required to swipe downward (ie, toward themself), which caused the image to expand as if the participant had "pulled" the image "toward" themself. When the frame was in portrait orientation, the participant was instructed to swipe upward (ie, away from themself), which caused the image to shrink until it disappeared as if they had "pushed" it "away." If the participant swiped in the wrong direction, a red "X" was displayed to inform them that they made an error. Additional technical details regarding image display (including image size, swipe movement criterion, rate of image size change after a swipe response, and interstimulus interval) are reported in the Australian New Zealand Clinical Trials Registry [44]. See Figure 1 for an example of the app's display during the ApBM task.

Figure 1. Example of the approach bias modification (ApBM) training task, with approach (left) and avoid (right) stimuli pictured. The user's score is displayed above each image, and their progress through the ApBM task is displayed in the bar along the bottom of each display.



Following the display of the instructions, participants completed 10 practice trials (including 5 images in portrait frames and 5 images in landscape frames, in random order) to familiarize them with the task before commencing the first session of ApBM. Each session consisted of 156 trials, comprised of 13 presentations of each image. For each alcohol image, 12 of the 13 presentations were framed in portrait orientation, and 1 presentation was framed in landscape orientation. This was reversed for positive images, whereby 12 of the 13 presentations of each positive image were framed in landscape orientation, while 1 presentation was framed in portrait orientation. Thus, participants were supposed to push away 92.3% of alcohol images and pull 92.3% of positive images toward themselves. If participants made the incorrect response, they were informed that it was an error, but the trial was not repeated.

To increase engagement and encourage participants to respond both quickly and accurately, the task was gamified with a scoring system. Each time the participant swiped an image in the correct direction, they were awarded 10 points. Additionally, they scored "bonus points" for correct responses if their response was fast enough. They received 30 bonus points (yielding a total of 40 points for that trial) if they swiped correctly and within 500 ms of image onset, 20 bonus points (ie, 30 points total) if they swiped correctly within 501 to 1000 ms, and 10 bonus points (ie, 20 points total) if they responded correctly within 1001 to 1500 ms. Correct responses that were slower than 1500 ms following image onset earned only 10 points. If they swiped an image incorrectly (ie, swiped down for portrait or swiped up for landscape), they lost 100 points regardless of their RT. Participants' scores were displayed on the screen as they performed the task. Upon completion of the task, the final point score was displayed. On the second and subsequent sessions, each participant's previous session score and the score of their highest-scoring session were displayed prior to commencing the task, to encourage them to beat their previous score.

Procedure

Individuals interested in participating in the study were directed by social media and online advertising to an online survey hosted by Qualtrics. Study information was displayed along with the option to provide consent to participate. Those who agreed to participate proceeded to a survey that screened for eligibility and collected additional information, including alcohol problem severity and craving (ie, demographic questionnaire, the screening question confirming whether they wished to reduce or cease drinking, AUDIT, SDS, and CEQ-F). Those screened as eligible were required to provide their mobile phone number in order to receive a link via SMS to download SWiPE from the Apple or Google Play Store. Upon first opening SWiPE, they were prompted to provide information about their past-month and past-week alcohol use. Participants were then prompted to select their alcohol-related and positive images and then proceeded to the first session of ApBM. Each session of ApBM was immediately preceded and followed by a VAS craving rating. If a participant's postsession VAS score was 90 or above after any session, contact details for a national addiction helpline service were displayed.

Participants were prompted by app notifications to complete a minimum of 2 ApBM sessions each week for 4 weeks. In addition, every 7 days, participants were prompted to report the number of standard drinks consumed on each day of the past week. At the end of the 4-week training protocol, participants were prompted to complete a second online survey that included the CEQ-F, SDS, and uMARS. Participants who completed this posttraining survey were given the option to provide their contact details to be in a draw to win 1 of 10 supermarket gift vouchers valued at Aus $100. At 4 weeks after completing training, participants were prompted to complete a final 1-month follow-up questionnaire that assessed past-month and past-week alcohol consumption. Participants were required to complete the follow-up within 48 hours of the prompt being sent for data to be treated as valid. This study was approved by the Monash University Human Research Ethics Committee (project number: 21393).

Statistical Analysis

The primary outcomes regarding feasibility were the number of sessions completed and the proportion of participants who completed 8 sessions of ApBM within 4 weeks of commencing

using the app. The primary outcome for alcohol use was the number of days of alcohol use in the past 7 days (primary time point 4 weeks after commencing the app). Secondary outcomes included uMARS scores (to assess acceptability); number of participants recruited (to assess feasibility); additional alcohol-related outcomes including number of days of alcohol use in the past 28 days, total standard drinks consumed in the past 7 days, SDS score, CEQ-F (and subscale) scores, and single-item craving VAS ratings; and session metrics including trial error rates, RTs, and session durations.

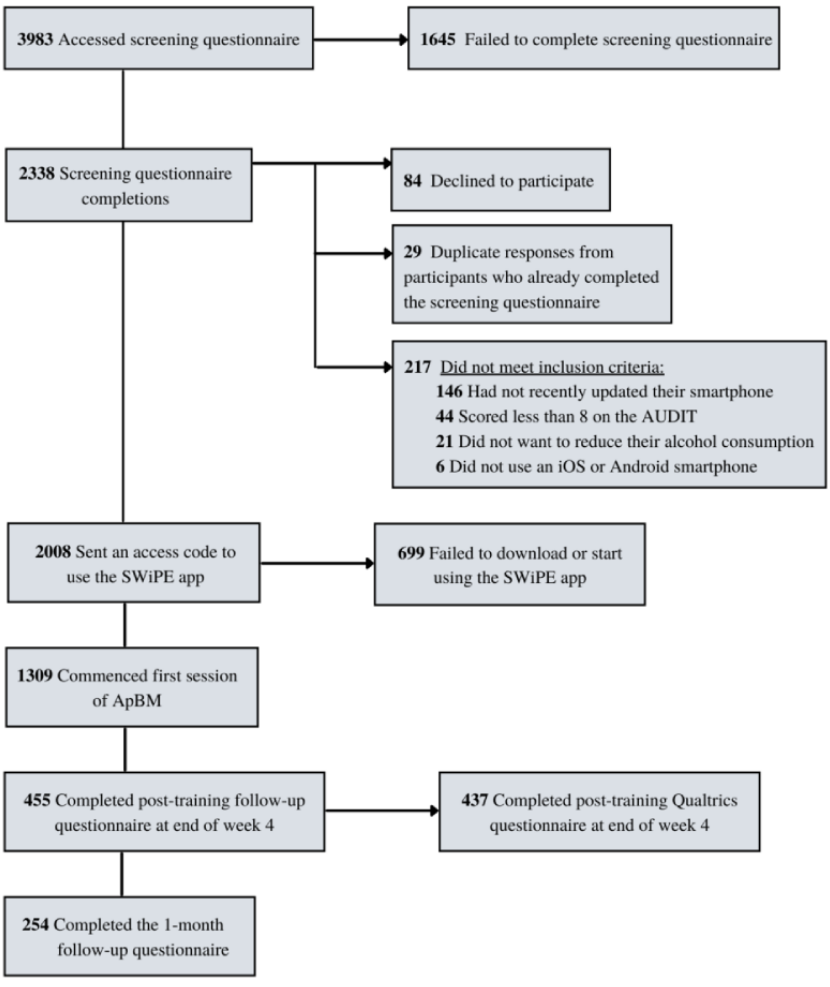
Feasibility and acceptability outcomes, as well as session metrics, were assessed using descriptive data. Changes in alcohol consumption, craving, and SDS scores were analyzed using paired samples *t* tests (in which 2 time points were compared) or repeated measures analyses of variance (RMANOVA) in which 3 or more time points were compared in the same model. To assess possible sources of outcome bias, we conducted analyses comparing characteristics of participants who provided versus those who did not provide outcome data posttraining or at follow-up. These were conducted using independent samples *t* tests for continuous variables (ordinal variables such as days of alcohol use in the past week were treated as continuous since all had at least 8 categories) or Pearson chi-square for categorical variables. To analyze whether number of ApBM sessions moderates the effect of time on past-week drinking and alcohol craving outcomes, we conducted RMANOVA analyses with number of sessions included as a covariate and tested the interaction between number of sessions and time. Descriptive data were analyzed using SPSS version 27 and Microsoft Excel, and inferential analyses were conducted in SPSS version 27. Assuming similar effect sizes for alcohol-related outcomes as that reported by Laurens et al [29] (ie, a 0.36 SD reduction in number of standard drinks per week between pre- and posttraining assessments), we calculated that 119 participants would provide 90% power to detect significant changes using α=.05. As such, we anticipated that the target sample size of 500 would provide ample statistical power to detect main effects of the expected magnitude, even with substantially higher rates of loss to follow-up than anticipated.

Results

Recruitment

Recruitment was open for 6 months (August 29, 2020 to February 28, 2021), and during this time, we recruited 1309 participants who met the eligibility criteria, downloaded SWiPE, and commenced at least 1 session of ApBM (see Figure 2).

Figure 2. Recruitment and participation flow chart. ApBM: approach bias modification; AUDIT: Alcohol Use Disorders Identification Test.



Sample Characteristics

Demographic characteristics of the sample are shown in Table 1. The mean age was slightly older, and the sample had a higher proportion of female participants than typical samples recruited from Australian alcohol treatment settings [19,45]. The distribution of the sample between metropolitan, regional, and remote areas corresponded approximately to the Australian general population (of whom 72% live in major cities, 18% in inner regional areas, 8.2% in outer regional areas, 1.2% in remote areas, and 0.8% in very remote areas [46]).

Table 1. Demographic characteristics of the sample at baseline (n=1309).

Characteristics	Values
Age (years), range	18-75
Age (years), mean (SD)	47.0 (10.0)
Gender, n (%)	
Female	758 (57.9)
Male	538 (41.1)
Other	13 (1.0)
State/territory[a], n (%)	
Australian Capital Territory	44 (3.4)
New South Wales	357 (27.4)
Northern Territory	40 (3.1)
Queensland	274 (21.0)
South Australia	89 (6.8)
Tasmania	49 (3.8)
Victoria	311 (23.8)
Western Australia	141 (10.8)
Remoteness category[b], n (%)	
Major city	864 (66.4)
Inner regional	293 (22.5)
Outer regional	128 (9.8)
Remote	10 (0.8)
Very remote	6 (0.5)
Phone type, n (%)	
Android	498 (38.0)
iPhone	811 (62.0)

[a]Data regarding state/territory were missing for 4 participants, and percentages are therefore calculated with a denominator of 1305.

[b]The Australian Bureau of Statistics classifies areas of Australia as “major cities,” “inner regional,” “outer regional,” “remote,” and “very remote” and publishes information regarding which postcodes are located in which remoteness category [46]. Postcode data were missing for 8 participants, and remoteness category percentages were therefore calculated with a denominator of 1301.

The sample’s alcohol use and related characteristics at baseline are shown in Table 2. Despite being recruited from the general community, with the large majority of participants not being in treatment, several indicators suggested high severity of alcohol use and AUD. Both mean AUDIT and SDS scores were above cut-offs that indicate likely dependence (AUDIT >20 [47]; SDS >3 [48]). Indeed, 59.4% (778/1309) of participants scored at least 20 on the AUDIT, and 98.2% (1284/1309) scored at least 3 on the SDS. Participants’ mean alcohol consumption in the week prior to commencing ApBM was nearly 4 times higher than the 10 standard drink per week limit recommended by the National Health and Medical Research Council for minimizing long-term risk of alcohol-related disease [49].

Table 2. Alcohol use, dependence, treatment, and craving at baseline (n=1309).

Variable	Values
Number of drinking days (past week), range	0-7
Number of drinking days (past week), mean (SD)	5.3 (1.9)
Number of drinking days (past 28 days), range	0-28
Number of drinking days (past 28 days), mean (SD)	20.7 (6.7)
Number of standard drinks (past week), range	0-221
Number of standard drinks (past week), mean (SD)	37.4 (24.2)
AUDIT[a] score, range	8-40
AUDIT score, mean (SD)	21.2 (6.5)
SDS[b] score[c], range	0-15
SDS score, mean (SD)	7.9 (3.0)
CEQ-F[d] score[c], range	0.2-9.9
CEQ-F score, mean (SD)	4.4 (2.0)
Currently accessing treatment for AUD[e], n (%)	
Yes	117 (8.9)
No	1192 (91.1)
Alcohol goal, n (%)	
Reduce drinking	1102 (84.2)
Cease drinking completely	207 (15.8)

[a]AUDIT: Alcohol Use Disorders Identification Test.

[b]SDS: Severity of Dependence Scale.

[c]Due to missing data, SDS and CEQ-F score statistics come from 1307 participants.

[d]CEQ-F: Craving Experience Questionnaire frequency scale.

[e]AUD: alcohol use disorder.

Feasibility

The target sample size of 500 participants was recruited within the first 26 days, 7 times faster than anticipated by our hypothesis of 500 recruits within 6 months. Participants completed between 1 and 27 sessions (median 5, IQR 2-9). Participants completed 98.6% (7632/7744) of sessions that were commenced (ie, only 1.4% of sessions that were commenced were not completed), indicating that participants were able to complete sessions without disruption. However, contrary to our hypothesis that at least 60% would complete the 8 sessions, this was only the case for 31.2% (409/1309) of participants.

Participants' mean number of errors per session was 3.9 (ie, 2.5% of the 156 trials per session), although this was highly skewed (SD 5.5), with the median number of errors per session being 2.25 (1.4% of trials); 95.0% (1244/1309) of participants averaged less than 11.5 errors per session (ie, an average error rate of less than 7.4 of the 156 trials delivered per session).

Analysis of RTs was conducted, excluding participants with average RTs over 3 seconds as these data are likely to be polluted by trials where the participant was distracted from the task for long periods of time (eg, left the phone unattended part way through the session) or repeatedly distracted over many trials. This resulted in exclusion of data for 0.8% (11/1309) of the participants for alcohol trials and 1.4% (18/1309) of the participants for positive trials. The mean of the participants' average RT was 816.3 (SD 173.3) ms for alcohol trials and 849.3 (SD 203.2) ms for positive images. Examining RTs averaged over both alcohol and positive trials for participants with valid data for both categories (n=1282), only 1 participant (0.1%) achieved a mean RT within the highest reward category (RT<500 ms), 1099 (85.7%) averaged an RT in the second-highest reward category (500<RT<1000), 272 (21.2%) averaged an RT within the third reward category (1000<RT<1500), and 10 (0.8%) had an average RT in the range that did not yield reward points (RT>1500).

Acceptability

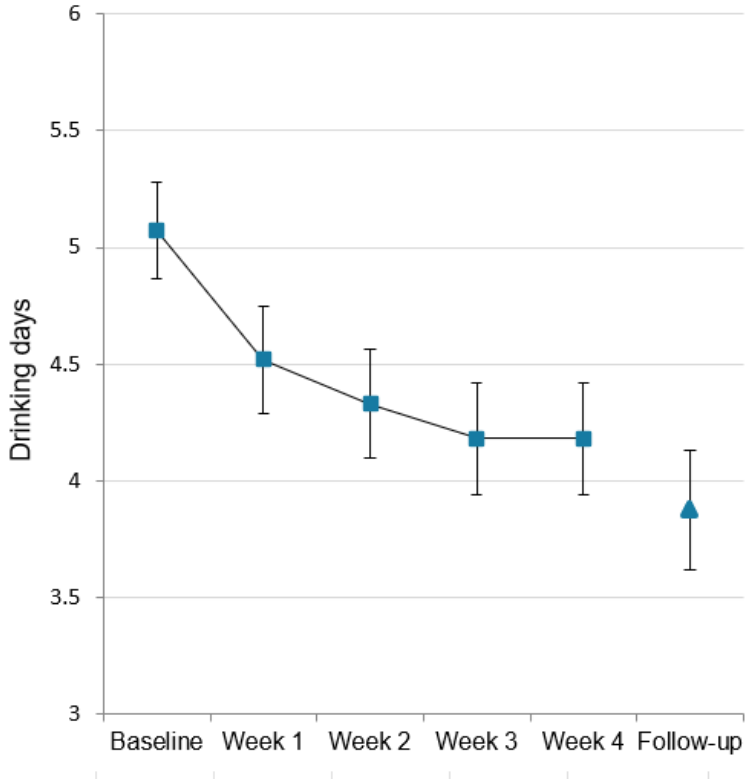
Mean uMARS scores were 4.4 (SD 0.5) for functionality, 4.2 (SD 0.5) for aesthetics, and 3.4 (SD 0.8) for subjective quality. Thus, mean scores were above 3 for all subscales, suggesting generally positive assessments of SWiPE's acceptability among participants who completed posttraining ratings. Indeed, of 429 participants completing uMARS ratings, 417 (97.2%) gave scores greater than 3 on the functionality subscale, 414 (96.5%) gave scores greater than 3 on the aesthetics subscale, and 293 (68.3%) gave scores greater than 3 on the subjective quality subscale.

Preliminary Effectiveness

Past-Week Drinking Days

Mean number of past-week drinking days at baseline and in each of weeks 1 through 4 among those with complete data for all time points (n=359) is shown in Figure 3. Tests of within-subjects contrasts showed a significant linear effect of time ($F_{1,358}$=57.39; P<.001; η^2_p=0.14), indicating that drinking days tended to decrease over time. The quadratic effect of time was also significant ($F_{1,358}$=18.86; P<.001; η^2_p=0.05), consistent with the smaller week-to-week reductions in drinking days with increasing time apparent in Figure 3. Bonferroni-adjusted pairwise comparisons between weeks showed that drinking days were significantly lower in all weeks of the intervention than they were at baseline (all P<.001). Additionally, mean drinking days in week 3 (P=.002) and week 4 (P=.01) were lower than in week 1 of training. As week 4 was the primary outcome time point, a supplementary paired samples t test was conducted comparing baseline (mean 5.1, SD 2.0 days) to week 4 (mean 4.2, SD 2.3 days) in all participants who provided data at both of these time points (n=455). This 18% reduction in weekly drinking days confirmed a robust reduction in the frequency of use (t_{454}=7.87; P<.001; Cohen d=0.37). In addition, 9.7% (44/455) of participants reported no alcohol days in the final week of training. Paired t tests conducted with participants in the "likely alcohol-dependent" range based on AUDIT score (>20) and who were not receiving treatment indicated that drinking days reduced significantly between baseline (n=207; mean 5.3, SD 1.9) and week 4 (mean 4.4, SD 2.2; t_{206}=5.82; P<.001; Cohen d=0.40) and between baseline (n=112; mean 5.4, SD 1.8) and the 1-month follow-up (mean 3.9, SD 2.7; t_{111}=5.94; P<.001; Cohen d=0.56).

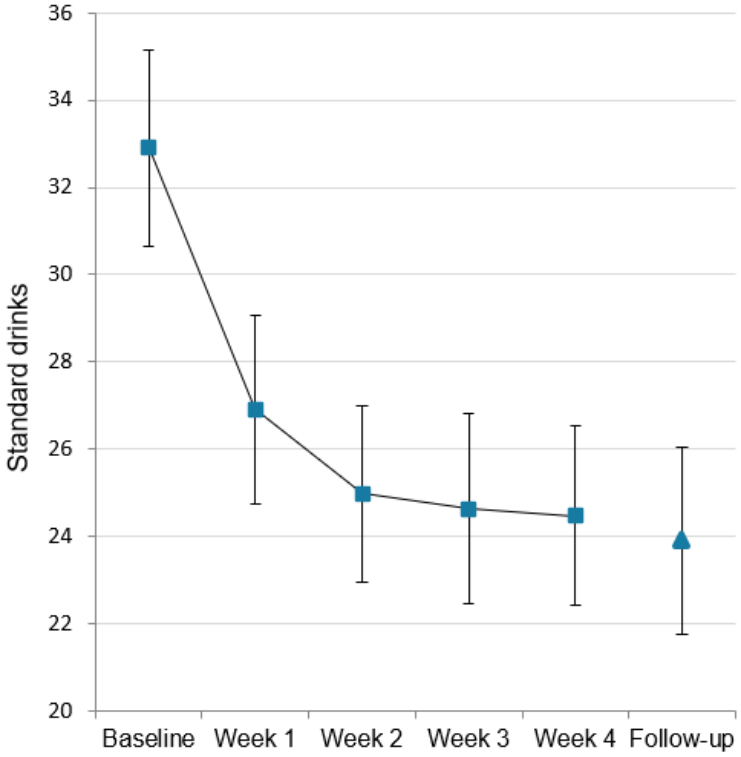
Figure 3. Mean past-week days of alcohol use at baseline and during each week of the intervention for participants with complete data for all 5 time points (n=359) and the 252 who completed the 1-month follow-up. Error bars show 95% CIs of the mean.



Past-Week Standard Drinks

Mean number of past-week standard drinks at baseline and in each week of the intervention among those with complete data is shown in Figure 4. Tests of within-subjects contrasts showed a significant linear effect of time ($F_{1,358}$=64.91; P<.001; η^2_p=0.15), indicating that alcohol consumption tended to decrease over time. The quadratic effect of time was also significant ($F_{1,358}$=30.8; P<.001; η^2_p=0.08), consistent with a deceleration in change over time. Bonferroni-adjusted pairwise comparisons between weeks showed that the number of standard drinks consumed was significantly lower in all weeks of the intervention than at baseline (all P<.001). Additionally, mean standard drinks in week 3 (P=.03) and week 4 (P=.03) were lower than in week 1 of training. A supplementary paired samples t test comparing baseline (mean 32.8, SD 22.1 standard drinks) to week 4 (mean 24.7, SD 20.1 standard drinks) in participants who provided data at both of these time points confirmed a robust decrease in weekly alcohol consumption by an average of 25% (t_{454}=8.58; P<.001; Cohen d=0.40). Among participants in the likely alcohol-dependent range (ie, AUDIT score >20) who were not currently receiving treatment, paired t tests indicated that standard drinks reduced significantly between baseline (n=207; mean 41.7, SD 24.2) and week 4 (mean 30.8, SD 22.1; t_{206}=6.79; P<.001; Cohen d=0.47) and between baseline (n=111; mean 39.0, SD 18.8) and the 1-month follow-up (mean 30.2, SD 23.4; t_{110}=3.80; P<.001; Cohen d=0.36).

Figure 4. Mean number of past-week standard drinks at baseline and during each week of the intervention for participants with complete data for all 5 time points (n=359) and the 252 who completed the 1-month follow-up. Error bars show 95% CIs of the mean.



Additional Secondary Alcohol-Related Outcomes

Among participants with complete data for drinking days over the 4-week intervention period (n=359), mean past-month drinking days declined from 20.4 (SD 6.6) to 17.2 (SD 8.1; t_{358}=8.84; *P*<.001; Cohen *d*=0.47). Participants who completed the SDS at both time points showed a reduction in mean scores from 7.7 (SD 3.0) to 6.0 (SD 3.2; t_{435}=12.44; *P*<.001; Cohen *d*=0.60; see Figure 5). Mean CEQ-F total scores declined significantly from 4.5 (SD 2.0) to 2.8 (SD 1.8; t_{435}=19.4; *P*<.001; Cohen *d*=0.93; see Figure 6). Reductions were also significant for all CEQ-F subscales (intensity: t_{435}=23.2; *P*<.001; Cohen *d*=1.11; imagery: t_{435}=15.3; *P*<.001; Cohen *d*=0.73; intrusiveness: t_{435}=11.1; *P*<.001; Cohen *d*=0.53).

Figure 5. Mean Severity of Dependence Scale (SDS) scores at baseline and posttraining in participants with complete data at both time points (n=436). Error bars show 95% CIs of the mean.

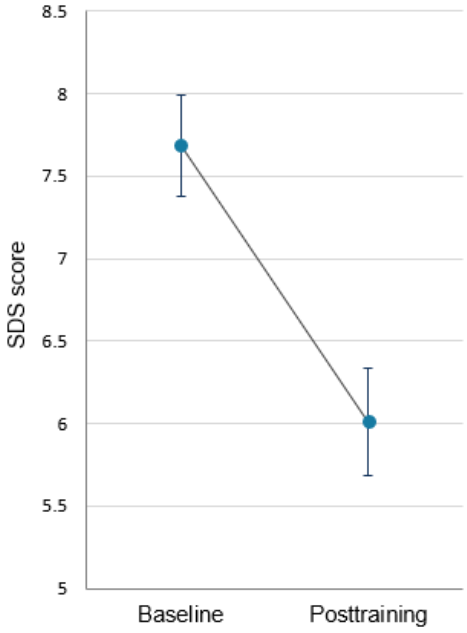
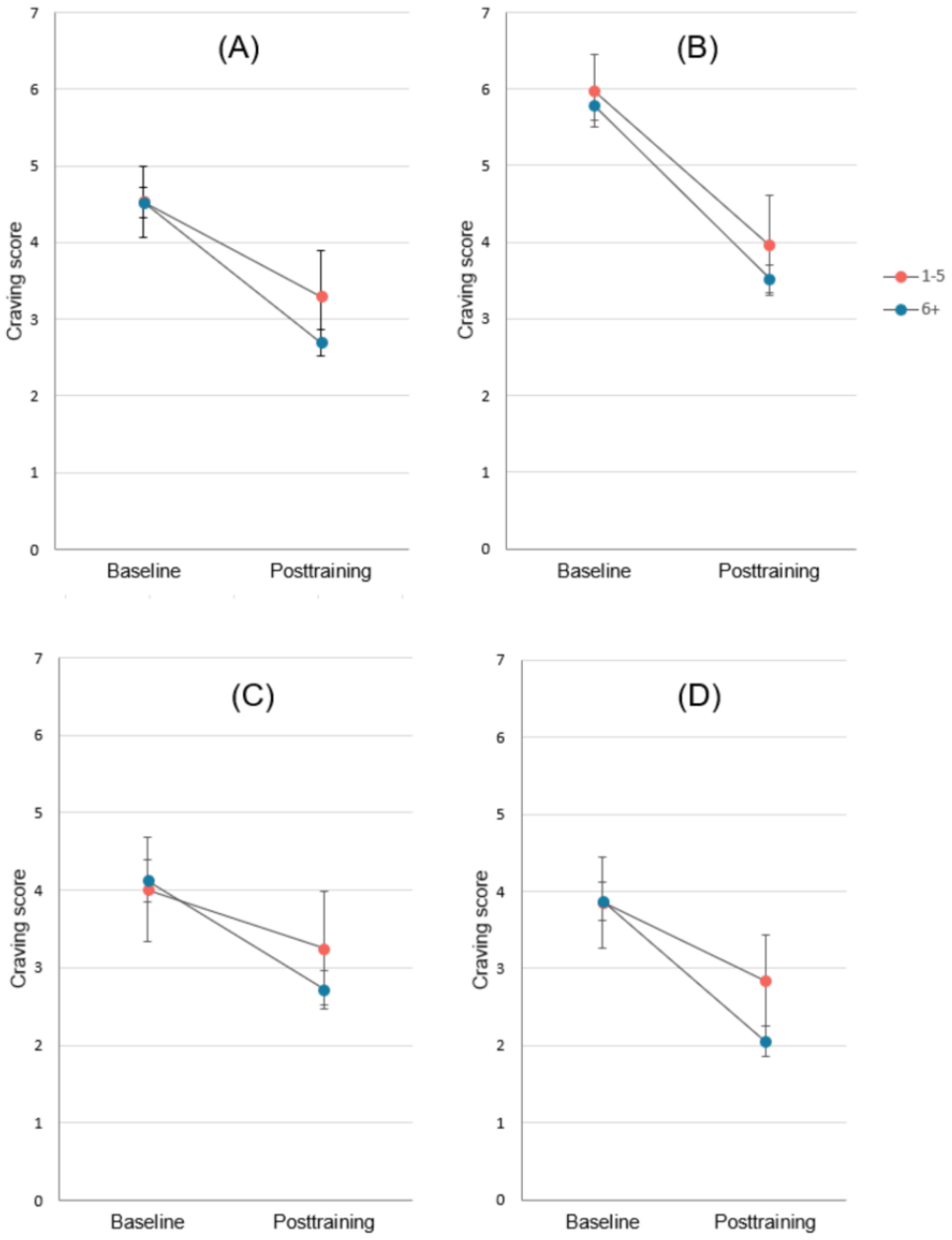


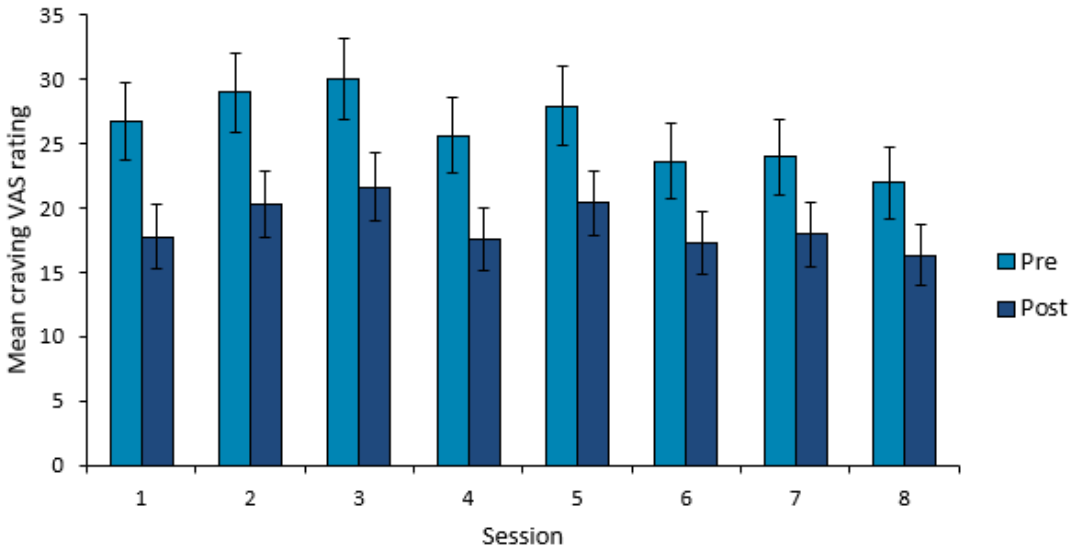
Figure 6. Mean Craving Experience Questionnaire – Frequency Scale (CEQ-F) scores at baseline and posttraining for participants with complete data for both time points who completed 1 to 5 sessions (n=50) and those who completed 6 or more sessions (n=386): (A) total, (B) intensity subscale, (C) intrusiveness subscale, (D) imagery subscale. Error bars show 95% CIs of the mean.



Further evidence for declines in craving come from the single-item craving VAS that was administered before and after each session. Analysis of these ratings across the first 8 sessions among 380 participants with complete data, with session (1, 2, 3, 4, 5, 6, 7, and 8) and timing (presession vs postsession) as separate within-subjects factors, showed main effects for both factors, as well as a significant interaction. Figure 7 suggests that the significant linear effect of session ($F_{1,379}$=10.41; P=.001; η^2_p=0.03) results from a slight tendency for craving ratings to decline in later sessions, relative to earlier ones. There was also a significant quadratic effect of session ($F_{1,379}$=5.54; P=.02; η^2_p=0.01), perhaps reflecting the tendency for craving ratings to increase over the first few sessions, before then declining.

As is highly evident in Figure 7, the strong effect of timing ($F_{1,379}$=295.93; P<.001; η^2_p=0.44) reflects mean craving ratings being lower posttraining relative to pretraining across all sessions. This effect of timing significantly interacted with the linear effect of session ($F_{1,379}$=27.07; P<.001; η^2_p=0.07). Separate analyses of presession and postsession ratings suggested that this interaction was due to the linear effect of session being larger for presession ratings ($F_{1,418}$=21.16; P<.001; η^2_p=0.05) than for postsession ratings ($F_{1,379}$=4.23; P=.04; η^2_p=0.01), likely due to a combination of presession craving declining over time and floor effects for the even lower postsession ratings.

Figure 7. Mean craving visual analogue scale (VAS) ratings before and after each of the first 8 sessions of training among participants with complete data for all sessions (n=380). Error bars show 95% CIs of the mean.



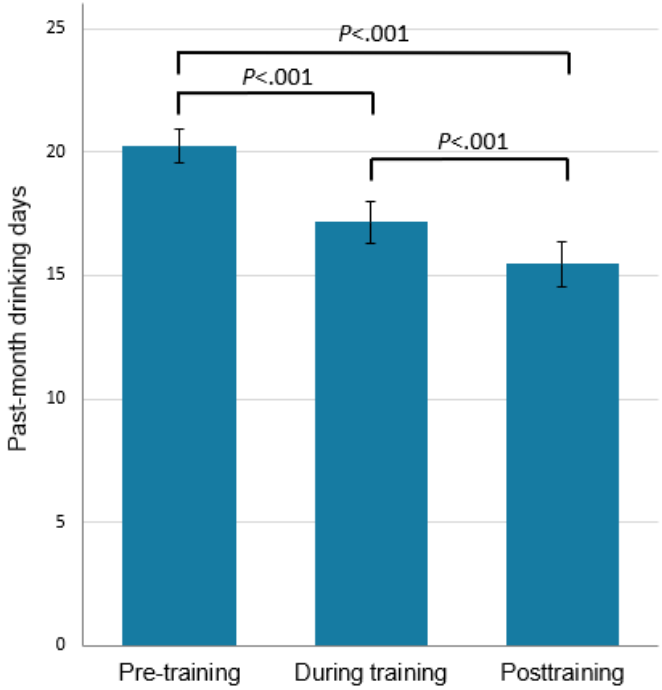
Moderation of Posttraining Outcomes by Number of Completed Sessions

To test whether changes between baseline and week 4/posttraining past-week alcohol use and cravings were associated with the number of sessions completed, we conducted additional RMANOVA analyses in which the interaction term between time (baseline vs week 4/posttraining) and sessions completed was included in the model. Tests of the interaction term indicated that number of sessions completed did not significantly moderate the effect of time on number of past-week drinking days ($F_{1,453}$=1.33; P=.25; η^2_p=0.003) or past-week standard drinks ($F_{1,453}$=1.23; P=.27; η^2_p=0.003). However, it significantly moderated CEQ-F total ($F_{1,434}$=8.97; P=.003; η^2_p=0.02). To better understand this interaction, we classified participants based on whether they completed 1 to 5 or ≥6 sessions (since 5 sessions was the median number completed in the whole sample and 6 sessions is a typical intervention in residential alcohol treatment settings) and tested a RMANOVA with this binary categorization of sessions completed as a between-groups factor. This showed a significant interaction between completion of 6 sessions and time ($F_{1,434}$=4.32; P=.04; η^2_p=0.01), which is depicted in Figure 6. Analyses of CEQ-F subscales suggested this interaction was present for the imagery ($F_{1,434}$=6.02; P=.01; η^2_p=0.01) and intrusiveness: ($F_{1,434}$=9.91; P=.002; η^2_p=0.02) subscales but was not significant for the intensity subscale ($F_{1,434}$=2.70; P=.10; η^2_p=0.006).

Alcohol Use at Follow-up

The 1-month follow-up survey was completed by 254 participants at 28 to 30 days after the end of the 4-week intervention period (252 of whom provided data regarding past-week and past-month drinking days and 251 of whom provided data regarding past-week standard drinks). Mean drinking days at follow-up were 3.9 (SD 2.5) in the past week and 15.9 (SD 8.8) across the previous 4 weeks. Mean standard drinks consumed in the past week were 23.9 (SD 20.7). Of the 254 participants, 26 participants (10.4%) reported no alcohol use in the past week, and 19 (7.5%) reported no alcohol use in the past month. All alcohol consumption outcomes represented significant reductions relative to baseline (past-week drinking days: t_{251}=8.07; P<.001; Cohen d=0.51; past-month drinking days: t_{251}=8.45; P<.001; Cohen d=0.53; past-week standard drinks: t_{250}=5.86; P<.001; Cohen d=0.37). Interestingly, for participants with complete week 4 and follow-up data, past-week (t_{227}=3.38; P<.001; Cohen d=0.22) and past-month drinking days (t_{198}=4.69; P<.001; Cohen d=0.33) were also lower at follow-up than in week 4, suggesting continued improvements in the month following the end of the intervention, although number of past-week standard drinks was not reduced between these time points (t_{227}=–0.14; P=.89; Cohen d=–0.009). The mean numbers of past-month drinking days before, during, and after the intervention among those with complete data for each time point (n=199) are shown in Figure 8. Tests of within-subjects contrasts showed a significant and strong linear effect of time ($F_{1,198}$=66.12; P<.001; η^2_p=0.25), indicating that drinking days decreased during and after training. The quadratic effect of time was also significant ($F_{1,198}$=4.73; P=.031; η^2_p=0.023). Bonferroni-adjusted pairwise comparisons between the 3 time points showed that drinking days were significantly lower during and after training, compared with baseline, and that drinking days also significantly decreased during training compared with after training (all P<.001; see Figure 8). The reductions between baseline and follow-up in past-week drinking days ($F_{1,250}$=0.59; P=.44; η^2_p=0.002), past-month drinking days ($F_{1,250}$=0.01; P=.91; η^2_p<0.001), or past-week standard drinks ($F_{1,249}$=0.43; P=.51; η^2_p=0.002) were not significantly moderated by number of sessions completed.

Figure 8. Mean past-month days of alcohol use in the 28 days before, during, and after the intervention for participants with complete data for all 3 time points (n=199). Error bars show 95% CIs of the mean, and horizontal bars display significant Bonferroni-adjusted pairwise comparisons.



Differences Between Posttraining Assessment Completers and Those Lost to Follow-up

Participants (455/1309, 34.8%) completed week 4 assessments of past-week alcohol use within the app, and 33.4% (437/1309) responded to the invitation to complete posttraining questionnaires (SDS, CEQ-F, and uMARS) in the online survey. Participants who completed week 4 alcohol use assessments within the app had a significantly higher mean age than those who did not (49.0 years vs 45.9 years; t_{1307}=–5.42; *P*<.001) and had completed a substantially higher mean number of ApBM sessions (9.7 vs 3.8; t_{1307}=–33.96; *P*<.001). These participants generally showed less severe alcohol use in terms of mean AUDIT scores (20.4 vs 21.6; t_{1307}=3.37; *P*<.001), SDS scores (7.6 vs 8.0; t_{1306}=2.40; *P*=.02), past-week drinking days (5.1 vs 5.4; t_{1307}=3.71; *P*<.001), and past-week standard drinks (32.8 vs 39.8; t_{1307}=4.98; *P*<.001). In addition, those who provided 4-week outcome data within the app were more likely to use an Android phone (197/455, 43.3% vs 301/854, 35.2%; χ^2_1=8.18; *P*=.004). Participants who completed week 4 alcohol use assessments within the app did not significantly differ from noncompleters in terms of gender, remoteness, past-month drinking days, CEQ-F scores, proportion who wanted to reduce versus cease drinking, or proportion currently attending treatment for AUD (data not shown). Comparisons of those who completed the online questionnaires to those who did not generally revealed the same pattern in terms of which differences were significant, with the exception that SDS score and phone type did not significantly differ in this comparison (data not shown).

Comparisons of those who completed the 1-month follow-up revealed a similar pattern, in that they had an older mean age (47.6 years vs 46.2 years; t_{1307}=–2.08; *P*=.04), had completed a substantially higher mean number of ApBM sessions (10.3 vs. 4.8; t_{1307}=–22.26; *P*<.001), and tended to drink fewer standard drinks in the week before baseline (32.2 vs 38.6; t_{1307}=3.81; *P*<.001). They were also more likely than those lost to follow-up to use an Android phone (118/254, 46.5% vs 380/1055, 36.0%; χ^2_1=9.46; *P*=.002). However, they did not differ significantly from those lost to follow-up in terms of gender; remoteness category; whether they wanted to reduce or cease drinking; whether or not they were in treatment at baseline; baseline AUDIT, SDS, or CEQ-F total score; or past-week or past-month drinking days (data not shown).

Discussion

Principal Findings

This study is the first to examine the feasibility, acceptability, and effectiveness of a personalized, gamified ApBM smartphone app. We found support for the preliminary effectiveness of SWiPE, where participants significantly reduced their standard drinks, drinking days, cravings, and dependence severity at the end of the 4-week training period. Reductions in drinking days and standard drinks not only were maintained during the month after training but also slightly improved even further over time. The results also supported SWiPE's acceptability, with large majorities of participants providing "positive" scores on the uMARS. Feasibility of SWiPE's potential for implementation, either in further trials or practice, was strongly supported by the fact that we reached our target sample size within 1 month of commencing recruitment. However, only 31.2% of the sample completed the recommended minimum of 8 sessions. We discuss these findings in the context of the broader literature in the following sections.

Feasibility

Hypothesis 1 was partially accepted in that we exceeded our recruitment target within 26 days, supporting the feasibility of smartphone-delivered ApBM. Laurens et al [29] reported similar findings, recruiting their sample within 13 days, while Crane et al [27] achieved their recruitment target in less than 2 months. These findings highlight the widespread demand for alcohol reduction apps and the eagerness for people drinking at hazardous levels to try novel approaches such as ApBM. We also found strong evidence for the feasibility of the task itself, with participants completing 98.6% of the sessions they commenced, with a median error rate of 1.4%, which suggests that the task was not overly burdensome, fatiguing, or difficult. However, hypothesis 1 was partially rejected in that fewer than 60% completed the recommended 8 sessions, suggesting the intended "training dose" was not feasible for a large majority of participants in its current form (ie, without additional features or incentives).

Acceptability

Hypothesis 2, which concerned SWiPE's acceptability, was accepted in that mean scores on the uMARS subscales were above 3, indicating "good" for the Functionality and Aesthetics subscales and "acceptable" for the Subjective Quality subscale. Laurens et al [29] reported that users rated their ApBM smartphone app—"Breindebaas"—as moderately satisfactory on the Client Satisfaction Questionnaire (mean score of 20.9 out of 30); their users criticized the lack of personalization and game elements in the task. Interestingly, free-text user feedback from our posttest completers praised the options for personalization of stimuli in the ApBM task (particularly the option to use photos from one's phone library to represent positive values or goals, such as family, friends, and holiday destinations). Given the overall high performance of participants, with most (86%) having mean RTs within the second-highest reward category (501-1000 ms), making the task more challenging may even increase engagement further, though dynamically adjusting to the individual participant's performance may be optimal, and, in this regard, we recommend further exploration of adaptive difficulty paradigms in the future development of gamified ApBM smartphone applications. Nonetheless, based on the uMARS findings and free-text comments, we would recommend that future studies include options for personalization of stimuli and engaging gamification paradigms in order to increase acceptability of ApBM smartphone apps.

Preliminary Effectiveness

Hypothesis 3 was accepted as there were significant reductions in standard drinks, drinking days, craving, and severity of dependence. The reductions in frequency and quantity of alcohol consumption are consistent with those reported for the Breindebaas app by Laurens et al [29], where participants were encouraged to complete 2 ApBM sessions for 3 weeks leading to an almost identical effect size for reductions in past-week standard drinks. Taken together, these findings suggest that smartphone-delivered ApBM holds promise. However, since controlled trials of delivering ApBM online have found equivalent reductions in active ApBM when compared with sham training and the only prior RCT of an alcohol-reduction app that included ApBM reported null findings [28], it remains necessary to establish the efficacy of smartphone-delivered ApBM in RCTs. This is a particularly worthwhile investment given the greater convenience, flexibility, and accessibility (eg, notification reminders and immediacy of access) that can be offered via smartphones relative to web-based platforms (eg, via a PC or laptop).

Hypothesis 4 was partially accepted. Although there was no clear association between "training dose" (number of sessions completed) and the degree to which participants' alcohol use was reduced, SWiPE was associated with a reduction in craving, both in the short-term (ie, immediately after a session) and over the duration of the training program. Additionally, it was notable that the effect size (Cohen *d*) for change in craving between baseline and week 4 was much larger than for other outcomes. Interestingly, we also observed a significant moderation effect of number of training sessions on reductions in total CEQ-F score as well as the imagery and intrusiveness subscales. The imagery subscale requires participants to imagine alcohol's taste, smell, sensation, and how they would picture it, and the intrusiveness scale requires them to reflect on how difficult it is to avoid thinking about alcohol (eg, "how often was it hard to think about anything else"). This is perhaps unsurprising given that SWiPE requires the user to repeatedly view the drinks that they personally regularly consume and practice avoiding them. However, it is important to acknowledge that greater craving reductions in heavier users of SWiPE could also reflect a greater motivation or commitment to reducing one's alcohol use rather than the "dose" of ApBM itself. Nonetheless, this observed association combined with the significant reduction in VAS craving scores immediately after each session and over time suggests SWiPE may be effective at reducing craving. We also observed significant reductions in participants' severity of alcohol dependence, which is encouraging given the high proportion in the dependent range on the SDS (98%) and AUDIT (60%) on study entry, despite our intention to recruit participants in the hazardous or harmful drinking range.

Limitations

These promising results on the preliminary effectiveness and acceptability of SWiPE must be interpreted in the context of the study design and limitations. Although we exceeded our recruitment target, only 33.4% completed the posttraining assessment, and 19.4% completed the 1-month survey. High attrition rates are common in mobile health (mHealth) intervention research, particularly in the absence of monetary incentivization for follow-up completion. The app included prompts (app notifications) to remind participants to complete assessments, yet the rate of participants providing primary outcome data in our study was similar to the 27% in the study by Crane et al [27] and 38% in the study by Laurens et al [29]. As such, it is important to acknowledge the potential attrition bias and overinflation of positive outcomes. It is possible that those who completed follow-ups were more committed to reduce their alcohol use and therefore both engaged more with SWiPE (which we observed) and achieved greater reductions in alcohol consumption, craving, and dependence severity. For similar reasons, the acceptability ratings of the task may be biased

toward painting a more positive picture (ie, participants who had a positive assessment of the app may have been more likely to remain engaged enough in the study to complete the posttraining acceptability questionnaire). The lower number of training sessions among those who did not complete the posttraining follow-up suggests the outcomes reported may not be entirely representative of the larger population who engaged in SWiPE. Future studies could reduce the risk of bias posed by high attrition rates by offering incentives for completing follow-ups and engaging in more assertive attempts to contact participants whose follow-ups are overdue. The observed reductions in alcohol consumption, severity of dependence, and craving could also be attributed in part to the Hawthorne effect [50], where participants may have reported reductions in alcohol consumption because of their awareness of being observed in the context of a research study [51]. However, we expect that the absence of personally identifiable data and the anonymity afforded by online self-report methods increased the likelihood of accurate reporting [52].

Another limitation is the reliance on self-reported consumption data, which is always likely to have some degree of inaccuracy (eg, due to poor recall). However self-reported alcohol use is the gold standard in mHealth interventions and alcohol intervention research more broadly [53]. In-person biometric measures to confirm self-report were beyond the scope of the current study given its national focus. We modelled the assessments closely on the computerized 7-day timeline follow-back assessment used by Simons et al [54], which showed good concordance with other measures of alcohol use, and our visual display of standard drink equivalents within the app, when reporting consumption, may have improved the accuracy.

Despite the aforementioned limitations, the study findings advance the ApBM literature by being the first ApBM study to personalize the “avoid” images by using those representing participants’ preferred alcoholic beverages and brands and also the first to personalize the “approach” images to reflect personally meaningful or goal-related behaviors (eg, family, hobbies). Although the positive findings could reflect a potential “dual-target” approach (ie, dampening alcohol associations and reinforcing positive ones), future research would benefit from exploring the extent of reduced alcohol approach bias and increased approach bias to positive cues and the degree to which these changes account for reduced alcohol use.

Conclusion

Evidence of SWiPE’s feasibility, high acceptability ratings, and multiple indicators of its preliminary effectiveness in terms of reduced alcohol consumption, frequency and quantity, dependence score, and craving are encouraging and suggest an RCT is now warranted. When using SWiPE, consumption (drinking days and standard drinks) decreased significantly. As such, SWiPE may be a useful public health tool given the large number of people drinking at risky levels and in line with the prevention paradox [51,55] where the majority of alcohol-related harm can be attributed to this population (owing to the sheer number of them). Nonetheless, this should not detract from the finding that SWiPE could also be a useful intervention for those with more severe alcohol problems (given that significant reductions were reported among those in the likely-dependent range who were not in treatment when using SWiPE). Establishing its efficacy is a critical next step, as its low cost, ease of implementation, high accessibility, and scalability mean SWiPE could address a significant gap between the demand for treatment and the availability of addiction treatment services [6]. Although we aimed to recruit a sample of individuals drinking at hazardous levels, results (on both the AUDIT and SDS) indicated that the majority of the sample were alcohol-dependent, further demonstrating the critical need for treatment interventions that are available outside of treatment services (particularly given that only 8.9% were currently accessing treatment). Importantly, SWiPE extends neuroscience-informed interventions beyond the laboratory and treatment service environment, ensuring ApBM is an accessible, easy-to-use tool for the broader community. SWiPE has the potential to deliver a “just-in-time” intervention during periods of heightened vulnerability (ie, events, days, and times associated with drinking), by reducing the impulsive subconscious drivers of drinking and enabling people to instead make more conscious, goal-aligned decisions around their alcohol use.

Acknowledgments

This project is supported by a grant from the Australian Rechabite Foundation (ARF). We would like to express our gratitude to the Association of Participating Service Users members who participated in the focus group that helped to refine the alcohol and positive image sets and to the participants in the second focus group who tested the app and provided feedback on its functioning and content. We feel that their perspective, arising from their lived experience, has improved the quality of this study. We also thank Katherine Mroz for her assistance in conducting the first focus group and Samuel Campbell for technical assistance that made that focus group possible. We acknowledge ANT Development Studios for programming the SWiPE app. We thank Paul Sanfilippo and John Reynolds for assisting with calculations and simulations of statistical power. DIL is supported by a National Health and Medical Research Council (NHMRC) Leadership Fellowship (ID 1196892).

Conflicts of Interest

VM and HP are Directors and shareholders of Cognitive Training Solutions Pty Ltd, which is currently exploring commercialization of the SWiPE app. DIL has provided consultancy advice to Lundbeck and Indivior and has received travel support and speaker honoraria from Astra Zeneca, Camurus, Indivior, Janssen, Lundbeck, Shire, and Servier. These organizations do not stand to

benefit from this project. DIL has been an investigator on an untied education grant from Sequirus, unrelated to the current work. The authors have no other competing interests to declare.

References

1. Bonomo Y, Norman A, Biondo S, Bruno R, Daglish M, Dawe S, et al. The Australian drug harms ranking study. J Psychopharmacol 2019 Jul;33(7):759-768. [doi: 10.1177/0269881119841569] [Medline: 31081439]
2. Nutt DJ, King LA, Phillips LD, Independent Scientific Committee on Drugs. Drug harms in the UK: a multicriteria decision analysis. Lancet 2010 Nov 06;376(9752):1558-1565. [doi: 10.1016/S0140-6736(10)61462-6] [Medline: 21036393]
3. Global status report on alcohol and health 2018. World Health Organization. 2018 Sep 27. URL: https://www.who.int/publications/i/item/9789241565639 [accessed 2021-11-26]
4. National Drug Strategy Household Survey 2019. Australian Institute of Health and Welfare. 2020 Jul 16. URL: https://www.aihw.gov.au/reports/illicit-use-of-drugs/national-drug-strategy-household-survey-2019/contents/summary [accessed 2021-11-26]
5. Alcohol and other drug treatment services in Australia: 2018-2019. Australian Institute of Health and Welfare. 2020 Jun 26. URL: https://www.aihw.gov.au/reports/alcohol-other-drug-treatment-services/alcohol-other-drug-treatment-services-2018-19/contents/summary [accessed 2021-11-26]
6. Ritter A, Chalmers J, Gomez M. Measuring unmet demand for alcohol and other drug treatment: the application of an Australian population-based planning model. J Stud Alcohol Drugs Suppl 2019 Jan;Sup 18:42-50 [FREE Full text] [Medline: 30681948]
7. Wallhed Finn S, Bakshi A, Andréasson S. Alcohol consumption, dependence, and treatment barriers: perceptions among nontreatment seekers with alcohol dependence. Subst Use Misuse 2014 May;49(6):762-769. [doi: 10.3109/10826084.2014.891616] [Medline: 24601784]
8. Mobile Nation 2019: The 5G Future. Deloitte Access Economics. 2019. URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/Economics/deloitte-au-economics-mobile-nation-2019-080419.pdf [accessed 2021-11-26]
9. Colbert S, Thornton L, Richmond R. Smartphone apps for managing alcohol consumption: a literature review. Addict Sci Clin Pract 2020 May 07;15(1):17 [FREE Full text] [doi: 10.1186/s13722-020-00190-x] [Medline: 32381062]
10. Robinson TE, Berridge KC. The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain Res Brain Res Rev 1993;18(3):247-291. [doi: 10.1016/0165-0173(93)90013-p] [Medline: 8401595]
11. Lubman DI, Yücel M, Pantelis C. Addiction, a condition of compulsive behaviour? Neuroimaging and neuropsychological evidence of inhibitory dysregulation. Addiction 2004 Dec;99(12):1491-1502. [doi: 10.1111/j.1360-0443.2004.00808.x] [Medline: 15585037]
12. Wiers RW, Gladwin TE, Hofmann W, Salemink E, Ridderinkhof KR. Cognitive bias modification and cognitive control training in addiction and related psychopathology. Clinical Psychological Science 2013 Jan 09;1(2):192-212. [doi: 10.1177/2167702612466547]
13. Carter BL, Tiffany ST. Meta-analysis of cue-reactivity in addiction research. Addiction 1999 Mar;94(3):327-340. [Medline: 10605857]
14. Schneekloth TD, Biernacka JM, Hall-Flavin DK, Karpyak VM, Frye MA, Loukianova LL, et al. Alcohol craving as a predictor of relapse. Am J Addict 2012 Nov;21 Suppl 1:S20-S26. [doi: 10.1111/j.1521-0391.2012.00297.x] [Medline: 23786506]
15. Sinha R, Fox HC, Hong KA, Hansen J, Tuit K, Kreek MJ. Effects of adrenal sensitivity, stress- and cue-induced craving, and anxiety on subsequent alcohol relapse and treatment outcomes. Arch Gen Psychiatry 2011 Sep;68(9):942-952 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.49] [Medline: 21536969]
16. Martin Braunstein L, Kuerbis A, Ochsner K, Morgenstern J. Implicit alcohol approach and avoidance tendencies predict future drinking in problem drinkers. Alcohol Clin Exp Res 2016 Sep;40(9):1945-1952 [FREE Full text] [doi: 10.1111/acer.13151] [Medline: 27421061]
17. Cox WM, Hogan LM, Kristian MR, Race JH. Alcohol attentional bias as a predictor of alcohol abusers' treatment outcome. Drug Alcohol Depend 2002 Dec 01;68(3):237-243. [doi: 10.1016/s0376-8716(02)00219-3] [Medline: 12393218]
18. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
19. Manning V, Garfield JBB, Staiger PK, Lubman DI, Lum JAG, Reynolds J, et al. Effect of cognitive bias modification on early relapse among adults undergoing inpatient alcohol withdrawal treatment: a randomized clinical trial. JAMA Psychiatry 2021 Feb 01;78(2):133-140 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.3446] [Medline: 33146693]
20. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 2011 Apr;22(4):490-497 [FREE Full text] [doi: 10.1177/0956797611400615] [Medline: 21389338]
21. Wiers RW, Rinck M, Kordts R, Houben K, Strack F. Retraining automatic action-tendencies to approach alcohol in hazardous drinkers. Addiction 2010 Feb;105(2):279-287. [doi: 10.1111/j.1360-0443.2009.02775.x] [Medline: 20078486]

22. Manning V, Staiger PK, Hall K, Garfield JBB, Flaks G, Leung D, et al. Cognitive bias modification training during inpatient alcohol detoxification reduces early relapse: a randomized controlled trial. Alcohol Clin Exp Res 2016 Sep;40(9):2011-2019. [doi: 10.1111/acer.13163] [Medline: 27488392]
23. Rinck M, Wiers RW, Becker ES, Lindenmeyer J. Relapse prevention in abstinent alcoholics by cognitive bias modification: Clinical effects of combining approach bias modification and attention bias modification. J Consult Clin Psychol 2018 Dec;86(12):1005-1016. [doi: 10.1037/ccp0000321] [Medline: 30507226]
24. Haber P, Lintzeris N, Proude E, Lopatko O. Guidelines for the Treatment of Alcohol Problems. Australian Government Department of Health and Ageing. 2009 Jun. URL: https://www.health.gov.au/sites/default/files/guidelines-for-the-treatment-of-alcohol-problems_0.pdf [accessed 2021-11-26]
25. Dawson DA, Li T, Grant BF. A prospective study of risk drinking: at risk for what? Drug Alcohol Depend 2008 May 01;95(1-2):62-72 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.12.007] [Medline: 18243584]
26. Kuntsche E, Kuntsche S, Thrul J, Gmel G. Binge drinking: Health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):976-1017. [doi: 10.1080/08870446.2017.1325889] [Medline: 28513195]
27. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
28. Garnett C, Michie S, West R, Brown J. Updating the evidence on the effectiveness of the alcohol reduction app, Drink Less: using Bayes factors to analyse trial datasets supplemented with extended recruitment. F1000Res 2019;8:114 [FREE Full text] [doi: 10.12688/f1000research.17952.2] [Medline: 31354942]
29. Laurens MC, Pieterse ME, Brusse-Keizer M, Salemink E, Ben Allouch S, Bohlmeijer ET, et al. Alcohol avoidance training as a mobile app for problem drinkers: longitudinal feasibility study. JMIR Mhealth Uhealth 2020 Apr 14;8(4):e16217 [FREE Full text] [doi: 10.2196/16217] [Medline: 32286235]
30. Manning V, Mroz K, Garfield JBB, Staiger PK, Hall K, Lubman DI, et al. Combining approach bias modification with working memory training during inpatient alcohol withdrawal: an open-label pilot trial of feasibility and acceptability. Subst Abuse Treat Prev Policy 2019 Jun 06;14(1):24 [FREE Full text] [doi: 10.1186/s13011-019-0209-2] [Medline: 31171005]
31. Stacy AW, Wiers RW. Implicit cognition and addiction: a tool for explaining paradoxical behavior. Annu Rev Clin Psychol 2010;6:551-575 [FREE Full text] [doi: 10.1146/annurev.clinpsy.121208.131444] [Medline: 20192786]
32. Garfield JBB, Piercy H, Arunogiri S, Lubman DI, Campbell SC, Sanfilippo PG, et al. Protocol for the methamphetamine approach-avoidance training (MAAT) trial, a randomised controlled trial of personalised approach bias modification for methamphetamine use disorder. Trials 2021 Jan 06;22(1):21 [FREE Full text] [doi: 10.1186/s13063-020-04927-6] [Medline: 33407781]
33. Manning V, Piercy H, Garfield JBB, Lubman DI. Personalized approach bias modification smartphone app ("SWIPE") to reduce alcohol use among people drinking at hazardous or harmful levels: protocol for an open-label feasibility study. JMIR Res Protoc 2020 Aug 14;9(8):e21278 [FREE Full text] [doi: 10.2196/21278] [Medline: 32795989]
34. Kopetz C, MacPherson L, Mitchell AD, Houston-Ludlam AN, Wiers RW. A novel training approach to activate alternative behaviors for smoking in depressed smokers. Exp Clin Psychopharmacol 2017 Feb;25(1):50-60. [doi: 10.1037/pha0000108] [Medline: 28150972]
35. Wiers RW, Becker D, Holland RW, Moggi F, Lejuez CW. Cognitive motivational processes underlying addiction treatment. In: Kopetz CE, Lejuez CW, editors. Addictions: A social psychological perspective. New York, NY: Routledge/Taylor & Francis Group; 2016:201-236.
36. Wiers RW, Zerhouni O, den Uyl T, Boffo M. Cognitive bias modification in the treatment of addiction. In: Verdejo-Garcia A, editor. Cognition and addiction: A researcher's guide from mechanisms towards interventions. Cambridge, MA: Academic Press; 2020:231-241. [doi: 10.1016/b978-0-12-815298-0.00017-4]
37. Wiers RW, Boffo M, Field M. What's in a trial? On the importance of distinguishing between experimental lab studies and randomized controlled trials: the case of cognitive bias modification and alcohol use disorders. J Stud Alcohol Drugs 2018 May;79(3):333-343. [Medline: 29885138]
38. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
39. 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]
40. Gossop M, Marsden J, Stewart D. Dual dependence: assessment of dependence upon alcohol and illicit drugs, and the relationship of alcohol dependence among drug misusers to patterns of drinking, illicit drug use and health problems. Addiction 2002 Feb;97(2):169-178. [doi: 10.1046/j.1360-0443.2002.00028.x] [Medline: 11860388]
41. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]

42. May J, Andrade J, Kavanagh DJ, Feeney GFX, Gullo MJ, Statham DJ, et al. The craving experience questionnaire: a brief, theory-based measure of consummatory desire and craving. Addiction 2014 May;109(5):728-735. [doi: 10.1111/add.12472] [Medline: 24400950]
43. Standard drinks guide. Australian Government Department of Health. URL: https://www.health.gov.au/health-topics/alcohol/about-alcohol/standard-drinks-guide [accessed 2021-11-26]
44. "SWIPE": A Personalised brain-training app to reduce alcohol cravings and consumption. Australian New Zealand Clinical Trials Registry. URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12620000638932p [accessed 2021-11-26]
45. Lubman DI, Garfield JBB, Manning V, Berends L, Best D, Mugavin JM, et al. Characteristics of individuals presenting to treatment for primary alcohol problems versus other drug problems in the Australian patient pathways study. BMC Psychiatry 2016 Jul 19;16:250 [FREE Full text] [doi: 10.1186/s12888-016-0956-9] [Medline: 27435013]
46. Australian Institute of Health and Welfare. 2019 Oct 22. URL: https://www.aihw.gov.au/reports/rural-remote-australians/rural-remote-health/contents/profile-of-rural-and-remote-australians [accessed 2021-11-26]
47. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: the Alcohol Use Disorders Identification Test: guidelines for use in primary health care. World Health Organization. 2001 Nov 18. URL: https://www.who.int/publications/i/item/audit-the-alcohol-use-disorders-identification-test-guidelines-for-use-in-primary-health-care [accessed 2021-11-26]
48. Lawrinson P, Copeland J, Gerber S, Gilmour S. Determining a cut-off on the Severity of Dependence Scale (SDS) for alcohol dependence. Addict Behav 2007 Jul;32(7):1474-1479. [doi: 10.1016/j.addbeh.2006.09.005] [Medline: 17081703]
49. National Health and Medical Research Council, Australian Research Council, Universities Australia. Australian guidelines to reduce health risks from drinking alcohol. National Health and Medical Research Council. 2020. URL: https://www.nhmrc.gov.au/about-us/publications/australian-guidelines-reduce-health-risks-drinking-alcohol [accessed 2021-11-26]
50. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 03;7:30 [FREE Full text] [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]
51. McCambridge J, Day M. Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. Addiction 2008 Feb;103(2):241-248. [doi: 10.1111/j.1360-0443.2007.02080.x] [Medline: 18199302]
52. McCambridge J, Wilson A, Attia J, Weaver N, Kypri K. Randomized trial seeking to induce the Hawthorne effect found no evidence for any effect on self-reported alcohol consumption online. J Clin Epidemiol 2019 Apr;108:102-109. [doi: 10.1016/j.jclinepi.2018.11.016] [Medline: 30458263]
53. Carreiro S, Chai PR, Carey J, Lai J, Smelson D, Boyer EW. mHealth for the detection and intervention in adolescent and young adult substance use disorder. Curr Addict Rep 2018 Jun;5(2):110-119 [FREE Full text] [doi: 10.1007/s40429-018-0192-0] [Medline: 30148037]
54. Simons JS, Wills TA, Emery NN, Marks RM. Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms. Addict Behav 2015 Nov;50:205-212 [FREE Full text] [doi: 10.1016/j.addbeh.2015.06.042] [Medline: 26160523]
55. Kreitman N. Alcohol consumption and the preventive paradox. Br J Addict 1986 Jun;81(3):353-363. [doi: 10.1111/j.1360-0443.1986.tb00342.x] [Medline: 3461846]

Abbreviations

ApBM: approach bias modification
AUD: alcohol use disorder
AUDIT: Alcohol Use Disorders Identification Test
CEQ-F: Craving Experience Questionnaire – Frequency Scale
mHealth: mobile health
NHMRC: National Health and Medical Research Council
RCT: randomized controlled trial
RMANOVA: repeated measures analysis of variance
RT: reaction time
SDS: Severity of Dependence Scale
uMARS: Mobile Application Rating Scale – User Version
VAS: visual analogue scale

Edited by G Eysenbach; submitted 18.06.21; peer-reviewed by M Laurens, M Lotto, M Rinck; comments to author 14.07.21; revised version received 19.07.21; accepted 24.10.21; published 10.12.21.

Please cite as:
Manning V, Piercy H, Garfield JBB, Clark SG, Andrabi MN, Lubman DI
A Personalized Approach Bias Modification Smartphone App ("SWiPE") to Reduce Alcohol Use: Open-Label Feasibility, Acceptability, and Preliminary Effectiveness Study
JMIR Mhealth Uhealth 2021;9(12):e31353
URL: https://mhealth.jmir.org/2021/12/e31353
doi:10.2196/31353
PMID:34890355

Original Paper

Perceptions of Factors Influencing Engagement With Health and Well-being Apps in the United Kingdom: Qualitative Interview Study

Dorothy Szinay[1], MSc; Olga Perski[2], PhD; Andy Jones[3], PhD; Tim Chadborn[4], PhD; Jamie Brown[2,5], PhD; Felix Naughton[1], PhD

[1]School of Health Sciences, University of East Anglia, Norwich, United Kingdom
[2]Department of Behavioural Science and Health, University College London, London, United Kingdom
[3]Norwich Medical School, University of East Anglia, Norwich, United Kingdom
[4]Behavioural Insights, Public Health England, London, United Kingdom
[5]SPECTRUM Consortium, London, United Kingdom

Corresponding Author:
Dorothy Szinay, MSc
School of Health Sciences
University of East Anglia
Norwich Research Park
Earlham Road
Norwich, NR4 7TJ
United Kingdom
Phone: 44 1603593064
Email: d.szinay@uea.ac.uk

Abstract

Background: Digital health devices, such as health and well-being smartphone apps, could offer an accessible and cost-effective way to deliver health and well-being interventions. A key component of the effectiveness of health and well-being apps is user engagement. However, engagement with health and well-being apps is typically poor. Previous studies have identified a list of factors that could influence engagement; however, most of these studies were conducted on a particular population or for an app targeting a particular behavior. An understanding of the factors that influence engagement with a wide range of health and well-being apps can inform the design and the development of more engaging apps in general.

Objective: The aim of this study is to explore user experiences of and reasons for engaging and not engaging with a wide range of health and well-being apps.

Methods: A sample of adults in the United Kingdom (N=17) interested in using a health or well-being app participated in a semistructured interview to explore experiences of engaging and not engaging with these apps. Participants were recruited via social media platforms. Data were analyzed with the framework approach, informed by the Capability, Opportunity, Motivation–Behaviour (COM-B) model and the Theoretical Domains Framework, which are 2 widely used frameworks that incorporate a comprehensive set of behavioral influences.

Results: Factors that influence the capability of participants included available user guidance, statistical and health information, reduced cognitive load, well-designed reminders, self-monitoring features, features that help establish a routine, features that offer a safety net, and stepping-stone app characteristics. Tailoring, peer support, and embedded professional support were identified as important factors that enhance user opportunities for engagement with health and well-being apps. Feedback, rewards, encouragement, goal setting, action planning, self-confidence, and commitment were judged to be the motivation factors that affect engagement with health and well-being apps.

Conclusions: Multiple factors were identified across all components of the COM-B model that may be valuable for the development of more engaging health and well-being apps. Engagement appears to be influenced primarily by features that provide user guidance, promote minimal cognitive load, support self-monitoring (capability), provide embedded social support (opportunity), and provide goal setting with action planning (motivation). This research provides recommendations for policy makers, industry, health care providers, and app developers for increasing effective engagement.

(JMIR Mhealth Uhealth 2021;9(12):e29098) doi:10.2196/29098

KEYWORDS

behavior change; health apps; mHealth; smartphone app; framework analysis; COM-B; TDF; user engagement; motivation; usability; engagement; mobile phone

Introduction

Background

Smoking, physical inactivity, inadequate diet, and excessive alcohol consumption are the main risk factors for noncommunicable diseases that are responsible for >56.9 million deaths worldwide [1]. People with mental health problems often have poor physical health and vice versa [2,3]. A range of interventions has been developed to reduce the burden of ill health. The integration of multimedia technologies within the health care domain has led to the development of interventions that are delivered digitally via mobile phones, wearable devices, and smartphone apps. Smartphone apps are constantly available to the user and therefore act as portable tools for the delivery of easily accessible health and well-being interventions [4]. There is early evidence of the effectiveness of apps for physical inactivity [5-8], weight loss [7,9,10], alcohol reduction in nondependent drinkers [11], and mental health promotion [12]. Health apps are also considered a cost-effective solution [7,13] and have the potential to increase access for hard-to-reach populations that are resistant or unable to seek face-to-face support, for instance, because of stigma or geographical barriers [14].

Engagement is a necessary component for the effectiveness of a health or well-being app. Engagement with health and well-being apps can be defined as "(1) the extent (e.g. amount, frequency, duration, depth) of usage and (2) a subjective experience characterised by attention, interest and affect" [15]. However, it has been argued that measuring *effective engagement* is more important than simply the time spent on an app and the frequency of use [16]. Yardley et al [16] define *effective engagement* with a smartphone app as involving 2 components: the first is the intensity of engagement that is necessary for achieving desired outcomes, with sustained app engagement over a period of weeks, months, or even years (referred to as *microengagement*). However, microengagement alone is not sufficient for behavior change [16]. The Yardley et al [16] model also emphasizes engagement with the broader behavior change process and goals (ie, *macroengagement*), which is considered separate from, although intimately linked with, microengagement. On the basis of this distinction of microengagement and macroengagement with health and well-being apps, some factors may relate more to the former or the latter, with microengagement influencing macroengagement and vice versa. For example, engagement may be affected by common contextual factors, such as personal (eg, their interest), environmental (eg, where the engagement occurs and the individual's lifestyle), or social context (eg, family or culture). Owing to the complexity of engagement, researchers recognize that it is difficult to define what constitutes *good* or *sufficient* engagement. Some individuals may require a longer period of engagement with an app than others for the desired behavior change to occur.

Despite the promise of health apps, engagement tends to be poor [17,18]. For example, a mobile consumer report found that for medical, health, and fitness apps, only 20% of users use the app 1 day after installation and only 8% after 7 days of installation [19]. A panel-based analysis systematically examined use patterns in 93 mental health apps and found that the median app retention rates at 15 and 30 days of installation were 3.9% and 3.3%, respectively [18].

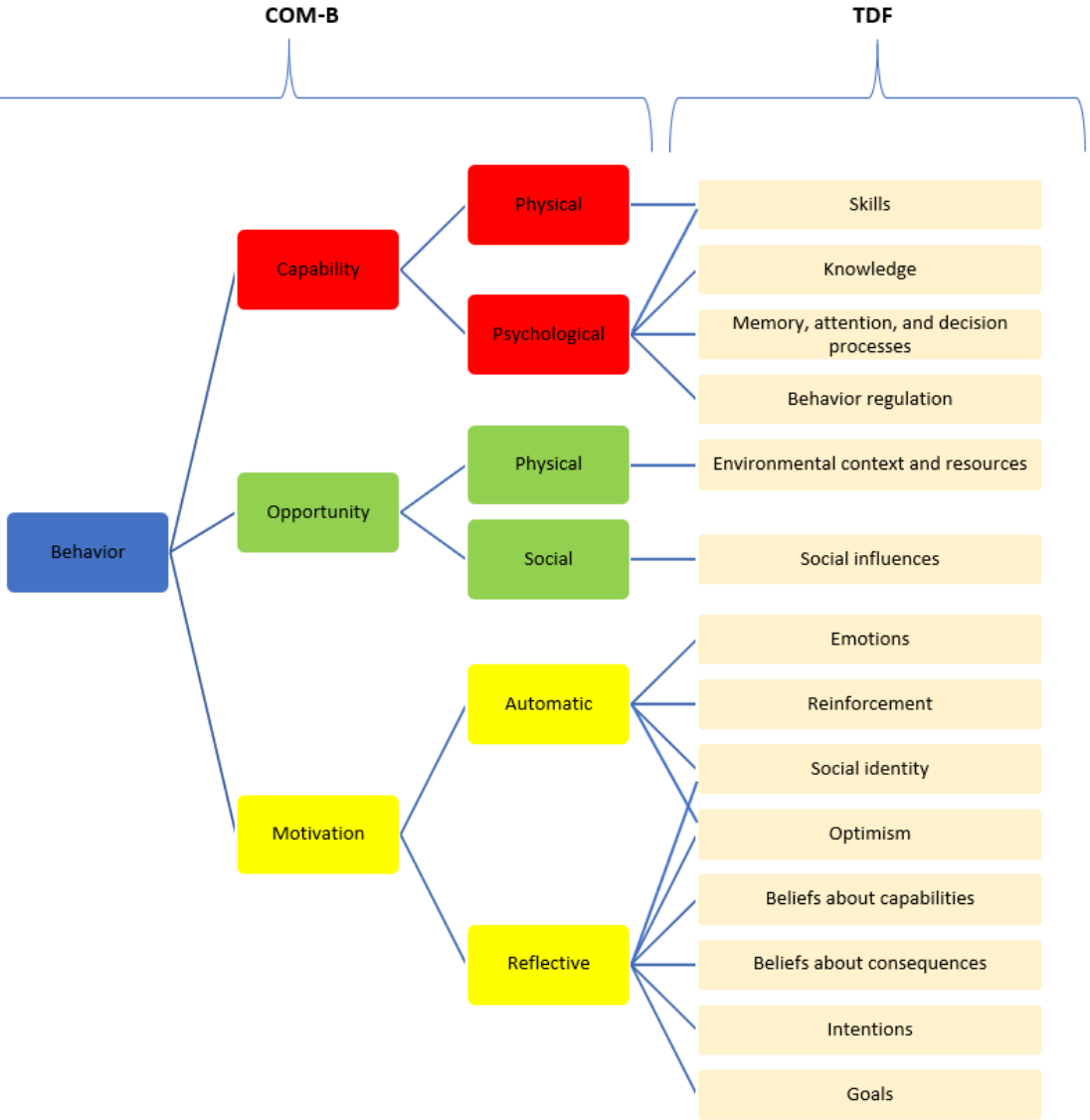
There is growing literature on the factors influencing engagement with health and well-being smartphone apps. In our recent review of 41 studies, we identified 26 factors that are important for the uptake of and engagement with such apps [20]. In addition to a wide range of behavior change techniques (eg, self-monitoring and goal setting) [21,22], several other factors were identified as influential, including the role of health care professionals in the promotion and recommendation of health apps [23] and embedded professional support [24]. The latter was found to be particularly important for certain behaviors (ie, alcohol reduction, suicide prevention, anxiety, and self-harm), with stand-alone apps considered insufficient by users and clinicians [14]. In an assessment of 93 mental health apps, daily minutes of engagement were higher for apps that included peer support (median 35.1, IQR not applicable; n=2) and coping strategies such as mindfulness and meditation (median 21.5, IQR 15) compared with apps that incorporated self-monitoring or psychoeducational features (median range 3.53-8.32) [18]. Few qualitative studies have been undertaken to explore the factors that affect engagement with health and well-being apps. These undertaken studies have focused on specific populations or behaviors. Existing studies have focused on weight loss behaviors and alcohol reduction and have found that the specific content of health information messages [17,22], personalization of app content [25], and the user's demographics [22] are some of the factors deemed to be important for engagement with weight loss and alcohol reduction apps. Findings from these studies highlight that the specific context in which apps are developed and used will influence user engagement. To date, most studies have investigated the features of health apps that are desirable by a certain population, and little is known about the factors deemed important for engagement with a wider range of health and well-being apps [20]. These studies suggest that the context in which apps are developed and used might often be behavior or population specific. Therefore, this research intends to address this gap by exploring the views on the big 4 public health priority behaviors related to prevention (smoking, alcohol consumption, physical activity, and diet) and mental health. The findings from this study may inform future app development to improve user engagement with apps that target health promotion. The findings may also be particularly useful for stakeholders in public health

to inform the development of interventions to promote engagement with evidence-based health and well-being apps, for example, directly contributing to the long-term plan of England's National Health Service to become *digital first*.

Theoretical Framework

The Capability, Opportunity, Motivation–Behaviour (COM-B) model [26] provides a broad framework for understanding the factors influencing user engagement. According to the COM-B model, behavior (eg, app uptake and engagement) arises from the interaction between the individual's physical (eg, app skills) and psychological (eg, knowledge of using an app) capability, physical (eg, features of the app) and social (eg, recommendations for an app) opportunity, and automatic (eg, feedback received from an app) and reflective (eg, user's self-confidence in using an app) motivation. The Theoretical Domains Framework (TDF) [27] is a synthesis of 33 theories and 128 psychological constructs and includes 14 domains that can be mapped under the 3 main components of the COM-B model. Taken together, the COM-B and TDF provide a detailed theoretical framework that allows the careful consideration of factors influencing engagement with health and well-being apps (Figure 1 [20]).

Figure 1. A visual representation of how the Theoretical Domains Framework can be mapped onto the components of the Capability, Opportunity, Motivation–Behaviour model [20]. COM-B: Capability, Opportunity, Motivation–Behaviour; TDF: Theoretical Domains Framework.



Aim

We aim to investigate people's experiences and reasons for engaging or not engaging with health and well-being apps using qualitative interviews and map the identified factors onto the COM-B model and the TDF.

Methods

Study Design

This qualitative study used semistructured interviews and was designed and reported in line with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Multimedia Appendix 1) [28]. The study protocol was

preregistered on the Open Science Framework [29]. This study was part of a larger project investigating both the uptake of and engagement with health and well-being apps, and the findings on uptake are published elsewhere [30].

Ethical approval was obtained from the Faculty of Medicine and Health Sciences Ethics Committee at the University of East Anglia (reference number 201819–089).

Participants and Recruitment

Participants were recruited through social media. Recruitment through Facebook is known to be an effective way to reach adults interested in health and well-being apps [31]. It is fast and cost-effective and has been found to provide better representation and improved participant selection compared with traditional recruitment methods [31]. Eligible participants for this study (1) were aged ≥18 years, (2) were able to give consent, (3) owned a smartphone, (4) would consider using a smartphone app to change their behavior in the future, and (5) could travel for an interview. Purposive sampling was used to ensure the diversity of the sample (age, gender, ethnicity, educational level, and employment) [32]. Invitations for interview were sent to 38 participants, of whom 14 (37%) individuals did not respond, and 6 (16%) subsequently canceled for personal reasons, leaving a total of 18 individuals. The recruitment and interviews took place in batches of 3 or 4, and the recruitment was stopped when data saturation was reached.

Measures

To determine eligibility and describe the sample, data were collected on (1) age; (2) gender; (3) ethnicity, which was measured using the Office for National Statistics' index; (4) level of education; (5) employment status; (6) whether they had ever used a health or well-being app; (7) whether they currently use a health or well-being app; (8) the last time they downloaded an app; and (9) frequency of app use.

Procedure

The participants read the information sheet available on the web. Those who expressed interest in participating were required to fill a web-based screening questionnaire to assess their eligibility (Multimedia Appendix 2). The questionnaire was hosted by the *Jisc Online Surveys* (JISC) software. Participants meeting the inclusion criteria received a comprehensive participant information sheet via email and were invited for an interview during which written consent was obtained. Participants received a US $27.50 (UK £20) gift voucher for their participation.

The interviews were conducted by a female researcher (DS) between July 2019 and August 2019. The interviews took place face-to-face in Norwich, England, at the University of East Anglia (17/18, 94%) or at the participant's home (1/17, 6%). A participant was not included in this study because they had no previous experience of using health or well-being apps. No one else was present during the interviews. The interviews lasted between 26 and 63 minutes. Semistructured interview techniques were used to elicit data on participants' experiences of and views on engagement with health and well-being apps. The final topic guide was informed by feedback from key stakeholders, including patient and public involvement representatives and domain experts from Public Health England. A think-aloud task (reported elsewhere [30]) was followed by several questions regarding participants' experiences of engagement with apps (see Multimedia Appendix 3 for the topic guide).

Data Analysis

The interviews were audio recorded and transcribed verbatim. A framework analysis approach was applied to analyze the data. This followed the stages of familiarization, identification of the thematic framework, indexing, charting, mapping, and interpretation [33]. For pragmatic reasons, a second author (OP) independently coded 3 randomly selected transcripts (representing approximately 15% of the transcripts). The deductive framework was based on the TDF. Through repeated discussions between the first and second authors, the deductive thematic framework was refined iteratively, and discrepancies were resolved through discussions with another author (FN). DS completed indexing using QSR NVivo 12 and charting. During charting, responses were clustered based on the thematic framework. This was followed by mapping and interpretation, during which the data were examined to identify patterns. To increase the trustworthiness of the findings, peer debriefing by the University College London Tobacco and Alcohol Research Group, which has extensive experience in the application of the COM-B model and TDF in health research, was used to ensure the accuracy of data interpretation and data analysis. Peer debriefing is a form of analytical triangulation in which researchers who are not directly involved in the study are prompted to provide input and critical opinions on various aspects of a project [34]. The use of the TDF in the deductive framework analysis approach was particularly useful for coding the results under several factors, which may otherwise have been overlooked. It was expected to explore a large number of factors because the TDF has 14 constructs, as opposed to other well-known methods. However, researchers were aware that the findings would not be coded under all available constructs. The constructs under which no findings were coded were omitted from the *Results* section.

External Validity

Member checking was conducted to ensure the trustworthiness of the results and further minimize researcher bias [35]. After the interview, participants were told that the researchers might contact them to share the findings and ask their views on the findings. Of the 18 participants, 6 (30%) randomly selected participants were contacted and invited to provide feedback on the summary of the findings and the conclusion. The purpose was to investigate whether participants agreed with the interpretation of the results and whether they felt that their opinions were captured and appropriately presented. Incentives were not offered for their input. Of the 6 participants who were approached, 2 (33%) participants responded and indicated that they agreed with the interpretation.

Reflexivity

DS, a PhD candidate, conducted the interviews after receiving extensive training in qualitative research methodology and kept field notes and a research journal during data collection. There

was no prior relationship established between DS and the participants. The coauthors had experience in mixed methods research and the application of the COM-B model and the TDF. Efforts to establish a good rapport with the participants were made throughout the study. The interviews were individually adapted to the flow of discussion made by each participant. Most participants (15/18, 83%) stated that they wished to find out more about the findings of the research. DS also shared her research interests with the participants after the interview.

Results

Participant Characteristics

A total of 18 adults (mean age 43, SD 14 years; range 21-68 years) were recruited, of whom 9 (50%) were female, 14 (78%) were White British, 13 (72%) were employed full time, and 8 (44%) had a college degree or higher. Of the 18 participants, 11 (61%) participants reported currently using at least one health or well-being app at the time of the interview. Of the 18 participants, 3 (17%) participants expressed their intention to change 1 behavior, and all the other participants were interested in changing >1 behavior (eg, losing weight, being more active, and managing their mood). Of the 18 participants, 1 (6%) participant had never used health apps before and did not wish to express their views on engagement; therefore, the findings of this paper are based on the views and experiences of the remaining 17 participants about their engagement with health and well-being apps (Multimedia Appendix 4).

Factors Influencing Engagement With Health and Well-Being Apps

An overview of the factors mapped under the constructs of the TDF and components of the COM-B can be found in Table 1. All relevant data were coded under 71% (10/14) of the constructs of the TDF. There were no data that could not be coded under any of the constructs of the TDF.

Table 1. Perception of factors influencing engagement with health apps.

COM-B[a] model component, TDF[b] construct, and factors	Description
Psychological capability	
Knowledge	
User guidance	Instructions on how to effectively use a health app
Statistical information	A visual or numerical summary of progress or quantification of the behavior
Health information	Educational information related to health and well-being aspects
Memory, attention, and decision processes	
Reduced cognitive load	The app is not too time-consuming, is easy to use, and requires minimal input
Reminders	Preferably customizable notification-type messages
Behavior regulation	
Self-monitoring	The ability of the app to support self-regulation of the target behavior
Routines	The ability to support routine or habit formation
Safety netting	Retaining the app for a potential upcoming event in the future
Stepping-stone	App as a first step in the behavior change process
Physical opportunity	
Environmental resources	
Tailoring	Innovative features, adaptability, and interactive and 2-way communication between the app and user
Social opportunity	
Social influences	
Peer support	Including social interaction with users with similar needs within the app or within their community; a choice to connect to social media platforms, competitions, and challenges with others or with themselves
Social support (practical)	Possibility to contact health professionals and practitioners within the app
Reflective motivation	
Beliefs about capabilities	
Self-confidence	Perceived capability to change one's behavior using an app
Goals	
Goal setting	Establishing what the user would like to achieve
Action planning	Establishing how the user would like to achieve set goals
Beliefs about consequences	
Commitment	The level of commitment while engaging with an app to change the behavior and achieve set goals
Automatic motivation	
Reinforcement	
Feedback	Feedback regarding the user's performance
Rewards	Tangible (eg, objects and discounts) and intangible (eg, badges and certificates) rewards in response to the user's effort and gamification elements
Encouragement	Additional ways to provide reinforcement (eg, encouraging messages)
Emotions	
Positive emotions	Triggered by the included user guidance, statistical information, additional health information, embedded professional support, community networking possibilities, tracking features, and rewards
Negative emotions	Triggered by lack of user guidance, invasive push notifications, cognitive overload, and unrevealed in-app costs
Mixed emotions	Triggered by reminders

[a]COM-B: Capability, Opportunity, Motivation–Behaviour.

[b]TDF: Theoretical Domains Framework.

Capability to Engage With Health and Well-Being Apps

Knowledge

Many participants perceived their knowledge on how to use an app and the embedded statistical and health information as an important influence on their engagement with an app. We inferred this from the desire that many people reported for clear *user guidance* and, in some cases, for help on how to increase their capability to perform a behavior (eg, demonstration of the behavior). A participant explained that they had stopped using an app in the past owing to "insufficient guidance on how to use it" (Participant 8):

> *So this is where I start getting, well why are you asking me these questions if you're not going to let me carry on with it and that's where I start getting confused, going back, not really understanding where I need to go from here.* [Participant 15]

Furthermore, the importance of *statistical information* about their progress and achievements was reported by most participants:

> *It's nice to see your progress on a graph and it's just very clear. It's a single screen, you have icons for all the activities that you've done during the day.* [Participant 6]

In addition, most participants expressed the need for relevant and comprehensive *health information*:

> *Knowledge is key.* [Participant 14]

Several participants stated that having embedded educational articles would help them to build knowledge and understand and manage their behavior better. Not getting enough health information was reported as the main reason for a participant to look for a different app:

> *It's got to have the information that I want and have it easily accessible.* [Participant 2]

Memory, Attention, and Decision Processes

Participants perceived *reduced cognitive load* and customizable notification-type *reminders* as factors that positively affect their capability to engage with an app. All participants described favoring apps with *reduced cognitive load*. This included apps with limited complexity, less data input, and a limited number of available features to choose from.

A participant suggested that an app should apply a multilevel approach with "a light version of an app and then enhanced" (Participant 15). They described that an app might have a simple version for basic users with no registration and minimum data input and a more advanced version with all features available for power users.

Several participants expressed that a time-consuming app would be immediately deleted:

> *A mood tracker is something I probably wouldn't use because it looks like it would require a lot of data of me putting in and typing it on to stuff.* [Participant 7]

Although push notifications were considered more or less annoying, many participants described *reminders* as being particularly useful. A participant described that not being reminded to engage with an app led him to disengage:

> *Because I wasn't reminded, I stopped using it. And I think that's really important.* [Participant 1]

However, a few participants who reported not finding notifications useful stated that they would immediately turn the reminders off or delete the app:

> *I'm sure there are many apps I've deleted because of reminders.* [Participant 7]

Others have suggested that *reminders* might cause harm. For example, a participant described uninstalling a smoking cessation app as the reminders periodically reminded them about their addiction, thus serving as a prompt that induced cigarette cravings. However, 12% (2/17) of participants proposed that opting in to receive reminders would be desirable instead of opting out. In addition, a participant suggested that human-like reminders in the form of SMS text messages would be less likely ignored and would create the perception of a human touch within the app:

> *I think text messages would work better because I don't ignore my text messages and my WhatsApp messages because there's real people connected to those, you know? ... if I could think of an ideal it would be a text message that kind of asked you a question and you replied and it felt like it was a human being.* [Participant 6]

Behavior Regulation

Participants perceived that *self-monitoring*, established *routines*, and *safety netting* and *stepping-stone* characteristics of the app would enhance their engagement with the app.

All described *self-monitoring* features as important in behavior regulation, even when there is no particular goal set or when achieving the goal shows a delay:

> *Monitoring, really because the goal is probably going to go a bit by the wayside because work has been too busy and life has changed and lots of stuff has happened this year. So I'm behind my goal but I still use it as a monitor.* [Participant 17]

Some participants reported that a daily *routine* of using an app would make engagement with it more accessible and continuous. Of the 17 participants, 2 (12%) participants described how using a weight management app for a week was necessary for them to get into a routine and helped them stay engaged after that. However, one of them explained that it was difficult to use the app in the beginning, although after a few days, it got easier.

A number of participants explained that they perceived physical activity apps as *stepping-stones* to physical activity services, with the app acting as an intermediate tool in behavior change. Of the 17 participants, 2 (12%) participants described that an app helped them to get enough experience and practice home workouts that boosted their confidence to eventually sign up for a gym membership:

> *You can just literally do it at home [fitness app] until you feel I suppose a bit more confident to go out and join [the gym]*. [Participant 10]

Many of the participants described apps as a *safety netting* tool (eg, relapse prevention). Several participants reported a tendency to reengage with a weight management app periodically and when necessary to regulate their weight, for example, before or after a holiday season or an important upcoming event because the app had helped them achieve their goals in the past:

> *I think I have periodically come back to it and thought "no it worked before, it'll work again.*" [Participant 13]

Opportunity to Engage With Health and Well-Being Apps

Environmental Resources

Participants perceived that *tailoring* the technology was a factor that would influence sustained engagement. Many participants expressed the need for features that would create a better physical opportunity to engage with an app and a more personalized experience during the engagement. Many participants described seeking to engage with apps that provide 2-way communication, which can adapt to the person's needs based on how they interact with such tools. Several participants mentioned the inclusion of innovative features. These features comprised embedded artificial intelligence (AI) to receive health-related advice and tailored content; facial recognition; and recognition of nonverbal cues for better outcomes in physical activity, for example, correcting posture; and using the phone's camera to provide nutritional data of cooked food:

> *If it's smart, as well. Has it got a little bit of artificial intelligence built into the background? Is it using my data? Is it saying "do you know what? Actually, you've done really well this week, you've used the app this amount of times. How are you feeling?"* [Participant 2]

A participant described that the lack of novelty of an app would lead them to disengage with it. In contrast, another reported the opposite—they would feel put off if they would need to learn new features:

> *It's no good downloading an app and then six months later looking at that app and it's still the same, that would stop me.* [Participant 14]

> *If something's working we want it to stay as it is, we don't want it to change, and even if there are improvements to it, if it's new it can kind of put people off in a way.* [Participant 13]

Syncing with wearables or other additional devices was described as desirable by many.

Social Influences

Peer support and social support (practical) were perceived by participants as factors that may sustain engagement with an app. Several participants perceived networking within a web-based community as necessary peer support. Some described that sharing and exchanging experiences with others would encourage and motivate them in their journey. Others suggested anonymity for users and moderation of discussions to avoid "misinformation" (Participant 12):

> *I like the idea that it's round the clock support, because so very often with mental health issues it's kind of 2 o'clock in the morning that they are the worst, and that is when you need to talk to somebody, and the idea of having a community who you don't have to explain how you're feeling sounds really good.* [Participant 11]

Embedded social media to share their progress with others was reported as a useful feature only by a few participants who were using physical activity or weight management apps. However, a couple of participants highlighted that this feature should be optional. Physical activity and weight management app users also described challenges and competitions as motivating and fun:

> *There's challenges, which will help you with your weight loss, your fruit and vegetable intake, the exercise challenges that you can do, either with yourself or your friends, which are good for motivation.* [Participant 15]

All participants expressed their preference for an app that would offer built-in professional support, such as health practitioners, coaches, and dieticians (for *social support* and *practical support*). A participant with an existing medical condition described the need for health practitioner support within an app. In addition, 12% (2/17) of participants described that built-in support would help with accountability, and 6% (1/17) of participants indicated that they would be willing to pay to access an app with in-built support. Another participant commented that the embedded professional support was the best feature of a mental health app they were using:

> *Yeah if you could sort of talk to a healthcare professional in that app I think that would be better, because then they would have the up to date I suppose treatments and methods so that you know you're not going on old information.* [Participant 10]

> *I: If you would need to say just one thing that is the best in the app, what would that be? P: The support.* [Participant 11]

Motivation to Engage With Health and Well-Being Apps

Beliefs About Capabilities

Apps were perceived by several participants as useful tools to enhance their *self-confidence* in changing their behavior. A participant described that the community networking opportunities further helped her *self-confidence* and motivated her to use the app:

> *The app made me feel more confident in doing it, even it was just basic home exercises.* [Participant 7]

Goals

Goal setting and *action planning* were perceived as key factors for sustained engagement and motivators of behavior change. *Goal setting* was reported as being valuable by all participants

in addressing behavior change; however, half of the participants described the need for *action planning* features to help them achieve their set goals:

> *I'd want something which was a bit more than press one button every day to say you haven't smoked; it was great for the first 10 minutes of using the app because I got all this information about "wow thousands of pounds and the health benefits," and then after that it was literally just press this button to say you haven't smoked, and that wasn't really enough for me.* [Participant 13]

Beliefs About Consequences

Several participants expressed that their level of *commitment* to achieve their goal shaped the level of engagement with the app they used:

> *The app, the initial—the main reason you're on that app is to get your result of what you want to achieve, what you want to do to help you stay on track.* [Participant 9]

Reinforcement

Many participants perceived *feedback*, *rewards*, and *encouragement* as automatic motivational factors that may sustain engagement with an app. A number of participants expressed that they needed continuous *feedback* to reinforce their continuous use:

> *I think an app that might give you feedback, a notification, that would keep me entertained and would keep my level of focus and wanting to continue with it.* [Participant 3]

Intangible *rewards* (ie, badges and certificates) were described as another form of reinforcement by several participants for motivating them and as *"nice"* (Participant 14) or something to *"show off"* (Participant 5). However, some other participants described intangible rewards as *irrelevant*. They reported that the tangible rewards they received in the past, including cinema tickets, lower insurance premiums, and loyalty points that can be exchanged for objects or a free water bottle, provided better motivation to engage with the app than intangible ones. In addition, a few participants expressed the need for *encouragement* in the form of motivational messages:

> *In this context, so badges, you earn nine of 24 badges so far. For me a little bit irrelevant actually, what are you going to do with it, there's other reasons why you're quitting, not to get the badges.* [Participant 16]

Emotions

Participants expressed positive emotions regarding available user guidance, statistical information, additional health information, embedded professional support, the possibility for community networking, self-monitoring features, and rewards. However, negative emotions were expressed for the lack of user guidance, invasive push notifications, and cognitive overload. Finally, feedback on reminders was person-dependent and triggered mixed feelings across participants.

Discussion

Principal Findings

This study applied the COM-B and the TDF to explore users' views about factors that influence engagement with health and well-being apps. We found that *knowledge*, such as user guidance and statistical information; *memory, attention, and decision processes*, such as reduced cognitive load; *environmental resources*, expressed by the tailored technology; and *social influences*, referred to as peer and professional support, are the most important factors for participants' engagement.

Many factors identified in this study are consistent with those in previous literature. Previous research has found that engagement with health apps is greatly influenced by factors affecting users' capabilities, including different types of knowledge (user guidance, statistical information, and health information) [20,36], reduced cognitive load, reminders, and self-monitoring features [20,22,37]. These factors could be targeted during app development updates of existing apps to improve user engagement. In line with previous findings, reminders were not found to be universally useful [20]. A possible explanation is that reminders may be behavior-dependent and person-dependent. Some participants reported that they had stopped engaging with a health app because they were not reminded to continue using it, whereas others tended to ignore or delete the apps that sent reminders. This research is the first to identify a novel factor, the perception of certain apps as *stepping-stones* to more intensive behavior change. For example, a home-based workout app or a walking app could seek to provide enough self-efficacy and competence for an individual to join a gym or start using a running app. An explicit *stepping-stone* approach could be a useful addition for apps targeting behaviors that are harder to achieve because of negative emotions, such as embarrassment, shame, or pressure, including those targeting sedentary behavior. This novel finding shows that sustained engagement is not always necessary to support desired health and well-being outcomes through additional behavior change activities.

Engagement is further influenced by the users' physical opportunities, such as tailored technology, and social opportunities and peer support, including community networking, embedded social media and social competitions, and professional support [20,24,25,37]. Some users would want the app to be based on machine learning opportunities and on 2-way interaction with users. The adaptable nature of an app and the provision and level of AI included may also play a key role in engagement. These factors may be harder to include once an app is developed; therefore, it might be important to consider these aspects during the development process. Indeed, such tailored technology may be the most important aspect to consider. For example, although there may be financial considerations precluding the provision of personal, professional support within an app, this service may be developed using AI. These forms of technological personalized models in health behaviors such as nutrition or smoking, including machine learning models, have been suggested to aid the process of

making decisions about diet and food [38]. However, AI has not yet been found in diet monitoring apps [39]. A randomized controlled trial found that participants allocated to an advanced version of a smoking cessation app with an AI chatbot had 107% higher engagement with the app and over twice the odds of being abstinent at 1-month follow-up compared with participants using the standard version of the app [40]. Furthermore, timely AI-based behavior change support received just in time may further increase behavior change. Although unguided interventions can be effective, having professional support within an app tends to increase effective engagement [16]. Simple interventions that do not require professional support can be more widely disseminated and are more cost-effective than those with embedded professional support [16].

Users' reflective motivation, including beliefs in their capabilities (self-confidence), consequences (commitment), and goals (goal setting and action planning), is essential for engagement. Although the first 2 factors are harder to address because they are within-person factors, the latter can be easily implemented as features of the app. A possible way to increase self-confidence and commitment is perhaps to address these within the app by using quizzes or articles [41], (eg, for commitment, "How to stay on track to achieve your goal?") or check-in messages using AI [40].

Emotions are considered as automatic motivation factors and are a powerful driver of behaviors that affect adherence, for example, engagement with a health app [42]. It is noteworthy that we did not identify emotions directly influencing engagement, or we failed to identify them. However, we found evidence that the other factors affected participants' emotions. Appealing features, such as statistical and health information, embedded peer and professional support, and tracking features and rewards, triggered positive emotions. In contrast, a lack of user guidance, invasive notifications, and cognitive load triggered negative emotions. A better understanding of how the presence or absence of specific features affect participants' emotions may be useful for the development of new apps or the refinement of existing apps, which, consequently, may lead to better engagement with health apps.

Strengths and Limitations

A strength of this study is the methodology used. First, to assure that the research was as relevant and meaningful as possible, the study protocol was developed with policy maker and patient and public involvement representatives in the design of the topic guide. Second, the research was further informed by well-established theoretical models; the COM-B and the TDF and peer debriefing were used to help data interpretation and data analysis [34]. Third, the purposive sampling technique allowed the recruitment of a diverse sample regarding gender, educational level, and employment status. Finally, member checking was conducted, which is a technique used to establish the credibility of the findings by sending a brief summary of the findings to randomly selected participants [35].

This study has several limitations. The recruitment of a sample of participants with more diverse demographics might have identified additional factors that are important for engagement. Several participants were not using health or well-being apps at the time of the interviews and had not downloaded any health and well-being apps during the 6 months before the interview. This may have led to limitations associated with the challenges of retrospective recall. Although the research aimed to recruit a heterogenous sample to capture a wide with *big 4* public health priority behaviors related to prevention (smoking, alcohol consumption, physical activity, and diet) and mental health apps, a homogenous sample may have allowed for a more in-depth understanding of engagement with apps for specific behaviors. The study only included participants who considered using a smartphone app to change their behavior in the future. Including participants who have used health and well-being apps in the past but are less receptive to using them now may have provided additional perspectives on factors influencing app engagement. The findings may be influenced by the intention-behavior gap, with participants reporting on factors perceived as important for changing their behavior through an app. However, this does not mean that they would act on their intention. An example of this is the finding that many participants wanted access to a web-based community. Although web-based communities typically suffer from the *90-9-1 principle*, whereby the content in web-based communities is generated by 1% of the members, with 9% editing or modifying it and 90% being passive observers [43], this may not be the case with a closed community built to support behavior change, where individuals are seeking support from each other. In addition, the meaning of the term *engagement* was not explicitly defined during the interview when individuals shared their experiences and views of engagement. Their interpretation of engagement is likely a mixture of microengagement and macroengagement, and a distinction between these 2 levels of engagement was not considered when interpreting the findings. Finally, of the 6 participants who were contacted, only 2 (33%) responded to the request for external validation. Therefore, it is not clear whether the other participants disregarded our request or disagreed with our interpretation.

Implications and Future Research

This research provides insight for stakeholders in public health, policy makers, and developers of apps that target disease prevention and health promotion. Our findings may also be used to inform the development of interventions aiming to promote engagement with evidence-based health and well-being apps. In the United Kingdom, this aligns with the priorities of the National Health Service's long-term plan (ie, *digital first*).

Our main finding is centered around providing the necessary support for increased engagement with health apps. We found that embedded professional support may have a substantial impact on engagement, although it may not be beneficial for all health behaviors. Embedded social support may be particularly important for some behaviors that are more likely to be complex and require intensive support to maintain engagement. These behaviors are the ones that require reassurance, guidance, or emotional support [16], such as apps targeting substance misuse or those developed to improve mental health. Although it is not always feasible to develop an app with embedded professional support, there might be different ways outside of the app to address this need. For instance, there may be a way to provide support within community-based care to

assist with the uptake of health apps and with the progress of or potential barriers to engagement. Another way to mitigate the absence of embedded professional support is to investigate the potential efficacy of advanced computational techniques, such as AI, to mimic the support provided by health care professionals (eg, in the form of chatbots or other types of conversational agents). There is an urgent need for more research on the optimal type (eg, technology-mediated or *blended*) and timing of support needed within various health and well-being smartphone apps.

To better meet users' needs, the design of apps would ideally be informed by a user-centered and iterative development process, supported by mixed methods research, including in-depth interviews. As app engagement is generally greater in those with higher socioeconomic status [44], involving individuals with lower socioeconomic status is particularly important [16]. Furthermore, people who are directly affected by the digital divide or digital exclusion and who may struggle to benefit from health apps due to lack of skills or low digital literacy could be targeted by offering app-use tutoring. Although this may require investment or relocation of resources within community health care settings, it may increase the reach of health apps and lead to a greater public health benefit. Furthermore, we noted a possible tension between users wanting the app to be easy to use (which may be facilitated by providing user guidance) but at the same time not too time-consuming. As the provision of user guidance helps individuals with limited technological skills, we believe that such features should still be prioritized. Undoubtedly, finding the balance between producing an app with all features necessary for behavior change to occur and ensuring that the app is intuitive enough will pose a challenge for app developers.

In addition, more experimental research would help us to better understand the effects and potential interactions among the engagement factors identified in this study, including usability (ease of use), reminders, embedded support, rewards, and goal management. Textbox 1 provides a summary of considerations to help app developers and commissioners design interventions to increase effective engagement. These factors are structured around COM-B and TDF.

Textbox 1. Considerations for policy makers, industry, health care providers, and app developers for maximizing engagement with health and well-being smartphone apps.

Capability, Opportunity, Motivation–Behaviour model components and considerations for policy makers, health app portal providers, and app developers

Capability

- Provide user guidance on how to use an app, visual or numerical summary of progress, and evidence-based additional health information related to the behavior targeted by the app
- Minimize the time required to use the app where possible
- Provide customizable reminders that users could opt out from
- Provide the option to self-monitor features
- Promote safety netting and relapse prevention features such as the possibility to restart or reengage with the app later
- Promote a routine for engagement with an app, for example, highlighting the role that routine may play in the effectiveness of an app

Opportunity

- Collaborate with interaction design experts and end users to enhance the aesthetics of apps
- Provide the possibility for community networking within the app and linking to social media as an optional feature to share progress where appropriate
- Offer the possibility for social competition and challenges where appropriate
- Consider the provision of embedded professional support and, if this is not feasible, providing offline one-to-one support with the uptake of and the engagement with health apps; this may improve motivational factors, such as commitment, self-confidence, and perceived competence of engaging with a health app
- We advise that exploration be made for where engagement enhancement could be made with appropriate and proportionate machine learning, artificial intelligence, or other forms of learning systems.

Motivation

- Develop a time-efficient app that would require as much engagement as is required to achieve the desired outcome; this might be different for different behaviors
- Include reinforcement in the form of feedback, encouraging messages, and rewards
- Offer intangible rewards, such as certificates or badges
- Offer tangible rewards that can be converted to discount in other places (eg, health insurance providers, pharmacies, or sports parks)
- Include goal setting and action planning features on how to achieve set goals (when applicable)
- Take into account the user's emotions about certain features by involving users in the development and update of health apps as the lack of some features could provoke strong negative emotions such as disappointment and might lead to rapid disengagement

Conclusions

People perceive their capability to engage with an app as an important influence on their sustained engagement with it. This perception was inferred from people's desire for apps to contain clear user guidance, require less cognitive load, and support easy self-monitoring. Tailored technology and peer and professional support may influence users' opportunity to engage with an app, and goal setting with action planning may play a key role in the motivation to engage with an app.

Acknowledgments

The authors thank all participants for their contribution to the study. The authors would like to acknowledge the support for the development of a topic guide received from the patient and public representatives (the Deputy Director of Public Health England [PHE] Digital, PHE Strategy and Planning, and PHE Strategy and Innovation lead). The authors are grateful to the University College London Tobacco and Alcohol Research Group for their expert opinions on data analysis.

Authors' Contributions

DS, FN, AJ, TC, and JB conceptualized the study design. DS wrote the study protocol, with contributions from FN, AJ, TC, and JB. All the authors commented on the topic guide. DS conducted stakeholder communication with patient and public involvement representatives and Public Health England. DS undertook recruitment of participants, data collection, data analysis, interpretation, and report writing. OP double-coded a proportion of the transcripts. DS, OP, and FN finalized the final thematic framework. DS prepared the manuscript. All the authors read, commented on, and contributed to the final manuscript.

Conflicts of Interest

JB has received unrestricted research funding to study smoking cessation from pharmaceutical companies that manufacture smoking cessation medications. JB, FN, OP, and DS are unpaid members of the scientific committee for the Smoke Free app and have no financial interest in the app. DS is funded through a PhD studentship, provided jointly by Public Health England and the University of East Anglia. OP receives salary support from the Cancer Research UK (C1417/A22962).

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 151 KB - mhealth_v9i12e29098_app1.pdf]

Multimedia Appendix 2
Screening questionnaire.
[PDF File (Adobe PDF File), 45 KB - mhealth_v9i12e29098_app2.pdf]

Multimedia Appendix 3
Topic guide.
[PDF File (Adobe PDF File), 135 KB - mhealth_v9i12e29098_app3.pdf]

Multimedia Appendix 4
Characteristics of the included participants.
[PDF File (Adobe PDF File), 105 KB - mhealth_v9i12e29098_app4.pdf]

References

1. Global status report on noncommunicable diseases. World Health Organisation. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf;jsessionid=CCD448C9B5F5912D0A1FABED56A20901?sequence=1 [accessed 2020-06-01]
2. Buttery AK, Mensink GB, Busch MA. Healthy behaviours and mental health: findings from the German Health Update (GEDA). Eur J Public Health 2015 Apr;25(2):219-225. [doi: 10.1093/eurpub/cku094] [Medline: 25061231]
3. Vaingankar JA, Chong SA, Abdin E, SivaKumar FD, Chua BY, Sambasivam R, et al. Understanding the relationships between mental disorders, self-reported health outcomes and positive mental health: findings from a national survey. Health Qual Life Outcomes 2020 Mar 04;18(1):55 [FREE Full text] [doi: 10.1186/s12955-020-01308-0] [Medline: 32131837]
4. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
5. Flores MG, 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;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]

6. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
7. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1) [FREE Full text] [Medline: 27034992]
8. Nour M, Chen J, Allman-Farinelli M. Efficacy and external validity of electronic and mobile phone-based interventions promoting vegetable intake in young adults: systematic review and meta-analysis. J Med Internet Res 2016 Apr 08;18(4):e58 [FREE Full text] [doi: 10.2196/jmir.5082] [Medline: 27059765]
9. Lee S, Lindquist R. A review of technology-based interventions to maintain weight loss. Telemed J E Health 2015 Mar;21(3):217-232. [doi: 10.1089/tmj.2014.0052] [Medline: 25692454]
10. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, 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]
11. Choo CC, Burton AA. Mobile phone apps for behavioral interventions for at-risk drinkers in australia: literature review. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e18 [FREE Full text] [doi: 10.2196/mhealth.6832] [Medline: 29439946]
12. Husain I, Spence D. Can healthy people benefit from health apps? Br Med J 2015 Apr 14;350:h1887. [doi: 10.1136/bmj.h1887] [Medline: 25873345]
13. Coughlin S. On the need for research-tested smartphone applications for reducing exposures to known or suspected breast carcinogens in work and home environments. J Environ Health Sci 2015;1(4). [doi: 10.15436/2378-6841.15.e004]
14. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
15. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
16. 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]
17. Kim Y, Oh B, Shin H. Effect of mhealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 21;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]
18. 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]
19. Mobile customer retention report. Braze. 2016. URL: https://www.braze.com/blog/app-customer-retention-spring-2016-report/ [accessed 2020-06-01]
20. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 Mar 23 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
21. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
22. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2017 Jan 01:1357633X17692722. [doi: 10.1177/1357633X17692722] [Medline: 28181859]
23. Hickey E, McMillan B, Mitchell C. Practitioners should embrace, not ignore, health apps. Br Med J 2015 May 07;350:h2336. [doi: 10.1136/bmj.h2336] [Medline: 25953643]
24. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
25. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: a mixed-methods study. Digit Health 2018;4:2055207618785841 [FREE Full text] [doi: 10.1177/2055207618785841] [Medline: 31463077]
26. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Intreventions. London, UK: Silverback Publishing; 2014.
27. 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]

28. 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]
29. Szinay D, Naughton F, Jones A, Chadborn T, Brown J, Perski O. A qualitative study exploring people's perception of factors influencing the uptake and use of health and wellbeing smartphone apps. OSF. 2019. URL: https://osf.io/jrkd3/ [accessed 2021-12-04]
30. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of health and well-being apps and curated app portals: think-aloud and interview study. JMIR Mhealth Uhealth 2021 Apr 27;9(4):e27173 [FREE Full text] [doi: 10.2196/27173] [Medline: 33904827]
31. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
32. Benoot C, Hannes K, Bilsen J. The use of purposeful sampling in a qualitative evidence synthesis: a worked example on sexual adjustment to a cancer trajectory. BMC Med Res Methodol 2016 Feb 18;16(1):21 [FREE Full text] [doi: 10.1186/s12874-016-0114-6] [Medline: 26891718]
33. Ritchie J, Lewis J, Nicholls CM, Ormston R. Qualitative research practice: a guide for social science students and researchers. Choice Reviews Online 2003 Nov 01;41(03). [doi: 10.5860/choice.41-1319]
34. Given LM. The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, CA: SAGE Publications Inc; 2008:1-1072.
35. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 2016 Jun 22:1802-1811. [doi: 10.1177/1049732316654870] [Medline: 27340178]
36. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Intervent 2014 Oct;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
37. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
38. Verma M, Hontecillas R, Tubau-Juni N, Abedi V, Bassaganya-Riera J. Challenges in personalized nutrition and health. Front Nutr 2018;5:117 [FREE Full text] [doi: 10.3389/fnut.2018.00117] [Medline: 30555829]
39. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
40. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
41. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mHealth app: process evaluation of the carrot rewards app. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
42. Lerner JS, Li Y, Valdesolo P, Kassam KS. Emotion and decision making. Annu Rev Psychol 2015 Jan 03;66:799-823. [doi: 10.1146/annurev-psych-010213-115043] [Medline: 25251484]
43. van MT. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
44. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]

Abbreviations

AI: artificial intelligence
COM-B: Capability, Opportunity, Motivation–Behaviour
PHE: Public Health England
TDF: Theoretical Domains Framework

Edited by L Buis; submitted 26.03.21; peer-reviewed by L Morrison, A Graham, S Haug, I Gottgens; comments to author 09.08.21; revised version received 13.09.21; accepted 11.11.21; published 16.12.21.

Please cite as:
Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F
Perceptions of Factors Influencing Engagement With Health and Well-being Apps in the United Kingdom: Qualitative Interview Study
JMIR Mhealth Uhealth 2021;9(12):e29098
URL: https://mhealth.jmir.org/2021/12/e29098
doi:10.2196/29098
PMID:34927597

Review

The Application of Human-Centered Design Approaches in Health Research and Innovation: A Narrative Review of Current Practices

Irene Göttgens[1*], MSc; Sabine Oertelt-Prigione[1*], MD, PhD

Department of Primary and Community Care, Radboud University Medical Center, Nijmegen, Netherlands

[*]all authors contributed equally

Corresponding Author:
Irene Göttgens, MSc
Department of Primary and Community Care
Radboud University Medical Center
Postbus 9101
Nijmegen, 6500 HB
Netherlands
Phone: 31 0243613110
Email: irene.gottgens@radboudumc.nl

Abstract

Background: Human-centered design (HCD) approaches to health care strive to support the development of innovative, effective, and person-centered solutions for health care. Although their use is increasing, there is no integral overview describing the details of HCD methods in health innovations.

Objective: This review aims to explore the current practices of HCD approaches for the development of health innovations, with the aim of providing an overview of the applied methods for participatory and HCD processes and highlighting their shortcomings for further research.

Methods: A narrative review of health research was conducted based on systematic electronic searches in the PubMed, CINAHL, Embase, Cochrane Library, Web of Science, PsycINFO, and Sociological Abstracts (2000-2020) databases using keywords related to *human-centered design*, *design thinking* (DT), and *user-centered design* (UCD). Abstracts and full-text articles were screened by 2 reviewers independently based on predefined inclusion criteria. Data extraction focused on the methodology used throughout the research process, the choice of methods in different phases of the innovation cycle, and the level of engagement of end users.

Results: This review summarizes the application of HCD practices across various areas of health innovation. All approaches prioritized the user's needs and the participatory and iterative nature of the design process. The design processes comprised several design cycles during which multiple qualitative and quantitative methods were used in combination with specific design methods. HCD- and DT-based research primarily targeted understanding the research context and defining the problem, whereas UCD-based work focused mainly on the direct generation of solutions. Although UCD approaches involved end users primarily as testers and informants, HCD and DT approaches involved end users most often as design partners.

Conclusions: We have provided an overview of the currently applied methodologies and HCD guidelines to assist health care professionals and design researchers in their methodological choices. HCD-based techniques are challenging to evaluate using traditional biomedical research methods. Previously proposed reporting guidelines are a step forward but would require a level of detail that is incompatible with the current publishing landscape. Hence, further development is needed in this area. Special focus should be placed on the congruence between the chosen methods, design strategy, and achievable outcomes. Furthermore, power dimensions, agency, and intersectionality need to be considered in co-design sessions with multiple stakeholders, especially when including vulnerable groups.

(JMIR Mhealth Uhealth 2021;9(12):e28102) doi:10.2196/28102

KEYWORDS

human-centered design; design thinking; user-centered design; design-based research; methodology; review; mobile phone

Introduction

Background

Health systems are experiencing a progressive imbalance between available resources and increasing needs. The world population is growing, and the incidence of chronic diseases is rising; however, the funds allocated to health care are limited [1,2]. The need to provide optimized, individualized, and person-centered care is growing. Addressing these competing needs and complex problems requires novel and creative approaches for the development of health care solutions. Design approaches to health care promise to aid the development of innovative, effective, and person-centered solutions to health challenges, supporting the realization of a future for health care that is preventative, personalized, and participatory in nature [3,4]. Different medical disciplines are increasingly applying human-centered design (HCD) to a range of complex questions, from process optimization to product design and social innovation [5-7]. HCD is often described as an iterative, collaborative, and people-centered approach for designing products, services, and systems and is argued to be particularly well-suited for solving complex challenges [8]. In recent years, a growing number of health care professionals have applied HCD to develop person-centered health care solutions in collaboration with patients [9]. For example, the Department of Obstetrics and Gynecology at Mayo Clinic used HCD to develop a new prenatal care model designed to demedicalize a healthy pregnancy experience [10]. By enabling women to meaningfully participate in the process through the use of self-measurement tools, their levels of engagement, sense of control, confidence, and reassurance significantly increased. Another example is the nurse-led quality improvement project at Kaiser Permanente Northern California. HCD principles were used for a patient-centered approach to improve inpatient pain management. The experiences of frontline nurses, patients, and managers were collected, evaluated, and applied to improve the care experience of patients and the work experience of care providers [11].

However, the application of HCD beyond the design sector and its adoption in health research is still in its infancy [4,12]. The number of HCD studies that describe a full project cycle is limited, and even fewer publications focus on the evaluation of research projects that use HCD [13]. A recent scoping review on the application of HCD in global health provided a first overview of its application and health outcomes in public health. The review concluded that increased methodological rigor in the application and reporting of HCD is needed to allow for more acceptance and integration of design practices into research and development [13,14]. However, currently, there is no integral collection of HCD approaches and methods used in the development of health innovations. We performed this review to fill this gap.

HCD evolved from the collaborative design movement and covers a range of overlapping collaborative processes and techniques such as, and not limited to, participatory design, ethnography, cocreation, contextual design, co-design, and empathic design. These processes share several principles: the active involvement of users, an iterative design process, and the organization of multidisciplinary teamwork [15-17]. The term HCD, as a collaborative multimethod approach, is often used interchangeably with terms such as *design thinking* (DT) or *user-centered design* (UCD) because of their similar design philosophies. DT is an approach that prioritizes developing empathy for users, working in collaborative multidisciplinary teams, and using an iterative process with *rapid prototyping* techniques for potential solutions [18]. Similarly, UCD, although deeply rooted in human-computer interactions, is described as both a philosophy and a set of methods in which end users actively influence and are involved in the design process [13]. As these principles are akin to those of HCD, this review includes both DT and UCD as variations that apply HCD principles to further explore their similarities and differences.

Objective

In this review, we systematically explore the following question: how is HCD, and the closely related approaches of DT and UCD, applied in the development of innovations for health research? We specifically focus on the applied research methodologies and design methods used throughout the study. We investigate the level of engagement of end users during the HCD design processes. As a result, we provide an overview of the current application practices of HCD in health research and a practice-oriented collection of the used design methods to aid future researchers in their choice of methodology.

Methods

Overview

A total of 2 librarians, 1 from medical sciences and 1 from social sciences, assisted with the development of a search strategy and the selection of the appropriate databases. Our research included health research related to biomedical, nursing, and allied health and public health sciences. We performed multiple test runs to optimize the search strategy before the first search in July 2019. A final search was performed in August 2020 to update the included publications. The protocol for this review can be found in Multimedia Appendix 1.

Search Strategy

We performed electronic searches in the following databases: PubMed, CINAHL, Embase, Cochrane Library, Web of Science, PsycINFO, and Sociological Abstracts. Gray literature searches were not included. We searched for studies in the English language that were published between 2000 and 2020. For medical databases, the following terms were used: *Human-centered* OR *Human-centred* OR *User-centered* OR *User-centred* AND *Design OR approach* OR *Design thinking.* For nonmedical databases, the following search terms were added: *Health* OR *Medic* OR *Clinic*. The exact search algorithms per database can be found in Multimedia Appendix 2.

Eligibility Criteria

We included health research studies that applied HCD, UCD, or DT; focused on the development process of a health innovation; and provided a detailed description of the design process, which included the applied process steps or phases,

the applied design methods per process step or phase, and a description of the involved design team and end users. We excluded studies if they did not focus on the design process and did not provide a detailed description of the design processes and the HCD, DT, or UCD methods used in the study. No specific criteria were formulated related to the end user population.

We conceptualized a *health innovation* as it is applied within the context of health research according to the World Health Organization concept of "Health innovation identifies new or improved health policies, systems, products and technologies, and services and delivery methods that improve people's health and wellbeing."

Screening and Data Extraction

We downloaded relevant papers on the Endnote bibliographic software (Clarivate Analytics) and removed duplicates. We then uploaded the Endnote database with the remaining papers on Rayyan, a web application that supports the initial screening of publication titles and abstracts [19]. A total of 2 reviewers independently screened the titles and abstracts for inclusion eligibility and subsequently screened the full-text articles independently for inclusion. We resolved disagreements through discussions. To determine the level of agreement, both Cohen κ value and the percentage of agreement were calculated.

Data Retrieval and Analysis

We conducted a stepwise analysis of the included publications, focusing on (1) *study characteristics*, including *design phases and methods*, (2) *level of end user involvement*, and (3) *quality assessment.*

Study Characteristics

We extracted the following data from each article: year of publication, first author, title of the study, aim of the study, end user of the innovation, type of innovation, study design, design approach, design approach reference, design process phases, applied research and design methods, and the design-based problem-solving strategy.

For the classification of the applied *qualitative and quantitative research and design methods*, research methods were defined as "methods traditionally used within scientific research, oriented towards understanding" and design methods were defined as "methods not traditionally used with scientific research, oriented towards action or solution creation for defined problems" [20,21]. These distinctions were made based on the discussions between the authors. To define the design-based problem-solving strategy, we used the categories of problem-focused strategy (PFS) versus solution-focused strategy (SFS). Studies that use a PFS aim to define or reframe the problem before formulating possible solutions. Studies that use an SFS approach focus on the development of a predefined solution, investing little time in defining or reframing the problem [22].

Level of Involvement of the End User

To define the level of engagement of the end user, we adopted a modified framework proposed by Druin [23], which was originally used to categorize the participating role of children in a design process. The participating roles were *users*, *testers*, *informants*, or *design partners*, with increased levels of involvement for each role. *Users* help researchers and designers understand the problem context and user needs. The role of *testers* builds upon this role by including end users as part of the initial or functional prototype testing. In the role of *informants*, end users are involved during various stages of the design process, and they contribute to idea generation and provide feedback on the initial and functional prototypes. In the role of *design partners*, end users are considered equal partners of the design team and are involved at all stages of the design process and fully included during the decision-making processes.

Quality Assessment

We assessed the quality of reporting and analysis of the study designs using the Mixed Methods Appraisal Tool (MMAT), which allows for the appraisal of studies for literature reviews that include qualitative, quantitative, and mixed methods studies [24]. As most HCD studies apply a multimethod approach, we considered this tool fit for purpose. The MMAT contains 2 general screening questions and 5 study design–specific criteria for assessing quantitative and qualitative studies. For mixed methods studies, we applied both sets of criteria, in addition to 5 specific mixed methods criteria. The scores per item could vary between *yes* (criterion is met), *no* (criterion is not met), and *can't tell* (paper did not report appropriate information to rate this criterion).

One of the authors first performed the data retrieval and conducted the stepwise analysis described above. Subsequently, both authors reviewed and discussed the results.

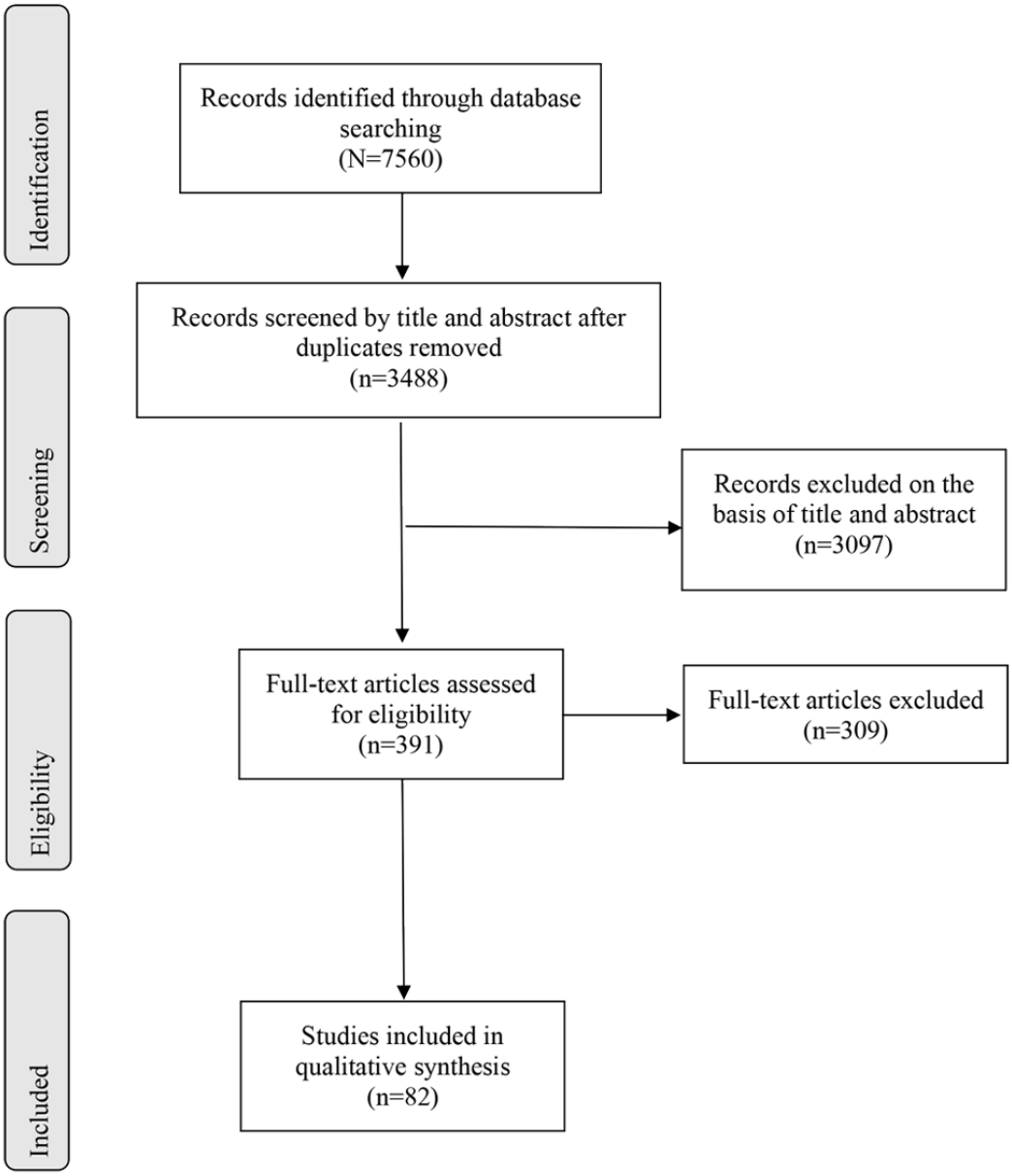
Results

Overview

In the following sections, we have provided an overview of the literature search results and the study characteristics of the included studies. Subsequently, several aspects of the studies have been highlighted, including the applied design theories, guidelines, strategies, and design process steps. Furthermore, we evaluated the applied research and design methods and the role in which end users were involved throughout the studies.

Our literature search identified 7560 records. Of the 7560 papers, after the removal of 4072 (53.86%) duplicates and exclusions on the basis of abstract for 3097 (40.97%) papers and full text for 309 (4.09%) papers, 82 (1.08%) articles were included in the final analysis (Figure 1). Interrater agreement on the inclusion and exclusion of the studies was 96%, with Cohen κ=0.81.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the screening process.



Study Characteristics

Of the 82 included papers, 68 (83%) were published between 2015 and 2020. Most studies originated from the United States (34/82, 41%), the Netherlands (7/82, 9%), the United Kingdom (6/82, 7%), and Canada (6/82, 7%). Most studies focused on patients as the end users and developed health innovations with a focus on improving patient care. The most common type of study design was a mixed methods strategy (47/82, 57%), that is, a combination of qualitative, quantitative, and design methods. Of the 82 studies, 33 (40%) combined only qualitative methods with design methods. A detailed overview of the study characteristics is presented in Table 1.

Table 1. Characteristics of the included studies.

Author	Country	Title	End user population	Innovation type	Design approach	Study design
Bae et al [25]	Korea	Development of a user-centered health information service system for depressive symptom management	Patients who experience depression	Web-based system	User-centered system development	Mixed methods
Birnie et al [26]	Canada	ICanCope PostOp: user-centered design of a smartphone-based app for self-management of postoperative pain in children and adolescents	Children and adolescents who have recently undergone any type of day surgery	Pain self-management app	UCD[a]	Mixed methods
Brox et al [27]	Norway	User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design	Seniors	Serious game	UCD	Mixed methods
Cairns et al [28]	United Kingdom	Rethinking the foam cosmesis for people with lower limb absence	People with lower limb absence	Foam cosmesis for prosthetic limbs	User-centered product design	Mixed methods
Carey-Smith et al [29]	United Kingdom	A user-centered design process to develop technology to improve sleep quality in residential care homes	Older people with sleep or wake pattern disturbance	Sleep improvement technology	UCD	Qualitative
Caro et al [30]	Mexico	FroggyBobby: an exergame to support children with motor problems practicing motor coordination exercises during therapeutic interventions	Children with motor coordination problems	Exergames for children with motor problems	UCD	Qualitative
Catalani et al [31]	Kenya	A clinical decision support system for integrating tuberculosis and HIV care in Kenya: a human-centered design approach	HIV clinical care providers	Clinical shared decision support system	HCD[b]	Mixed methods
Cawood et al [32]	New Zealand	Creating the optimal workspace for hospital staff using human centered design	Hospital staff	Nonclinical workspaces	HCD	Qualitative
Civan-Hartzler et al [33]	United States	Bringing the field into focus: user-centered design of a patient expertise locator	Survivors of breast cancer	Patient expertise locator for web-based health communities	UCD	Qualitative
Connelly et al [34]	United States	Development of an ecological momentary assessment mobile app for a low-literacy, Mexican American population to collect disordered eating behaviors	Mexican-American women	Patient experiences assessment app	User-centered, iterative design	Mixed methods
Crespin et al [35]	Canada	Feasibility of adapting the fundamentals of laparoscopic surgery trainer box to endoscopic skills training tool	Surgeons and gastroenterologists	Laparoscopic surgery training box	UCD	Mixed methods
Curtis et al [36]	United Kingdom	Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app	Parents of children with weight management problems	Healthy eating app	UCD	Mixed methods
Dabbs de Vito et al [37]	United States	User-centered design and interactive health technologies for patients	Patients with a lung transplant	Personal health tracking app	UCD	Mixed methods

Author	Country	Title	End user population	Innovation type	Design approach	Study design
Das and Svanaes [38]	Norway	Human-centered methods in the design of an eHealth solution for patients undergoing weight loss treatment	Patients undergoing weight loss treatment	eHealth solution for weight loss treatment	HCD	Mixed methods
Davies et al [39]	United Kingdom	Recommendations for developing support tools with people suffering from chronic obstructive pulmonary disease: co-design and pilot testing of a mobile health prototype	People with COPD[c]	Mobile app for COPD self-management	User-centered, iterative design	Mixed methods
Dijkstra et al [40]	The Netherlands	Development of ehome, a mobile instrument for reporting, monitoring, and consulting drug-related problems in home care: human-centered design study	Home care nurses, general practitioners, and pharmacists	e-home solution for monitoring and consulting	HCD	Mixed methods
Eberhart et al [41]	United States	Using a human-centered design approach for collaborative decision-making in pediatric asthma care	Parents and children who are dealing with asthma management in a lower income environment	Physical decision-making aids	HCD	Qualitative
Erol Barkana and Açik [42]	Turkey	Improvement of design of a surgical interface using an eye tracking device	Surgeons who perform kidney tumor cryoablations	Eye-tracking device	UCD	Qualitative
Erwin et al [43]	United States	Development of a framework and tool to facilitate cost-of-care conversations with patients during prenatal care	Patients receiving prenatal care	Conversation framework	HCD	Qualitative
Ettinger et al [44]	South Africa	Building quality mHealth[d] for low resource settings	Community health care workers	mHealth app to inform clinical decision-making	HCD	Mixed methods
Fabri et al [45]	United Kingdom	Using design thinking to engage autistic students in participatory design of an online toolkit to help with transition into higher education	Students with autism	Web-based toolkit	DT[e]	Mixed methods
Farinango et al [46]	Colombia	Human-centered design of a personal health record system for metabolic syndrome management based on the ISO 9241-210:2010 standard	Individuals at risk for metabolic syndrome	Personal health record system	HCD	Mixed methods
Ferris and Shepley [47]	United States	The design of neonatal incubators: a systems-oriented, human-centered approach	Infants, medical practitioners, and family members	Neonatal incubators	HCD	Qualitative
Foley et al [48]	United States	Primary care women's health screening: a case study of a community engaged human centered design approach to enhancing the screening process	Women receiving health screening in primary care	Health screening tool	HCD	Qualitative
Fortuna et al [49]	United States	Adapting a psychosocial intervention for smartphone delivery to middle-aged and older adults with serious mental illness	Middle-aged and older adults with serious mental illnesses	Mobile app for medical and psychiatric self-management	UCD	Qualitative

Author	Country	Title	End user population	Innovation type	Design approach	Study design
Furberg et al [50]	United States	A digital decision support tool to enhance decisional capacity for clinical trial consent: design and development	People diagnosed with fragile X syndrome and clinicians	Tablet-based decision support tool	UCD	Qualitative
Gačnik et al [51]	Slovenia	User-centered app design for speech sound disorders interventions with tablet computers	Children with speech-language pathology	App for speech sound disorder therapy	UCD	Mixed methods
Garvelink et al [52]	Canada	Development of a decision guide to support the elderly in decision making about location of care: an iterative, user-centered design	Older adults and their informal caregivers	Decision guide (physical)	UCD	Mixed methods
Garvelink et al [53]	Canada	Deciding how to stay independent at home in later years: development and acceptability testing of an informative web-based module	Seniors with loss of autonomy	Interactive website	UCD	Qualitative
Garvin et al [54]	United States	Descriptive usability study of CirrODS: clinical decision and workflow support tool for management of patients with cirrhosis	Clinicians caring for patients with cirrhosis	Clinical decision and workflow support tool (digital)	UCD	Mixed methods
Garzo et al [55]	France	Design and development of a gait training system for Parkinson's disease	People with Parkinson disease	Gait training app	UCD	Mixed methods
Gaynor et al [56]	United States	A user-centered, learning asthma smartphone application for patients and providers	People with asthma	Mobile app for asthma self-management	UCD	Qualitative
Gill et al [57]	Canada	Feasibility and acceptability of a mobile technology intervention to support post abortion care (The FACTS[f] study phase II) after surgical abortion: user-centered design	Women who underwent an abortion	Web-based intervention for postabortion care support	UCD	Mixed methods
Giunti et al [58]	Spain	More stamina, a gamified mHealth solution for persons with multiple sclerosis: research through design	Young adults who have been diagnosed with multiple sclerosis	mHealth solution	UCD	Qualitative
Godinho et al [59]	Portugal	Improving accessibility of mobile devices with EasyWrite	Motor-disabled persons who experience text-entry difficulties when using mobile devices	Text-entry method for mobile devices	User-centered approach	Mixed methods
Gould et al [60]	United States	Development and refinement of educational materials to help older veterans use VA[g] mental health mobile apps	Older veterans	Educational material for mobile mental health apps	UCD	Mixed methods
Green et al [61]	United States	Tracking care in the emergency department	Emergency department physicians	Emergency department tracking board	UCD	Qualitative
Griffin et al [62]	United States	Creating an mHealth app for colorectal cancer screening: user-centered design approach	People at risk for colorectal cancer aged ≥50 years	mHealth screening solution	UCD	Mixed methods

Author	Country	Title	End user population	Innovation type	Design approach	Study design
Grossman et al [63]	United States	Leveraging patient-reported outcomes using data visualization	Patients with heart failure and health care providers for patients with heart failure	Data visualization	UCD	Mixed methods
Hafiz et al [64]	Denmark	The internet-based cognitive assessment tool: system design and feasibility study	Patients with unipolar and bipolar disorder	Web-based cognitive assessment tool	UCD	Mixed methods
Hardy et al [65]	United Kingdom	How inclusive, user-centered design research can improve psychological therapies for psychosis: development of SlowMo	People who fear harm from others	Digital solution for psychological therapy	UCD	Qualitative
Harte, R. [66]	Ireland	Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users	Older adults with fall risk	Mobile app for fall risk detection	HCD	Mixed methods
Hartlzer et al [67]	United States	Design and feasibility of integrating personalized PRO[h] dashboards into prostate cancer care	Patients following prostate cancer treatment	Patient dashboard	HCD	Mixed methods
Herschman et al [68]	Canada	Development of a smartphone app for adolescents with lupus: a collaborative meeting-based methodology inclusive of a wide range of stakeholders	Adolescents with lupus	Mobile app for adolescents	UCD	N/A[i]
Horsky and Ramelson [69]	United States	Development of a cognitive framework of patient record summary review in the formative phase of user-centered design	Clinicians	Patient record summary review	UCD	Qualitative
Huberty et al [70]	United States	Development and design of an intervention to improve physical activity in pregnant women using Text4baby	Pregnant women	SMS text messaging	UCD	Mixed methods
Isenberg et al [71]	United States	An advance care plan decision support video before major surgery: a patient- and family-centered approach	Patients who are preparing for major surgery	Advance care planning decision support video	HCD	Mixed methods
Johnston et al [72]	United States	Designing and testing a web-based interface for self-monitoring of exercise and symptoms for older adults with COPD	Older adults with COPD	Web-based interface for self-monitoring of exercise	UCD	Mixed methods
Lan Hing Ting et al [73]	France	Examining usage to ensure utility: co-design of a tool for fall prevention	Older adults with fall risk	Balance assessment tool	HCD	Mixed methods
Luna et al [74]	Argentina	User-centered design improves the usability of drug-drug interaction alerts: experimental comparison of interfaces	Physicians	Drug–drug interaction alert system	UCD	Mixed methods
Ma, Wu and Chang [75]	Taiwan	A new design approach of user-centered design on a personal assistive bathing device for hemiplegia	Patients with stroke and hemiplegia	Personal assistive bathing device	UCD	Qualitative

Author	Country	Title	End user population	Innovation type	Design approach	Study design
Madrigal-Cadavid et al [76]	Colombia	Design and development of a mobile app of drug information for people with visual impairment	People with visual impairment	Mobile app for drug information	UCD	Qualitative
Marker and Monzon [77]	United States	Iterative development of a web-based intervention for families of young children with type 1 diabetes: DIPPer academy	Parents of children with type 1 diabetes	Web-based intervention	UCD	Mixed methods
Marko-Holguin et al [78]	United States	A two-way interactive text messaging application for low-income patients with chronic medical conditions: design-thinking development approach	Patients with low income and chronic medical conditions	Interactive SMS text messaging app	DT	Mixed methods
Martin et al [79]	Ireland	A qualitative study adopting a user-centered approach to design and validate a brain computer interface for cognitive rehabilitation for people with brain injury	People with brain injury	Brain-computer interface	UCD	Qualitative
McGinn et al [80]	Ireland	A human-oriented framework for developing assistive service robots	People with disabilities	Assistive service robot	HCD	Qualitative
McMullen et al [81]	United States	Designing for impact: identifying stakeholder-driven interventions to support recovery after major cancer surgery	Patients who recover from major cancer surgery	Web-based educational platform for patients	UCD	Qualitative
Melnick et al [82]	United States	Patient-centered decision support: formative usability evaluation of integrated clinical decision support with a patient decision aid for minor head injury in the emergency department	Emergency department physicians	Electronic clinical decision support	UCD	Mixed methods
Nunez-Nava et al [83]	Colombia	Human-centered development of an online social network for metabolic syndrome management	People with metabolic syndrome	Web-based social network	HCD	Mixed methods
Person et al [84]	Tanzania	Community co-designed schistosomiasis control interventions for school-aged children in Zanzibar	School-aged children	Intervention to reduce schistosomiasis transmission	HCD	Qualitative
Petersen, and Hempler [85]	Denmark	Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study	People with newly diagnosed type 2 diabetes	Mobile app for newly diagnosed patients with type 2 diabetes	DT	Qualitative
Ragouzeos et al [86]	United States	*Am I OK?* using human centered design to empower rheumatoid arthritis patients through patient reported outcomes	Patient with rheumatoid arthritis	Dashboard to display PROs	HCD	Qualitative

Author	Country	Title	End user population	Innovation type	Design approach	Study design
Ray et al [87]	United States	Computerized clinical decision support system for emergency department–initiated buprenorphine for opioid use disorder: user-centered design	Emergency department physicians	Computerized clinical decision support system	UCD	Qualitative
Rothgangel et al [88]	The Netherlands	Design and development of a telerehabilitation platform for patients with phantom limb pain: a user-centered approach	Patients with phantom limb pain	Tele-rehabilitation platform	UCD	Mixed methods
Salmon et al [89]	Congo	Alternative ultrasound gel for a sustainable ultrasound program: application of human centered design	Local clinicians who use point of care ultrasound	Alternative ultrasound gel	HCD	Mixed methods
Schild et al [90]	Germany	A digital cognitive aid for anesthesia to support intraoperative crisis management: results of the user-centered design process	Anesthesiologists	Digital cognitive aid for intraoperative crisis management	UCD	Mixed methods
Sedlmayr et al [91]	Germany	User-centered design of a mobile medication management	People who use medication	Mobile interface for medication management	UCD	Mixed methods
Seeber et al [92]	Germany	A design thinking approach to effective vaccine safety communication	Parents and babies	Effective vaccine safety communication	DT	Qualitative
Sonney et al [93]	United States	Applying human-centered design to the development of an asthma essentials kit for school aged children and their parents	School-aged children and their parents who deal with asthma management	Asthma essential kit	HCD	Qualitative
Srinivas et al [94]	United States	Context-sensitive ecologic momentary assessment: application of user-centered design for improving user satisfaction and engagement during self-report	Middle-aged women with obesity	Patients' experiences assessment app	UCD	Mixed methods
Stevens et al [95]	The Netherlands	The development of a patient-specific method for physiotherapy goal setting: a user-centered design	Physiotherapists and patients	A new method for goal setting	UCD	Qualitative
Taylor et al [96]	United States	User-centered development of a web-based preschool vision screening tool	Parents of preschool-aged children with amblyopia	Web-based vision screening tool	UCD	Mixed methods
Timmerman et al [97]	The Netherlands	Cocreation of an ICT[j]-supported cancer rehabilitation application for resected lung cancer survivors: design and evaluation	Health care professionals and survivors of lung cancer	ICT-supported cancer rehabilitation program	UCD	Mixed methods
Tucker Edmonds et al [98]	United States	Creation of a decision support tool for expectant parents facing threatened periviable delivery: application of a user-centered design approach	Prospective parents	Decision support tool	UCD	N/A

Author	Country	Title	End user population	Innovation type	Design approach	Study design
van der Weegen et al [99]	The Netherlands	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	People with chronic disease	Mobile monitoring and feedback tool	UCD	Qualitative
Vechakul et al [100]	United States	Human-centered design as an approach for place-based innovation in public health: a case study from Oakland, California	Citizens of Castlemont neighborhood	Novel programs to reduce inequities in infant mortality rates	HCD	Qualitative
Vermeulen et al [101]	The Netherlands	User-centered development and testing of a monitoring system that provides feedback regarding physical functioning to elderly people	Older adults	Mobile interface for a monitoring system	User-centered development process	Mixed methods
Vilardaga et al [102]	United States	User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness	People with serious mental illnesses who smoke	Smoking cessation app	UCD	Mixed methods
Wachtler et al [103]	Australia	Development of a mobile clinical prediction tool to estimate future depression severity and guide treatment in primary care: user-centered design	People with depressive symptoms	App for improvement of treatment allocation for depression	UCD	Qualitative
Willard et al [104]	The Netherlands	Development and testing of an online community care platform for frail older adults in The Netherlands: a user-centered design	Frail older adults	Web-based community platform	UCD	Mixed methods
Woodard et al [105]	United States	The Pathways fertility preservation decision aid website for women with cancer: development and field testing	Women survivors of cancer	Decision aid website for young women with cancer	UCD	Mixed methods
Wysocki et al [106]	United States	A web-based coping intervention by and for parents of very young children with type 1 diabetes: user-centered design	Parents of young children with type 1 diabetes	Web-based coping resource	UCD	Qualitative

[a]UCD: user-centered design.

[b]HCD: human-centered design.

[c]COPD: chronic obstructive pulmonary disease.

[d]mHealth: mobile health.

[e]DT: design thinking.

[f]FACTS: factors affecting combination trial success.

[g]VA: veterans affairs.

[h]PRO: patient-reported outcome.

[i]N/A: not applicable.

[j]ICT: information and communication technology.

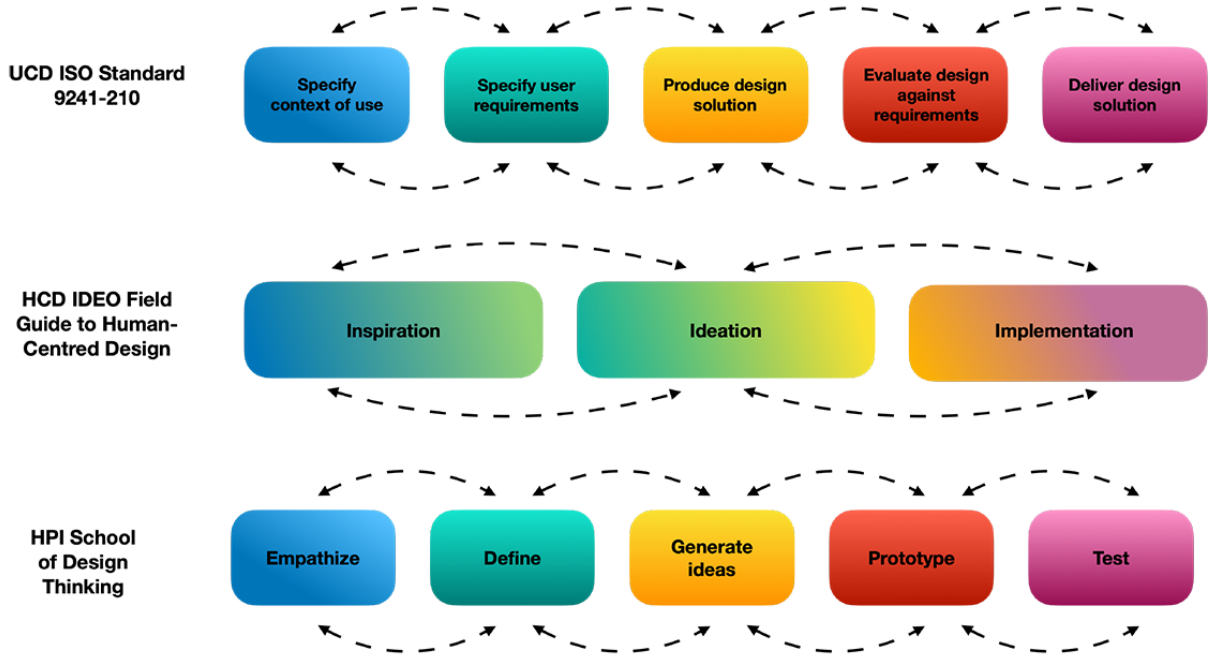
Design Theories and Methodologies

This review explores the various applications of HCD approaches, including HCD, UCD, and DT. Of the 82 studies, HCD was used in 21 (26%) studies, whereas 4 (4%) studies applied a DT approach. Most (57/82, 70%) used a UCD approach. All approaches prioritized the users' needs and the participatory and iterative nature of the design process. Some HCD definitions included a focus on a multiple stakeholder or system perspective, whereas some UCD definitions aimed at

increasing usability or user friendliness of the solution. These design approaches are generally characterized by the use of different standards or models.

A total of 3 standards or models were frequently mentioned in the studies and used as references. These models overlap in their attempt to classify the distinct phases of the design process but operationalize the steps differently. The UCD ISO Standard 9241-210 for HCD of interactive systems encompasses a 5-phase design process including (1) understanding and specifying the context of use, (2) specifying user requirements, (3) producing design solutions, (4) evaluating design against requirements, and (5) delivering design solutions that meet user requirements. The HCD IDEO Field Guide to Human-Centered Design and the Hasso Plattner Institute (HPI) School of Design Thinking models are characterized by different versions of a similar 3-phase design process: (1) inspiration, (2) ideation, and (3) implementation. Studies that applied DT worked with a multiphase approach that included versions of the following phases: (1) empathizing with stakeholders, (2) defining the problem, (3) generating ideas for solutions, (4) prototyping the solutions, and (5) testing the solutions. In Figure 2, we have

Figure 2. Illustration of human-centered design processes. HCD: human-centered design; HPI: Hasso Plattner Institute; UCD: user-centered design.



illustrated how the different approaches to the HCD process align.

Of the 82 articles identified, 57 (70%) applied a UCD approach, 21 (26%) used HCD, and 4 (5%) used DT. In 17% (14/82) of the studies, the concepts of HCD and UCD were referred to interchangeably; of these 14 studies, 9 (64%) studies referred to the use of the ISO 9241-210 standard. In the 5% (4/82) of studies that applied DT, the concept was used interchangeably with HCD in all cases. These studies referred to the IDEO Field Guide to Human-Centered Design or the HPI School of Design Thinking Guide as standards. For clarity, we have continued to report the results of the HCD and DT studies and UCD studies separately in this review.

Design Strategies and Methods

Of the 82 studies, 74 (90%) applied an SFS versus 8 (10%) applied a PFS to drive the design process. Thus, most design studies focused on directly generating solutions or developing a specific predefined solution. Only a minority used design-based methods to define the problem and selectively gather information before proceeding to solution development. Of the 74 studies that applied an SFS, 55 (74%) applied the UCD approach. Of the 8 studies that applied a PFS, 6 (75%) applied an HCD and DT approach. Overall, HCD and DT appears to be the preferred approach for problem-driven strategies, whereas UCD is generally applied for solution-driven strategies.

The design processes comprised several design cycles during which multiple qualitative and quantitative methods were used in combination with specific design methods. Of the 82 studies, 47 (57%) applied a mixed methods approach, and 33 (40%) applied qualitative methodology. A synthesis of the methods used in the different phases of the included studies is presented in Table 2 (details about the described design methods can be found in Multimedia Appendix 3). The first design phase—understanding the context—was often characterized by the use of a limited range of design-based methods. During the second and third phases—problem specification and idea

generation—a broader range of design methods was used in different studies. In the fourth phase—testing of solutions—the range of design methods was reduced again. Some design-based methods were applied in multiple phases of the process, for example, personas, intervention mapping, or the Wizard of Oz technique; however, most were uniquely used in a single phase.

Overall, qualitative methods or mixed methods were mostly used in the first and last phases of the design process to understand user needs or to evaluate user experiences. In the first phase of the process, qualitative methods such as interviews and observations as well as literature reviews were commonly used to understand the problem context. In later stages, the use of methods diverges based on the type of foreseen solution, for example, digital or nondigital solutions. Quantitative methods were used to either support qualitative findings during the first phase of the process or as an evaluation instrument in the later design phases.

Table 2. Meta-analysis of applied research and design methods.

Design phase	Qualitative methods	Quantitative methods	Design methods
Understanding the context	• Literature review • Observations • Expert meetings • Delphi technique • Diary studies	Surveys (not specified)	• Storytelling • Metaphors • Personas • Experience mapping
Specify the problem or user need	• Focus groups • Interviews • Delphi technique • Contextual inquiry • Observations • Critical incident technique	• Context assessments • Needs assessments • Surveys (not specified)	• Participatory workshop • Personas • Use case scenarios • Decision matrix • MoSCoW[a] method • House of quality analysis • Goal, question, metric approach • Roleplay • User journey mapping • Intervention mapping • System mapping • Low functional prototype • Use case diagram
Generate ideas and design solutions	• Observations • Interviews • Focus groups • Literature review	• Usability surveys • Feasibility surveys • Surveys (not specified)	• Brainstorm • Round Robin Concept Ideation • Voting • Round table discussions • Sketching • Visual mind maps • Idea or concept voting • Storyboarding • User narratives • Use case scenarios • Low functional prototyping • High functional prototyping • Intervention mapping • Heuristic evaluation • Task analysis • SWOT[b] or competitor analysis • User journey map • Wizard of Oz method • Card sorting • Weekly sprints • Think-aloud techniques
Test solutions	• Interviews • Observations • Focus groups • EMA[c]	• Usability surveys • Feasibility surveys • Viability assessments • EMA • Surveys (not specified)	• Low functional prototyping • High functional prototyping • Roleplay • Story boarding • Card sorting • Simulations • Intervention mapping • Cognitive walkthrough • Brainstorm (general) • Heuristic evaluation • Workflow evaluation • Participatory workshop • Wizard of Oz method • Value versus effort matrix • Think-aloud techniques

[a]MoSCoW: must have, should have, could have, won't have.

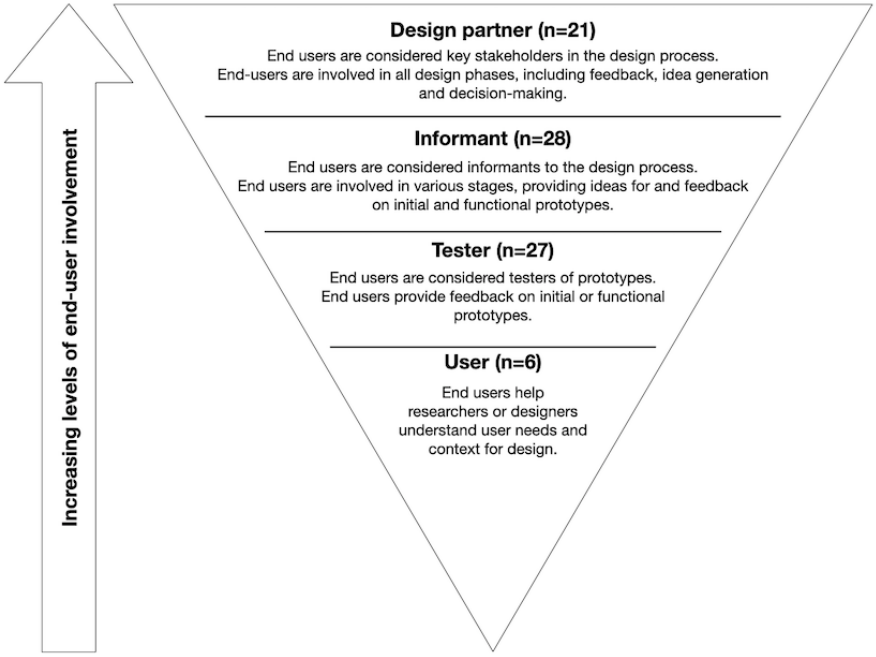
[b]SWOT: strengths, weaknesses, opportunities, and threats.

[c]EMA: ecological momentary assessment.

End User Involvement

In 6 studies (UCD 5/6, 83%; HCD and DT 1/6, 17%), the end users were actively involved as *users*, that is, as information sources but not as active participants in the design process. In 27 (UCD 21/27, 78%; HCD and DT 6/27, 22%) studies, the end users participated as *testers*; that is, they were involved in the first and last phases of the design process as testers of the developed solutions. In 28 (UCD 22/28, 79%; HCT and DT 6/28, 21%) studies, the end users were involved as *informants*. Here, end users were involved in various phases of the process and asked for input on the design prototypes, such as sketches and low-fidelity prototypes. Participation as *design partners*, that is, as contributors to all phases and being involved in the decision-making process, was identified in 21 (UCD 9/21, 43%; HCD and DT 12/21, 57%) studies (Figure 3). Although UCD approaches involved end users primarily in the role of tester (21/57, 37%) and informant (22/57, 39%), HCD and DT approaches involved end users as design partners in 48% (12/57) of the studies.

Figure 3. Levels of end user involvement during human-centered design processes.



Quality Assessment of the Studies

Using the MMAT, 16% (13/82) of the included studies met ≥1 MMAT reporting criteria, based on the study type. The remaining studies had to be rated as unclear on all MMAT reporting criteria. An overview of the quality assessment results can be found in Multimedia Appendix 4 [25-106]. The biggest limitation to the quality assessment was the lack of uniformity in reporting and the broad extent of the design studies that needed to be captured in limited words for publication. In fact, most studies used multiple research and design cycles and generally offered limited details about the applied methodology.

Discussion

Principal Findings

In this review, we explored how different HCD approaches, including DT and UCD, were applied for the development of innovations in health research. Overall, the concepts of HCD and DT, and HCD and UCD, were used interchangeably in 22% (18/82) of the included studies. This applied to all studies that referred to HCD and DT; however, UCD was defined as a standalone entity in 84% (48/57) of the papers that used this approach. Most of the studies using HCD and UCD interchangeably referred to the ISO 9241-210 standard. This aligns with the theoretical framework pursued by the studies, that is, a problem-driven versus a solution-driven strategy. DT- and HCD-based studies commonly engaged in understanding the underlying problem and focused on a broad range of health, social, or medical topics. They often included a focus on human values and a multistakeholder or systems perspective. Instead, UCD-based approaches focused primarily on the direct identification of a solution and were mostly used in health technology innovation. They often focused on human factors to increase the usability or user friendliness of the solution. The limitations of this functional approach in promoting human interests have been previously described as a potential shortcoming of UCD [107].

It has been reported that designers who use a problem-driven design strategy produce solutions with the best balance between quality and creativity [22]. However, in this review, 90% (74/82) of the included studies used a solution-driven strategy. Although the evaluation of solutions can be used to further define the design problem, this was not an objective of the included studies. Their solution-driven approach generally focused on generating a large number of ideas and solutions, potentially leaving the initial design problem ill-defined and ignoring the relationships between various stakeholders. However, health care innovation

could significantly benefit from problem-driven design processes, especially from the perspective of resource efficiency. Innovation in health care is characterized by a development or implementation cost trade-off. Therefore, it is critical that the most impactful innovations be prioritized based on a critical understanding of the underlying problem [108].

HCD in health research is often perceived as a single unitary method, as emphasized by the reference to a single practitioner guideline in the included studies. However, in this review, we found that the application of HCD entails a wide array of design methods and techniques that can be used selectively and that are dependent on the specific design case. Design methods diverge from the traditional methods of academic research as they are primarily oriented toward action or solution of defined problems rather than toward theory and hypotheses building. To date, little is known about their effectiveness according to evidence-based medical standards. The creation of a new product, system, or service to improve health might be considered an outcome from a design perspective but would not be considered a health outcome from a scientific perspective [13,14]. In the literature, a scientific method is described as a strategy to understand the nature of a phenomenon, whereas a design method is a strategy to invent things of value. According to this distinction, science is analytical and design is constructive and it is therefore difficult to assess both methods according to the same standard [20]. However, according to Frey and Dym [109], many of the validation techniques found in medicine can be used for the validation of design methods. For example, where medicine uses animal models and clinical trials to test medical treatments, detailed simulations and controlled field experiments of design methods could be developed for the explicit purpose of evaluating design methodologies [109].

This logical, empirical approach toward the evaluation of design methods fits well with, for example, the field of engineering design, which is based on mathematical modeling, as it is most appropriate for closed, objective problems that can lead to binary (yes or no) answers. However, HCD approaches often address open, complex problems that involve both objective and subjective elements without a single *correct* answer. For design methods addressing open, complex problems, a relativist validation approach that gradually builds confidence in the usefulness of the methods can be considered a more appropriate paradigm [110]. A relativist approach to design claims no absolute objectivity for methods or models; however, it assumes that a valid method or model is only one of the many possible ways of measuring or describing a real situation. In a relativist approach to design methods, validity becomes a matter of practical use and contextual functionality rather than formal and universal accuracy. The validity of design methods becomes a contextual, semiformal, and conversational process, because establishing models of usefulness is a conversational matter [111]. It is important to note that a relativist approach toward the evaluation of design methods does not antagonize the logical, empirical approach toward the evaluation of scientific research methods used in HCD processes.

There is an ongoing demand for the development of a *design science* with systematic and formalized design methods that adhere to the values of the empirical scientific method: objectivity, rationality, and universalism [112,113]. Scientific design methods have been developed in engineering and computer science; however, there is limited evidence that the systematic use of design practices leads to measurable and reproducible results in health research [112]. Design researchers themselves still debate whether design conforms to a scientific activity or represents an academic discipline with a rigorous culture of its own [20,113]. As a result, critical appraisal and best practice selections of design methods in health research remain challenging.

In this review, the diverse reporting formats challenged our ability to assess the quality of the studies from an evidence-based perspective. Although initial guidelines have been proposed to improve the reporting of design studies in health research, this is still an area that is in development [14]. The guidelines by Bazzano et al [14] represent the first detailed overview of reporting items for health research that includes design approaches. Although we acknowledge that this reporting guideline is an important first step toward improving transparency, evaluability, and wider dissemination of design approaches in health research, it is, however, debatable whether the application of these guidelines is feasible in the context of health research manuscripts. The level of detail that the Bazzano [14] guidelines propose implies that the design research component should be reported as a standalone article, separate from the connected empirical studies. Most of the design studies included in this review offered limited details about their multimethod design cycles, possibly because of the word count limits that most scientific journals apply. It would be almost impossible to describe a multimethod design process in adequate detail and also effectively report on the research and design outcomes in a single manuscript. Applying the Bazzano [14] guidelines with rigor is likely to result in the reporting of separate design cycles across multiple manuscripts, and essential findings for the design process might appear fragmented or be lost among reports that are published separately.

However, it could be argued that the separate publication of multiple waves of data collection in design research is preferable for both researchers and reviewers to support the validity, reliability, and reproducibility of design-based health research. Rather than aiming for complex integrated manuscripts, multiple publications would allow researchers to report in more detail on both their methods and findings and also allow for easier critical appraisal and quality assessment by reviewers. In addition to traditional research articles, innovative publication formats such as registered reports could be used to submit design research protocols and results that are judged on their methodological robustness rather than the potential novelty of the findings [114]. We recommend registering the design research protocols in a research registry to address the issue of potential fragmented data publication. This would allow for systematic referencing to previous design activities, even when their results have not been published.

The active engagement of stakeholders is one of the key principles of the HCD approach. Stakeholders can be defined as “individuals, organizations or communities that have a direct interest in the process and outcomes of a project, research or policy endeavor” [115]. In health care innovation, the

engagement of diverse stakeholders is essential to the development of a shared agenda for responsible innovation and for the cocreation of social value [116]. However, a multistakeholder innovation process brings about several challenges. HCD practitioners acknowledge the challenge of equitably including the experience and expertise of all participants in the design process. Although the importance of creative interdisciplinary collaborations between various disciplines in health care is increasing, it is still a relatively new and complex phenomenon [117]. Each stakeholder brings their own motivations, attitudes, priorities, and incentives to the process, and such differences will influence the cocreative space and interpersonal interactions. HCD practitioners should critically reflect on the participatory methods that they intend to apply, considering the possible contribution of each participant in the design process to facilitate the effective use of their expertise and experiences [16]. This is particularly important when working with vulnerable patient groups or health care professionals with limited time to participate in co-design sessions [118].

An earlier study suggested that HCD processes can rely too much on anecdotal evidence of key stakeholders who might not fully understand what they want and need [31]. However, a more strategic application of HCD aims to identify themes that describe people's deeper needs and values rather than their wishes and desires and uses those themes to inform the creation of innovative strategies and solutions [119]. Field studies with the use of qualitative methods, such as observations, to study key stakeholders and their activities in their own environments could offer a valid alternative [120].

In addition, it is essential for HCD practitioners to take power dimensions and the agency of different stakeholders into account, especially during co-design sessions. To achieve inclusive design processes, intersectional aspects should be considered for stakeholder engagement and methodological choices, such as gender identity, class, sexuality, geography, age, and disability and ability [121,122]. Reflective project planning aids and frameworks for involving patients and the public in research and design projects should be used to guarantee meaningful engagement of stakeholders and facilitate democratic design processes [123,124].

Study Limitations

At present, MMAT is the most comprehensive tool available for appraising multimethod studies [125]. Although the MMAT is a tool that allows for the critical appraisal of most common types of study designs, the tool seems less appropriate for HCD, DT, and UCD because of the inclusion of multiple research and design cycles and the often-limited word space to describe the applied methodologies and methodological choices in detail. To our knowledge, there is no appropriate tool available for the critical appraisal of design studies in health research.

In this review, we have only reviewed articles that described the complete development processes of a health innovation. This criterion might have limited the inclusion of studies that describe the complete process through multiple publications. For example, in a few studies, the authors referred to future studies in which they expressed the intention to test a designed solution in a randomized controlled trial. Those studies were not included in this review. Furthermore, no selection criteria for the end user populations were applied. This might have influenced the choice for the use of particular design methods, as design researchers need to take intersectional aspects into account, as mentioned in the *Discussion*. Although this was not the main objective of this review, future research could focus on the application of design methods and their suitability for specific stakeholder populations in health care.

In addition, our search strategy was limited to scientific databases related to biomedical, nursing, and allied health and public health sciences, and gray literature was not included. Disciplines that publish design research related to health systems outside this scope were not considered in our searches. Finally, the existence of different design methods and models with principles related to HCD and the interchangeable use of these terms in the literature made it challenging to scope and perform a fully systematic search.

Conclusions

A wide variety of design practices and methods such as HCD, DT, and UCD are increasingly being applied in health research. In our analysis, HCD- and DT-based projects tended to primarily follow integrated and problem-driven approaches, whereas UCD-based projects engaged in more functional and solution-driven approaches. Most of these design studies used mixed methods approaches, combined qualitative and quantitative research with design methods, and frequently referred to the following 3 design guides: the IDEO Field Guide to Human-Centered Design, the HPI School of Design Thinking Guide, and the ISO Standard 9241-210.

The increasing use of design-based approaches such as HCD and DT and UCD in health research subjects them to evaluation according to traditional biomedical standards. However, the analytic approach of the scientific method versus the constructive approach of the design method impedes the assessment of both methods according to the same standard. To address the validation of design methods, a relativist validation approach that gradually builds confidence in the usefulness of methods could be considered a more appropriate paradigm for design methods, particularly those that are concerned with subjective elements of the design process.

Specific standards for reporting HCD practices in health and biomedical research have been developed in recent years. However, these reporting standards remain challenging to apply for single design research papers because of the extensiveness of multimethod design processes in combination with customary word limits in biomedical publications. Separate publications detailing the multiple waves of data collection in design research might be preferable for both researchers and reviewers to support the validity, reliability, and reproducibility of design-based health research. In addition, innovative publication formats such as registered reports could be used to submit design research protocols and results that are judged on their methodological robustness rather than the potential novelty of the findings. Furthermore, future research on HCD approaches in health should focus on the development of an HCD practitioner guideline for stakeholder engagement that takes stakeholder

roles, experiences, expertise, agency, and power dimensions into account.

Acknowledgments

The authors thank the librarians OY Chan (Radboud University Medical Center Library) and S van Putten (Radboud University Library) for their continued support in formulating and testing the search strategy for this review. The authors also thank M Wijnen (Radboud University Medical Center; Department of Primary and Community Care; Nijmegen, the Netherlands) for her contribution during literature screening.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Review protocol.
[DOC File , 52 KB - mhealth_v9i12e28102_app1.doc]

Multimedia Appendix 2
Search strategy.
[DOC File , 31 KB - mhealth_v9i12e28102_app2.doc]

Multimedia Appendix 3
Descriptions of the design method included in the review.
[DOC File , 53 KB - mhealth_v9i12e28102_app3.doc]

Multimedia Appendix 4
Mixed Methods Appraisal Tool checklist.
[DOC File , 274 KB - mhealth_v9i12e28102_app4.doc]

References

1. Lehnert T, Heider D, Leicht H, Heinrich S, Corrieri S, Luppa M, et al. Review: health care utilization and costs of elderly persons with multiple chronic conditions. Med Care Res Rev 2011 Aug 03;68(4):387-420. [doi: 10.1177/1077558711399580] [Medline: 21813576]
2. McPhail S. Multimorbidity in chronic disease: impact on health care resources and costs. Risk Manag Healthc Policy 2016 Jul;Volume 9:143-156. [doi: 10.2147/rmhp.s97248]
3. Matheson GO, Pacione C, Shultz RK, Klügl M. Leveraging human-centered design in chronic disease prevention. Am J Prev Med 2015 Apr;48(4):472-479. [doi: 10.1016/j.amepre.2014.10.014] [Medline: 25700655]
4. Chan K. A design thinking mindset beyond the public health model. World Med Health Policy 2018;10(1):111-119. [doi: 10.1002/wmh3.253]
5. Vesely R. Applying 'design thinking' to health care organizations. Health Facil Manag 2017 Mar;30(3):10-11. [Medline: 29490448]
6. Valentine L, Kroll T, Bruce F, Lim C, Mountain R. Design thinking for social innovation in health care. Design J 2017 Oct 17;20(6):755-774. [doi: 10.1080/14606925.2017.1372926]
7. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
8. Zuber CD, Moody L. Creativity and innovation in health care: tapping into organizational enablers through human-centered design. Nurs Adm Q 2018;42(1):62-75. [doi: 10.1097/NAQ.0000000000000267] [Medline: 29194334]
9. Kalaichandran A. Design thinking for doctors and nurses. New York Times. 2017. URL: https://www.nytimes.com/2017/08/03/well/live/design-thinking-for-doctors-and-nurses.html [accessed 2021-11-20]
10. de Mooij MJ, Hodny RL, O'Neil DA, Gardner MR, Beaver M, Brown AT, et al. OB nest: reimagining low-risk prenatal care. Mayo Clin Proc 2018 Apr;93(4):458-466. [doi: 10.1016/j.mayocp.2018.01.022] [Medline: 29545005]
11. Trail-Mahan T, Heisler S, Katica M. Quality improvement project to improve patient satisfaction with pain management: using human-centered design. J Nurs Care Qual 2016;31(2):105-113. [doi: 10.1097/NCQ.0000000000000161] [Medline: 26447343]
12. van der Bijl-Brouwer M. The challenges of human-centred design in a public sector innovation context. Design Research Society. 2016. URL: https://dl.designresearchsociety.org/drs-conference-papers/drs2016/researchpapers/125/ [accessed 2021-11-20]

13. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: a scoping review of applications and contexts. PLoS One 2017 Nov;12(11):e0186744 [FREE Full text] [doi: 10.1371/journal.pone.0186744] [Medline: 29091935]
14. Bazzano AN, Yan SD, Martin J, Mulhern E, Brown E, LaFond A, et al. Improving the reporting of health research involving design: a proposed guideline. BMJ Glob Health 2020 Feb 16;5(2):e002248 [FREE Full text] [doi: 10.1136/bmjgh-2019-002248] [Medline: 32133198]
15. Leiying PU. Effect of co-design approach on working efficiency and satisfaction in user support design processes - Co-Design Approach: A Collective Design Method for Improving Working Efficiency and Satisfaction in User Support Design Processes. Master graduation thesis, University of Twente, The Netherlands. 2019. URL: https://essay.utwente.nl/79455/1/Pu_MA_BMS.pdf [accessed 2021-11-21]
16. Steen M. Tensions in human-centred design. CoDesign 2011 Mar;7(1):45-60. [doi: 10.1080/15710882.2011.563314]
17. 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]
18. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
20. Cross N, Naughton J, Walker D. Design method and scientific method. Design Stud 1981 Oct;2(4):195-201. [doi: 10.1016/0142-694x(81)90050-8]
21. Galle P, Kroes P. Science and design: identical twins? Design Stud 2014 May;35(3):201-231. [doi: 10.1016/j.destud.2013.12.002]
22. Kruger C, Cross N. Solution driven versus problem driven design: strategies and outcomes. Design Stud 2006 Sep;27(5):527-548. [doi: 10.1016/j.destud.2006.01.001]
23. Druin A. The role of children in the design of new technology. Behav Inform Technol 2002 Jan;21(1):1-25. [doi: 10.1080/01449290110108659]
24. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
25. Bae J, Wolpin S, Kim E, Lee S, Yoon S, An K. Development of a user-centered health information service system for depressive symptom management. Nurs Health Sci 2009 Jun;11(2):185-193. [doi: 10.1111/j.1442-2018.2009.00454.x] [Medline: 19519707]
26. 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]
27. Brox E, Konstantinidis ST, Evertsen G. User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]
28. Cairns N, Corney J, Murray K, Moore-Millar K, Hatcher GD, Zahedi S, et al. Rethinking the foam cosmesis for people with lower limb absence. Prosthet Orthot Int 2018 Apr;42(2):223-227 [FREE Full text] [doi: 10.1177/0309364617708650] [Medline: 28521576]
29. Carey-Smith B, Evans N, Orpwood R. A user-centred design process to develop technology to improve sleep quality in residential care homes. Technol Disabil 2013 Mar 25;25(1):49-58. [doi: 10.3233/tad-130368]
30. Caro K, Tentori M, Martinez-Garcia AI, Zavala-Ibarra I. FroggyBobby: an exergame to support children with motor problems practicing motor coordination exercises during therapeutic interventions. Comput Hum Behav 2017 Jun;71:479-498. [doi: 10.1016/j.chb.2015.05.055]
31. Catalani C, Green E, Owiti P, Keny A, Diero L, Yeung A, et al. A clinical decision support system for integrating tuberculosis and HIV care in Kenya: a human-centered design approach. PLoS One 2014 Aug;9(8):e103205 [FREE Full text] [doi: 10.1371/journal.pone.0103205] [Medline: 25170939]
32. Cawood T, Saunders E, Drennan C, Cross N, Nicholl D, Kenny A, et al. Creating the optimal workspace for hospital staff using human centred design. Intern Med J 2016 Jul 12;46(7):840-845. [doi: 10.1111/imj.13124] [Medline: 27405891]
33. Civan-Hartzler A, McDonald DW, Powell C, Skeels MM, Mukai M, Pratt W. Bringing the field into focus: user-centered design of a patient expertise locator. Proc SIGCHI Conf Hum Factor Comput Syst 2010 Apr;2010:1675-1684 [FREE Full text] [doi: 10.1145/1753326.1753577] [Medline: 21243114]
34. Connelly K, Stein KF, Chaudry B, Trabold N. Development of an ecological momentary assessment mobile app for a low-literacy, Mexican American population to collect disordered eating behaviors. JMIR Public Health Surveill 2016 Jul 14;2(2):e31 [FREE Full text] [doi: 10.2196/publichealth.5511] [Medline: 27418020]
35. Crespin OM, Okrainec A, Kwong AV, Habaz I, Jimenez MC, Szasz P, et al. Feasibility of adapting the fundamentals of laparoscopic surgery trainer box to endoscopic skills training tool. Surg Endosc 2018 Jun 2;32(6):2968-2983. [doi: 10.1007/s00464-018-6154-8] [Medline: 29611046]

36. 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]
37. De Vito Dabbs A, 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]
38. Das A, Svanæs D. Human-centred methods in the design of an e-health solution for patients undergoing weight loss treatment. Int J Med Inform 2013 Nov;82(11):1075-1091. [doi: 10.1016/j.ijmedinf.2013.06.008] [Medline: 23886482]
39. Davies A, Mueller J, Hennings J, Caress A, Jay C. Recommendations for developing support tools with people suffering from chronic obstructive pulmonary disease: co-design and pilot testing of a mobile health prototype. JMIR Hum Factors 2020 May 15;7(2):e16289 [FREE Full text] [doi: 10.2196/16289] [Medline: 32410730]
40. Dijkstra NE, Sino CG, Heerdink ER, Schuurmans MJ. Development of eHOME, a mobile instrument for reporting, monitoring, and consulting drug-related problems in home care: human-centered design study. JMIR Hum Factors 2018 Mar 07;5(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.8319] [Medline: 29514771]
41. Eberhart A, Slogeris B, Sadreameli S, Jassal M. Using a human-centered design approach for collaborative decision-making in pediatric asthma care. Public Health 2019 May;170:129-132. [doi: 10.1016/j.puhe.2019.03.004] [Medline: 31035123]
42. Barkana DE, Açık A, Duru DG, Duru AD. Improvement of design of a surgical interface using an eye tracking device. Theor Biol Med Model 2014 May 07;11 Suppl 1(Suppl 1):S4 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S4] [Medline: 25080176]
43. Erwin K, Fitzpatrick V, Norell S, Gilliam M. Development of a framework and tool to facilitate cost-of-care conversations with patients during prenatal care. Ann Intern Med 2019 May 07;170(9_Supplement):S62. [doi: 10.7326/m18-2207]
44. Ettinger KM, Pharaoh H, Buckman RY, Conradie H, Karlen W. Building quality mHealth for low resource settings. J Med Eng Technol 2016 Aug 25;40(7-8):431-443. [doi: 10.1080/03091902.2016.1213906] [Medline: 27560713]
45. Fabri M, Andrews PC, Pukki HK. Using design thinking to engage autistic students in participatory design of an online toolkit to help with transition into higher education. J Assist Technol 2016;10(2):102-114. [doi: 10.1108/JAT-02-2016-0008]
46. Farinango C, Benavides J, Cerón J, López D, Álvarez R. Human-centered design of a personal health record system for metabolic syndrome management based on the ISO 9241-210:2010 standard. J Multidiscip Healthc 2018 Jan;Volume 11:21-37. [doi: 10.2147/jmdh.s150976]
47. Ferris TK, Shepley MM. The design of neonatal incubators: a systems-oriented, human-centered approach. J Perinatol 2013 Apr 28;33 Suppl 1(S1):24-31. [doi: 10.1038/jp.2013.11] [Medline: 23536027]
48. Foley KA, Shelton J, Richardson E, Smart N, Smart-McMillan C, Mustakem O, et al. Primary care women's health screening: a case study of a community engaged human centered design approach to enhancing the screening process. Matern Child Health J 2019 Nov 27;23(11):1446-1458. [doi: 10.1007/s10995-019-02802-8] [Medline: 31250241]
49. Fortuna KL, Lohman MC, Gill LE, Bruce ML, Bartels SJ. Adapting a psychosocial intervention for smartphone delivery to middle-aged and older adults with serious mental illness. Am J Geriatr Psychiatry 2017 Aug;25(8):819-828 [FREE Full text] [doi: 10.1016/j.jagp.2016.12.007] [Medline: 28169129]
50. Furberg RD, Ortiz AM, Moultrie RR, Raspa M, Wheeler AC, McCormack LA, et al. A digital decision support tool to enhance decisional capacity for clinical trial consent: design and development. JMIR Res Protoc 2018 Jun 06;7(6):e10525 [FREE Full text] [doi: 10.2196/10525] [Medline: 29875084]
51. Gačnik M, Starčič AI, Zaletelj J, Zajc M. User-centred app design for speech sound disorders interventions with tablet computers. Univ Access Inf Soc 2017 Jul 1;17(4):821-832. [doi: 10.1007/s10209-017-0545-9]
52. 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 Jul 19;2(1):26 [FREE Full text] [doi: 10.1186/s40900-016-0040-0] [Medline: 29062524]
53. 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]
54. Garvin JH, Ducom J, Matheny M, Miller A, Westerman D, Reale C, et al. Descriptive usability study of CirrODS: clinical decision and workflow support tool for management of patients with cirrhosis. JMIR Med Inform 2019 Jul 03;7(3):e13627 [FREE Full text] [doi: 10.2196/13627] [Medline: 31271153]
55. Garzo A, Silva PA, Garay-Vitoria N, Hernandez E, Cullen S, De Cock VC, et al. Design and development of a gait training system for Parkinson's disease. PLoS One 2018 Nov 12;13(11):e0207136 [FREE Full text] [doi: 10.1371/journal.pone.0207136] [Medline: 30418993]
56. Gaynor M, Schneider D, Seltzer M, Crannage E, Barron ML, Waterman J, et al. A user-centered, learning asthma smartphone application for patients and providers. Learn Health Syst 2020 Jul;4(3):e10217 [FREE Full text] [doi: 10.1002/lrh2.10217] [Medline: 32685685]
57. Gill RK, Ogilvie G, Norman WV, Fitzsimmons B, Maher C, Renner R. Feasibility and acceptability of a mobile technology intervention to support postabortion care (The FACTS Study Phase II) after surgical abortion: user-centered design. JMIR Hum Factors 2019 Oct 10;6(4):e14558 [FREE Full text] [doi: 10.2196/14558] [Medline: 31603429]

58. Giunti G, Mylonopoulou V, Rivera Romero O. More stamina, a gamified mHealth solution for persons with multiple sclerosis: research through design. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e51 [FREE Full text] [doi: 10.2196/mhealth.9437] [Medline: 29500159]
59. Godinho R, Condado PA, Zacarias M, Lobo FG. Improving accessibility of mobile devices with EasyWrite. Behav Inform Technol 2014 Dec 03;34(2):135-150. [doi: 10.1080/0144929x.2014.981584]
60. Gould CE, Loup J, Scales AN, Juang C, Carlson C, Ma F, et al. Development and refinement of educational materials to help older veterans use VA mental health mobile apps. Prof Psychol Res Pr 2020 Aug;51(4):414-423 [FREE Full text] [doi: 10.1037/pro0000354] [Medline: 34465939]
61. Green R, Buckley A, Scott B. Tracking care in the emergency department. Ergon Des 2015 Nov 11;23(4):5-12. [doi: 10.1177/1064804615572626]
62. Griffin L, Lee D, Jaisle A, Carek P, George T, Laber E, et al. Creating an mhealth app for colorectal cancer screening: user-centered design approach. JMIR Hum Factors 2019 May 08;6(2):e12700 [FREE Full text] [doi: 10.2196/12700] [Medline: 31066688]
63. Grossman L, Feiner S, Mitchell E, Creber R. Leveraging patient-reported outcomes using data visualization. Appl Clin Inform 2018 Jul 01;9(3):565-575 [FREE Full text] [doi: 10.1055/s-0038-1667041] [Medline: 30068012]
64. Hafiz P, Miskowiak KW, Kessing LV, Jespersen AE, Obenhausen K, Gulyas L, et al. The internet-based cognitive assessment tool: system design and feasibility study. JMIR Form Res 2019 Jul 26;3(3):e13898 [FREE Full text] [doi: 10.2196/13898] [Medline: 31350840]
65. Hardy A, Wojdecka A, West J, Matthews E, Golby C, Ward T, et al. How inclusive, user-centered design research can improve psychological therapies for psychosis: development of SlowMo. JMIR Ment Health 2018 Dec 05;5(4):e11222 [FREE Full text] [doi: 10.2196/11222] [Medline: 30518514]
66. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. JMIR Mhealth Uhealth 2017 May 30;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
67. Hartzler AL, Izard JP, Dalkin BL, Mikles SP, Gore JL. Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. J Am Med Inform Assoc 2016 Jan;23(1):38-47 [FREE Full text] [doi: 10.1093/jamia/ocv101] [Medline: 26260247]
68. Herschman J, Kasenberg T, Levy D, Ruth N, Taberner C, Kaufman M, et al. Development of a smartphone app for adolescents with lupus: a collaborative meeting-based methodology inclusive of a wide range of stakeholders. Rev Panam Salud Publica 2014;35(5-6):471-476. [Medline: 25211579]
69. Horsky J, Ramelson HZ. Development of a cognitive framework of patient record summary review in the formative phase of user-centered design. J Biomed Inform 2016 Dec;64:147-157 [FREE Full text] [doi: 10.1016/j.jbi.2016.10.004] [Medline: 27725292]
70. Huberty J, Rowedder L, Hekler E, Adams M, Hanigan E, McClain D, et al. Development and design of an intervention to improve physical activity in pregnant women using Text4baby. Transl Behav Med 2016 Jun 25;6(2):285-294 [FREE Full text] [doi: 10.1007/s13142-015-0339-7] [Medline: 27356999]
71. Isenberg SR, Crossnohere NL, Patel MI, Conca-Cheng A, Bridges JF, Swoboda SM, et al. An advance care plan decision support video before major surgery: a patient- and family-centred approach. BMJ Support Palliat Care 2018 Jun 18;8(2):229-236. [doi: 10.1136/bmjspcare-2017-001449] [Medline: 29555812]
72. Johnston SK, Nguyen HQ, Wolpin S. Designing and testing a web-based interface for self-monitoring of exercise and symptoms for older adults with chronic obstructive pulmonary disease. Comput Inform Nurs 2009;27(3):166-174 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c1d] [Medline: 19411946]
73. Ting K, Dessinger G, Voilmy D. Examining usage to ensure utility: co-design of a tool for fall prevention. IRBM 2020 Oct;41(5):286-293. [doi: 10.1016/j.irbm.2020.03.001]
74. Luna DR, Lede DA, Otero CM, Risk MR, de Quirós F. User-centered design improves the usability of drug-drug interaction alerts: experimental comparison of interfaces. J Biomed Inform 2017 Feb;66:204-213 [FREE Full text] [doi: 10.1016/j.jbi.2017.01.009] [Medline: 28108211]
75. Ma M, Wu F, Chang R. A new design approach of user-centered design on a personal assistive bathing device for hemiplegia. Disabil Rehabil 2007 Jul 30;29(14):1077-1089. [doi: 10.1080/09638280600949712] [Medline: 17612994]
76. Madrigal-Cadavid J, Amariles P, Pino-Marín D, Granados J, Giraldo N. Design and development of a mobile app of drug information for people with visual impairment. Res Social Adm Pharm 2020 Jan;16(1):62-67. [doi: 10.1016/j.sapharm.2019.02.013] [Medline: 30853509]
77. Marker AM, Monzon AD, Goggin KJ, Clements MA, Patton SR. Iterative development of a web-based intervention for families of young children with type 1 diabetes: DIPPer Academy. Clin Pract Pediatr Psychol 2019 Mar;7(1):20-30 [FREE Full text] [doi: 10.1037/cpp0000263] [Medline: 31662953]
78. Marko-Holguin M, Cordel SL, Van Voorhees BW, Fogel J, Sykes E, Fitzgibbon M, et al. A two-way interactive text messaging application for low-income patients with chronic medical conditions: design-thinking development approach. JMIR Mhealth Uhealth 2019 May 01;7(5):e11833 [FREE Full text] [doi: 10.2196/11833] [Medline: 31042152]

79. Martin S, Armstrong E, Thomson E, Vargiu E, Solà M, Dauwalder S, et al. A qualitative study adopting a user-centered approach to design and validate a brain computer interface for cognitive rehabilitation for people with brain injury. Assist Technol 2018 Jul 14;30(5):233-241. [doi: 10.1080/10400435.2017.1317675] [Medline: 28708963]
80. McGinn C, Cullinan MF, Culleton M, Kelly K. A human-oriented framework for developing assistive service robots. Disabil Rehabil Assist Technol 2018 Apr 24;13(3):293-304. [doi: 10.1080/17483107.2017.1328616] [Medline: 28537814]
81. McMullen C, Nielsen M, Firemark A, Price PM, Nakatani D, Tuthill J, et al. Designing for impact: identifying stakeholder-driven interventions to support recovery after major cancer surgery. Support Care Cancer 2018 Dec 6;26(12):4067-4076 [FREE Full text] [doi: 10.1007/s00520-018-4276-0] [Medline: 29876832]
82. Melnick ER, Hess EP, Guo G, Breslin M, Lopez K, Pavlo AJ, et al. Patient-centered decision support: formative usability evaluation of integrated clinical decision support with a patient decision aid for minor head injury in the emergency department. J Med Internet Res 2017 May 19;19(5):e174 [FREE Full text] [doi: 10.2196/jmir.7846] [Medline: 28526667]
83. Núñez-Nava J, Orozco-Sánchez PA, López DM, Ceron JD, Alvarez-Rosero RE. Human-centered development of an online social network for metabolic syndrome management. In: Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2016:100-104.
84. Person B, Knopp S, Ali SM, A'kadir FM, Khamis AN, Ali JN, et al. Community co-designed schistosomiasis control interventions for school-aged children in Zanzibar. J Biosoc Sci 2016 Jul 18;48(S1):56-73. [doi: 10.1017/s0021932016000067]
85. Petersen M, Hempler NF. Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):91 [FREE Full text] [doi: 10.1186/s12911-017-0493-6] [Medline: 28651639]
86. Ragouzeos D, Gandrup J, Berrean B, Li J, Murphy M, Trupin L, et al. "Am I OK?" using human centered design to empower rheumatoid arthritis patients through patient reported outcomes. Patient Educ Couns 2019 Mar;102(3):503-510 [FREE Full text] [doi: 10.1016/j.pec.2018.10.016] [Medline: 30446358]
87. Ray JM, Ahmed OM, Solad Y, Maleska M, Martel S, Jeffery MM, et al. Computerized clinical decision support system for emergency department-initiated buprenorphine for opioid use disorder: user-centered design. JMIR Hum Factors 2019 Feb 27;6(1):e13121 [FREE Full text] [doi: 10.2196/13121] [Medline: 30810531]
88. Rothgangel A, Braun S, Smeets R, Beurskens A. Design and development of a telerehabilitation platform for patients with phantom limb pain: a user-centered approach. JMIR Rehabil Assist Technol 2017 Feb 15;4(1):e2 [FREE Full text] [doi: 10.2196/rehab.6761] [Medline: 28582249]
89. Salmon M, Salmon C, Bissinger A, Muller MM, Gebreyesus A, Geremew H, et al. Alternative ultrasound gel for a sustainable ultrasound program: application of human centered design. PLoS One 2015 Aug 7;10(8):e0134332 [FREE Full text] [doi: 10.1371/journal.pone.0134332] [Medline: 26252003]
90. Schild S, Sedlmayr B, Schumacher A, Sedlmayr M, Prokosch H, St Pierre M, German Cognitive Aid Working Group. A digital cognitive aid for anesthesia to support intraoperative crisis management: results of the user-centered design process. JMIR Mhealth Uhealth 2019 Apr 29;7(4):e13226 [FREE Full text] [doi: 10.2196/13226] [Medline: 31033445]
91. Sedlmayr B, Schöffler J, Prokosch H, Sedlmayr M. User-centered design of a mobile medication management. Inform Health Soc Care 2019 Mar 05;44(2):152-163. [doi: 10.1080/17538157.2018.1437042] [Medline: 29504838]
92. Seeber L, Michl B, Rundblad G, Trusko B, Schnjakin M, Meinel C, et al. A design thinking approach to effective vaccine safety communication. Curr Drug Saf 2015 Apr 07;10(1):31-40. [doi: 10.2174/1574886310011504071054000] [Medline: 25859673]
93. Sonney J, Duffy M, Hoogerheyde LX, Langhauser E, Teska D. Applying human-centered design to the development of an asthma essentials kit for school-aged children and their parents. J Pediatr Health Care 2019 Mar;33(2):169-177. [doi: 10.1016/j.pedhc.2018.07.008] [Medline: 30228032]
94. Srinivas P, Bodke K, Ofner S, Keith NR, Tu W, Clark DO. Context-sensitive ecological momentary assessment: application of user-centered design for improving user satisfaction and engagement during self-report. JMIR Mhealth Uhealth 2019 Apr 03;7(4):e10894 [FREE Full text] [doi: 10.2196/10894] [Medline: 30942698]
95. Stevens A, Köke A, van der Weijden T, Beurskens A. The development of a patient-specific method for physiotherapy goal setting: a user-centered design. Disabil Rehabil 2018 Aug 13;40(17):2048-2055. [doi: 10.1080/09638288.2017.1325943] [Medline: 28504014]
96. Taylor DP, Bray BE, Staggers N, Olson RJ. User-centered development of a web-based preschool vision screening tool. AMIA Annu Symp Proc 2003:654-658 [FREE Full text] [Medline: 14728254]
97. 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]
98. Edmonds BT, Hoffman SM, Lynch D, Jeffries E, Jenkins K, Wiehe S, et al. Creation of a decision support tool for expectant parents facing threatened periviable delivery: application of a user-centered design approach. Patient 2019 Jun 29;12(3):327-337 [FREE Full text] [doi: 10.1007/s40271-018-0348-y] [Medline: 30488236]
99. 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 02;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]

100. Vechakul J, Shrimali BP, Sandhu JS. Human-centered design as an approach for place-based innovation in public health: a case study from Oakland, California. Matern Child Health J 2015 Dec 22;19(12):2552-2559. [doi: 10.1007/s10995-015-1787-x] [Medline: 26197732]
101. Vermeulen J, Neyens J, Spreeuwenberg M, van Rossum E, Sipers W, Habets H, et al. User-centered development and testing of a monitoring system that provides feedback regarding physical functioning to elderly people. Patient Prefer Adher 2013 Aug:843-854. [doi: 10.2147/ppa.s45897]
102. Vilardaga R, Rizo J, Zeng E, Kientz JA, 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]
103. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a mobile clinical prediction tool to estimate future depression severity and guide treatment in primary care: user-centered design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95 [FREE Full text] [doi: 10.2196/mhealth.9502] [Medline: 29685864]
104. Willard S, Cremers G, Man YP, van Rossum E, Spreeuwenberg M, de Witte L. Development and testing of an online community care platform for frail older adults in the Netherlands: a user-centred design. BMC Geriatr 2018 Apr 07;18(1):87 [FREE Full text] [doi: 10.1186/s12877-018-0774-7] [Medline: 29625562]
105. Woodard TL, Hoffman AS, Covarrubias LA, Holman D, Schover L, Bradford A, et al. The Pathways fertility preservation decision aid website for women with cancer: development and field testing. J Cancer Surviv 2018 Feb 15;12(1):101-114. [doi: 10.1007/s11764-017-0649-5] [Medline: 29034438]
106. Wysocki T, Pierce J, Caldwell C, Aroian K, Miller L, Farless R, et al. A web-based coping intervention by and for parents of very young children with type 1 diabetes: user-centered design. JMIR Diabetes 2018 Dec 17;3(4):e16 [FREE Full text] [doi: 10.2196/diabetes.9926] [Medline: 30559089]
107. Gasson S. Human-centered vs. user-centered approaches to information system design. J Inf Theory Appl 2003;5(2):29-46 [FREE Full text]
108. Soliman E, Mogefors D, Bergmann JH. Problem-driven innovation models for emerging technologies. Health Technol 2020 Jun 23;10(5):1195-1206. [doi: 10.1007/s12553-020-00450-5]
109. Frey DD, Dym CL. Validation of design methods: lessons from medicine. Res Eng Design 2006 May 25;17(1):45-57. [doi: 10.1007/s00163-006-0016-4]
110. Pedersen K, Emblemsvåg J, Bailey R, Allen JK, Mistree F. Validating design methods and research: the validation square. In: Proceedings of the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (ASME 2000). 2000 Presented at: International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (ASME 2000); September 10–13, 2000; Baltimore, Maryland, USA p. 379-390. [doi: 10.1115/DETC2000/DTM-14579]
111. Barlas Y, Carpenter S. Philosophical roots of model validation: two paradigms. Syst Dyn Rev 1990;6(2):148-166 [FREE Full text] [doi: 10.1002/sdr.4260060203]
112. Cross N. Designerly ways of knowing: design discipline versus design science. Design Issues 2001 Jul;17(3):49-55. [doi: 10.1162/074793601750357196]
113. You X, Hands D. A reflection upon Herbert Simon's vision of design in "The Sciences of the Artificial". Design J 2019 May 31;22(sup1):1345-1356. [doi: 10.1080/14606925.2019.1594961]
114. No authors listed. An innovative way to publish. Nature 2019 Jul 23;571(7766):447-447. [doi: 10.1038/d41586-019-02277-1] [Medline: 31337921]
115. Deverka PA, Lavallee DC, Desai PJ, Esmail LC, Ramsey SD, Veenstra DL, et al. Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. J Comp Eff Res 2012 Mar;1(2):181-194 [FREE Full text] [doi: 10.2217/cer.12.7] [Medline: 22707880]
116. Abrishami P, Repping S. Nurturing societal values in and through health innovations comment on "What health system challenges should responsible innovation in health address?". Int J Health Policy Manag 2019 Oct 01;8(10):613-615 [FREE Full text] [doi: 10.15171/ijhpm.2019.57] [Medline: 31657189]
117. Moirano R, Sánchez MA, Štěpánek L. Creative interdisciplinary collaboration: a systematic literature review. Think Skills Creat 2020 Mar;35:100626 [FREE Full text] [doi: 10.1016/j.tsc.2019.100626]
118. Langley J, Bec R, Wheeler G, Partridge R. The need for distributed co-design in healthcare contexts. In: Proceedings of the 5th European International Conference on Design4Health. 2018 Presented at: 5th European International Conference on Design4Health; September 4 – 6, 2018; Sheffield Hallam University, Sheffield, UK p. In Press URL: http://shura.shu.ac.uk/23417/3/Langley%20The%20need%20for%20distributed%20co-design%20in%20healthcare%20contexts.pdf
119. van der Bijl-Brouwer M, Dorst K. Advancing the strategic impact of human-centred design. Design Stud 2017 Nov;53:1-23 [FREE Full text] [doi: 10.1016/j.destud.2017.06.003]
120. Kujala S. User involvement: a review of the benefits and challenges. Behav Inform Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
121. Bauer GR. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. Soc Sci Med 2014 Jun;110:10-17 [FREE Full text] [doi: 10.1016/j.socscimed.2014.03.022] [Medline: 24704889]

122. Erete S, Israni A, Dillahunt T. An intersectional approach to designing in the margins. interactions 2018 Apr 23;25(3):66-69. [doi: 10.1145/3194349]
123. Pratt B. Sharing power in global health research: an ethical toolkit for designing priority-setting processes that meaningfully include communities. Int J Equity Health 2021 May 25;20(1):127 [FREE Full text] [doi: 10.1186/s12939-021-01453-y] [Medline: 34034747]
124. Greenhalgh T, Hinton L, Finlay T, Macfarlane A, Fahy N, Clyde B, et al. Frameworks for supporting patient and public involvement in research: systematic review and co-design pilot. Health Expect 2019 Aug 22;22(4):785-801 [FREE Full text] [doi: 10.1111/hex.12888] [Medline: 31012259]
125. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. J Clin Epidemiol 2019 Jul;111:49-59 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]

Abbreviations

DT: design thinking
HCD: human-centered design
HPI: Hasso Plattner Institute
MMAT: Mixed Methods Appraisal Tool
PFS: problem-focused strategy
SFS: solution-focused strategy
UCD: user-centered design

Edited by L Buis; submitted 21.02.21; peer-reviewed by A Bazzano, I Holeman, L Powell, S Ren, D Szinay; comments to author 10.05.21; revised version received 16.07.21; accepted 21.10.21; published 06.12.21.

Please cite as:
Göttgens I, Oertelt-Prigione S
The Application of Human-Centered Design Approaches in Health Research and Innovation: A Narrative Review of Current Practices
JMIR Mhealth Uhealth 2021;9(12):e28102
URL: https://mhealth.jmir.org/2021/12/e28102
doi:10.2196/28102
PMID:34874893

Original Paper

Using Personalized Anchors to Establish Routine Meditation Practice With a Mobile App: Randomized Controlled Trial

Chad Stecher[1], PhD; Mariah Sullivan[1], MA; Jennifer Huberty[1], PhD

College of Health Solutions, Arizona State University, Phoenix, AZ, United States

Corresponding Author:
Chad Stecher, PhD
College of Health Solutions
Arizona State University
550 North 3rd Str
Phoenix, AZ, 85004
United States
Phone: 1 602 496 0957
Email: chad.stecher@asu.edu

Abstract

Background: Physical and mental health benefits can be attained from persistent, long-term performance of mindfulness meditation with a mobile meditation app, but in general, few mobile health app users persistently engage at a level necessary to attain the corresponding health benefits. Anchoring or pairing meditation with a mobile app to an existing daily routine can establish an unconsciously initiated meditation routine that may improve meditation persistence.

Objective: The purpose of this study was to test the use of either personalized anchors or fixed anchors for establishing a persistent meditation app routine with the mobile app, Calm.

Methods: We conducted a randomized controlled trial and randomly assigned participants to one of 3 study groups: (1) a personalized anchor (PA) group, (2) fixed anchor (FA) group, or (3) control group that did not use the anchoring strategy. All participants received app-delivered reminder messages to meditate for at least 10 minutes a day using the Calm app for an 8-week intervention period, and app usage data continued to be collected for an additional 8-week follow-up period to measure meditation persistence. Baseline, week 8, and week 16 surveys were administered to assess demographics, socioeconomic status, and changes in self-reported habit strength.

Results: A total of 101 participants across the 3 study groups were included in the final analysis: (1) PA (n=56), (2) FA (n=49), and (3) control group (n=62). Participants were predominantly White (83/101, 82.2%), female (77/101, 76.2%), and college educated (ie, bachelor's or graduate degree; 82/101, 81.2%). The FA group had a significantly higher average odds of daily meditation during the intervention (1.14 odds ratio [OR]; 95% CI 1.02-1.33; *P*=.04), and all participants experienced a linear decline in their odds of daily meditation during the 8-week intervention (0.96 OR; 95% CI 0.95-0.96; *P*<.001). Importantly, the FA group showed a significantly smaller decline in the linear trend of their odds of daily meditation during the 8-week follow-up (their daily trend increased by 1.04 OR from their trend during the intervention; 95% CI 1.01-1.06; *P*=.03). Additionally, those who more frequently adhered to their anchoring strategy during the intervention typically used anchors that occurred in the morning and showed a significantly smaller decline in their odds of daily meditation during the 8-week follow-up period (1.13 OR; 95% CI 1.02-1.35; *P*=.007).

Conclusions: The FA group had more persistent meditation with the app, but participants in the FA or PA groups who more frequently adhered to their anchoring strategy during the intervention had the most persistent meditation routines, and almost all of these high anchorers used morning anchors. These findings suggest that the anchoring strategy can create persistent meditation routines with a mobile app. However, future studies should combine anchoring with additional intervention tools (eg, incentives) to help more participants successfully establish an anchored meditation routine.

Trial Registration: ClinicalTrials.gov NCT04378530; https://clinicaltrials.gov/ct2/show/NCT04378530

(JMIR Mhealth Uhealth 2021;9(12):e32794) doi:10.2196/32794

KEYWORDS

mindfulness; meditation; mobile meditation app; behavioral persistence; habit formation; randomized controlled trial; mental health; physical health; app engagement; routine

Introduction

Mindfulness meditation is an evidence-based health behavior regimen that can produce a wide range of physical and mental health benefits, such as reduced blood pressure and decreased symptoms of anxiety, depression, and insomnia [1-3]. However, similar to other health behaviors, such as physical activity, the benefits of mindfulness meditation are primarily experienced after persistent, long-term performance [4-7]. Research has shown that increases in meditation frequency, duration, and long-term performance are all associated with greater health benefits among both clinical and general populations of adults [8-10]. Additionally, many of the proposed mechanisms for the benefits of meditation include biological changes, such as altering brain morphology, which happen over time through persistent meditation performance [11,12].

A diverse set of barriers exist to persistently performing mindfulness meditation, including structural (eg, financial and access-related)-, social (eg, stigma and peer-support)-, and individual (eg, impatience and motivation)-level factors. These multifaceted barriers have been shown to inhibit persistent mindfulness meditation practices and the persistent performance of other health-promoting behaviors [13-16], and thus novel behavioral interventions are still needed to help individuals attain the benefits from the long-term performance of healthy behaviors.

Mindfulness meditation has been successfully adapted for mobile phone apps, which helps to address several of the common structural and social barriers to persistent meditation. Mindfulness meditation apps are easily accessible, scalable, and cost-effective, improving individuals' access to meditation instruction and education [17,18]. Numerous commercial meditation apps are available to the public, and to date, the 2 leading apps are Headspace and Calm with 65 and 200 million downloads, respectively [3,19,20]. Although access to these popular apps is not free (roughly US $70 for an annual subscription), the cost is significantly lower than that of in-person, guided meditations. Additionally, employers are increasingly providing free access to meditation apps to their employees to help them improve their mental health and workplace productivity [18].

Interventions using commercial meditation apps have proven to be feasible and have demonstrated small- to medium-sized effects in reducing symptoms of depression and anxiety and increasing life satisfaction and positive affect [17,21]. Despite the accessibility and popularity of commercial meditation apps, app-based meditation persistence rates are low [22-24]. For example, a recent review found that adherence to app-based meditation interventions can be as low as 24% [25], and in the real world (ie, not in a research study), only 2% of health app users persistently engage at a level necessary to attain the corresponding health benefits [25-27].

Although mindfulness meditation apps have addressed several important structural and social barriers, the low persistence among app users might result from a lack of successful strategies for overcoming common individual-level barriers (eg, impatience and motivation) to persistent meditation. Behavioral economics and psychology research has demonstrated that individual-level barriers are significant determinants of nonpersistent (ie, only short-term) health behavior change, even after structural and social barriers have been overcome [13,28-30]. This has also been documented in the mobile health app literature: despite the popularity and ability of meditation apps to improve mental health, sustained engagement among mobile health app subscribers is low [4,24,25,31]. Moreover, a recent systematic review and meta-analysis of mental health app interventions reported app participation consistently decreased over time [4]. Therefore, novel strategies are needed to address individual-level barriers and help individuals increase and maintain their use of mindfulness meditation apps.

One strategy for overcoming individual-level barriers to mindfulness meditation app use may be the development of a meditation routine. Psychology research has shown that behaviors consistently performed in response to the same contextual (or environmental) cue become routinized, meaning they are completed with little or no cognitive effort [32,33]. One successful strategy for establishing a new routine is anchoring or pairing the new behavior to an existing routine that is already executed with very little cognitive effort [34-36]. For example, one might pair his or her daily meditation with their existing routine of an afternoon walk in order to routinize an afternoon meditation practice. Existing anchoring interventions have successfully established these reflexive or automatic routines for smoking cessation [37] and medication adherence [38,39]. However, the success of anchoring interventions has so far been limited to simple behaviors, such as drinking water or taking medications. Additionally, anchoring has largely only been effective for participants with high initial intrinsic motivation [40-44], so it is still unknown whether anchoring can help an individual successfully establish a persistent meditation app routine.

Furthermore, there are important design considerations in anchoring interventions that have not been rigorously tested in the literature, such as how to optimally select a participant's anchor. Research has shown that personalization is an important component to many other health interventions [45-51]; however, the theory of contextually cued routines is new for most people, so it may be difficult for participants to identify their own (ie, personalized) existing routine that can serve as an effective anchor for a new meditation routine. It has also been shown that daily routines most frequently occur in the morning [44,52], and recent research on circadian rhythms has suggested that routinization may be easier in the morning [53]. Thus, the purpose of this study was to test the efficacy of using a personalized anchor versus having an anchor assigned in the morning (ie, fixed) for successfully establishing a persistent

meditation app routine using the mobile app Calm. These 2 intervention groups (ie, personalized vs fixed anchors) received app-delivered reminder messages of their anchoring strategy for an 8-week period, and the persistence of the meditation routine over the subsequent 8 weeks was compared between these 2 groups and a control group that did not use the anchoring strategy for daily meditation. We hypothesized that the personalized anchor group would be the most persistent over the 8-week follow-up period and that both intervention groups would have significantly greater meditation persistence relative to the control group.

Methods

Recruitment

A randomized controlled trial was conducted between July 2020 and March 2021 with an 8-week intervention period, an 8-week follow-up period, and survey assessments at baseline, week 8, and week 16. The Institutional Review Board at Arizona State University approved this study (STUDY00011788), and all participants provided consent electronically prior to participating in the survey. This study design was preregistered on ClinicalTrials.gov (NCT04378530) and was funded by Arizona State University. The CONSORT file is available in Multimedia Appendix 1.

Study recruitment took place from July 2020 to August 2020. Participants were paying subscribers to the Calm app who were identified as not having already formed a daily meditation routine. Specifically, subscribers were eligible if they had subscribed to the Calm app after January 2020, had not completed a meditation session with the app in the past 30 days, and did not report practicing meditation with or without the app for more than 60 minutes in 1 month over the past 6 months. Additionally, new subscribers were eligible if they could read and understand English, were willing to be randomized, and were between 18 and 60 years old (see Textbox 1 for a full list of study eligibility criteria). Eligible subscribers were identified by Calm and invited to participate in the study via email. The email contained a brief overview of the study and a link to a short eligibility survey, and Qualtrics software was used to verify that participants satisfied all remaining study eligibility criteria. Eligible participants were then automatically directed to read and electronically sign an informed consent document in Qualtrics. Consenting participants were then contacted by the research team via email to complete the baseline questionnaire in Qualtrics. Once they completed the questionnaire, participants were randomized to 1 of 3 study groups using a predetermined allocation list generated on Randomizer.org by a researcher not involved in the participant assignment. Participants were then assigned to a study group based on the allocation list and the order in which they were enrolled in the study.

Textbox 1. Eligibility criteria.

Inclusion criteria

- 18-60 years of age
- Purchased Calm after January 2020
- Inactive: have not used app in the past 30 days
- Own an iOS/Android smartphone
- Own home internet or unlimited data plan
- Able to read and understand English
- Willing to be randomized

Exclusion criteria

- Report practicing mindfulness meditation >60 min in 1 month within the last 6 months
- Any meditation sessions with app in the past last 30 days
- Currently reside outside the USA

Intervention

Participants were randomized into a personalized anchor (PA) group, fixed anchor (FA) group, or control group (CG). Participants in this study used their own paid Calm accounts to access the app during the study. After completing the baseline survey, participants were sent a link to watch an instructional video that provided information about the benefits of meditating 10 minutes per day and study group–specific instructions on how to participate in the study. For those in the PA group, the video instructed participants to select an existing routine to which they would anchor their 10 minutes of daily meditation practice. The PA group's instructional video emphasized the importance of selecting a consistently occurring daily routine that could reliably be followed by 10 or more minutes of meditation and provided clear examples of such existing routines (eg, "After I finish my coffee in the afternoon" or "After I finish breakfast in the morning"). For those in the FA group, participants were instructed to use a fixed anchor provided by the research team to which they would anchor their 10 minutes of daily meditation practice. The anchor provided was the following: "After I finish in the bathroom (brushing teeth, removing mouth guard, etc.) in the morning, I will meditate for at least 10 minutes." Participants in the CG were given information about the mental health benefits of meditating for at least 10 minutes per day and instructed to complete 10

minutes of daily meditation but were not given any instruction on how or when to meditate. Participants were blinded to the other intervention protocols and did not know what intervention component was the focus of this study. To verify participants' comprehension of their study group–specific instructions, participants completed a 3-question comprehension quiz in Qualtrics and were given unlimited chances to answer each question correctly. Once all questions were correctly answered, participants were emailed with a start date for their intervention and they were provided with a written copy of the study instructions.

During the 8-week intervention period, all participants received a daily app–delivered reminder message (ie, push notification) to either meditate for at least 10 minutes or to meditate for at least 10 minutes using their anchor. Messages were randomly delivered at either 8 AM, 1 PM, or 6 PM (ie, a 33.3% chance of receiving the daily message at 1 of the 3 possible times), with adjustments made for participants' time zone. The message content was also randomized with a 50% chance of receiving 1 of 2 message types. The first message type included study group–specific reminders reinforcing participants' use of either their personalized or fixed anchors, or reminding the control group to meditate. The second message type was evenly randomized between reminders to use 3 motivational tools in the Calm app: mood check-ins, the meditation activity tracker, or the in-app daily reminder tool. The success of each type, timing, and sequence of daily supports was evaluated based on both participants' daily app usage data and ecological momentary assessments collected via SMS text messages once per evening (8 PM) during the 8-week intervention. The results from this microrandomized trial on the effectiveness of different daily reminder messages are not reported in this paper, and it is important to note that this microrandomized trial study design meant that each message type, timing, and sequence were randomly delivered across all study groups; thus, the sequence of messages would not bias our analysis of the overall study group differences in meditation persistence during this study.

Participants were initially instructed to use their anchors (PA and FA groups) and meditate for 10 minutes per day (all groups) for 8 weeks. After 8 weeks, participants were emailed a postintervention survey to complete and were encouraged to continue meditating but were not given further instructions. Participants were emailed again at the end of the 8-week follow-up period and given a final questionnaire to complete.

Surveys

The baseline, postintervention, and final questionnaires were all completed in Qualtrics. Participants were asked to respond using "A little bit," "Neutral," "Quite a bit," or "A lot" to the following 3 questions about the COVID-19 pandemic: "To what extent do you feel the COVID-19 pandemic has affected your mental health?", "To what extent do you feel the COVID-19 pandemic has affected your physical health?", and "To what extent do you feel the COVID-19 pandemic has affected your stress?" Participants also completed the Self-Report Behavioral Automaticity Index (SRBAI) on each survey to assess the strength of their meditation habit (ie, self-reported habit strength) [54]. The SRBAI contains 4 items scored on a 5-point Likert scale from 1 "Strongly disagree" to 5 "Strongly agree" in response to statements like "Daily meditation is something I do automatically," where a higher sum of item scores indicates a stronger habit. The SRBAI has a Cronbach's α of $\geq$.81 and was designed using discriminant content validity while preserving strong predictive validity [54]. Each survey also asked participants to rate their overall health as either "Poor," "Fair," "Good," "Very Good," or "Excellent." On the baseline survey, participants answered questions on their demographic and socioeconomic characteristics.

Outcomes

The primary outcome measure for this study was a binary measure of any daily meditation over the 16-week study, which was derived from participants' Calm app usage data provided by the Calm analytics team. Specifically, we used minute-level data on the time of day and duration of meditation sessions with the Calm app to construct an indicator variable equal to 1 if a participant completed any minutes of meditation on a given day, and 0 otherwise. To study how our intervention impacted meditation persistence, we examined how the odds of performing any daily meditation changed over time both during and after the intervention. The app usage data were also used to construct an indicator variable equal to 1 if a participant completed any minutes of meditation within 1 hour of the typical time that their personalized anchor was reported to occur (this typical time was collected when the PA group selected their anchor) or when the fixed anchor was expected to occur (8 AM). This measure of temporally consistent meditation was used to study participants' adherence to their anchoring strategy during and after the intervention. The secondary outcome of interest was the change in SRBAI between the study groups.

Statistical Analysis

A total sample size of 150 participants (study group sizes of 50) was targeted based on our available resources, and our expected statistical power was informed by prior interventions using the Calm app [3,55,56]. Assuming a small-to-medium-effect size of 0.20, study group sizes of 50 yielded a statistical power of $1–\beta$ =.76 for detecting study group x day–level differences in linear models of our repeated daily outcome (any meditation minutes) over the 16-week study at α=.05 (calculated using GLIMMPSE [57]).

Participants' demographic, socioeconomic, and health characteristics were compared across the 3 study groups to confirm that the randomization was effective using the Kruskal-Wallis nonparametric tests of equality (Table 1).

The primary outcome measuring the odds of any daily meditation was analyzed using panel logistic regression models with participant-level random effects. Aggregate study group differences in the primary outcome were estimated using separate indicator variables for the PA and FA groups, where the CG was the omitted reference group, and differences in the primary outcome over time were estimated using interaction terms between each study group indicator variable and a daily time trend. Two modeling approaches for the daily time trend were used: (1) a single linear time trend over the full 16-week study and (2) a piecewise linear trend with a breakpoint after

the 8-week intervention (ie, daily reminder messages) being withdrawn. The same panel logistic model with random effects was estimated for an outcome variable indicating whether participants performed any minutes of meditation within 1 hour of the expected time of their anchor (referred to as "anchored meditations"). These models were estimated as intention-to-treat analyses that used daily Calm app data for all participants who were retained in the study.

Table 1. Participant characteristics by study group.

Characteristic	Control, n (%) (N=37)	Fixed anchor, n (%) (N=27)	Personalized anchor, n (%) (N=37)	Two-sided *P* value[a]
Black	0 (0.00)	1 (3.70)	3 (8.11)	.21
Asian/Arab	2 (5.41)	1 (3.70)	1 (2.70)	.84
White	33 (89.19)	22 (81.48)	28 (75.68)	.32
Bi- or multiracial	0 (0.00)	1 (3.70)	1 (2.70)	.54
Race: nonresponse	2 (5.41)	0 (0.00)	2 (5.41)	.47
Male	8 (21.62)	6 (22.22)	4 (10.81)	.38
Female	28 (75.68)	19 (70.37)	30 (81.08)	.61
Less than 20 k[b]	0 (0.00)	1 (3.70)	3 (8.11)	.20
21-40 k[b]	2 (5.41)	1 (3.70)	6 (16.22)	.15
41-60 k[b]	5 (13.51)	5 (18.52)	7 (18.92)	.80
61-80 k[b]	2 (5.41)	2 (7.41)	1 (2.70)	.69
81-100 k[b]	8 (21.62)	8 (29.63)	1 (2.70)	.01
More than 100 k[b]	19 (51.35)	10 (37.04)	17 (45.95)	.53
Married	26 (70.27)	12 (44.44)	17 (45.95)	.05
Partnered	2 (5.41)	6 (22.22)	3 (8.11)	.08
Single/divorced/widowed	9 (24.32)	9 (33.33)	17 (45.95)	.15
Graduate degree	24 (64.86)	15 (55.56)	15 (40.54)	.11
Bachelor's degree	6 (16.22)	10 (37.04)	12 (32.43)	.14
Less than a bachelor's	7 (18.92)	2 (7.41)	10 (27.03)	.14
Poor health	0 (0.00)	4 (14.81)	1 (2.70)	.02
Fair health	7 (18.92)	6 (22.22)	9 (24.32)	.85
Good health	12 (32.43)	10 (37.04)	13 (35.14)	.93
Very good health	13 (35.14)	4 (14.81)	12 (32.43)	.17
Excellent health	4 (10.81)	3 (11.11)	0 (0.00)	.12
Currently with depression	11 (29.73)	9 (33.33)	11 (29.73)	.94
COVID-19 stress	30 (83.33)	20 (74.07)	27 (77.14)	.66
COVID-19 mental health	24 (66.67)	21 (77.78)	24 (68.57)	.61
COVID-19 physical health	15 (41.67)	12 (44.44)	12 (34.29)	.69

[a]Two-sided *P* values are presented for Kruskal-Wallis nonparametric tests of equality for each measure of participants' characteristics across the 3 study groups.

[b]Income in US $.

In subgroup analyses, participants were split into high- and low-meditation subgroups based on their total number of days with any meditation during the 8-week intervention. The high-meditation subgroup was defined as those participants who meditated on 14 (the median number of days) or more of the intervention days. All other participants were placed in the low-meditation subgroup. These subgroups were created to test whether the success of the anchoring strategy differed based on the total number of meditations performed during the intervention. Participants in the PA and FA groups were also split according to the number of intervention days that they potentially meditated with the Calm app using their anchor during the intervention. Participants from the PA and FA groups were classified as high anchorers if they completed 12 (the

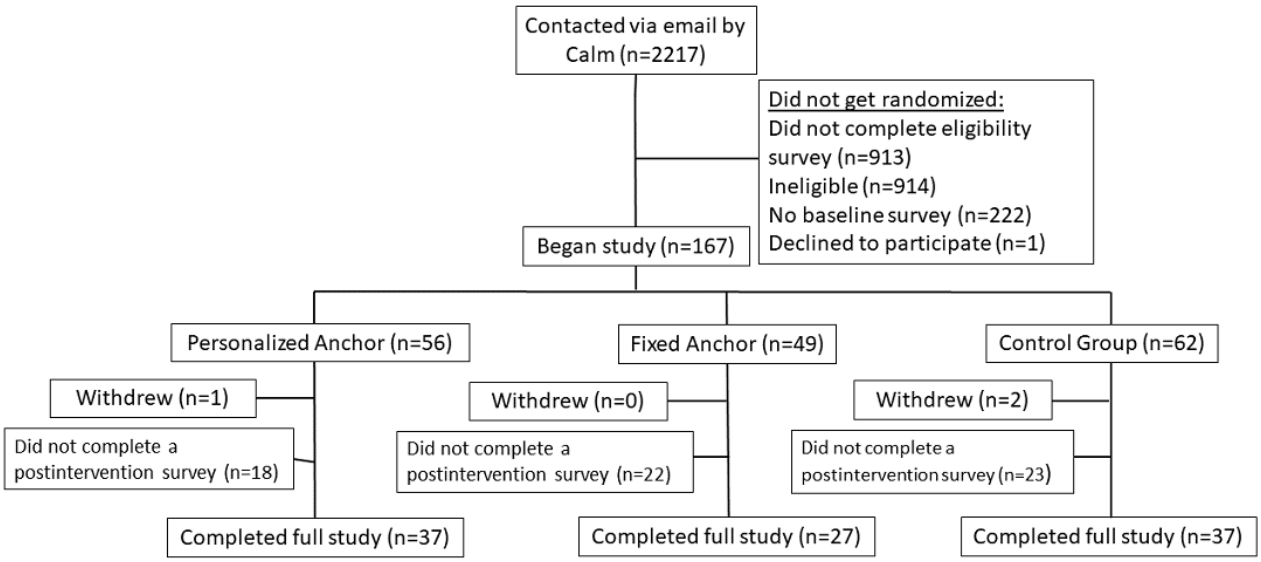
median number of days using one's anchor) or more meditations within 1 hour of the expected time of their anchor. All other participants in the PA and FA groups were considered low anchorers, and the CG did not use the anchoring strategy and so were not classified as either high or low anchorers. These additional subgroups were created to examine how the success of the anchoring strategy varied based on the number of anchored meditations during the intervention.

Study group differences in the SRBAI between the baseline and postintervention survey were analyzed using analysis of variance (ANOVA), and pairwise comparisons between the PA and FA groups and the CG used were analyzed with the *t* test. All statistical analyses were performed using Stata/MP (StataCorp) 16.1 for Windows. (Microsoft Corp).

Results

A total of 2217 Calm subscribers were emailed to participate in this study. Among those who completed the eligibility survey and were identified as eligible, 167 provided informed consent, completed the baseline survey, and were randomized into 1 of the 3 study groups: (1) the PA group (n=56), (2) the FA group (n=49), or the CG (n=62). Figure 1 is a flow diagram outlining participant enrollment, randomization, and retention.

Figure 1. Study flow diagram of participant enrollment and randomization.



After a few participants asked to withdraw (n=3), a total of 101 participants completed at least 1 postintervention survey (either week 8 or week 16) and were included in the final analysis. Due to the different attrition rates across study groups, the final analytical sample was not balanced in size across groups, limiting the statistical power of our analyses. However, Table 1 shows that the study groups were still balanced on most of the observed participant characteristics. Participants were predominantly White (83/101, 82.2%), female (77/101, 76.2%), college educated (ie, bachelor's or graduate degree; 82/101, 81.2%), and earned $81,000 per year and above (63/101, 62.4%). Additionally, only 26.7% (27/101) of participants reported "Poor" or "Fair" health, 76.2% (77/101) reported that COVID-19 has affected their stress either "Quite a bit" or "A lot," and 68.3% (69/101) reported that COVID-19 has affected their mental health either "Quite a bit" or "A lot." Importantly, there were few statistically significant differences between study groups at baseline (see Table 1). The only observable differences between study groups were in terms of marital status and the percent reporting "poor" health, where the 2 treatment groups (PA and FA) were less likely to be married and more likely to report poor health than was the control group. Given these differences across study groups, we included covariates for each of these characteristics in additional regression models presented in Multimedia Appendix 2.

Table 2 displays the study group differences in the daily odds of any meditation (our primary outcome). Specifically, Table 2 shows the exponentiated coefficients from panel logistic regression models estimated with participant-level random effects predicting the primary outcome among the full sample (column 1) and separately estimated among the high-meditation subgroup (column 2). The FA group had a significantly higher average odds of daily meditation during the intervention (1.14 odds ratio [OR]; 95% CI 1.02-1.33; *P*=.04), and all participants experienced a significant linear decline in their odds of daily meditation during the 8-week intervention (0.96 OR; 95% CI 0.95-0.96; *P*<.001). Additionally, the FA group showed a significantly smaller decline in the linear trend of their odds of daily meditation during the 8-week follow-up period (their daily trend increased by 1.04 OR from their trend during the intervention; 95% CI 1.01-1.06; *P*=.03 during the follow-up). A separate model was estimated that also included measures of participants' race, gender, education, marital status, health status, and an identifier for self-reporting being depressed, and these results are presented in Multimedia Appendix 2 and do not significantly differ from the model without these additional participant characteristics. To visualize these study group differences in our primary outcome, Figure 2 displays both the raw and predicted daily probability of any minutes of meditation for each study group based on the coefficient estimates from the full analytic sample shown in column 1 of Table 2.

Table 2. Treatment effects on the odds of daily meditation.

Independent variables:	All participants, OR[a] (95% CI)	High-meditation subgroup, OR (95% CI)
Fixed anchor	1.139 (1.019-1.326)**	1.081 (1.021-1.310)**
Personalized anchor	1.012 (0.270-3.860)	1.062 (0.884-1.276)
Days in study	0.960 (0.956-0.964)***	0.964 (0.957-0.970)***
Fixed anchor × days	0.989 (0.977-1.000)	1.001 (0.999-1.003)
Personalized anchor × days	0.993 (0.976-1.002)	0.987 (0.974-1.000) *
Days postintervention	0.992 (0.978-1.006)	0.985 (0.969-1.000) *
Fixed anchor × postintervention days	1.035 (1.013-1.057) ***	1.005 (0.978-1.032)
Personalized anchor × postintervention days	1.013 (0.990-1.036)	0.985 (0.975-1.000)*
Participant-day observations, n[b]	11,312	5712
Participants, n[b]	101	51

[a]OR: odds ratio.

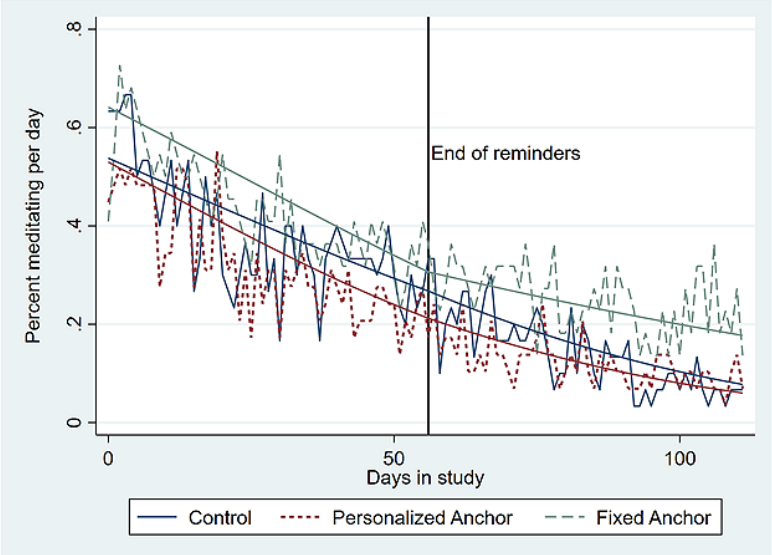
[b]Data in this row are expressed as integers and not odds ratio and CI.

*P<.10.

**P<.05.

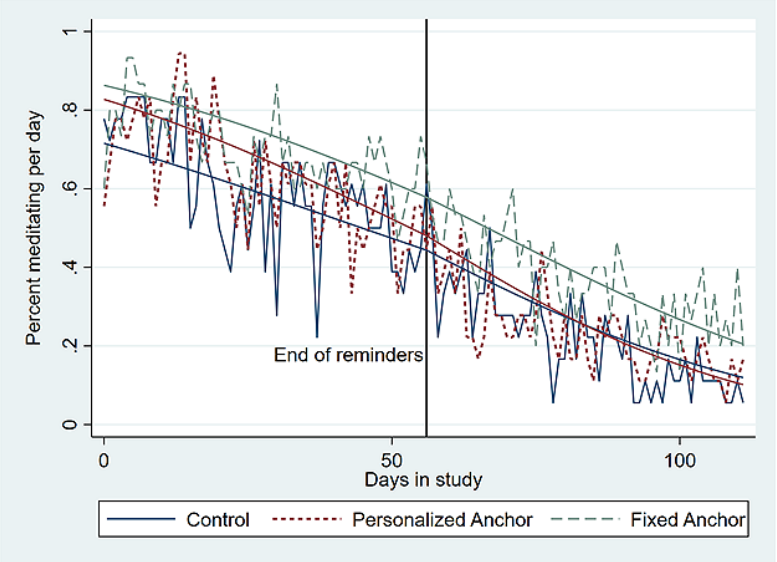
***P<.01.

Figure 2. Daily percent of participants who performed any minutes of meditation.



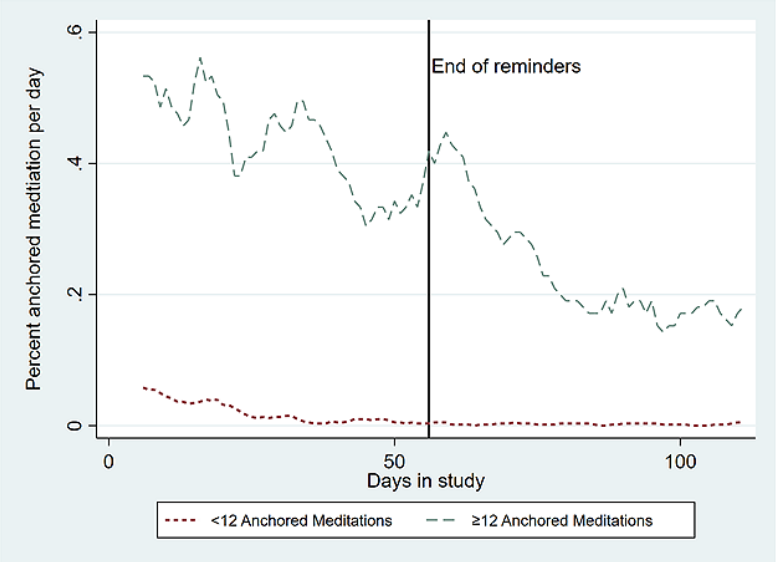
To examine if the anchoring strategy was more successful for more frequent meditators, the raw and predicted daily probability of any minutes of meditation among the high-meditation subgroup (n=51) in each study group was determined (Figure 3). The corresponding regression results in Table 2 show that among the high-meditation subgroup, the FA group still had a significantly higher average odds of daily meditation during the intervention (1.08 OR; 95% CI 1.02-1.31; P=.03), and all participants experienced a significant linear decline in their odds of daily meditation during the 8-week intervention (0.96 OR; 95% CI 0.96-0.97; P<.001). However, there was no statistically significant difference between study groups in the decline of daily odds of meditation during the 8-week follow-up among the high-meditation subgroup.

Figure 3. Daily percent of participants in the high-meditation subgroup who performed any minutes of meditation.



To visualize how the anchoring strategy impacted meditation persistence, Figure 4 plots the average daily percent of participants who completed any minutes of meditation within 1 hour of the expected time of their anchor (ie, anchored meditations) among the high anchorers (n=19) and separately among the low anchorers (n=45). The high-anchorer subgroup was composed of 13 participants from the FA group and 6 participants from the PA group, which demonstrates the relative success of using the fixed morning anchor versus allowing participants to select their own anchor. Additionally, 4 out of the 6 high anchorers from the PA group selected a morning anchor that occurred between 7 AM and 9 AM, which further suggests that morning anchors are the most likely to be successful. The trends in Figure 4 show that most participants (ie, the low anchorers) did not use their anchoring strategy beyond the first 4 weeks of the intervention but that anchored meditations remained fairly persistent among the high anchorers.

Figure 4. Daily percent of participants who performed any minutes of anchored meditation.



The stronger persistence in anchored meditations among the high anchorers was tested empirically and is shown in Table 3, which displays the panel logistic regression results from models predicting the odds of any minutes of meditation for the low anchorers and those in the CG or the odds of any anchored meditations among the high anchorers. This split outcome variable provided a more conservative test of the differences in meditation persistence between the high anchorers versus the low anchorers or the CG because all nonanchored meditations were not considered as evidence of meditation persistence for the high anchorers. The high anchorers had a significantly higher average odds of daily meditation during the intervention (34.68 OR; 95% CI 5.70-210.80; *P*=.008), and all participants experienced a significant linear decline in their odds of daily meditation during the 8-week intervention (0.96 OR; 95% CI 0.96-0.97; *P*<.001). Importantly, the high anchorers showed a significantly smaller decline in the linear trend of their odds of daily meditation during the 8-week follow-up period (their daily

trend increased by 1.13 OR from their trend during the intervention; 95% CI 1.02-1.35; *P*=.007 during the follow-up). A separate model was estimated for this split outcome that also included measures of participants' race, gender, education, marital status, health status, and an identifier for self-reporting being depressed, and these results are presented in Multimedia Appendix 2 and do not significantly differ from the model without these additional participant characteristics. Figure 5 displays the raw and predicted probability of this split outcome for high anchorers, low anchorers, and the CG.

Table 3. Effect of successfully anchoring on the odds of daily meditation.

Independent variables	Linear time trend, OR[a] (95% CI)	Piecewise linear trend, OR (95% CI)
<12 anchored meditations	0.613 (0.206-1.824)	0.793 (0.264-2.388)
≥12 anchored meditations	28.079 (4.773-165.201) ***	34.675 (5.704-210.796)***
Days in study	0.960 (0.956-0.964)***	0.964 (0.957-0.971)***
<12 anchored meditations × days	0.998 (0.993-1.004)	0.987 (0.977-1.002)
≥12 anchored meditations × days	1.002 (0.994-1.010)	0.994 (0.979-1.009)
Days postintervention	—[b]	0.992 (0.978-1.006)
<12 anchored meditations × postintervention days	—	1.067 (0.990-1.145)
≥12 anchored meditations × postintervention days	—	1.129 (1.019-1.351)***
Participant-day observations, n[c]	11,312	11,312
Participants, n[c]	101	101

[a]OR: odds ratio.

[b]Not included in the model.

[c]Data in this row are expressed as integers and not odds ratio and CI.

*P<.10.

**P<.05.

***P<.01.

Figure 5. Daily percent of participants who performed any minutes of meditation or any minutes of anchored meditations.

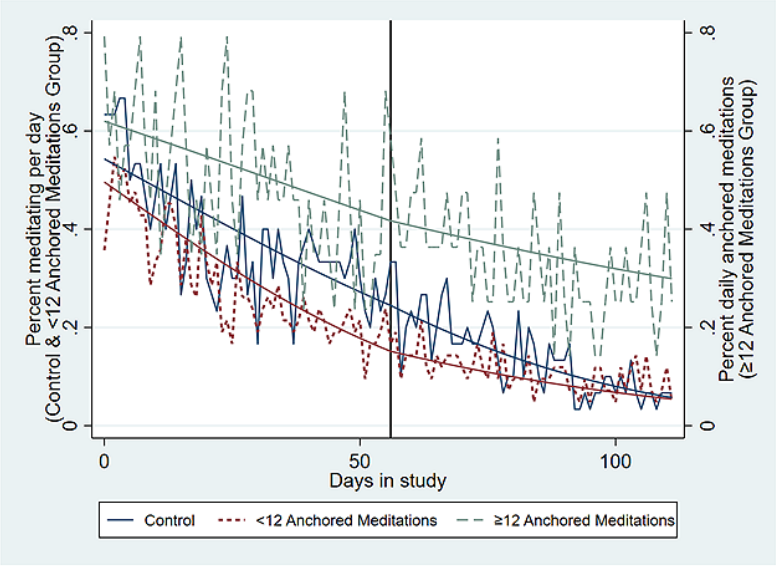
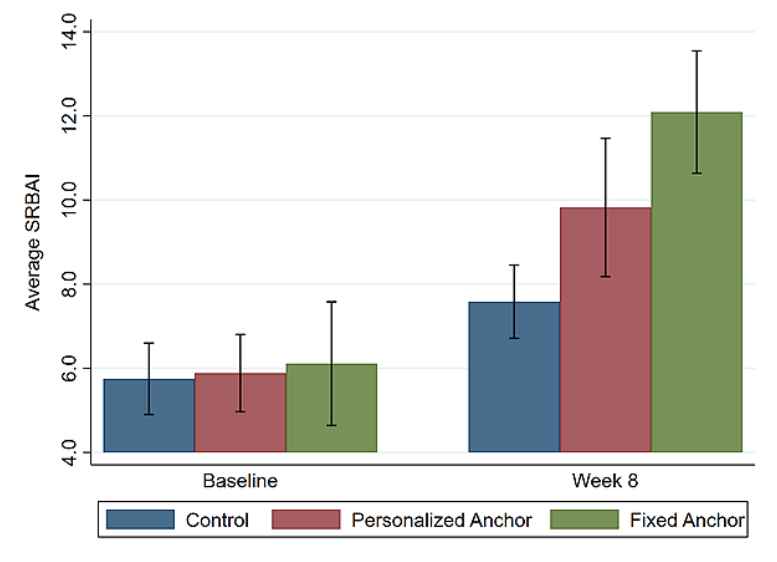


Figure 6 displays the average self-reported meditation habit strength, our secondary outcome, among the 3 study groups on the baseline and week 8 surveys. There was no statistically significant difference in self-reported habit strength between the study groups at baseline. Participants in the FA group reported a significantly higher increase in self-reported habit strength between baseline and week 8 than did the CG (4.56 greater SRBAI increase; 95% CI 1.46-7.66; *P*<.001), while the differences between the FA and PA groups and PA and CG were not statistically significant.

Figure 6. Average Self-Reported Behavioral Automaticity Index. SRBAI: Self-Reported Behavioral Automaticity Index.



Discussion

Principal Findings

This study tested the efficacy of using either PAs or FAs for establishing a persistent meditation app routine with the mobile app, Calm. Although the results found that all study groups (ie, PA, FA, and CG) experienced an equal decline in their daily odds of performing any minutes of meditation with the Calm app during the 8-week intervention, the FA group was significantly more persistent (ie, smaller daily decline in the odds of any meditation) during the 8-week follow-up period. Subgroup analyses revealed that performing a larger number of meditations during the intervention was not sufficient for displaying meditation persistence. Instead, the participants who were high anchorers during the intervention (ie, equal to or above the median number of meditations performed within 1 hour of the expected time of their anchor) showed the most persistent meditation routines during the follow-up period. These findings indicate that the anchoring strategy can create persistent meditation routines for some participants but that additional intervention tools are likely needed to help more participants successfully adhere to their anchored meditation routine.

The results in Table 2 and Figure 3 show that simply performing meditation on more days of the intervention was not associated with higher meditation persistence among any of the 3 study groups. Specifically, Figure 3 shows that despite high meditators (ie, participants who meditated on 14—the median number of days—or more of the intervention days) having an average daily probability of meditating roughly equal to 80% at the start of the study, the average high meditator in any of the 3 study groups displayed a steady decline in their daily probability of meditating. This observation stands in contrast to the expected role that high behavioral performance should have on behavioral persistence according to standard microeconomic theory of habit formation [58,59]. As high-meditators' level of meditation performance did not sufficiently increase their marginal utility for continuing their meditation behavior, the results suggest that meditation needs to be performed for a longer duration of time (ie, more days with any meditation) in order to form a persistent routine or that alternative theories of habit formation may be more appropriate for understanding persistent meditation routines.

Additionally, the results show that high anchorers were significantly more persistent in their daily meditation, and these findings were estimated using only the anchored meditations to measure persistence among the high-anchorer subgroup. In other words, high anchorers were not just more likely to meditate at any time of day, but this subgroup was more likely to meditate at a time that corresponded to their anchor than the control group or the low anchorers were to meditate at any time of day. This observation supports the theory that contextually cueing behaviors is one method for creating a persistent meditation routine [32,33]. As fewer than half (19/64, 30%) of the participants in either the FA or PA groups were high anchorers, these results also suggest that setting an anchoring strategy and receiving app-delivered reminder messages are not sufficient for helping all participants adhere to their anchoring strategy. Importantly, almost all of the high anchorers were participants who either selected a morning anchor or were given the fixed morning anchor, which indicates that meditating in the morning might be an important strategy for establishing persistent meditation routines and warrants further research.

The results from the SRBAI (ie, self-reported habit strength) show that those in the FA group experienced the largest increase in habit strength. However, there was an increase in self-reported habit strength among all study groups, including the CG. Additionally, self-reported habit strength increased on average for all study groups despite the clear decline in daily meditation performance observed in the objective app usage data. These trends highlight a potential limitation of this self-report habit strength measure: since habitual behaviors are theorized to be unconsciously initiated, individuals should not be able to recall their experience performing the behavior (in this case meditation). Thus, this measure may be capturing participants'

perceived self-efficacy or fluency for meditation [60], which suggests that simply being involved in a study and receiving information on the benefits of meditating daily for 10 or more minutes might have boosted participants' feeling of behavioral competence or self-efficacy for meditation. Therefore, although this self-reported habit strength was significantly greater among the FA participants who anchored meditation in the morning, these results should be interpreted with caution.

Prior Work

This study contributes to the existing literature testing anchoring interventions for health behaviors, which has already demonstrated the success of anchoring for establishing persistent smoking cessation routines [37,61] and medication adherence routines [38,39]. However, the anchoring approach has been less effective in other settings, such as demonstrating limited efficacy for improving diets [40,62,63], which suggests that the success of anchoring may vary depending on the behavioral complexity of the targeted new routine. Our study shows that anchoring can help to improve the persistence of meditation with a mobile app for some participants, but the success of anchoring was not universally experienced by all participants, and further research is needed to determine whether anchoring can be more effectively implemented to establish persistent meditation routines with a mobile app.

It is important to note that our design of the anchoring intervention was targeted toward establishing "instigation" habits as opposed to "execution" habits for daily meditation [64]. In other words, the suggested anchors were all chosen to help initiate meditation with the Calm app as opposed to helping participants continue to perform a given meditation session. This was because we hypothesized that continuing to perform a given meditation session is a relatively easier action since meditation is generally a passive behavior and most of the meditations with the app are timed, so users do not need to self-monitor the clock and their time meditating. Future studies should test the efficacy of anchoring interventions that target the execution component of daily meditation, which may help us understand how anchoring can be successfully applied to complex behaviors like daily meditation.

Finally, this study demonstrates that an 8-week intervention was not sufficiently long for even the high anchorers to form a meditation routine. Existing research has suggested that it takes anywhere from 18 to 254 days to successfully form a new routine [65], so our results help to increase the lower bound on this range for meditation routines. Additional research is needed to generate a more precise estimate of the average number of days of behavioral performance for successfully routinizing meditation with a mobile app.

Limitations

Although this was the first study to use personalized or fixed anchors for establishing a persistent meditation app routine with a consumer-based app (ie, Calm) and there were no unexpected events, there were still a number of limitations. First, we had a homogeneous, small sample size limiting the generalizability of our findings, particularly to other racial groups and people of different socioeconomic status. Second, our study targeted dormant users of Calm who had paid for an annual subscription but had not recently used the app, which again limits the generalizability of our results for other types of app users. Third, the daily app–delivered reminder messages appeared to be an ineffective method of boosting most participants' attention to and use of the anchoring strategy, so it is difficult to know whether a longer duration of intervention or increased intervention supports are necessary to increase adherence to the anchoring strategy and more rigorously test the efficacy of this intervention approach for establishing behavioral routines. Finally, a significant degree of study attrition from either withdrawals or missing survey data occurred during the intervention, which limited the statistical power of our analyses.

Conclusions

This study tested the efficacy of using either personalized or fixed anchors for establishing a persistent meditation app routine with the mobile app Calm. Participants given the FA of meditating in the morning were slightly more persistent during the 8-week follow-up period. Additionally, the participants who more frequently used their anchor during the intervention showed the most persistent meditation routines during the follow-up period, and almost all of these high anchorers were using morning anchors. Our findings suggest that using the anchoring strategy can create persistent morning meditation routines. However, future studies should combine anchoring with additional intervention tools (eg, incentives) to help more participants successfully establish an anchored meditation routine.

Conflicts of Interest

CS and MS have no conflicts to declare. JH discloses that she receives an annual salary from Calm and holds stock within the company. However, her salary and equity are not dependent upon the results of her research.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 405 KB - mhealth_v9i12e32794_app1.pdf]

Multimedia Appendix 2
Additional regression results.
[DOCX File , 22 KB - mhealth_v9i12e32794_app2.docx]

References

1. Tang Y, Hölzel BK, Posner MI. The neuroscience of mindfulness meditation. Nat Rev Neurosci 2015 Apr;16(4):213-225. [doi: 10.1038/nrn3916] [Medline: 25783612]
2. Tang Y, Rothbart MK, Posner MI. Neural correlates of establishing, maintaining, and switching brain states. Trends Cogn Sci 2012 Jun;16(6):330-337 [FREE Full text] [doi: 10.1016/j.tics.2012.05.001] [Medline: 22613871]
3. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273. [doi: 10.2196/14273]
4. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
5. Athanas A, McCorrison J, Smalley S, Price J, Grady J, Wehner P, et al. Association between improvement in baseline mood and long-term use of a mindfulness and meditation app: observational study. JMIR Ment Health 2019 May 08;6(5):e12617. [doi: 10.2196/12617]
6. Engen HG, Bernhardt BC, Skottnik L, Ricard M, Singer T. Structural changes in socio-affective networks: Multi-modal MRI findings in long-term meditation practitioners. Neuropsychologia 2018 Jul 31;116(Pt A):26-33. [doi: 10.1016/j.neuropsychologia.2017.08.024] [Medline: 28842274]
7. Lardone A, Liparoti M, Sorrentino P, Rucco R, Jacini F, Polverino A, et al. Mindfulness meditation is related to long-lasting changes in hippocampal functional topology during resting state: a magnetoencephalography study. Neural Plast 2018;2018:5340717 [FREE Full text] [doi: 10.1155/2018/5340717] [Medline: 30662457]
8. Fredrickson BL, Boulton AJ, Firestine AM, Van Cappellen P, Algoe SB, Brantley MM, et al. Positive emotion correlates of meditation practice: a comparison of mindfulness meditation and loving-kindness meditation. Mindfulness (N Y) 2017 Dec;8(6):1623-1633 [FREE Full text] [doi: 10.1007/s12671-017-0735-9] [Medline: 29201247]
9. Adams ZW, Sieverdes JC, Brunner-Jackson B, Mueller M, Chandler J, Diaz V, et al. Meditation smartphone application effects on prehypertensive adults' blood pressure: Dose-response feasibility trial. Health Psychol 2018 Sep;37(9):850-860 [FREE Full text] [doi: 10.1037/hea0000584] [Medline: 30010353]
10. Lacaille J, Sadikaj G, Nishioka M, Carrière K, Flanders J, Knäuper B. Daily mindful responding mediates the effect of meditation practice on stress and mood: the role of practice duration and adherence. J Clin Psychol 2018 Jan;74(1):109-122. [doi: 10.1002/jclp.22489] [Medline: 28759111]
11. Tang YY, Lu Q, Fan M, Yang Y, Posner MI. Mechanisms of white matter changes induced by meditation. Proc Natl Acad Sci U S A 2012 Jun 26;109(26):10570-10574 [FREE Full text] [doi: 10.1073/pnas.1207817109] [Medline: 22689998]
12. Shen H, Chen M, Cui D. Biological mechanism study of meditation and its application in mental disorders. Gen Psychiatr 2020;33(4):e100214 [FREE Full text] [doi: 10.1136/gpsych-2020-100214] [Medline: 32695961]
13. Volpp KG. Knocking down barriers to behavior change. NEJM Catalyst 2018;4(4):1. [doi: 10.1056/CAT.18.0140]
14. Petrovic D, de Mestral M, Bochud M, Bartley M, Kivimäki M, Vineis P, et al. The contribution of health behaviors to socioeconomic inequalities in health: A systematic review. Prev Med 2018 Aug;113:15-31. [doi: 10.1016/j.ypmed.2018.05.003] [Medline: 29752959]
15. Sriram U, Morgan EH, Graham ML, Folta SC, Seguin RA. Support and sabotage: a qualitative study of social influences on health behaviors among rural adults. J Rural Health 2018 Dec;34(1):88-97 [FREE Full text] [doi: 10.1111/jrh.12232] [Medline: 28045193]
16. Macinko J, Upchurch DM. Factors associated with the use of meditation, US adults 2017. J Altern Complement Med 2019 Sep;25(9):920-927. [doi: 10.1089/acm.2019.0206] [Medline: 31381366]
17. Gál É, Stefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
18. Kellen M, Saxena D. Calm my headspace: motivations and barriers for adoption and usage of meditation apps during times of crisis. 2020 Presented at: The 20th International Conference on Electronic Business; December 5-8, 2020; Hong Kong.
19. Tan G. Calm meditation app hits $2 billion value with lightspeed, Goldman Backing. Bloomberg. URL: https://www.bloomberg.com/news/articles/2020-12-08/calm-hits-2-billion-value-with-lightspeed-goldman-tpg-backing [accessed 2021-01-01]
20. Headspace and Snap Inc. team up for World Mental Health Day, launch new Headspace mini content to support Snapchatters' rising stress levels. Businesswire. 2020 Oct 09. URL: https://www.businesswire.com/news/home/20201009005034/en/Headspace-and-Snap-Inc.-Team-Up-for-World-Mental-Health-Day-Launch-New-Headspace-Mini-Content-To-Support-Snapchatters%E2%80%99-Rising-Stress-Levels [accessed 2021-01-01]
21. Huberty J, Eckert R, Larkey L, Kurka J, Rodríguez De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]

22. Forbes L, Gutierrez D, Johnson SK. Investigating adherence to an online introductory mindfulness program. Mindfulness 2017 Jul 14;9(1):271-282. [doi: 10.1007/s12671-017-0772-4]
23. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: A randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. Journal of Happiness Studies 2016;17:163-185. [doi: 10.1007/s10902-014-9589-1]
24. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
25. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
26. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
27. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. Journal of Medical Internet Research 2014;16(4):e109. [doi: 10.2196/jmir.3084] [Medline: 24735567]
28. Stecher C, Mukasa B, Linnemayr S. Uncovering a behavioral strategy for establishing new habits: Evidence from incentives for medication adherence in Uganda. J Health Econ 2021 May;77:102443. [doi: 10.1016/j.jhealeco.2021.102443] [Medline: 33831632]
29. Biddle SJH. Sedentary behavior at the individual level: Correlates, theories, and interventions. In: Leitzmann M, Jochem C, Schmid D, editors. Sedentary Behavior Epidemiology: Springer Series on Epidemiology and Public Health. New York: Springer; 2018:405.
30. Linnemayr S, Stecher C. Behavioral economics matters for HIV research: the impact of behavioral biases on adherence to antiretrovirals (ARVs). AIDS Behav 2015 Nov;19(11):2069-2075 [FREE Full text] [doi: 10.1007/s10461-015-1076-0] [Medline: 25987190]
31. Kerst A, Zielasek J, Gaebel W. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur Arch Psychiatry Clin Neurosci 2020 Mar;270(2):139-152. [doi: 10.1007/s00406-018-0974-3] [Medline: 30607530]
32. Rothman AJ, Gollwitzer PM, Grant AM, Neal DT, Sheeran P, Wood W. Hale and hearty policies: how psychological science can create and maintain healthy habits. Perspect Psychol Sci 2015 Nov;10(6):701-705. [doi: 10.1177/1745691615598515] [Medline: 26581721]
33. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]
34. Wood W, Neal DT. Healthy through habit: Interventions for initiating & maintaining health behavior change. Behavioral Science & Policy 2016;2(1):71-83. [doi: 10.1353/bsp.2016.0008]
35. Gollwitzer PM. Implementation intentions: Strong effects of simple plans. American Psychologist 1999;54(7):493-503. [doi: 10.1037/0003-066x.54.7.493]
36. Orbell S, Hodgkins S, Sheeran P. Implementation intentions and the theory of planned behavior. Pers Soc Psychol Bull 1997 Sep;23(9):945-954. [doi: 10.1177/0146167297239004] [Medline: 29506445]
37. Armitage CJ, Arden MA. How useful are the stages of change for targeting interventions? Randomized test of a brief intervention to reduce smoking. Health Psychol 2008 Nov;27(6):789-798. [doi: 10.1037/0278-6133.27.6.789] [Medline: 19025275]
38. O'Carroll RE, Chambers JA, Dennis M, Sudlow C, Johnston M. Improving adherence to medication in stroke survivors: a pilot randomised controlled trial. Ann Behav Med 2013 Dec;46(3):358-368. [doi: 10.1007/s12160-013-9515-5] [Medline: 23670112]
39. Brown I, Sheeran P, Reuber M. Enhancing antiepileptic drug adherence: a randomized controlled trial. Epilepsy Behav 2009 Dec;16(4):634-639. [doi: 10.1016/j.yebeh.2009.09.014] [Medline: 19864187]
40. Lally P, Chipperfield A, Wardle J. Healthy habits: efficacy of simple advice on weight control based on a habit-formation model. Int J Obes (Lond) 2008 Apr;32(4):700-707. [doi: 10.1038/sj.ijo.0803771] [Medline: 18071344]
41. Sheeran P, Orbell S. Implementation intentions and repeated behaviour: augmenting the predictive validity of the theory of planned behaviour. Eur. J. Soc. Psychol 1999 Mar;29(2-3):349-369. [doi: 10.1002/(sici)1099-0992(199903/05)29:2/3<349::aid-ejsp931>3.0.co;2-y]
42. 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]
43. Verplanken B, Faes S. Good intentions, bad habits, and effects of forming implementation intentions on healthy eating. Eur. J. Soc. Psychol 1999 Aug;29(5-6):591-604. [doi: 10.1002/(sici)1099-0992(199908/09)29:5/6<591::aid-ejsp948>3.0.co;2-h]

44. Beeken RJ, Croker H, Morris S, Leurent B, Omar R, Nazareth I, et al. Study protocol for the 10 Top Tips (10TT) trial: randomised controlled trial of habit-based advice for weight control in general practice. BMC Public Health 2012 Aug 16;12:667 [FREE Full text] [doi: 10.1186/1471-2458-12-667] [Medline: 22898059]
45. 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]
46. Noar SM, Harrington NG, Aldrich RS. The role of message tailoring in the development of persuasive health communication messages. Annals of the International Communication Association 2016 May 18;33(1):73-133 [FREE Full text] [doi: 10.1080/23808985.2009.11679085]
47. de Vries RAJ, Truong KP, Zaga C, Li J, Evers V. A word of advice: how to tailor motivational text messages based on behavior change theory to personality and gender. Pers Ubiquit Comput 2017 Jun 15;21(4):675-687. [doi: 10.1007/s00779-017-1025-1]
48. Rimer BK, Kreuter MW. Advancing tailored health communication: A persuasion and message effects perspective. Journal of Communication 2006;56(1):S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
49. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: A systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
50. Mayer JE, Fontelo P. Meta-analysis on the effect of text message reminders for HIV-related compliance. AIDS Care 2017 Apr;29(4):409-417 [FREE Full text] [doi: 10.1080/09540121.2016.1214674] [Medline: 27477580]
51. Bull FC, Kreuter MW, Scharff DP. Effects of tailored, personalized and general health messages on physical activity. Patient Education and Counseling 1999 Feb;36(2):181-192. [doi: 10.1016/s0738-3991(98)00134-7]
52. Sanders MJ, Van Oss T. Using daily routines to promote medication adherence in older adults. Am J Occup Ther 2013;67(1):91-99. [doi: 10.5014/ajot.2013.005033] [Medline: 23245787]
53. Fournier M, d'Arripe-Longueville F, Rovere C, Easthope CS, Schwabe L, El Methni J, et al. Effects of circadian cortisol on the development of a health habit. Health Psychol 2017 Nov;36(11):1059-1064. [doi: 10.1037/hea0000510] [Medline: 28650196]
54. 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]
55. Huberty J, Puzia ME, Green J, Vlisides-Henry RD, Larkey L, Irwin MR, et al. A mindfulness meditation mobile app improves depression and anxiety in adults with sleep disturbance: Analysis from a randomized controlled trial. Gen Hosp Psychiatry 2021;73:30-37 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.09.004] [Medline: 34537477]
56. Huberty J, Vranceanu AM, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the Calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]
57. GLIMMPSE: General Linear Mixed Model Power and Sample Size. Glimmpse. URL: https://glimmpse.samplesizeshop.org/ [accessed 2021-01-01]
58. Pollak RA. Habit formation and dynamic demand functions. Journal of Political Economy 1970 Jul;78(4, Part 1):745-763. [doi: 10.1086/259667]
59. Ryder H, Heal GM. Optimal Growth with Intertemporally Dependent Preferences. The Review of Economic Studies 1973 Jan;40(1):1-31. [doi: 10.2307/2296736]
60. Hagger MS, Rebar AL, Mullan B, Lipp OV, Chatzisarantis NLD. The subjective experience of habit captured by self-report indexes may lead to inaccuracies in the measurement of habitual action. Health Psychol Rev 2015;9(3):296-302. [doi: 10.1080/17437199.2014.959728] [Medline: 25189762]
61. McWilliams L, Bellhouse S, Yorke J, Lloyd K, Armitage CJ. Beyond "planning": a meta-analysis of implementation intentions to support smoking cessation. Health Psychol 2019 Dec;38(12):1059-1068. [doi: 10.1037/hea0000768] [Medline: 31414843]
62. Adriaanse MA, Vinkers CDW, De Ridder DT, Hox JJ, De Wit JBF. Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. Appetite 2011 Feb;56(1):183-193 [FREE Full text] [doi: 10.1016/j.appet.2010.10.012] [Medline: 21056605]
63. Tam L, Bagozzi RP, Spanjol J. When planning is not enough: the self-regulatory effect of implementation intentions on changing snacking habits. Health Psychol 2010 May;29(3):284-292. [doi: 10.1037/a0019071] [Medline: 20496982]
64. Gardner B, Phillips LA, Judah G. Habitual instigation and habitual execution: Definition, measurement, and effects on behaviour frequency. Br J Health Psychol 2016 Sep;21(3):613-630. [doi: 10.1111/bjhp.12189] [Medline: 26991427]
65. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur. J. Soc. Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]

Abbreviations

ANOVA: analysis of variance

CG: control group
FA: fixed anchor
OR: odds ratio
PA: personalized anchor
SRBAI: Self-Report Behavioral Automaticity Index

Edited by G Eysenbach; submitted 10.08.21; peer-reviewed by L Ellis; comments to author 03.09.21; revised version received 28.09.21; accepted 12.10.21; published 22.12.21.

Please cite as:
Stecher C, Sullivan M, Huberty J
Using Personalized Anchors to Establish Routine Meditation Practice With a Mobile App: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(12):e32794
URL: https://mhealth.jmir.org/2021/12/e32794
doi:10.2196/32794
PMID:34941558

Original Paper

Impact of a Mobile Telerehabilitation Solution on Metabolic Health Outcomes and Rehabilitation Adherence in Patients With Obesity: Randomized Controlled Trial

François Bughin[1*], MD; Gaspard Bui[1*], MD; Bronia Ayoub[1]; Leo Blervaque[1], PhD; Didier Saey[2], PhD; Antoine Avignon[3,4], MD, PhD; Jean Frédéric Brun[1], MD, PhD; Nicolas Molinari[5,6], PhD; Pascal Pomies[1], PhD; Jacques Mercier[1], MD, PhD; Fares Gouzi[1], MD, PhD; Maurice Hayot[1], MD, PhD

[1]PhyMedExp, Centre Hospitalier Universitaire Montpellier, University of Montpellier, Institut national de la santé et de la recherche médicale, Centre national de la recherche scientifique, Montpellier, France

[2]Centre de Recherche, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Québec, QC, Canada

[3]Institut Desbrest de Santé Publique, University of Montpellier, Institut national de la santé et de la recherche médicale, Montpellier, France

[4]Endocrinologie-Diabétologie-Nutrition, Centre Hospitalier Universitaire Montpellier, University of Montpellier, Montpellier, France

[5]Institut Montpelliérain Alexander Grothendieck, University of Montpellier, Centre national de la recherche scientifique, Montpellier, France

[6]Department of Medical Statistics and Epidemiology, Centre Hospitalier Universitaire Montpellier, University of Montpellier, Montpellier, France

[*]these authors contributed equally

Corresponding Author:
François Bughin, MD
PhyMedExp, Centre Hospitalier Universitaire Montpellier
University of Montpellier
Institut national de la santé et de la recherche médicale, Centre national de la recherche scientifique
Avenue Doyen Gaston Giraud
Montpellier, 34090
France
Phone: 33 467338284
Email: f-bughin@chu-montpellier.fr

Abstract

Background: Obesity is a major public health issue. Combining exercise training, nutrition, and therapeutic education in metabolic rehabilitation (MR) is recommended for obesity management. However, evidence from randomized controlled studies is lacking. In addition, MR is associated with poor patient adherence. Mobile health devices improve access to MR components.

Objective: The aim of this study is to compare the changes in body composition, anthropometric parameters, exercise capacity, and quality of life (QOL) within 12 weeks of patients in the telerehabilitation (TR) program to those of usual care patients with obesity.

Methods: This was a parallel-design randomized controlled study. In total, 50 patients with obesity (BMI>30 kg/m²) were included in a TR group (TRG) or a usual care group (UCG) for 12 weeks. Patients underwent biometric impedance analyses, metabolic exercise tests, actimetry, and QOL and satisfaction questionnaires. The primary outcome was the change in fat mass at 12 weeks from baseline. Secondary outcomes were changes in body weight, metabolic parameters, exercise capacity, QOL, patients' adhesion, and satisfaction.

Results: A total of 49 patients completed the study. No significant group × time interaction was found for fat mass (TRG: mean 1.7 kg, SD 2.6 kg; UCG: mean 1.2 kg, SD 2.4 kg; *P*=.48). Compared with the UCG, TRG patients tended to significantly improve their waist to hip ratios (TRG: −0.01 kg, SD 0.04; UCG: +0.01 kg, SD 0.06; *P*=.07) and improved QOL physical impact (TRG: +21.8, SD 43.6; UCG: −1.2, SD 15.4; *P*=.005). Significant time effects were observed for body composition, 6-minute walk test distance, exercise metabolism, sedentary time, and QOL. Adherence (95%) and satisfaction in the TRG were good.

Conclusions: In adults with obesity, the TR program was not superior to usual care for improving body composition. However, TR was able to deliver full multidisciplinary rehabilitation to patients with obesity and improve some health outcomes. Given the patients' adherence and satisfaction, pragmatic programs should consider mobile health devices to improve access to MR. Further studies are warranted to further establish the benefits that TR has over usual care.

Trial Registration: ClinicalTrials.gov NCT03396666; http://clinicaltrials.gov/ct2/show/NCT03396666

(JMIR Mhealth Uhealth 2021;9(12):e28242) doi:10.2196/28242

KEYWORDS

telerehabilitation; mHealth; rehabilitation; obesity; mobile phone

Introduction

Background

Obesity is a chronic disease defined by a BMI of >30 kg/m² in the context of increased fat mass (FM). It is currently a highly prevalent disorder and a major public health issue [1]. It is associated with increased morbidity and mortality [2], including metabolic comorbidities, disabilities, and impaired quality of life (QOL). The metabolic risk is worst in cases of FM increase and when abdominal visceral fat predominates [3]. Thus, waist circumference (WC)—a marker of intraabdominal fat [4]—and the waist to hip ratio (WHR) better predict metabolic (ie, insulin sensitivity and lipid profile) and cardiovascular complications than BMI [5].

Obesity results in an imbalance between energy intake and energy expenditure [6,7]. Creating a negative energy balance can induce or maintain weight loss in patients with obesity [8]. Thus, physical activity (PA) and nutrition interventions are the cornerstones of obesity treatment, improving weight, WC, FM, and health outcomes [9]. Aerobic training alone induces significant weight loss in individuals with obesity. Specifically, light to moderate intensity corresponding to the intensity of the maximum lipid oxidation (LIPOXmax) individually determined in patients [10] has demonstrated significant weight, WC and FM reduction [8], as well as benefits on body composition and biological parameters (cholesterol and blood glucose) [11]. However, the most efficient strategy in obesity combines exercise training with nutrition interventions and therapeutic education [12,13] in a multidimensional metabolic rehabilitation (MR) for at least 12 weeks [14,15]. Although scientific societies recommend MR for patients with obesity [16], the benefits of such interventions remain to be compared with usual care alone.

However, the delivery of MR in the clinical field is a complex issue, and population-based trials have shown poor patient adherence (large dropout rates [17] and poor attendance [18]). This large underutilization of MR [19] is also because of the financial cost of such programs [20]. In the field of pulmonary rehabilitation, a widely developed domain, such barriers limiting the access to and delivery of rehabilitation have been well-described [21]. Thus, trials testing the effects of MR versus usual care—even if positive—would have limited clinical relevance because it is poorly applicable in patients with obesity.

The barriers of access to MR can be waived by recent technological innovations in the field of mobile devices. Mobile health (mHealth) facilities (smartphone-based educational apps, web-based tools, SMS text messaging, PDA physiological status monitoring, and connected captors) improved the delivery of the components of rehabilitation when taken individually [22]. In patients with obesity, a 10-week web-based exercise program has shown a significant effect on patients' FM [23]. In addition, mHealth nutrition management or therapeutic education had significant effects on body weight (BW) and BMI in obesity [24,25]. Thus, because mHealth facilities deliver full MR, a telerehabilitation (TR) program could be more efficient than usual care in patients with obesity. In addition, this pragmatic research approach based on affordable tools could provide evidence for real-world MR.

Objectives

Therefore, we developed a mobile TR solution for patients with obesity and used it in a blended multidisciplinary MR combining exercise training at LIPOXmax intensity, nutritional intervention, and educational tools. The aim of this randomized controlled study is to compare the changes in body composition, anthropometric parameters, exercise capacity, and QOL within 12 weeks of the TR program versus usual care in patients with obesity. In addition, feasibility, patients' adherence satisfaction, and effects of this TR were assessed in the TR group (TRG).

Methods

Study Population

Adults aged 25-65 years with a BMI of ≥30 kg/m² were eligible for participation. The main exclusion criteria were participants with a contraindication for exercise training (such as unstable cardiovascular disease or musculoskeletal problems).

Study Design

This was a 12-week, prospective, parallel-group, randomized controlled trial. Individuals were recruited from consultations of the Physiology Department of the University Hospital of Montpellier (France) and from the general population with media advertisements. After a screening period of 12 months, interested patients were contacted by email or phone and were registered on the Aviitam health platform. They were scheduled for half-day baseline assessments. All participants provided written informed consent. The study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) ethical guidelines and the CONSORT of Electronic and mHealth Applications and Online Telehealth checklist [26]. The study was approved by the ethics committee (CPP Nord-Ouest IV, France; ClinicalTrials.gov identifier: NCT03396666).

Patients were admitted to the Physiology Department of the University Hospital of Montpellier (France) between January 2018 and November 2018. Baseline assessments included physical examination, bioimpedance, blood test, effort calorimetry, 6-minute walk test (6MWT), and self-questionnaires. Once baseline assessments were completed, participants were randomized to either a 12-week TR program or usual care. The randomization sequence was computer-generated using random blocks in an order unknown

to investigators. The list was established by a statistician and was only accessible by the personnel in charge of randomization. Although participants could not be blinded to their treatment, both programs were presented as active interventions.

All tests and evaluations of the study were performed at the same place for each group under the same conditions and with the same devices. All assessments made at baseline were realized at the end of follow-up by technicians blinded to group allocation.

Intervention and Control Groups

Patients from both groups had a specialized medical consultation with cardiopulmonary exercise testing and prescription of an adapted PA program at an intensity that elicited maximal lipid oxidation (called LIPOXmax). The patients in the usual care group (UCG) were advised to carry out their sessions independently, focusing on endurance PA sessions such as brisk walking, cycling, or swimming. Moreover, these patients received a booklet with different exercises and tips on PA and nutrition management. All patients were registered on the Aviitam website before and during the trial. Aviitam is a highly secure health record that allows the centralization, protection, and sharing of medical data with doctors. No restrictive diet was prescribed in either group.

The TR program is a multicomponent intervention available on smartphones (Figure 1) and the website (Figure 2). Patients received a package containing a smartphone (Archos with Android operating system) on which the TR Telemouv app was installed. A pedometer (Care Trackfit) and a heart rate monitor (Polar H7) were connected to the smartphone via Bluetooth. The patients received secure access codes for the app and website. They were trained in the use of the program and connected objects by a technician and received an instruction booklet to guide their first steps in the TR program. Telemouv solution contains 3 components: PA, nutritional, and educational programs. After 1 week, a PA teacher went to the patients' homes to install a connected bicycle ergometer (Care Fitness) and performed the first supervised session. The exercise training program included endurance sessions, muscle reinforcement, and posture and balance exercises. Regarding endurance exercises, patients were advised to increase the volume and intensity of the sessions to reach the weekly goal of 150 minutes, with a minimum of 3 sessions per week, which could combine sessions on the connected cycloergometer and walking sessions. Wearing a connected heart rate monitor was recommended during endurance sessions to reach the target heart rate (corresponding to LIPOXmax). For muscle strengthening, balance, and posture exercises, patients had access to video sessions and were sent to the mobile solution throughout the program. Moreover, patients could track their daily step counts to reach their individualized goals. Nutrition management tools included hunger and satiety questionnaires and a 24-hour food intake questionnaire. Moreover, patients received daily educational content about illness, nutrition, and the benefits of PA. Patients with TRG had 2 teleconsultations at 1 and 2 months. In addition, doctors also had access to a secure website with access to patient data from the TRG.

Figure 1. Telerehabilitation mobile app.

Figure 2. Telerehabilitation website.

Outcomes

The primary outcome was the modification of the amount of FM, expressed in kilograms between baseline and 12-week follow-up in the TRG versus those in the UCG.

The main secondary outcomes were changes from baseline in body composition indexes (fat-free mass [FFM], muscular mass, and mass muscular index), anthropometric parameters (weight, BMI, WC, and WHR), and metabolism during exercise (maximal fat oxidation [MFO] and power at LIPOXmax and at crossover). Other outcomes were changes from baseline in PA and sedentary levels, exercise capacity, biological parameters, and self-assessment QOL questionnaires. The usability of the solution and satisfaction with the TR program were also assessed.

Measures

Body composition; weight; height; and waist, hip, and neck circumferences were measured after 12 hours of fasting.

Bioelectrical Impedance Analyses

Participants' body composition was assessed by bioimpedance analysis with a 6 terminal impedance plethysmograph (Biacorpus RX4000 software, BodyComp 8.4). This device measures the total resistance of the body to an alternating electric current of 50 kHz [27,28]. Body FM and FFM were calculated for each segment of the body according to the manufacturer's database-derived disclosed equations and total water with published equations using the classical cylindrical model and Hanai mixture theory [29]. FM, FFM, and muscular mass were expressed in kilograms and as a percentage of total body mass. Muscle mass index was calculated as muscular mass/height² and expressed in kg/m².

Anthropometric Parameters

Height was measured to the nearest 0.5 cm using a standardized height gauge. BMI was calculated as weight (kg)/height² (m). Neck, chest, waist, and hip circumference measurements were obtained using standardized procedures. WHR was then calculated.

Metabolic Exercise Test

The participants performed an exercise test on an electromagnetically braked cycle ergometer (Ergoline Bosch 500, Ergoline) connected to a breath-by-breath device (COSMED Quark cardiopulmonary exercise testing, COSMED) for gas exchange measurements. The theoretical maximal aerobic power (W_{max} th) was calculated for all patients using Wasserman equations [30]. After a fasting period of 12 hours, participants underwent a standardized submaximal exercise test [31] consisting of five 6-minute submaximal steady-state workloads (set at 30%, 40%, 50%, and 60% of W_{max} th), with a calculation of carbohydrate and lipid oxidation rates from gas exchange measurements at steady state at the 5th to 6th minute

of every step according to the nonprotein respiratory quotient technique [32]. Fat oxidation rates were calculated using the following equation:

$$\text{Fat (mg/min)} = -1.7012 \times VCO_2 + 1.6946 \times VO_2$$

(gas volume expressed in mL/min)

where, VCO_2 is carbon dioxide output and VO_2 is oxygen uptake.

After smoothing the curves, we calculated 2 parameters representative of the balance between fat and carbohydrate oxidation: the crossover point, which is the point at which carbohydrate becomes the predominant fuel representing more than 70% of the total energy [33] and the LIPOXmax point, where lipid oxidation reaches a maximum. The MFO rate is defined as the highest observed use of fat as an energy source during oxidative metabolism and is expressed in mg/min.

6MWT Overview

The 6MWT was performed at the hospital on a plane surface in a 30-m–long covered corridor marked every 2 minutes. The tests were conducted according to the recommendations of the American Thoracic Society [34]. Heart rate and oxyhemoglobin saturation were recorded every minute, and dyspnea scores were measured on a Borg scale at the end of the test. The total distance was then recorded.

Questionnaires

QOL was assessed with a questionnaire for the general population (36-item short form survey, SF-36) and one specific to the population of patients with obesity (echelle qualité de vie, obésité et diététique [EQVOD]). The PA level was assessed using the Voorrips questionnaire (modified Baecke questionnaire). The questionnaire scored the past year's household activities, sports activities, and other physically active leisure time activities and gave an overall PA score. The participants were asked to describe the type of activity, hours per week spent on it, and the period of the year in which the activity was normally performed. All activities were classified according to posture and movement. This questionnaire provides a reliable and valid method for classifying the activity level of older participants as high, medium, or low. With this method, normal participants with scores <9.4 are classified as having low PA.

The SF-36 is a generic self-reported measure of health-related QOL comprising 36 questions across 8 domains (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health). Answers to each question are rated on a Likert-type scale and summed to produce a raw score that is transformed to a scale of 0 to 100, with higher scores indicating a better QOL [35]. SF-36 subscales were computed to generate 2 summary measures: the physical component summary and the mental component summary. EQVOD is a French, validated scale specific to obesity, derived from the Impact of Weight on QOL questionnaire and its short version Impact of Weight on QOL-Lite [36]. The EQVOD questionnaire was adapted to the sociocultural factors of obesity and its dietetic treatment in France. It is easy to self-administer. Usability was evaluated with the System Usability Scale (SUS) [37]. It is a 10-item questionnaire with 5 response options for the respondents. An SUS score of >68 would be considered above average, and anything <68 is below average.

Blood Test

A venous blood sample was obtained in the fasting state to measure the lipid profile, plasma glucose, insulinemia, and C-reactive protein.

Actigraphy

Participants wore a GT3X accelerometer (ActiGraph) on the nondominant wrist, programmed to collect data from the vertical axis in 15-second epochs and initialized using a normal filter (AGNorm). Accelerometers were worn for 7 days during all waking hours and removed for sleeping and during water-based activities. The minimum wearing criteria was ≥4 days, with ≥8 hours of wearing time each day [38]. In addition to the daily steps, daily sedentary time in minutes and daily time spent in moderate to very vigorous activity were extracted from actimetry according to the manufacturer's specifications using the Freedson cutoff and the software (Actilife) provided by the company.

Power Calculations and Statistical Analysis

The calculation of the number needed to treat is based on a hypothesis supported by the literature [8]. We expect a difference between the 2 arms of 1 kg of fat loss with a common SD of 1 kg. For an α threshold of .05 and a study power of 90%, the study included 22 patients in each group. Considering a possible 10% dropout rate, the study will need to include 25 patients by randomized arm to demonstrate an effect.

The baseline characteristics of the 2 groups were compared using the independent 1-tailed *t* test or Mann-Whitney U test according to the data distribution. The intra- and intergroup kinetics of changes for the variables under intervention were analyzed with linear mixed effects models including a time effect, a group effect, and the interaction between these effects as a fixed factor and a subject effect as a random factor, using the nmle package in R. In case of significant interaction effect, false discovery rate–adjusted post hoc tests were performed. Linear mixed effects assumptions were tested before each test. Per-group analysis of the effect of the intervention was performed in the TRG group using paired *t* tests. The effect size was also calculated using Cohen *d*. Spearman rank order or Pearson correlations, depending on the data distribution, were used to determine associations between continuous variables. The data were analyzed using R software (R 4.0.3; R Foundation for Statistical Computing) and plotted using Prism Software. Statistical significance was set at $P<.05$.

Results

Patients

Of the 140 screened patients, 50 were included in the study and underwent randomization. A total of 49 patients completed the study (Figure 3). One patient in the TRG discontinued the trial before the endpoint without a postbaseline assessment. The baseline characteristics of both groups were not statistically

different for any of the parameters assessed (Table 1). All severities were represented, as class I, II, and III obesity represented 44% (22/50), 36% (18/50), and 20% (10/50) of patients, respectively. A total of 90% (45/50) of patients had a low PA level, as defined by a Voorrips score of <9.4. None of the patients were currently medically treated for obesity. A total of 8 patients had diabetes, and 7 were treated for dyslipidemia.

Figure 3. Study participant flowchart.

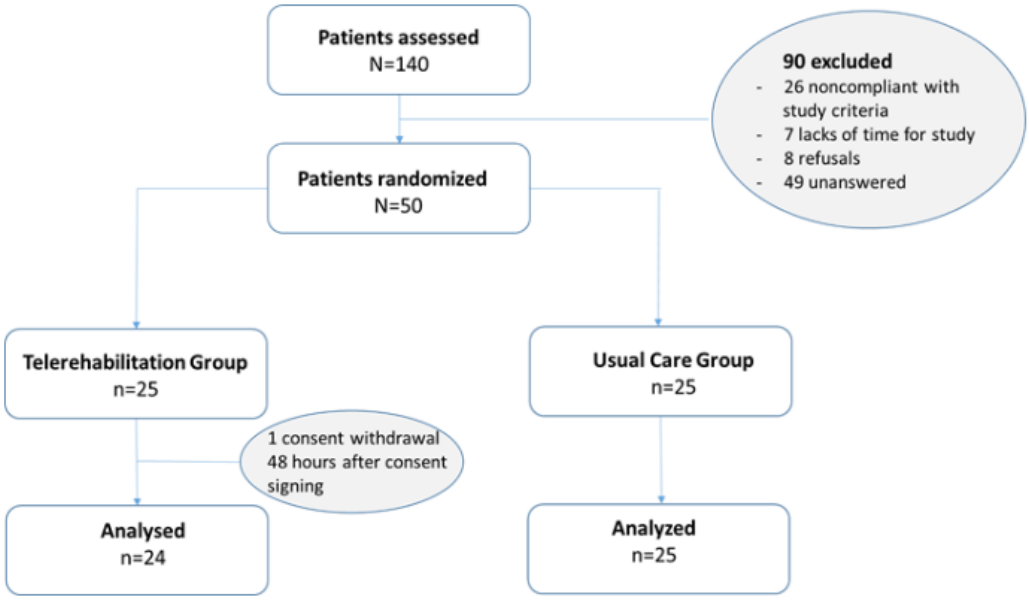


Table 1. Characteristics of participants.

Characteristic	Intervention group (n=25)	Control group (n=25)	*P* value
Male, n (%)	11 (44)	12 (48)	.78
Age (years), mean (SD)	51.2 (10.8)	53.3 (11.3)	.47
Active smoker, n (%)	5 (20)	1 (4)	.23
Voorrips score, mean (SD)	5.7 (3.0)	6.4 (3.0)	.37
Total blood cholesterol (mmol/L), mean (SD)	5.2 (1.1)	5.1 (1.1)	.16
Fasting blood glucose (mmol/L), mean (SD)	7.3 (3.0)	5.9 (1.1)	.26
Fat mass (kg), mean (SD)	44.8 (10.6)	43.6 (18.1)	.51
Body weight (kg), mean (SD)	105.7 (18.1)	104.4 (14.9)	.78
BMI (kg/m²), mean (SD)	36.2 (4.2)	36.82 (5.0)	.74
Waist circumference (cm), mean (SD)	120.1 (11.5)	120.4 (14.1)	.21
Waist to hip ratio, mean (SD)	0.96 (0.07)	0.97 (0.10)	.05
6-minute walk test distance (m), mean (SD)	511 (69.6)	514.8 (69.9)	.77
LIPOXmax[a] (W), mean (SD)	34.9 (11.6)	36.04 (10.3)	.87

[a]LIPOXmax: maximum lipid oxidation.

Effects of TR on Primary and Secondary Outcomes

No significant group or group × time interaction was found for the FM (Figure 2; Table 2). However, there was a significant time effect (*P*<.001), meaning that although not different between groups, an improvement in FM occurred in both study groups. An FM decrease was observed in the TRG (−1.7 kg, SD 2.6 kg; −4%, SD 6.2%) and in the in the UCG (−1.2 kg, SD 2.4 kg; −3%, SD 6.6%), with 48% (12/25) of patients improving the FM of >5% of initial values (Figure 4). Similarly, significant time effects, with no group × time interactions were observed for the FFM (%), muscle mass, 6MWT distance, crossover point, and power at the LIPOXmax and psychosocial component of the EQVOD (Table 2). Significant differences for a group × time interaction were found for the WHR (*P*=.07; Figure 5) and for the physical impact component of the EQVOD (*P*=.005; Figure 6), which was significantly increased in the TRG.

Table 2. Changes in primary and secondary outcomes between baseline and 12-week follow-up.

Outcome	Intervention group		Control group		*P* value (between group)		
	Baseline, mean (SD)	Follow-up, mean (SD)	Baseline, mean (SD)	Follow-up, mean (SD)	Group	Time	Interaction
Fat mass (kg)	44.80 (10.56)	43.18 (10.79)[a]	43.56 (12.19)	43.18 (13.18)[a]	.94	<.001	.48
Fat mass (%)	44.80 (10.56)	41.11 (7.50)[a]	41.88 (8.46)	41 (8.78)[a]	.86	<.001	.41
Fat-free mass (kg)	60.92 (13.35)	62.09 (14.08)	60.85 (11.75)	61.77 (12.03)	.95	.05	.45
Fat-free mass (%)	57.52 (7.17)	58.79 (7.52)	58.12 (8.46)	59 (8.78)	.84	<.001	.52
Muscle mass (kg)	26.99 (7.78)	28.23 (8.12)	27.33 (7.07)	27.98 (7.07)	.85	.02	.18
Muscle mass index (kg/m²)	9.10 (1.71)	9.49 (1.70)[a]	9.49 (1.83)	9.75 (1.95)	.39	.02	.26
Body weight (kg)	105.72 (18.06)	105.26 (19)	104.41 (14.86)	104.89 (16.69)	.89	.84	.41
BMI (kg/m²)	36.22 (4.15)	36.02 (4.40)	36.82 (5.00)	36.98 (5.72)	.54	.82	.41
Waist to hip ratio	0.96 (0.07)	0.95 (0.08)	0.97 (0.10)	0.99 (0.11)	.28	.56	.07
6-minute walk test distance (m)	511(70)	526 (71)	515 (70)	526 (67)	.90	.03	.75
Crossover point	63.96 (20.69)	72.96 (22.60)	66.48 (23.97)	71.16 (21.73)	.91	.002	.33
Power at crossover (W)	44.00 (17.16)	49.21 (16.52)	46.48 (14.21)	50.88 (18.10)	.67	<.001	.64
Power at LIPOXmax[b] (W)	34.92 (11.59)	36.88 (11.51)	36.04 (10.33)	38.12 (13.31)	.75	.02	.90
MFO[c] (mg/min)	270.5 (95.3)	298.4 (81.5)	301.5 (109.1)	303.9 (102.0)	.46	.14	.20
MFO (mg/min/kg FFM[d])	10.61 (4.08)	11.21 (3.11)	11.23 (3.64)	11.20 (3.45)	.70	.52	.47
SF-36[e] mental component	44.44 (12.43)	48.29 (10.06)	43.74 (11.97)	43.71 (12.89)	.48	.92	.13
SF-36 physical component	47.92 (7.31)	45.90 (8.46)	43.96 (9.23)	45.49 (10.18)	.26	.08	.34
EQVOD[f] physical impact	64 (17)	72 (16)[a]	72 (16)	72 (16)	.46	.004	.005
EQVOD psychosocial	68 (19)	75 (21)	64 (21)	66 (22)	.18	.01	.17

[a]*P*<.05 between baseline and follow-up (within group).

[b]LIPOXmax: maximum lipid oxidation.

[c]MFO: maximal fat oxidation.

[d]FFM: fat-free mass.

[e]SF-36: 36-item short form survey.

[f]EQVOD: echelle qualité de vie, obésité et diététique.

Figure 4. Relative change in fat mass after 12 weeks. TRG: telerehabilitation group; UCG: usual care group.

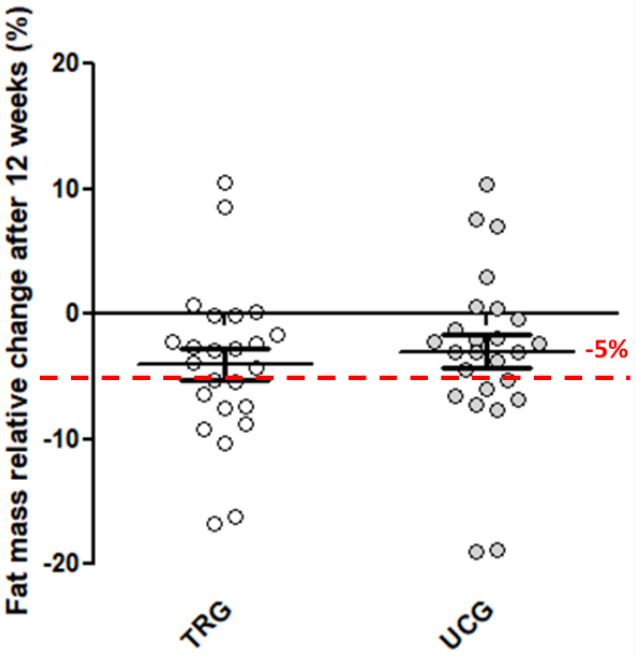


Figure 5. Waist to hip ratio change after 12 weeks (absolute). TRG: telerehabilitation group; UCG: usual care group.

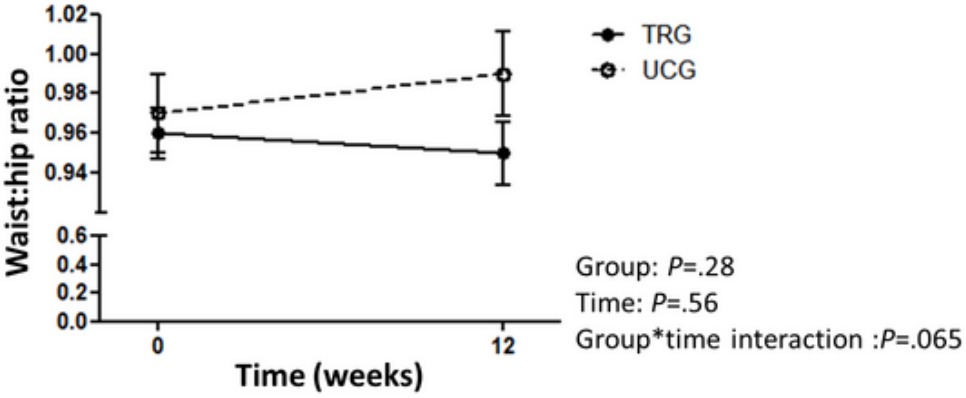
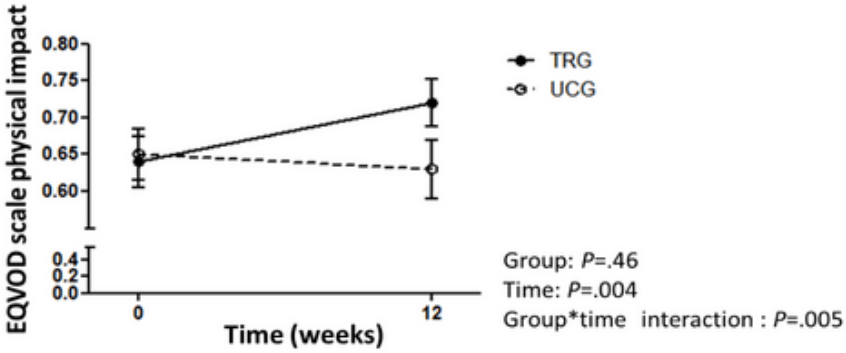


Figure 6. EQVOD scale scores for physical impact changes after 12 weeks (absolute). EQVOD: echelle qualité de vie, obésité et diététique; TRG: telerehabilitation group; UCG: usual care group.



Effect and Feasibility of the TR in Users

TR patients who completed the 12-week intervention performed an average of 30.5 (SD 16.7) sessions of PA and connected 283.5 (SD 193.4) times to the app and the website throughout the study. The SUS score averaged 65.3 (SD 15.02). A score of >68, indicating good usability of the system, was found in 46% (11/24) of patients. Only 1 patient found the program with poor usability, with an SUS score of <39. In the TRG, per-group analyses showed an improvement in body composition, exercise metabolism, and health-related QOL after the intervention. The FM (−1.7 kg, SD 2.6 kg; *P*=.004), FFM (+1.2%, SD 1.9%; *P*=.005), muscle mass (+0.9 kg, SD 2.0 kg; *P*=.03), crossover

point (+8.6, SD 14.0; *P*=.006), power at the crossover point and LIPOXmax (+5.8 W, SD 9.4 W, *P*=.006 and +2.48, SD 5.2 W, *P*=.03, respectively), SF-36 mental component (+3.53, SD 7.19; *P*=.04), and EQVOD physical and psychosocial components (+9.4, SD 14.1, *P*=.005 and +6.4, SD 11.7, *P*=.02, respectively) showed statistically significant improvements. Effect sizes ranged from small to moderate according to Cohen *d* (Table 3).

Table 3. Changes in parameters in the telerehabilitation group (paired *t* test and Cohen *d*).

Parameter	T0 (n=25), mean (SD)	T12 (n=25), mean (SD)	Difference (n=25), mean (SD)	*P* value	Effect size (Cohen *d*)	Magnitude
Fat mass (kg)	44.80 (10.56)	43.17 (10.79)	−1.70 (2.60)	.004	−0.65	Small
Fat mass (%)	42.48 (7.17)	41.11 (7.50)	−1.31 (1.94)	.003	−0.67	Moderate
Fat-free mass (kg)	60.92 (13.35)	62.09 (14.08)	0.86 (2.42)	.10	0.35	Moderate
Fat-free mass (%)	57.52 (7.17)	58.79 (7.52)	1.21 (1.89)	.005	0.64	Moderate
Muscle mass (kg)	26.99 (7.78)	28.23 (8.12)	0.94 (1.97)	.03	0.48	Small
Muscle mass index (kg/m²)	9.10 (1.71)	9.49 (1.70)	0.30 (0.61)	.03	0.49	Small
Body weight (kg)	105.72 (18.06)	105.26 (19.00)	−0.85 (2.82)	.16	−0.30	Small
BMI (kg/m²)	36.22 (4.15)	36.02 (4.40)	−0.31 (1.02)	.15	−0.31	Small
Waist to hip ratio	0.96 (0.07)	0.95 (0.08)	−0.01 (0.04)	.25	−0.24	Small
6-minute walk test distance (m)	511.00 (69.58)	526.46 (70.79)	13.88 (36.44)	.08	0.38	Small
Crossover point	63.96 (20.69)	72.96 (22.60)	8.58 (14.02)	.006	0.61	Moderate
Power at crossover point (W)	44.00 (17.16)	49.21 (16.52)	5.79 (9.43)	.006	0.61	Moderate
Power at LIPOXmax[a] (W)	34.92 (11.59)	36.88 (11.51)	2.38 (5.15)	.03	0.46	Small
Maximal fat oxidation (mg/min)	270.48 (95.25)	298.38 (81.45)	25.88 (73.12)	.10	0.35	Small
SF-36[b] mental component	44.44 (12.43)	48.29 (10.06)	3.53 (7.19)	.04	0.49	Small
SF-36 physical component	47.92 (7.31)	45.90 (8.46)	−1.44 (6.76)	.36	−0.21	Small
EQVOD[c] physical impact	63.70 (17.45)	72.25 (16.11)	9.44 (14.11)	.005	0.67	Moderate
EQVOD psychosocial	67.94 (19.32)	75.28 (21.30)	6.35 (11.66)	.02	0.54	Moderate

[a]LIPOXmax: maximum lipid oxidation.

[b]SF-36: 36-item short form survey.

[c]EQVOD: echelle qualité de vie, obésité et diététique.

Baseline and Intervention-Induced Change Correlations in Parameters

At baseline, univariate correlations between parameters were found in all patients with obesity. FM was correlated with BMI (*r*=0.850; *P*<.001). FFM and muscle mass were correlated with the crossover point (*r*=0.509, *P*<.001 and *r*=0.507, *P*<.001, respectively), LIPOXmax (*r*=0.495, *P*<.001 and *r*=0.469, *P*<.001, respectively), and MFO (*r*=0.365, *P*=.009 and *r*=0.34, *P*=.02, respectively). The 6MWT distance was correlated with muscle mass (*r*=0.316; *P*=.03), LIPOXmax (*r*=0.275; *P*=.05), and the SF-36 physical impact (*r*=0.301; *P*=.05). EQVOD's psychosocial score was correlated with FM (*r*=−0.338; *P*=.02) and WHR (*r*=0.281; *P*=.05). In addition, at the end of the 12-week trial, the change in BW was correlated with relative changes in FM (%) and FFM (%) in the TRG and UCG (TRG: *r*=0.598, *P*<.001 and *r*=0.670, *P*<.001, respectively; UCG: *r*=0.616, *P*=.01 and *r*=0.426, *P*=.04, respectively). In the whole population (N=49), BW change was inversely correlated with the 6MWT distance (*r*=−0.281; *P*=.05). In contrast, the number of training sessions and changes in WHR or FM (*P*=.56 and *P*=.26, respectively) and the number of connections and changes in WHR or FM (*P*=.86 and *P*=.69, respectively) were not significantly correlated.

Discussion

Principal Findings

This study was one of the first to propose the use of a mobile TR program using mHealth devices to deliver full MR in patients with obesity. Although our study did not show significant additional benefits versus usual care regarding the primary outcome (FM), there was a significant advantage regarding the domain of QOL and tendency for the WHR. In addition, per-group analysis indicated that the significant time effects on body composition, exercise capacity, PA behavior, and QOL were mainly because of the TRG. These effects occurred while the TR solution's adherence and usability were good during the 12-week trial duration.

Comparison of the Effects of TR Versus Usual Care in Patients With Obesity

Scientific societies have recommended that patients with obesity should benefit from a multidisciplinary program including

exercise, diet, and cognitive behavioral therapy [16]. However, this recommendation was not based on randomized controlled trial evidence, except for one study [39]. In patients with obesity, rehabilitation improvements in BW, exercise capacity, and comorbidities have not been compared with those of a control group [40-42]. Thus, our randomized controlled trial fills a gap in scientific knowledge regarding obesity treatment. Although our study did not show significant improvement in FM—the primary outcome of the study—with TR versus usual care, numerous observations have to be underlined. Among the secondary outcomes, we observed a significant time × group effect for the physical impact domain of the EQVOD. This result is in line with the improvement of SF-36–assessed QOL previously reported in patients with obesity by MR [43], particularly after 12 weeks [44]. There was also a tendency for the WHR in favor of the TRG versus UCG (P=.07), which is consistent with previous reports of the effect on WC induced by multidisciplinary rehabilitation [45] or 12-week exercise training [46]. This effect of the TR would be clinically relevant, given the critical role of the visceral fat accumulation in the patients' comorbidities [3,47]. As power calculation was performed on the basis of expected change in FM, further studies are required to confirm the effect of the TR on these secondary outcomes. Nonetheless, the mixed model showed significant time effects for most of the secondary outcomes (FM [kg, %], FFM [%], muscle mass [kg], muscle mass index, 6MWT distance [m], crossover point, intensity at the LIPOXmax, sedentary time [%], SF-36, and EQVOD), which requires per-group analyses to complete the interpretation.

Effects in the UCG

Detailed, structured PA counseling in daily life was provided to the UCG patients in line with the guidelines for obesity [48]. The intensity of the endurance exercises was set at LIPOXmax, an individual intensity determined on the metabolic exercise test, but there was no supervision. Therefore, UCG patients did not benefit from the metabolic effects reported after 8-12 weeks of supervised training at LIPOXmax intensity [8,49]. Similarly, patients also benefited from nutrition counseling through the Aviitam platform registration, but the nutritional intervention was not supervised. Thus, the metabolic effects were logically not significant and close to those reported in previous UCGs in obesity [50]. Altogether, this means that although optimized standard care with specific assessments, prescription, and counseling was provided to the UCG patients, its short-term impact on the patients' metabolism was limited.

Effects in the TRG

In contrast, patients in the TRG showed significant improvements in body composition (FM), exercise capacity (6MWT distance), exercise metabolism (intensity at the crossover point and LIPOXmax), health-related QOL (SF-36 and EQVOD), biological parameters (total and low-density lipoprotein cholesterol), and sedentary time in each group analysis (Multimedia Appendices 1 and 2). All these parameters constitute classical outcomes that are improved by multidisciplinary MR. The results showed internal validity because physiological correlations were found between study parameters at baseline (FFM and exercise metabolism, 6 MWT distance and muscle mass, LIPOXmax, and SF-36), and after 12 weeks. The magnitude of the BW loss after 3 months in the TRG (−2.82%, SD 2.81%) appeared to be limited. However, this is in line with the 6% BW loss after 12 months of nonsurgical clinical obesity services [17,39]. In addition, clinical benefits have been reported in randomized controlled trials reporting weight loss of ≤3%, when mediated by physical exercise [51], particularly regarding body composition [52]. Training combined with diet induced a 5.1% reduction in FM [53], and the 4% (SD 6%) FM decrease in our TRG was consistent with previous studies. The effect size for FM loss was medium (Cohen d>0.50) and reached 5% or more in 48% (12/25) of patients in the TRG. These effects appear to be particularly relevant, as FM loss appears to be the best predictor of physical functioning improvement during weight loss in patients with obesity [13]. In addition, FM loss has been associated with the improvement of systemic inflammation and lipid profiles [47,54]. The decrease in total and low-density lipoprotein cholesterol (Multimedia Appendix 2) in the TRG supports the clinical relevance of FM loss in the TRG. The absolute and relative increase in intensity at the crossover point was +17% (SD 27%), less than previously reported [49], but Cohen d was 0.644, which indicates a medium improvement exercise metabolism in patients with obesity. In addition, given that the minimally clinically important difference has been estimated from 2 to 4 points of the SF-36 physical component summary or mental component summary [55,56], 25% to 35% and 5% to 10% of the patients with obesity in the TRG experienced clinically significant improvements in the physical and mental components of QOL, respectively. Altogether, results in patients with obesity of the TRG support mild to medium effects of multidisciplinary TR, with benefits being clinically in a significant, relevant proportion of patients.

mHealth to Foster Adherence to MR in Patients With Obesity

TR has been studied in several conditions such as stroke [57], chronic obstructive pulmonary disease [58], cardiac diseases [59,60], diabetes [61], or neurodegenerative diseases [62]. To the best of our knowledge, this study is the first to test a TR solution in a population of patients with obesity. Current mHealth interventions for obesity have been limited to self-management, self-monitoring, PA, or nutrition or education alone [63]. A meta-analysis of mHealth in obesity has shown heterogeneous evidence of health outcomes [25]. However, establishing the impact of mHealth-based MR was not the aim of our study. In accordance with previous studies, our strategy was to develop an mHealth device to improve access to MR in patients with obesity. Attendance and dropout represent a critical issue in rehabilitation because previous studies have shown that the highest clinical benefits were seen in participants or patients with obesity with the highest attendance [64,65]. Accordingly, we found that the adherence of the 12-week program reached 95% in the TRG, with only 1 dropout. This is largely above the dropout rate reported in mHealth studies [66]. Dropout rates usually range from 43% to 62% over 6-24 months during multidisciplinary rehabilitation for patients with obesity [17]. This high attendance in the TRG was in line with the SUS score showing that the solution was acceptable to the patients. Thus,

our study showed that our TR solution succeeded in overcoming some of the barriers to face-to-face rehabilitation and appeared to be a relevant tool to deliver MR in patients with obesity.

Study Limitations

One limitation of the study was the lack of sufficient objective monitoring of the intervention in the TRG. Patients experienced difficulties, mainly secondary to connectivity defects with Bluetooth, to use heart rate monitors, pedometers, and cycle ergometers. Therefore, too few data were collected to monitor the intervention correctly. However, our study was a pragmatic trial, and all analyses were intention-to-treat analyses, which means that the impact of these missing data on the results was limited. Nonetheless, monitoring the intervention would have allowed for a better understanding of the effects of the TRG in patients with obesity, particularly in terms of the patients' phenotype in response to the intervention. In addition, the information technology firm that codeveloped the TR solution with our team did not have the opportunity to implement push notifications and provide pertinent feedback to patients. Finally, it is probable that patients need a longer and more stimulant intervention. Therefore, long-term intervention and the addition of human support, for example via videoconferencing, could address these limitations and improve the outcomes of TR programs. Altogether, the development of a TR solution remains an issue that must be addressed from the global perspective of our mHealth project.

Conclusions

In adults with obesity, our TR program was able to deliver full MR but did not demonstrate superiority to the usual care on body composition. Over a period of 12 weeks, it induced effects on most rehabilitation outcomes in patients with obesity (body composition, total cholesterol, and lipid oxidation during exercise). These effects were not significantly superior to those induced in our UCG. However, the excellent patient adherence to the TR constitutes an answer to the challenge of patient adherence to face-to-face rehabilitation programs. Pragmatic MR programs should consider mHealth devices to deliver interventions. In parallel to the continuous development of technological solutions, large-scale and long-term studies are needed to translate these technological promises into fully efficient interventions in the clinical field.

Acknowledgments

This was not an industry-supported study. This work was realized thanks to the financing of AxLR, société d'accélération du transfert de technologies (SATT) Occitanie Méditerranée.

Authors' Contributions

FB and MH were involved in the design of the work, investigation, data curation, funding acquisition, and writing. GB and BA were involved in the investigation and data curation. LB was involved in the formal analysis. DS was involved in the data curation. NM was involved in the methodology and formal analysis. FG was involved in the data curation, formal analysis, and writing. All authors contributed to and reviewed the final version of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Changes in activity level between baseline and 12-week follow-up in per-protocol population.
[PNG File , 27 KB - mhealth_v9i12e28242_app1.png]

Multimedia Appendix 2
Changes in biological parameters between baseline and 12-week follow-up in per-protocol population.
[PNG File , 46 KB - mhealth_v9i12e28242_app2.png]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 12070 KB - mhealth_v9i12e28242_app3.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. Global BMI Mortality Collaboration, Di Angelantonio E, Bhupathiraju S, Wormser D, Gao P, Kaptoge S, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet 2016 Dec 20;388(10046):776-786 [FREE Full text] [doi: 10.1016/S0140-6736(16)30175-1] [Medline: 27423262]

3. Després J. Body fat distribution and risk of cardiovascular disease: an update. Circulation 2012 Sep 04;126(10):1301-1313. [doi: 10.1161/CIRCULATIONAHA.111.067264] [Medline: 22949540]
4. Pouliot MC, Després JP, Lemieux S, Moorjani S, Bouchard C, Tremblay A, et al. Waist circumference and abdominal sagittal diameter: best simple anthropometric indexes of abdominal visceral adipose tissue accumulation and related cardiovascular risk in men and women. Am J Cardiol 1994 Mar 01;73(7):460-468. [doi: 10.1016/0002-9149(94)90676-9] [Medline: 8141087]
5. Kissebah AH, Vydelingum N, Murray R, Evans DJ, Hartz AJ, Kalkhoff RK, et al. Relation of body fat distribution to metabolic complications of obesity. J Clin Endocrinol Metab 1982 Feb;54(2):254-260. [doi: 10.1210/jcem-54-2-254] [Medline: 7033275]
6. Martinez JA. Body-weight regulation: causes of obesity. Proc Nutr Soc 2000 Aug;59(3):337-345. [doi: 10.1017/s0029665100000380] [Medline: 10997649]
7. Sellayah D, Cagampang FR, Cox RD. On the evolutionary origins of obesity: a new hypothesis. Endocrinology 2014 May;155(5):1573-1588. [doi: 10.1210/en.2013-2103] [Medline: 24605831]
8. Romain A, Carayol M, Desplan M, Fedou C, Ninot G, Mercier J, et al. Physical activity targeted at maximal lipid oxidation: a meta-analysis. J Nutr Metab 2012;2012:285395 [FREE Full text] [doi: 10.1155/2012/285395] [Medline: 22928092]
9. Chin S, Kahathuduwa CN, Binks M. Physical activity and obesity: what we know and what we need to know. Obes Rev 2016 Dec;17(12):1226-1244. [doi: 10.1111/obr.12460] [Medline: 27743411]
10. Amaro-Gahete FJ, Sanchez-Delgado G, Jurado-Fasoli L, De-la-O A, Castillo MJ, Helge JW, et al. Assessment of maximal fat oxidation during exercise: a systematic review. Scand J Med Sci Sports 2019 Jul;29(7):910-921. [doi: 10.1111/sms.13424] [Medline: 30929281]
11. Brun J, Romain A, Mercier J. Maximal lipid oxidation during exercise (Lipoxmax): from physiological measurements to clinical applications. Facts and uncertainties. Sci Sports 2011 Apr;26(2):57-71. [doi: 10.1016/j.scispo.2011.02.001]
12. 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. J Am Coll Cardiol 2014 Jul 01;63(25 Pt B):2985-3023 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.004] [Medline: 24239920]
13. Unick JL, Beavers D, Bond DS, Clark JM, Jakicic JM, Kitabchi AE, Look AHEAD Research Group. The long-term effectiveness of a lifestyle intervention in severely obese individuals. Am J Med 2013 Mar;126(3):236-42, 242.e1 [FREE Full text] [doi: 10.1016/j.amjmed.2012.10.010] [Medline: 23410564]
14. Baillot A, Romain AJ, Boisvert-Vigneault K, Audet M, Baillargeon JP, Dionne IJ, et al. Effects of lifestyle interventions that include a physical activity component in class II and III obese individuals: a systematic review and meta-analysis. PLoS One 2015;10(4):e0119017 [FREE Full text] [doi: 10.1371/journal.pone.0119017] [Medline: 25830342]
15. Ross R, Janssen I, Dawson J, Kungl A, Kuk JL, Wong SL, et al. Exercise-induced reduction in obesity and insulin resistance in women: a randomized controlled trial. Obes Res 2004 May;12(5):789-798. [doi: 10.1038/oby.2004.95] [Medline: 15166299]
16. Giusti V. [Obesity: epidemiology, socio-political implications and conventional management]. Ther Umsch 2019 Sep;76(3):117-121. [doi: 10.1024/0040-5930/a001071] [Medline: 31498046]
17. Brown TJ, O'Malley C, Blackshaw J, Coulton V, Tedstone A, Summerbell C, et al. Exploring the evidence base for Tier 3 weight management interventions for adults: a systematic review. Clin Obes 2017 Oct;7(5):260-272. [doi: 10.1111/cob.12204] [Medline: 28695579]
18. Dansinger ML, Gleason JA, Griffith JL, Selker HP, Schaefer EJ. Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: a randomized trial. JAMA 2005 Jan 05;293(1):43-53. [doi: 10.1001/jama.293.1.43] [Medline: 15632335]
19. Seida JC, Sharma AM, Johnson JA, Forhan M. Hospital rehabilitation for patients with obesity: a scoping review. Disabil Rehabil 2018 Jan;40(2):125-134. [doi: 10.1080/09638288.2016.1243163] [Medline: 27848247]
20. Lehnert T, Streltchenia P, Konnopka A, Riedel-Heller SG, König H. Health burden and costs of obesity and overweight in Germany: an update. Eur J Health Econ 2015 Dec;16(9):957-967. [doi: 10.1007/s10198-014-0645-x] [Medline: 25381038]
21. Rochester CL, Vogiatzis I, Holland AE, Lareau SC, Marciniuk DD, Puhan MA, ATS/ERS Task Force on Policy in Pulmonary Rehabilitation. An Official American Thoracic Society/European Respiratory Society Policy Statement: enhancing implementation, use, and delivery of pulmonary rehabilitation. Am J Respir Crit Care Med 2015 Dec 01;192(11):1373-1386. [doi: 10.1164/rccm.201510-1966ST] [Medline: 26623686]
22. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Mar 03;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]
23. Ballin M, Hult A, Björk S, Lundberg E, Nordström P, Nordström A. Web-based exercise versus supervised exercise for decreasing visceral adipose tissue in older adults with central obesity: a randomized controlled trial. BMC Geriatr 2020 May 12;20(1):173 [FREE Full text] [doi: 10.1186/s12877-020-01577-w] [Medline: 32398024]

24. Park S, Hwang J, Choi Y. Effect of mobile health on obese adults: a systematic review and meta-analysis. Healthc Inform Res 2019 Jan;25(1):12-26 [FREE Full text] [doi: 10.4258/hir.2019.25.1.12] [Medline: 30788177]
25. Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
26. Eysenbach G, CONSORT- E Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
27. Brun J, Guiraudou M, Mardemootoo C, Traoré A, Raingeard I, Chalançon A, et al. Validation de la mesure segmentaire de la composition corporelle en comparaison avec la DEXA : intérêt de la mesure de la masse grasse tronculaire. Sci Sports 2013 Jun;28(3):158-162. [doi: 10.1016/j.scispo.2012.09.002]
28. Ling CH, de Craen AJ, Slagboom PE, Gunn DA, Stokkel MP, Westendorp RG, et al. Accuracy of direct segmental multi-frequency bioimpedance analysis in the assessment of total body and segmental body composition in middle-aged adult population. Clin Nutr 2011 Oct;30(5):610-615 [FREE Full text] [doi: 10.1016/j.clnu.2011.04.001] [Medline: 21555168]
29. Jaffrin MY, Morel H. Body fluid volumes measurements by impedance: a review of bioimpedance spectroscopy (BIS) and bioimpedance analysis (BIA) methods. Med Eng Phys 2008 Dec;30(10):1257-1269. [doi: 10.1016/j.medengphy.2008.06.009] [Medline: 18676172]
30. Wasserman K, Hansen JE, Sue DY, Whipp BJ, Froelicher VF. Principles of exercise testing and interpretation. J Cardiopulmonary Rehab 1987;7(4):189. [doi: 10.1097/00008483-198704000-00014]
31. Pérez-Martin A, Dumortier M, Raynaud E, Brun JF, Fédou C, Bringer J, et al. Balance of substrate oxidation during submaximal exercise in lean and obese people. Diabetes Metab 2001 Sep;27(4 Pt 1):466-474. [Medline: 11547220]
32. Brun J, Varlet-Marie E, Jrme A, Mercier J. Measurement and Physiological Relevance of the Maximal Lipid Oxidation Rate During Exercise (LIPOXmax). In: An International Perspective on Topics in Sports Medicine and Sports Injury. London: InTech; 2012.
33. Brooks GA, Mercier J. Balance of carbohydrate and lipid utilization during exercise: the "crossover" concept. J Appl Physiol (1985) 1994 Jun;76(6):2253-2261. [doi: 10.1152/jappl.1994.76.6.2253] [Medline: 7928844]
34. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002 Jul 01;166(1):111-117. [doi: 10.1164/ajrccm.166.1.at1102] [Medline: 12091180]
35. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994 Jan;32(1):40-66. [doi: 10.1097/00005650-199401000-00004] [Medline: 8277801]
36. Kolotkin RL, Chen S, Klassen P, Gilder K, Greenway FL. Patient-reported quality of life in a randomized placebo-controlled trial of naltrexone/bupropion for obesity. Clin Obes 2015 Oct;5(5):237-244. [doi: 10.1111/cob.12108] [Medline: 26222044]
37. Brooke J. SUS: a 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996.
38. Demeyer H, Burtin C, Van Remoortel H, Hornikx M, Langer D, Decramer M, et al. Standardizing the analysis of physical activity in patients with COPD following a pulmonary rehabilitation program. Chest 2014 Aug;146(2):318-327 [FREE Full text] [doi: 10.1378/chest.13-1968] [Medline: 24603844]
39. Lih A, Pereira L, Bishay RH, Zang J, Omari A, Atlantis E, et al. A novel multidisciplinary intervention for long-term weight loss and glycaemic control in obese patients with diabetes. J Diabetes Res 2015;2015:729567 [FREE Full text] [doi: 10.1155/2015/729567] [Medline: 25950007]
40. McCombie L, Brosnahan N, Ross H, Bell‐Higgs A, Govan L, Lean ME. Filling the intervention gap: service evaluation of an intensive nonsurgical weight management programme for severe and complex obesity. J Hum Nutr Diet 2018 Nov 22;32(3):329-337. [doi: 10.1111/jhn.12611] [Medline: 30565772]
41. Atlantis E, Langford K, Piya M, Ho V, Skelsey K, Rickards L, et al. Physical capacity outcomes in patients with severe obesity after 12 months of physician-led multidisciplinary team care: a case series from a public hospital clinical obesity service. Clin Obes 2019 Dec;9(6):e12337. [doi: 10.1111/cob.12337] [Medline: 31475476]
42. Medveczky D, Kodsi R, Skelsey K, Grudzinskas K, Bueno F, Ho V, et al. Class 3 obesity in a multidisciplinary metabolic weight management program: the effect of preexisting type 2 diabetes on 6-month weight loss. J Diabetes Res 2020;2020:9327910 [FREE Full text] [doi: 10.1155/2020/9327910] [Medline: 32832564]
43. Khammassi M, Miguet M, O'Malley G, Fillon A, Masurier J, Damaso AR, et al. Health-related quality of life and perceived health status of adolescents with obesity are improved by a 10-month multidisciplinary intervention. Physiol Behav 2019 Oct 15;210:112549. [doi: 10.1016/j.physbeh.2019.05.010] [Medline: 31082444]
44. Lemstra ME, Rogers MR. Improving health-related quality of life through an evidence-based obesity reduction program: the Healthy Weights Initiative. J Multidiscip Healthc 2016 Mar 7;9:103-109 [FREE Full text] [doi: 10.2147/JMDH.S100693] [Medline: 27022273]

45. Budui S, Bigolin F, Giordano F, Leoni S, Berteotti M, Sartori E, et al. Effects of an intensive inpatient rehabilitation program in elderly patients with obesity. Obes Facts 2019;12(2):199-210 [FREE Full text] [doi: 10.1159/000497461] [Medline: 30928989]
46. Herring LY, Wagstaff C, Scott A. The efficacy of 12 weeks supervised exercise in obesity management. Clin Obes 2014 Aug;4(4):220-227. [doi: 10.1111/cob.12063] [Medline: 25826793]
47. Després JP, Pouliot MC, Moorjani S, Nadeau A, Tremblay A, Lupien PJ, et al. Loss of abdominal fat and metabolic response to exercise training in obese women. Am J Physiol 1991 Aug;261(2 Pt 1):E159-E167. [doi: 10.1152/ajpendo.1991.261.2.E159] [Medline: 1872379]
48. Kahan S, Manson JE. Obesity treatment, beyond the guidelines: practical suggestions for clinical practice. JAMA 2019 Apr 09;321(14):1349-1350. [doi: 10.1001/jama.2019.2352] [Medline: 30896727]
49. Bordenave S, Metz L, Flavier S, Lambert K, Ghanassia E, Dupuy A, et al. Training-induced improvement in lipid oxidation in type 2 diabetes mellitus is related to alterations in muscle mitochondrial activity. Effect of endurance training in type 2 diabetes. Diabetes Metab 2008 Apr;34(2):162-168. [doi: 10.1016/j.diabet.2007.11.006] [Medline: 18396088]
50. Ross R, Janssen I. Physical activity, total and regional obesity: dose-response considerations. Med Sci Sports Exerc 2001 Jun;33(6 Suppl):S521-7; discussion S528. [doi: 10.1097/00005768-200106001-00023] [Medline: 11427779]
51. Ross R, Hudson R, Stotz PJ, Lam M. Effects of exercise amount and intensity on abdominal obesity and glucose tolerance in obese adults: a randomized trial. Ann Intern Med 2015 Mar 03;162(5):325-334. [doi: 10.7326/M14-1189] [Medline: 25732273]
52. Ramírez-Vélez R, Izquierdo M, Castro-Astudillo K, Medrano-Mena C, Monroy-Díaz AL, Castellanos-Vega RD, et al. Weight loss after 12 weeks of exercise and/or nutritional guidance is not obligatory for induced changes in local fat/lean mass indexes in adults with excess of adiposity. Nutrients 2020 Jul 26;12(8):2231 [FREE Full text] [doi: 10.3390/nu12082231] [Medline: 32722652]
53. Millstein R. Measuring outcomes in adult weight loss studies that include diet and physical activity: a systematic review. J Nutr Metab 2014;2014:421423 [FREE Full text] [doi: 10.1155/2014/421423] [Medline: 25525513]
54. Sarin HV, Lee JH, Jauhiainen M, Joensuu A, Borodulin K, Männistö S, et al. Substantial fat mass loss reduces low-grade inflammation and induces positive alteration in cardiometabolic factors in normal-weight individuals. Sci Rep 2019 Mar 05;9(1):3450 [FREE Full text] [doi: 10.1038/s41598-019-40107-6] [Medline: 30837600]
55. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000 Jul;43(7):1478-1487 [FREE Full text] [doi: 10.1002/1529-0131(200007)43:7<1478::AID-ANR10>3.0.CO;2-M] [Medline: 10902749]
56. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001 Aug;45(4):384-391 [FREE Full text] [doi: 10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0] [Medline: 11501727]
57. Tchero H, Tabue Teguo M, Lannuzel A, Rusch E. Telerehabilitation for stroke survivors: systematic review and meta-analysis. J Med Internet Res 2018 Oct 26;20(10):e10867 [FREE Full text] [doi: 10.2196/10867] [Medline: 30368437]
58. Bernocchi P, Vitacca M, La Rovere MT, Volterrani M, Galli T, Baratti D, et al. Home-based telerehabilitation in older patients with chronic obstructive pulmonary disease and heart failure: a randomised controlled trial. Age Ageing 2018 Dec 01;47(1):82-88. [doi: 10.1093/ageing/afx146] [Medline: 28985325]
59. Dinesen B, Nielsen G, Andreasen JJ, Spindler H. Integration of rehabilitation activities into everyday life through telerehabilitation: qualitative study of cardiac patients and their partners. J Med Internet Res 2019 Apr 15;21(4):e13281 [FREE Full text] [doi: 10.2196/13281] [Medline: 30985284]
60. Subedi N, Rawstorn JC, Gao L, Koorts H, Maddison R. Implementation of telerehabilitation interventions for the self-management of cardiovascular disease: systematic review. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e17957 [FREE Full text] [doi: 10.2196/17957] [Medline: 33245286]
61. Duruturk N, Özköslü MA. Effect of tele-rehabilitation on glucose control, exercise capacity, physical fitness, muscle strength and psychosocial status in patients with type 2 diabetes: a double blind randomized controlled trial. Prim Care Diabetes 2019 Dec;13(6):542-548. [doi: 10.1016/j.pcd.2019.03.007] [Medline: 31014938]
62. Cotelli M, Manenti R, Brambilla M, Gobbi E, Ferrari C, Binetti G, et al. Cognitive telerehabilitation in mild cognitive impairment, Alzheimer's disease and frontotemporal dementia: A systematic review. J Telemed Telecare 2019 Feb;25(2):67-79. [doi: 10.1177/1357633X17740390] [Medline: 29117794]
63. Wang SH, Keenan BT, Wiemken A, Zang Y, Staley B, Sarwer DB, et al. Effect of weight loss on upper airway anatomy and the apnea-hypopnea index. The importance of tongue fat. Am J Respir Crit Care Med 2020 Mar 15;201(6):718-727 [FREE Full text] [doi: 10.1164/rccm.201903-0692OC] [Medline: 31918559]
64. MacLaughlin HL, Hall WL, Condry J, Sanders TA, Macdougall IC. Participation in a structured weight loss program and all-cause mortality and cardiovascular morbidity in obese patients with chronic kidney disease. J Ren Nutr 2015 Nov;25(6):472-479. [doi: 10.1053/j.jrn.2015.05.001] [Medline: 26143293]

65. Counterweight Project Team. Evaluation of the Counterweight Programme for obesity management in primary care: a starting point for continuous improvement. Br J Gen Pract 2008 Aug;58(553):548-554 [FREE Full text] [doi: 10.3399/bjgp08X319710] [Medline: 18682018]
66. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462. [doi: 10.3945/an.116.014100] [Medline: 28507010]

Abbreviations

6MWT: 6-minute walk test
BW: body weight
CONSORT: Consolidated Standards of Reporting Trials
EQVOD: echelle qualité de vie, obésité et diététique
FFM: fat-free mass
FM: fat mass
LIPOXmax: maximum lipid oxidation
MFO: maximal fat oxidation
mHealth: mobile health
MR: metabolic rehabilitation
PA: physical activity
QOL: quality of life
SUS: System Usability Scale
TR: telerehabilitation
TRG: telerehabilitation group
UCG: usual care group
WC: waist circumference
WHR: waist to hip ratio

Edited by L Buis; submitted 25.02.21; peer-reviewed by P Bernocchi, M Vitacca; comments to author 15.04.21; revised version received 10.06.21; accepted 17.09.21; published 06.12.21.

Please cite as:
Bughin F, Bui G, Ayoub B, Blervaque L, Saey D, Avignon A, Brun JF, Molinari N, Pomies P, Mercier J, Gouzi F, Hayot M
Impact of a Mobile Telerehabilitation Solution on Metabolic Health Outcomes and Rehabilitation Adherence in Patients With Obesity: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(12):e28242
URL: https://mhealth.jmir.org/2021/12/e28242
doi:10.2196/28242
PMID:34874887

Original Paper

Personalized Reminders for Immunization Using Short Messaging Systems to Improve Human Papillomavirus Vaccination Series Completion: Parallel-Group Randomized Trial

Chelsea S Wynn[1], MPH; Marina Catallozzi[1,2,3], MSCE, MD; Chelsea A Kolff[1,2], MPH; Stephen Holleran[1], BSc; Dodi Meyer[1,3], MD; Rajasekhar Ramakrishnan[1], EngScD; Melissa S Stockwell[1,2,3], MPH, MD

[1]Division of Child and Adolescent Health, Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, United States

[2]Department of Population and Family Health, Mailman School of Public Health, Columbia University Irving Medical Center, New York, NY, United States

[3]NewYork-Presbyterian Hospital, New York, NY, United States

Corresponding Author:
Melissa S Stockwell, MPH, MD
Division of Child and Adolescent Health
Department of Pediatrics
Columbia University Irving Medical Center
622 W 168th St
VC 417
New York, NY, 10032
United States
Phone: 1 212 342 5732
Email: mss2112@cumc.columbia.edu

Abstract

Background: Completion rates among adolescents who initiate the human papillomavirus (HPV) vaccine 3-dose series are low. SMS text message vaccine reminders are effective, but less is known about the best types for HPV series completion or the ability to assess and target vaccine decision-making stage.

Objective: The aim of this study is to compare the effectiveness of HPV vaccine series completion in minority adolescents who received precision and educational versus conventional SMS text message reminders.

Methods: Enrolled parents of adolescents aged 9-17 years who received the first HPV vaccine dose at 1 of the 4 academic-affiliated community health clinics in New York City were randomized 1:1 to 1 of the 2 parallel, unblinded arms: precision SMS text messages (which included stage-targeted educational information, next dose due date, and site-specific walk-in hours) or conventional SMS text messages without educational information. Randomization was stratified according to gender, age, and language. The primary outcome was series completion within 12 months. In post hoc analysis, enrollees were compared with concurrent nonenrollees and historical controls.

Results: Overall, 956 parents were enrolled in the study. The precision (475 families) and conventional (481 families) SMS text message arms had similarly high series completion rates (344/475, 72.4% vs 364/481, 75.7%). A total of 42 days after the first dose, two-thirds of families, not initially in the preparation stage, moved to preparation or vaccinated stage. Those in either SMS text message arm had significantly higher completion rates than nonenrollees (708/1503, 47.1% vs 679/1503, 45.17%; $P<.001$). Even after removing those needing only 2 HPV doses, adolescents receiving any SMS text messages had higher completion rates than historical controls (337/2823, 11.93% vs 981/2823, 34.75%; $P<.001$). A population-wide effect was seen from 2014 to 2016, above historical trends.

Conclusions: SMS text message reminders led to timely HPV vaccine series completion in a low-income, urban, minority study population and also led to population-wide effects. Educational information did not provide an added benefit to this population.

Trial Registration: ClinicalTrials.gov NCT02236273; https://clinicaltrials.gov/ct2/show/NCT02236273

(JMIR Mhealth Uhealth 2021;9(12):e26356) doi:10.2196/26356

KEYWORDS

text messaging; mobile reminders; human papillomavirus; adolescent; text reminders; vaccine completion; vaccine decision-making; vaccine education; transtheoretical model; mobile phone; smartphone; mHealth; mobile health; minority health

Introduction

Background

Adolescents, particularly minority adolescents, are not adequately protected against human papillomavirus (HPV) and its potential sequelae, including cancer and genital warts. Despite the highly efficacious vaccine being recommended for all adolescents, completion rates among those who initiate the series are low [1]. Nationwide, only 71.5% of adolescents aged 13-17 years initiate the HPV vaccine series, and half (54.2%) have received all needed doses [1]. Adherence to the recommended HPV vaccine dosing schedule is also exceedingly low; one study found that of 9- to-16-year-olds who had initiated HPV vaccination, only 28% completed the then-recommended 3-dose series within 1 year [2]. Caregiver-decided vaccination delays can significantly contribute to the spread of infectious diseases in adolescents [3]. This is a particularly salient factor to counter for HPV vaccination, as HPV infection not only carries short-term infection risk but also long-term chronic disease risk.

Health information technology interventions that link communication technologies, such as SMS text messaging, with electronic health record data offer low-cost, scalable opportunities to foster vaccination and other preventive care behaviors [4]. Currently, there is a growing body of work supporting the feasibility, acceptability, and efficacy of eHealth and mobile health (mHealth) technologies [5,6], particularly within the realm of adolescent and child health promotion [7,8].

Studies have demonstrated the effectiveness of mHealth SMS text message interventions on vaccination coverage and timeliness at levels in line with other forms of reminder or recall, particularly in low-income populations for whom other forms of reminder recall have been less successful [9-13]. SMS text message reminders have been found to increase HPV vaccine uptake in various populations, particularly in the adolescent cohort [11,14-21]. However, their effect has not been as robust as needed. One potential advantage of SMS text message interventions, which has not been well investigated, is the ability to provide precision messages. Such tailored messages may promote increased engagement with SMS text message reminders and, in turn, positively impact HPV vaccination completion rates.

Objectives

In this study, we compare the impact of precision SMS text message on HPV vaccine series completion (tailored vaccine health literacy–promoting information targeted to the family's stage of vaccine decision-making) with conventional SMS text message reminders in a pragmatic randomized trial. The transtheoretical model of behavior change guided the tailoring of SMS text messages.

Methods

Overview

This trial was conducted in 4 community health clinics affiliated with the New York-Presbyterian Hospital Ambulatory Care Network and Columbia University between December 2014 and December 2017. These practices provide care for primarily publicly insured Latino patients. The Vaccines for Children program provides free vaccines for nearly all the patients at the study sites, and all the study sites allow walk-ins for second and third HPV vaccine doses without an appointment. All vaccinations given at the study sites are documented in the New York-Presbyterian Hospital immunization registry**,** which extracts information about vaccinations directly from the provider order entry module of the electronic health record, making data accurate for HPV vaccines administered at clinical sites. The registry also synchronizes data with the New York Citywide Immunization Registry (CIR), which is a population-based registry. The New York City Public Health Law requires documentation for all vaccinations administered to those aged ≤18 years old be submitted to CIR [22], which captures >85% of vaccines administered in New York City and 93% of vaccines from the Vaccines for Children program.

Recruitment followed a 2-pronged enrollment process. First, nurses at the study sites provided families with a recruitment card and information sheet. Families interested in being contacted could put a cell phone number on the card, which was then left for the research assistants. Those who did not want to be contacted could also indicate this as such. Of the 547 families from whom a recruitment card was collected, only 9 (1.6%) indicated that they wished not to be contacted. Permission was received to contact the families of all adolescents who received their first HPV vaccine dose at one of the study sites during the enrollment period for whom a card was not left or for whom the nurse did not have time to give a card to assess eligibility and interest.

Parents or legal guardians involved in the study had to meet the following eligibility criteria: (1) have a child aged between 9 and 17 years who received their first HPV vaccine at study clinics, (2) own a cell phone with SMS text message capabilities, (3) have English or Spanish literacy, (4) plan to remain in New York City for the next 12 months, and (5) have not been previously contacted to enroll with a different child. Children down to 9 years were included as the vaccine was licensed down to those aged 9 years. There was no compensation for enrollment in the study.

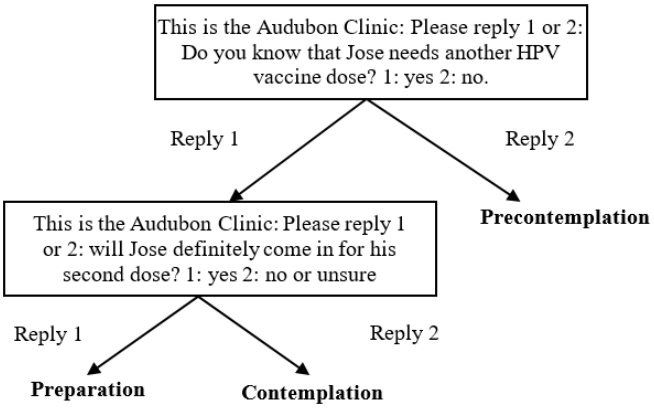
After each enrollment, the project coordinator used an adaptive electronic randomization algorithm to parallel randomize all participants with a 1:1 allocation ratio stratified by (1) clinic site, (2) adolescent's gender, (3) adolescent age group (9-14 years vs 15-17 years), and (4) parental language. The scheme was adaptive in that the assignment of each family was to the arm that would keep the allocation ratios in the 4 strata the

family belonged to (site, sex, age, and language) closer to 1:1. The statistician and analyst were blinded to the arm assignments.

We designed the messages (*precision* and *conventional*) using information gathered during our previous studies, the relevant literature, and expertise in SMS text messaging, HPV vaccination, adolescent medicine, health literacy, and the community. Messages were then pretested with 20 parents iteratively until no new message changes were made. Parents participating in this pretest received a round trip New York City Metrocard.

For families receiving precision SMS text message reminders, an automated, in-house SMS text messaging software program first used a short cascade of questions based on the transtheoretical model to assess the family's stage of decision-making regarding vaccination (Figure 1).

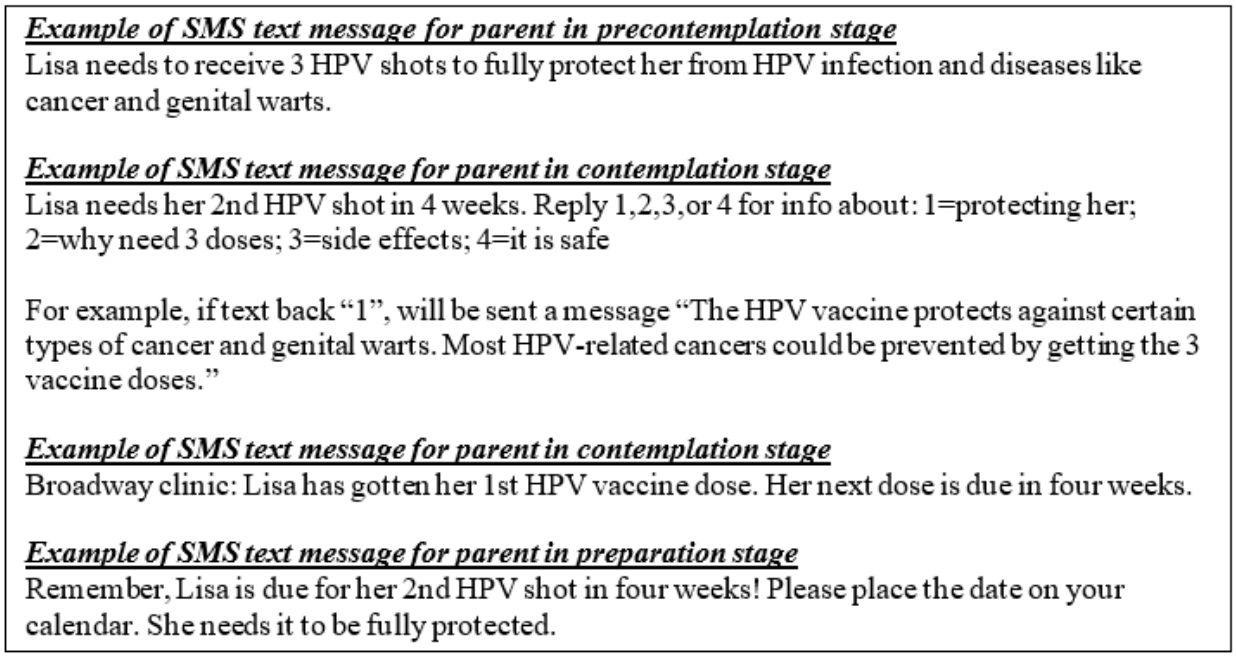
Figure 1. Transtheoretical model stage allocation in the day 21 message. HPV: human papillomavirus.



This was sent on day 21 after administration of the first HPV vaccine dose. On the basis of the person's response, the platform automatically placed them into the correct stage and proceeded to send educational information targeted at that stage of decision-making. Parents who were in the *precontemplation* stage were unaware that their adolescents needed a second vaccination or when it was due. Those in *contemplation* knew their child or adolescent needed a dose but might still have had questions such as vaccine efficacy, side effects, and safety. Finally, those in *preparation* were planning to have their adolescents continue the vaccination series but might not have known where or when to access care. The program was monitored by the project coordinator.

Parents in each stage received different messages (Figure 2). For example, messages for parents in *precontemplation* first notified them that their adolescents were in need of subsequent doses and why those doses were needed, whereas messages for parents in the *contemplation* stage provided information needed to answer any remaining questions they might have had regarding vaccination. For some messages were responsive to user input, such that parents were able to self-tailor the content by texting back indicators for which items they wanted more information about. For those in the *preparation* stage, the messages provided information regarding where and when to walk-in for vaccination. Parents in the other 2 stages also received information regarding where and when to walk-in for vaccination. Families also had 2 additional instances, on days 33 and 40, where their current stage was assessed, if they were not already in the preparation stage. On the basis of responses to these messages, families could switch into a different stage track of messages.

Figure 2. Precision and conventional arm SMS text message examples. HPV: human papillomavirus.



Conventional SMS Text Message Reminders

Parents not randomized to the precision text message arm received conventional text reminders notifying them when the next dose was due. These messages did not include vaccine health literacy–promoting information and were similar to those used in our previous adolescent studies (Figure 2).

Message Frequency

Informational messages for both arms began on day 28. This was designed so that, if a family reacted to the message and came in to be vaccinated, it would not be before the 28-day minimal required interval between the first and second dose; the intervention began before the HPV recommendation changed from 3 to 2 doses for younger populations [23]. Subsequent messages were sent on days 35, 42, 49, and 56 post vaccination for both study arms. These dates were chosen because of their proximity to the initial vaccination date; the second dose was due on day 56 (counting from the day the first dose was administered). We selected reminder message send dates at the time they were due and included an additional 1 week (63 days), 2 weeks (70 days), 4 weeks (84 days), and 6 weeks (98 days) *overdue* reminders. This series of 5 messages (days 28, 35, 42, 49, and 56) was selected based on the protocol from our previous trial in which a median of 5 messages was needed for a family to bring a child in for vaccination and was well tolerated by parents with very few stop requests [10]. Booster messages were also sent on days 63, 70, 84, and 98 post vaccination. Once or if an adolescent received the second dose, parents began receiving their next set of messages 28 days before the due date of the next dose, and then followed the same timing as with the second dose messages (days 28, 21, 14, 7, and 0 before the third dose).

Messages were sent in English or Spanish, based on parent preference. On the basis of our previous study [24], which identified parental preferences, messages were designed to include the child's name and stating that it was being sent on behalf of the clinic. The recipient of the messages was the parent, rather than the child, based on previous preferences elicited in this population. Families stopped receiving messages on their original schedule when the required dose was abstracted from the hospital's immunization registry, which included synchronized data from the CIR, as described earlier.

In October 2016, the Centers for Disease Control and Prevention (CDC) Advisory Committee on Immunization Practices recommended a change to 2 doses of the HPV vaccine series (0 and 6 months) for those children or adolescents who start the series under the age of 15 years. There was no change in those who started the series when they were aged ≥15 years. In response, the following alterations were made: parents of adolescents <15 years when first vaccinated, who were already enrolled, and had not yet received their second dose, received an SMS text message update informing them that their child now only needed 2 doses 6 months apart and that we would text them when the second dose was due. These parents then received an updated series of messages consistent with the new 2-dose recommendation in terms of both phrasing and timing. Parents of adolescents who either initiated the series (1) at the age of ≥15 years or (2) at <than 15 years but had already received their second dose less than 6 months after the first, stayed on the original protocol because the adolescent was still in need of 3 doses. Any parents of already-enrolled adolescents who were aged <15 years when first vaccinated and received their second dose at least 5 months after their first were considered to have completed the series under the new recommendations. These parents received an SMS text message letting them know that, because their child's second shot was at least 5 months after the first, they had now completed the series.

Measures and Analysis

On the basis of previous baseline data, 50.9% (245/481) of patients in the *conventional* group were expected to receive 3 doses in 12 months. With a sample size of 956, we were powered at 80% to detect a 9% difference in completion rate (the primary outcome) between arms to be statistically significant at P=.05. The primary outcome measure was timely HPV vaccine series completion within 12 months (operationalized as the receipt of 2 or 3 doses, based on age and enrollment date, and accounting for the 2016 CDC guideline change). Vaccine completion was extracted from local vaccine sources.

HPV vaccine completion rates were compared for all adolescents of participant parents at the end of a 12-month observation period starting at the receipt of their first HPV vaccine dose. All primary analyses were conducted on an intention-to-treat basis. Completion rates in the 2 randomized groups were compared using 2-tailed chi-squared tests at a significance level of P<.05. Multiple logistic regressions were performed to assess any potential differences in receipt among demographic groups.

Chi-square tests were conducted as a post hoc analysis to compare both intervention arms with concurrent nonenrollees (n=1503) who received their first vaccine dose during the study period, as well as with historical controls (n=2823; first dose administered 2011-2013). Intervention-arm adolescents who received 2 doses per the new guideline were removed to achieve comparability for the historical analysis. This study was approved by the Columbia University Irving Medical Center Institutional Review Board. Personalized Reminders for Immunization Using Short Messaging is registered with ClinicalTrials.gov with identifier NCT02236273.

Role of the Funding Source

The Agency for Healthcare Research and Quality funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Overview

In total, we screened 1593 adolescents who received their first HPV vaccine dose at the study sites. The majority (1454/1593, 91.3%) were able to be contacted. Of the contacted patients, only 2% (29/1454) of families did not have a cell phone with SMS text messaging, and 11.1% (161/1454) of were excluded based on other exclusion criteria. Of the eligible families, most (956/1264, 75.63%) were enrolled (Figure 3).

The arms were well-balanced; 481 families were randomized to the usual care arm and 475 to the intervention arm. Most of the adolescents in the enrolled families were aged ≤14 years (880/956, 92.1%). Half of them (478/956, 50%) were female, and most of them (903/956, 94.5%) were publicly insured. Two-thirds (614/956, 64.2%) of parents/caregivers were primarily Spanish speaking, and 59.9% (573/956) had a high school education or less (Table 1).

Figure 3. Study enrollment.

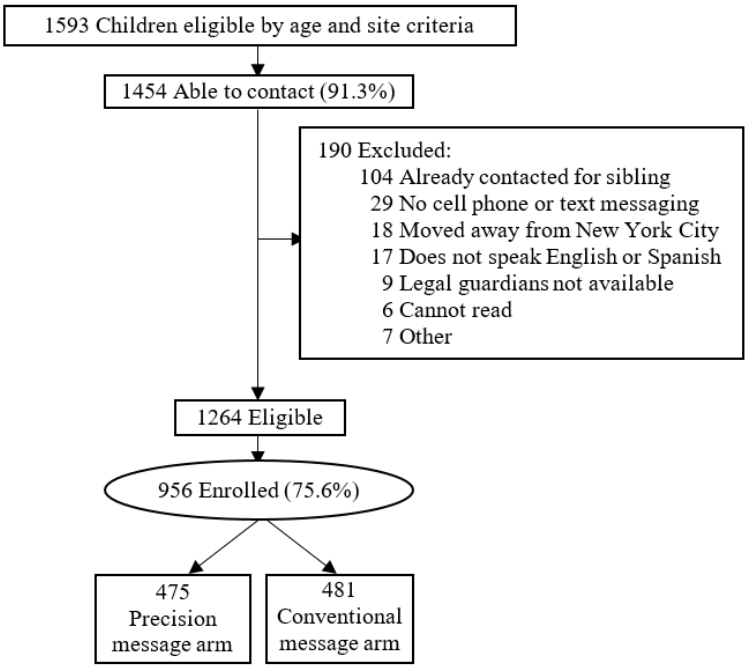


Table 1. Characteristics of the study participants.

Characteristics	Total (n=956), n (%)	Precision message arm (n=475), n (%)	Conventional message arm (n=481), n (%)	*P* value
Age (years)				.95
<14	880 (92)	437 (92)	443 (92.1)	
15-17	76 (8)	38 (8)	38 (7.9)	
Sex				.56
Female	478 (50)	242 (51)	236 (49.1)	
Language				.88
Spanish	614 (64.2)	304 (64)	310 (64.5)	
Site				.90
Clinic 1	238 (24.9)	121 (25.5)	117 (24.3)	
Clinic 2	239 (25)	114 (24)	125 (26)	
Clinic 3	296 (31)	147 (30.9)	149 (31)	
Clinic 4	183 (19.1)	93 (19.6)	90 (18.7)	
Insurance				.30
Public	903 (94.5)	445 (93.7)	458 (95.2)	
Parental education				.64
<High school	219 (22.9)	114 (24)	105 (21.9)	
Finished high school	354 (37.1)	170 (35.8)	184 (38.3)	
>High school	382 (40)	191 (40.2)	191 (39.8)	

Movement Through Stages

Overall, 12,000 messages were sent. Of those randomized to the intervention arm, 1 family received their second dose early, and therefore, did not receive second dose messages; a second family requested to stop the program before the messages started. This left 473 of the 475 families randomized to the intervention arm eligible to receive messages. Most families (428/473, 90.5%) in the precision reminder arm received the day 21 message. There were technical issues for 45 families (45/473, 9.5%), as some messages were undelivered because of user service disruption. Of those who received the messages, two-thirds (288/428, 67.3%) of families responded to stage-assessment messages: 52.6% (225/428) were in preparation, 10.3% (44/428) contemplation, and 4.4% (19/428) precontemplation. The remaining families including the 32.7% (140/428) who did not respond and 10.5% (45/428) for whom there were technical difficulties remained in precontemplation.

On day 33, there were 215 families randomized to the precision arm that was either not in preparation or had not yet been vaccinated. Of those, 60% (129/215) responded to either day 33 or 40 messages, 54.4% (117/215) were automatically switched to preparation, 3.7% (8/215) remained in precontemplation, and 1.4% (3/215) moved into contemplation. The remaining stayed in the stage they had been in previously.

By day 42, 72.7% (344/473) of the intervention families were in preparation, with 47.6% (225/473) being there at day 21, and an additional 25.2% (119/473) who moved there through prompts. An additional 34 who had not been in preparation at the beginning of the study had already been vaccinated by day 42, resulting in 79.7% (377/473) either being in preparation or already vaccinated by day 42. Overall, 13.5% (64/473) of families did not respond to the stage questions.

The movement was similar for the third dose. Overall, 336 families randomized to the precision message arm needed a third dose of the vaccine, including those for whom a third dose was needed based on the year and age at first dose. We received replies from half (181/336, 53.9%) of the participants. Of those who were not in the preparation stage at the beginning of the third dose cycle, half were able to be moved into preparation within 2 weeks before the next dose was due.

Receipt of HPV Vaccination

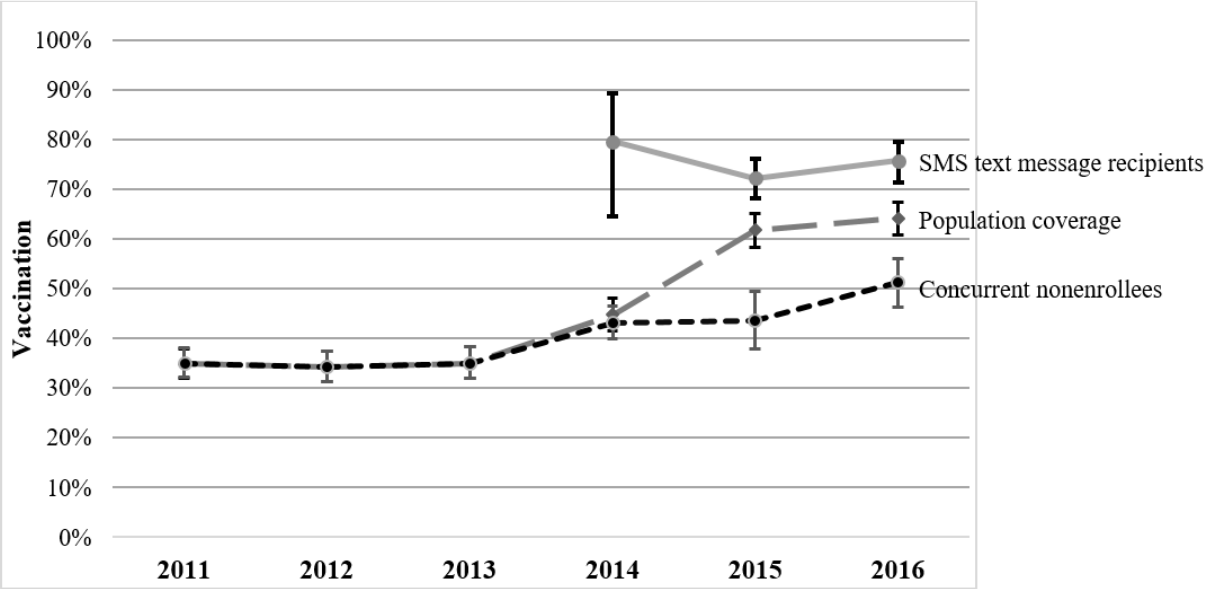
Of those receiving SMS text messages, Spanish-speaking parents (adjusted rate ratio 1.17, 95% CI 1.08-1.27) had an increased rate of timely completion; ≥15 years old had a decreased completion rate (adjusted rate ratio 0.95, 95% CI 0.93-0.98). No differences existed in terms of sex, education, or insurance.

Overall, both the precision SMS text message (344/473, 72.7%) and conventional (364/481, 75.7%) arms had very high timely series completion rates within 12 months, which did not significantly differ (*P*=.25). We found a significant difference in completion rate for those who responded to the first day 21 intervention messages (n=291) versus those who did not respond (n=153; 219/291, 75.3% vs 100/153, 65.4%; *P*=.04).

In a post hoc analysis, those in either SMS text message arm had a significantly higher completion rate than the nonenrollees. In addition, even after removing those who only needed 2 doses to complete the series, those in the SMS text messaging arms

had higher rates than the historical controls (n=2823). Enrollees were more likely to speak Spanish (614/917, 66.9% vs 830/1444, 57.5%; *P*<.001) and were more likely to have a child aged <15 years (880/956, 92.1% vs 1300/1503, 86.49%; *P*<.001) than nonenrollees, but there were no differences in the percentage of children enrolled who were male (479/956, 50.1% vs 795/1503, 52.9%; *P*=.19). In the enrollee comparison to historical controls, those enrolled versus historical controls were less likely to be male (479/956, 50.1% vs 1749/2823, 62%; *P*<.001). There was no difference in the percentage of participants who spoke Spanish (614/917, 66.9% vs 1788/2754, 64.9%; *P*=.35). Ultimately, a population-wide effect on HPV vaccination series completion was seen during the years of the study 2014-2016, above historical trends (Figure 4).

Figure 4. Human papillomavirus vaccine series completion within 12 months of initiation by year of initiation.



Discussion

Principal Findings

In this study, the addition of educational information in SMS text messages targeted to the stage of caregiver vaccine decision-making did not provide additional benefits in this low-income, urban, minority population. However, those receiving SMS text message reminders had more timely series completion than historical controls and nonenrollees. SMS text messaging also led to population-level effects that far exceeded historical trends, illustrating the potential impact of such reminders if implemented. Although ad hoc analyses, the very large differences in completion between enrollees and nonenrollees as well as historical controls adds further support. Together, our findings suggest that SMS text message reminders could be used to increase the likelihood that caregivers will follow through with HPV vaccination series completion and could also be used to combat vaccination delay.

Of those in the intervention group, more individuals who responded to any message on day 21 received the required doses than those who did not. However, in this study population, education information in the SMS text messages did not provide additional benefits over conventional SMS text messages without educational information. Studies have shown that prolonged exposure to similarly themed messages may lead to *reactance* or active resistance against the health behavior the health message advocates [25,26]. In addition, for parents with an unfavorable attitude toward vaccines, educational interventions have been found to reduce the intention to vaccinate [27]. For the subset of our patient population who failed to interact with the SMS text messages, additional educational messages may have acted as a mental deterrent to bringing in their children for subsequent doses. It is possible that for certain populations, changing to a different modality is needed. These messages were based on formative work and pretesting with patients and their families, which is critical in the development of mHealth interventions [28-31]; however, future work should also potentially include an intentional exploration with the target populations of unintended impacts of messages.

Despite this finding, our study demonstrates how SMS text message reminders could interrupt a common pathway to vaccination delay and vaccine series incompletion. Vaccination delay, particularly for HPV vaccination, is often studied as an active decision by caregivers [32]; however, given the increase in timeliness and the lack of impact of targeted vaccine-readiness information shows that delay is often not an active decision but rather a circumstantial effect based on other factors (eg, forgetfulness) that can be mitigated with timely reminders. SMS text message reminders work as a *call to action* and help prompt caregivers that would otherwise vaccinate their children but may fail to bring their child back for vaccination because of other factors. Further research could be conducted to explore the benefit of this technology for direct youth use, as an increasing number of health interventions are targeted at

adolescent self-use to encourage increases in their health autonomy [33].

In this study, we also demonstrated the ability to use SMS text messages to assess a family's stage of vaccine decision-making and move them along the stages of the transtheoretical model. Most families (409/473, 86.5%) responded to at least 1 message prompt, and two-thirds of families who were not in preparation at the first assessment were either in preparation or vaccinated by 42 days after the first dose. Similarly, half of the families of adolescents who needed 3 doses who were not in preparation at the first assessment for that dose were either in preparation or vaccinated by the time the third dose was due. Although ultimately such an in-depth, precision intervention may not have been needed for this population, it does lay the foundation for using SMS text messaging both to assess a person's stage of decision-making and to intervene to move them through to a possible behavior change. Several SMS text messaging interventions that have targeted stages of change have been tested and found to be effective in encouraging health behavior change, namely with physical activity [34-36], smoking cessation [37,38], and diabetes care management [39]. However, many of these studies were conducted internationally, and none have addressed HPV vaccine uptake. Our study contributes to the growing body of knowledge on SMS text messages targeted to the stage of change and presents a novel understanding of SMS text message efficacy in increasing HPV vaccination completion in adolescents. These findings may be particularly applicable given the increased levels of vaccine hesitancy in caregivers in the wake of acute COVID-19 activity. Although app-based interventions offer a number of benefits, they require users to have higher levels of technological literacy than SMS text message–based interventions. SMS text messages are preset on mobile phones and require virtually no instruction for use when receiving messages outside general literacy. The results of this study, along with prior vaccine SMS text message reminder research, underscore the sustained role SMS text messaging can still play in providing a digital precision SMS text message health approach to behavioral interventions, even in the modern mobile use landscape.

During this study, the CDC changed their recommendations for the number of doses of HPV vaccine needed by adolescents who initiated the series before the age of 15 years. An unintended benefit of this study was the demonstration of the ability of SMS text messages to facilitate rapid communication with families to inform them of the CDC schedule changes. Such ability extends the possibilities for health care providers to notify families when they need them to either take or not take a certain action. If sites had to call families to tell them not to come in, it would have required extensive personnel time. Conversely, it would have been frustrating to families if they had not been notified and had showed up too early. This real-time notification can be beneficial for both health care providers and public health practitioners. Investigating modalities and best practices of remote pediatric clinic communication with caregivers is particularly needed as we rethink health care communication and adolescent care in the wake of the height of the COVID-19 pandemic [40,41].

Limitations

Our study has several limitations. This study took place in a single medical system that serves a primarily low-income, Latino, urban community, who may be particularly sensitive to SMS text message interventions. These findings may not be generalizable to other settings. Recommendations also changed during the intervention period. Most of the study population received the first dose before the new CDC guidelines were implemented, when 3 doses were needed. However, these were accounted for in the analyses. In addition, vaccine administration has been underreported. However, all administered vaccines are documented in the electronic health record, including synchronization with the New York CIR, which has an excellent capture rate. Therefore, underreporting of vaccinations is likely low; underreporting would also have affected the intervention and usual care groups similarly.

Conclusions

Despite these aspects of the study, our findings lend strength to the growing body of evidence showing that mHealth or eHealth interventions such as SMS text message reminders can be used to tangibly promote child and adolescent health [40], particularly in the realm of HPV vaccination, in which outcomes are consistently substandard to national goals. We also demonstrate the efficacy of SMS text message reminders in a low-income, tight knit, and connected minority community, which helps answer the call to improve upon digital approaches that "address disparities in access to care related to race and ethnicity, socioeconomic status" [40]. SMS text message reminders function as an accessible, easy-to-use, low-cost remote intervention that can be rapidly deployed, although information detailing the economic impact and cost-effectiveness of these interventions should be evaluated in future studies.

SMS text message reminders led to timely HPV vaccine series completion in our study population, which led to population-wide effects. Although education information did not provide added benefit in this very responsive population, we also demonstrated the feasibility of using SMS text messages to both identify a family's stage of vaccine decision-making, move them further down the pathway to behavior change, and possibly decrease HPV vaccine delay. In the face of health information changes, SMS text messages also helped facilitate real-time and remote communication of these changes to caregivers, which is needed in our post pandemic clinical pediatric landscape.

Acknowledgments

The authors wish to thank the EzVac team, NewYork-Presbyterian Hospital for its support of the EzVac Immunization Information System, NewYork-Presbyterian Hospital Ambulatory Care Network, and Nadira Fernandez, Stewin Camargo, Angela Barrett, and Luis Alba.

Conflicts of Interest

All authors declare no competing financial interests. MSS was a coinvestigator on a grant from the Pfizer Medical Education Group for a different investigator-initiated study but received no financial support.

References

1. Elam-Evans LD, Yankey D, Singleton JA, Sterrett N, Markowitz LE, Williams CL, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13-17 years - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Aug 21;69(33):1109-1116 [FREE Full text] [doi: 10.15585/mmwr.mm6933a1] [Medline: 32817598]
2. Widdice LE, Bernstein DI, Leonard AC, Marsolo KA, Kahn JA. Adherence to the HPV vaccine dosing intervals and factors associated with completion of 3 doses. Pediatrics 2011 Jan;127(1):77-84 [FREE Full text] [doi: 10.1542/peds.2010-0812] [Medline: 21149425]
3. Damnjanović K, Graeber J, Ilić S, Lam WY, Lep Z, Morales S, et al. Parental decision-making on childhood vaccination. Front Psychol 2018 Jun 13;9:735 [FREE Full text] [doi: 10.3389/fpsyg.2018.00735] [Medline: 29951010]
4. Iribarren SJ, Brown W, Giguere R, Stone P, Schnall R, Staggers N, et al. Scoping review and evaluation of SMS/text messaging platforms for mHealth projects or clinical interventions. Int J Med Inform 2017 Dec;101:28-40. [doi: 10.1016/j.ijmedinf.2017.01.017] [Medline: 28347445]
5. 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 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
6. 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]
7. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered ehealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940. [doi: 10.2196/10940] [Medline: 30026178]
8. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
9. Ahlers-Schmidt CR, Chesser A, Hart T, Paschal A, Nguyen T, Wittler RR. Text messaging immunization reminders: feasibility of implementation with low-income parents. Prev Med 2010;50(5-6):306-307 [FREE Full text] [doi: 10.1016/j.ypmed.2010.02.008] [Medline: 20178813]
10. Stockwell MS, Kharbanda EO, Martinez RA, Vargas CY, Vawdrey DK, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. JAMA 2012 Apr 25;307(16):1702-1708. [doi: 10.1001/jama.2012.502] [Medline: 22535855]
11. O'Leary ST, Lee M, Lockhart S, Eisert S, Furniss A, Barnard J, et al. Effectiveness and cost of bidirectional text messaging for adolescent vaccines and well care. Pediatrics 2015 Dec;136(5):e1220-e1227. [doi: 10.1542/peds.2015-1089] [Medline: 26438703]
12. Jacobson VJ, Jacobson RM, Coyne-Beasley T, Asafu-Adjei JK, Szilagyi PG. Patient reminder and recall interventions to improve immunization rates. Cochrane Database Syst Rev 2018 Dec 18;1:CD003941. [doi: 10.1002/14651858.CD003941.pub3] [Medline: 29342498]
13. Kharbanda EO, Stockwell MS, Fox HW, Andres R, Lara M, Rickert VI. Text message reminders to promote human papillomavirus vaccination. Vaccine 2011 Mar 21;29(14):2537-2541. [doi: 10.1016/j.vaccine.2011.01.065] [Medline: 21300094]
14. Stephens AB, Wynn CS, Stockwell MS. Understanding the use of digital technology to promote human papillomavirus vaccination - A RE-AIM framework approach. Hum Vaccin Immunother 2019;15(7-8):1549-1561 [FREE Full text] [doi: 10.1080/21645515.2019.1611158] [Medline: 31158064]
15. Rand CM, Brill H, Albertin C, Humiston SG, Schaffer S, Shone LP, et al. Effectiveness of centralized text message reminders on human papillomavirus immunization coverage for publicly insured adolescents. J Adolesc Health 2015 May;56(5 Suppl):S17-S20. [doi: 10.1016/j.jadohealth.2014.10.273] [Medline: 25863549]
16. Kempe A, O'Leary S, Shoup J, Stokley S, Lockhart S, Furniss A, et al. Parental choice of recall method for HPV vaccination: a pragmatic trial. Pediatrics 2016 Mar;137(3):e20152857 [FREE Full text] [doi: 10.1542/peds.2015-2857] [Medline: 26921286]

17. Aragones A, Bruno D, Ehrenberg M, Tonda-Salcedo J, Gany F. Parental education and text messaging reminders as effective community based tools to increase HPV vaccination rates among Mexican American children. Preventive Medicine Reports 2015 Aug;2(2):554-558. [doi: 10.1016/j.pmedr.2015.06.015] [Medline: 26844117]
18. Morris J, Wang W, Wang L, Peddecord KM, Sawyer MH. Comparison of reminder methods in selected adolescents with records in an immunization registry. J Adolesc Health 2015 May;56(5 Suppl):S27-S32. [doi: 10.1016/j.jadohealth.2015.01.010] [Medline: 25863551]
19. Rand CM, Vincelli P, Goldstein NP, Blumkin A, Szilagyi PG. Effects of phone and text message reminders on completion of the human papillomavirus vaccine series. J Adolesc Health 2017 Jan;60(1):113-119. [doi: 10.1016/j.jadohealth.2016.09.011] [Medline: 27836533]
20. Bar-Shain DS, Stager MM, Runkle AP, Leon JB, Kaelber DC. Direct messaging to parents/guardians to improve adolescent immunizations. J Adolesc Health 2015 May;56(5 Suppl):S21-S26. [doi: 10.1016/j.jadohealth.2014.11.023] [Medline: 25863550]
21. Tull F, Borg K, Knott C, Beasley M, Halliday J, Faulkner N, et al. Short message service reminders to parents for increasing adolescent human papillomavirus vaccination rates in a secondary school vaccine program: a randomized control trial. J Adolesc Health 2019 Jul;65(1):116-123 [FREE Full text] [doi: 10.1016/j.jadohealth.2018.12.026] [Medline: 30879881]
22. Immunization Laws and Regulations. New York State Department of Health. URL: https://www.health.ny.gov/prevention/immunization/laws_regs.htm [accessed 2021-11-18]
23. CDC recommends only two HPV shots for younger adolescents: CDC Newsroom Releases. Centers For Disease Control and Prevention. URL: https://www.cdc.gov/media/releases/2016/p1020-HPV-shots.html [accessed 2020-08-10]
24. Hofstetter AM, Vargas CY, Kennedy A, Kitayama K, Stockwell MS. Parental and provider preferences and concerns regarding text message reminder/recall for early childhood vaccinations. Prev Med 2013 Aug;57(2):75-80. [doi: 10.1016/j.ypmed.2013.04.007] [Medline: 23624252]
25. Kim S, So J. How message fatigue toward health messages leads to ineffective persuasive outcomes: examining the mediating roles of reactance and inattention. J Health Commun 2018;23(1):109-116. [doi: 10.1080/10810730.2017.1414900] [Medline: 29272208]
26. So J, Kim S, Cohen H. Message fatigue: conceptual definition, operationalization, and correlates. Commun Monographs 2016 Nov 02;84(1):5-29. [doi: 10.1080/03637751.2016.1250429]
27. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 Apr;133(4):e835-e842. [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
28. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
29. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
30. 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]
31. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Dec 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
32. Bianco A, Mascaro V, Zucco R, Pavia M. Parent perspectives on childhood vaccination: how to deal with vaccine hesitancy and refusal? Vaccine 2019 Feb 08;37(7):984-990. [doi: 10.1016/j.vaccine.2018.12.062] [Medline: 30655175]
33. Radovic A, Badawy SM. Technology use for adolescent health and wellness. Pediatrics 2020 May;145(Suppl 2):S186-S194. [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
34. Lau P, Pitkethly A, Leung B, Lau E, Wang J. The intervention effect of SMS delivery on Chinese adolescent's physical activity. Int J Environ Res Public Health 2019 Mar 04;16(5):787 [FREE Full text] [doi: 10.3390/ijerph16050787] [Medline: 30836685]
35. Horner GN, Agboola S, Jethwani K, Tan-McGrory A, Lopez L. Designing patient-centered text messaging interventions for increasing physical activity among participants with type 2 diabetes: qualitative results from the text to move intervention. JMIR Mhealth Uhealth 2017 Apr 24;5(4):e54 [FREE Full text] [doi: 10.2196/mhealth.6666] [Medline: 28438728]
36. Agboola S, Jethwani K, Lopez L, Searl M, O'Keefe S, Kvedar J. Text to move: a randomized controlled trial of a text-messaging program to improve physical activity behaviors in patients with type 2 diabetes mellitus. J Med Internet Res 2016 Nov 18;18(11):e307 [FREE Full text] [doi: 10.2196/jmir.6439] [Medline: 27864165]
37. Chahar P, Mohanty VR, Aswini YB. Designing and validation of text messages for m-Health intervention for tobacco cessation and management. Indian J Cancer 2018;55(4):390-393 [FREE Full text] [doi: 10.4103/ijc.IJC_569_17] [Medline: 30829276]

38. Kong J, Ha Y. Effects of a smoking cessation program including telephone counseling and text messaging using stages of change for outpatients after a myocardial infarction. J Korean Acad Nurs 2013 Aug;43(4):557-567. [doi: 10.4040/jkan.2013.43.4.557] [Medline: 24071760]
39. A text messaging approach to behavioural change, tailored using the transtheoretical model, in the primary prevention of type 2 diabetes mellitus in Asian Indians. Imperial College London, United Kingdom,. URL: https://spiral.imperial.ac.uk/handle/10044/1/32146 [accessed 2021-11-14]
40. Badawy SM, Radovic A. Digital approaches to remote pediatric health care delivery during the covid-19 pandemic: existing evidence and a call for further research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
41. Serlachius A, Badawy SM, Thabrew H. Psychosocial challenges and opportunities for youth with chronic health conditions during the covid-19 pandemic. JMIR Pediatr Parent 2020 Oct 12;3(2):e23057 [FREE Full text] [doi: 10.2196/23057] [Medline: 33001834]

Abbreviations

CDC: Centers for Disease Control and Prevention
CIR: Citywide Immunization Registry
HPV: human papillomavirus
mHealth: mobile health

Edited by L Buis; submitted 15.02.21; peer-reviewed by S Badawy, J Cornelius, C Rios-Bedoya; comments to author 07.04.21; revised version received 09.07.21; accepted 13.10.21; published 27.12.21.

Please cite as:
Wynn CS, Catallozzi M, Kolff CA, Holleran S, Meyer D, Ramakrishnan R, Stockwell MS
Personalized Reminders for Immunization Using Short Messaging Systems to Improve Human Papillomavirus Vaccination Series Completion: Parallel-Group Randomized Trial
JMIR Mhealth Uhealth 2021;9(12):e26356
URL: https://mhealth.jmir.org/2021/12/e26356
doi:10.2196/26356
PMID:34958306

Original Paper

Long-Term Effectiveness of a Decision Support App (Pink Journey) for Women Considering Breast Reconstruction Surgery: Pilot Randomized Controlled Trial

Su-Ying Fang[1], PhD; Pin-Jun Lin[2], MSc; Yao-Lung Kuo[3,4], MD, PhD

[1]Department of Nursing, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[2]Department of Nursing, DaYeh University, Changhua, Taiwan

[3]Department of Surgery, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[4]Department of Surgery, National Cheng Kung University Hospital, Tainan, Taiwan

Corresponding Author:
Su-Ying Fang, PhD
Department of Nursing
College of Medicine
National Cheng Kung University
Number 1, University Road
Tainan, 701
Taiwan
Phone: 886 62353535 ext 6271
Fax: 886 62377550
Email: suying@ncku.edu.tw

Abstract

Background: Various kinds of breast reconstruction (BR) options, including implants and autologous, and surgery techniques, including traditional and endoscope assisted, can be used to perform surgery. All options have their own advantages and disadvantages. Women decide on an option depending on the values and preferences they emphasize. Lacking knowledge about BR or having decision difficulties during the treatment decision process makes women experience more decision regret, psychological distress, and poor body image. Delivering decision support with a values clarification exercise using eHealth approaches would be beneficial for patient outcomes.

Objective: This study aims to examine the effects of a decision support app on decision-making quality and psychological morbidity for women considering BR surgery.

Methods: This randomized controlled trial included women who were over 20 years of age and were newly diagnosed with breast cancer and candidates for mastectomy. Women having an option for breast conservation were excluded. After being referred from the outpatient physician, the women provided consent and completed the baseline assessment. Women allocated to the control group (CG) received usual care and were provided with a pamphlet with information about types of surgery and the advantages and disadvantages of different surgery types. Women allocated to the intervention group (IG) were given the same pamphlet and guided to use the Pink Journey app to support their decision. Then they were also prompted to discuss the opinions with their significant others. Finally, the decision-making process of using the app was printed out for women that they could take home. Decision conflict, anxiety, and depression were measured at baseline. At 1 week after the intervention (T1) and at 1 month (T2), 8 months (T3), and 12 months (T4) after surgery, the women completed decision conflict, decision regret, anxiety, depression, and body image scales. An intention-to-treat analysis was performed.

Results: From February 2018 to July 2019, 96 women were randomly assigned to the CG (n=48) or the IG (n=48). Results revealed that body image distress declined significantly for the IG but increased for the CG. The interaction of time and group also reached significance, indicating a significant decrease in body image distress from baseline in the IG compared with the CG after the 12th month (T4) follow-up (β=–2.25, standard error=1.01, *P*=.027). However, there was no significant difference in decision conflict (*P*=.21-.87), decision regret (*P*=.44-.55), anxiety (*P*=.26-.33), and depression (*P*=.20-.75), indicating that the decrease in these outcomes in the IG was not greater than those in the CG.

Conclusions: Although we found no effect on decision conflict, decision regret, anxiety, and depression, a decision aid that combines surgery information and values clarification can help women reduce their body image distress.

Trial Registration: ClinicalTrials.gov NCT04190992; https://clinicaltrials.gov/ct2/show/NCT04190992

(JMIR Mhealth Uhealth 2021;9(12):e31092) doi:10.2196/31092

KEYWORDS

breast cancer; breast reconstruction surgery; decision aid; decision support; mHealth; app; women

Introduction

Breast cancer is the most prevalent cancer type in the world [1]. It is also the most common cancer among females in Taiwan [2]. Although breast conservative surgery (BCS) is now a standard treatment for early stage breast cancer, mastectomy rates in women eligible for BCS are increasing, with reports indicating that 35.5%-40% of women with breast cancer undergo mastectomy [3,4]. For women undergoing mastectomy, the change in appearance can lead to various types of psychological adjustment problems, including body image discomfort, psychological distress, anxiety, and depression [5-7]. Breast reconstruction (BR) has now become an option for women to restore their appearance. One cohort study revealed that the rates of BR increased from 11.6% in 1998 to 36.4% in 2011 (P<.001 for trend) [4].

BR can be performed immediately after a mastectomy or delayed according to each woman's preferred timing after all required treatments have been completed [8]. Furthermore, various kinds of BR options (including implants and autologous) and surgical techniques (including traditional and endoscope assisted) can be used to perform the surgery [9]. All options have their own advantages and disadvantages. Women decide on an option depending on the values and preferences they emphasize [10]. Because of a new diagnosis and the nature of complex medical treatments involved, women feel stressed when making decisions related to surgery. Although a recent review study revealed that women are satisfied with their new breasts and reported low regret after receiving BR surgery [11], many felt surprised and perceived the reconstructed breasts to be unnatural, unreal, and unequal, and that the outcome was different from their original expectations before surgery [12,13]. Indeed, some women also reported high levels of decision regret after undergoing BR [14]. One recent study reported that patients undergoing BR preferred only a mastectomy that reflected a discordance with preferences [15]. Other studies documented the idea that if women lack BR knowledge or have decision difficulties during their treatment decision process, they experience more decision regret, psychological distress, and poor body image [16-18]. Helping women to make appropriate decision in accordance with their own values would be beneficial for their psychological well-being after surgery.

According to the Ottawa Decision Support Framework (ODSF), decision support needs to cover the provision of treatment/disease information, clarification of personal values, and assessment of support resources [19]. Furthermore, the decision aid (DA) following ODSF to support a treatment decision has to be validated to be helpful in improving knowledge, decreasing decision conflict, and increasing the consistency between the chosen option and personal values [20]. However, studies examining the effects of decision aid on helping women make BR surgery decision remain limited [21]. These studies found that the effect of intervention on decision conflict may occur within a short period, but the effect on decision regret may be delayed and occur much longer after an intervention. The effect of decision aid on other psychological indicators such as anxiety, depression, and body image was rare and needs to be further explored [21].

Computer-based DAs, including CD-ROM, computerized multimedia programs, and websites, were validated to perform better than paper-based DAs due to their potential for wide use by patients [22]. A recent review also documented that using electronically delivered decision support with a values clarification exercise would be beneficial to patient outcomes [23]. Given that smartphone devices and downloaded apps are more convenient than other devices with or without an internet connection [24], the aim of this study is to examine the effects of a decision support app on decision-related outcomes and psychological indicators including body image, depression, and anxiety for women considering BR surgery.

Methods

Study Design

This 1:1 randomized controlled parallel-arm trial with permuted block randomization that compares pamphlet + app with pamphlet alone was performed in Taiwan. The protocol was registered with ClinicalTrials.gov (NCT04190992), and the process of app development was published previously [25]. This study was approved by the Institution Review Board of National Cheng Kung University (B-ER-106-072) and was performed in accordance with the ethical principles of the Declaration of Helsinki. None of the data collected contained identifiable information; data were kept locked in the office of the first author (S-YF).

Participants

Women were eligible for participation if they were (1) 20 years of age or older, (2) newly diagnosed with breast cancer and candidates for mastectomy, and (3) able to read and speak Taiwanese or Mandarin. Women were excluded if they had an option for BCS, reconstruction following a lumpectomy, reported active psychiatric illness, or severe cognitive illnesses that would prevent full participation. They were enrolled from February 2018 to July 2019 and completed their last follow-up in February 2021.

Randomization, Blinding, and Procedure

Women were referred to the study by an outpatient physician. After signing informed consent, women completed the baseline assessment. Consenting women were randomized using online automated randomization software (Create a Randomization List [26]) to determine group allocation. Permutated block

randomization (allocation ratio 1:1) was performed to maintain equal sample sizes. An independent research assistant generated the allocation sequence and prepared 136 numbered, opaque, sealed envelopes with assignments to be equally distributed between the 2 study groups. The researcher (S-YF) opened 1 envelope for each participant in the order in which she received a message or call from the interviewer (PJ-L) indicating that the participants were ready for randomization. The participants were not blinded to their allocation. At 1 week after intervention (T1) and at 1 month (T2), 8 months (T3), and 12 months (T4) after surgery, the women completed a follow-up questionnaire during their routine clinic visits.

Intervention Versus Usual Care

Women allocated to the control group (CG) received usual care from health care providers. They were also provided with a pamphlet with information about types of surgery, including mastectomy, implant-based BR, and autologous BR, and the advantages and disadvantages of different surgery types. Women allocated to the intervention group (IG) were given the same pamphlet and were further guided to use the Pink Journey app to support their decision [25]. Women first saw a video that is compatible with the content of the pamphlet and available in 2 languages (Chinese and Taiwanese), with selection depending on participant's preference (Figure 1). Next, they were coached on how to use a values calcification exercise that elicited them to think about 10 possible factors that they were concerned about and to rank their concerns. They were then also prompted to discuss the opinions with their significant others. Finally, the decision-making process of using the app was printed out for women that they could take home. Detailed information about the Pink Journey app was published elsewhere [25].

Figure 1. Two languages of the education video.

Instruments

Baseline data collection (T0) included demographic data and clinical data from the medical records of the patients. The following 5 instruments using a paper format were also administered at T0-T4 (ie, at baseline, 1 week, and 1, 8, and 12 months).

The Decision Conflict Scale (DCS) with 16 items developed by O'Connor [27] assesses the perception of uncertainty in information, values, or support for surgery options. The items were summed, divided by 16, and multiplied by 25. According to the user manual, scores below 25 were associated with more certainty of their decision; scores exceeding 37.5 indicated a greater feeling of uncertainty about their decision. This scale was also validated for Chinese women with surgery decisions related to breast cancer [28]. The Cronbach α coefficient for this scale in this study was .93.

The understanding of medical information was evaluated using the subscale of Involvement in the BR Decision-Making Process scale. This subscale with 6 items assesses perception of medical information about surgery and provides us information about women's understanding of BR. The scale uses a 5-point Likert scale (1=*strongly disagree* to 5=*strongly agree*), where the higher the score, the greater the amount of information women believed they had obtained. This scale had good construct validity and good internal and test–retest reliability [18]. The Cronbach α coefficient for this scale in this study was .88.

The Decision Regret Scale (DRS) contains 5 items that assess distress or remorse after a surgery decision. The items were summed, divided by 5, and then multiplied by 25. This scale was also validated for Chinese women with surgery decisions related to breast cancer [28]. The Cronbach α coefficient for this scale in this study was .90.

The Body Image Scale (BIS) with 10 items was developed by Hopwood et al [29]. The scale uses a 4-point Likert scale (0=*not at all* to 3=*very much*), with total scores ranging from 0 to 30. Higher scores indicate greater body image distress. This scale has been widely used in numerous countries and in many

languages in samples of patients with cancer [5]. The Cronbach α coefficient for this scale in this study was .92.

The Hospital Anxiety and Depression Scale (HADS) includes 14 items, of which 7 measure anxiety and 7 measure depression. All items are scored using a 0-3-point scale, with higher scores indicating more depressive symptoms. Cut-off scores are 8 and 11 to categorize the severity of anxiety and depressive symptoms, respectively. Values of over 8 indicate possible anxiety and depression, whereas values of 11 or above indicate probable anxiety and depression. The Chinese version of the scale has been widely used with good reliability and validity for women with breast cancer [5].

Statistical Analysis

Descriptive statistics were used to describe baseline and background demographic data. The chi-square test for categorical variables and independent *t* test for continuous variables were used to examine homogeneity between groups and assess covariates. An intention-to-treat analysis was conducted using a mixed effects model analysis using SPSS (version 24; IBM, Inc.) with significance set at $P<.05$. A mixed effects model that included the study group, a categorical indicator of time, and the interaction between groups and times was generated after controlling for covariates (with a significant interaction indicating that compared with the CG, the intervention effects change over time). An autoregressive covariance structure analyzed changes among the time points and residual maximum likelihood to estimate the fixed effects. Missing data were not imputed due to low attrition rate 16/96 (17%), but restricted maximum likelihood estimation was used for data management.

Results

Study Flow and Participant Characteristics

A total of 104 women were referred by a physician and completed the baseline questionnaires. Eight women dropped out after the pretest because they were rejected for surgery (n=1); scheduled for delayed (n=1), prophylactic (n=2), oncoplastic BR (n=3); or had a psychiatric disease (n=1; Figure 2). The remaining 96 women then were randomly assigned to either the CG or the IG. At T4, 72 women had provided complete data for each time point.

Among these women, 25 received neoadjuvant chemotherapy before surgery. A total of 38/96 (40%) women received chemotherapy, and 19/96 (20%) received radiotherapy following surgery. At baseline, there were no between-group differences in terms of demographic (Table 1) and disease-/treatment-related (Table 2) characteristics except that the women in the IG were younger than those in the CG ($P=.01$). In addition, women in both groups had similar preoperative appearance satisfaction ($P=.15$), anxiety ($P=.09$), and depression ($P=.09$) scores. Given the significant difference in age at diagnosis ($P=.01$) between women in the 2 groups, age may have played a role in BR decision and body image concerns, so the mixed effects analyses were adjusted for this variable.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants through each stage of the randomized controlled trial.

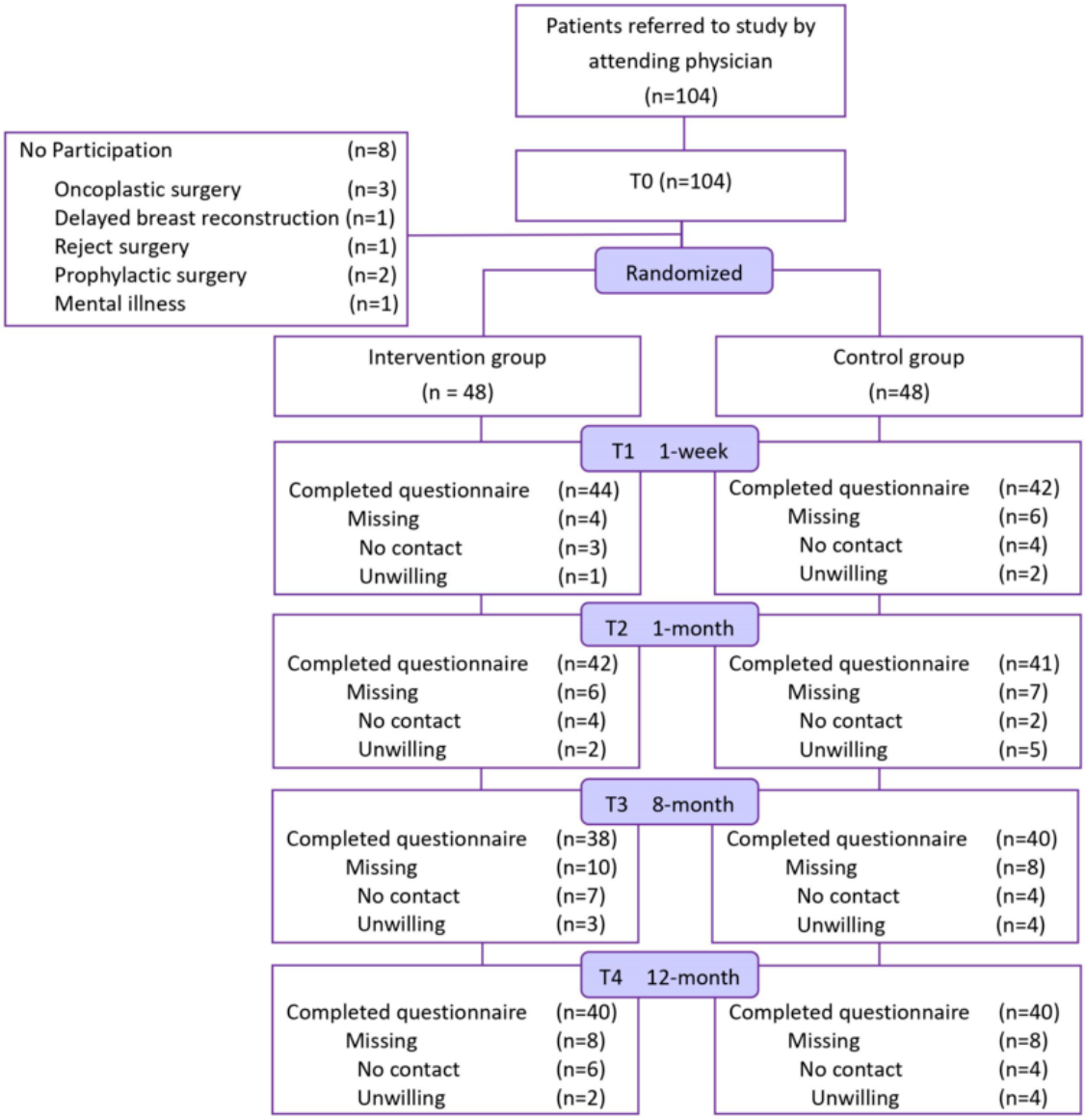


Table 1. Demographic characteristics of the participants (N=96).

Characteristic	Total (N=96)	Intervention group (n=48)	Control group (n=48)	*P* value[a]
Age (years)				*.011*
Mean (SD)	48.81 (8.22)	46.71 (8.19)	50.92 (7.77)	
Range	27-71	27-71	32-68	
Age groups (years), n (%)				
≤44	28 (29)	19 (40)	9 (19)	
45-64	61 (64)	26 (54)	35 (73)	
≥65	7 (7)	3 (6)	4 (8)	
BMI (kg/m^2)				.66
Mean (SD)	22.68 (3.27)	22.53 (3.09)	22.83 (3.46)	
Range	17.31-36.57	17.31-30.67	18.83-36.57	
Education level (years), n (%)				.32
<9	17 (18)	5 (10)	12 (25)	
9-12	26 (27)	12 (25)	14 (29)	
≧13	53 (55)	31 (65)	22 (46)	
Relationship status (% partner), n (%)	71 (74)	33 (69)	38 (79)	.25
Employment status (% employed), n (%)	66 (69)	34 (71)	32 (66)	.66
Monthly household income (US $), n (%)				.297
Lower class (<1000)	19 (20)	7 (15)	12 (25)	
Middle class (1001-1666)	27 (28)	11 (23)	16 (33)	
Middle high (1667-3333)	24 (25)	14 (29)	10 (21)	
High (>3333)	18 (19)	12 (25)	6 (13)	
Unknown	8 (8)	4 (8)	4 (8)	
Private insurance, n (%)	80 (83)	42 (88)	38 (79)	.27

[a]Statistical significance.

Table 2. Disease-related characteristics of the participants after surgery (N=96).

Characteristics	Total (N=96)	Intervention group (n=48)	Control group (n=48)	*P* value
Tumor size (cm), n (%)				.63
<1	11 (11)	5 (10)	6 (13)	
1-3	49 (51)	22 (46)	27 (46)	
>3	21 (22)	12 (25)	9 (19)	
Missing	15 (16)	9 (19)	6 (13)	
Breast cancer stage, n (%)				.54
0-I	45 (47)	23 (48)	22 (46)	
II-III	34 (35)	15 (31)	19 (40)	
Missing	17 (18)	10 (21)	7 (15)	
Neoadjuvant chemotherapy, n (%)				.96
Yes	25 (26)	12 (25)	13 (27)	
No	57 (59)	27 (56)	30 (63)	
Missing	14 (15)	9 (19)	5 (10)	
Chemotherapy, n (%)				.73
Yes	38 (40)	17 (35)	21 (44)	
No	40 (42)	22 (46)	18 (38)	
Missing	18 (19)	9 (19)	9 (19)	
Radiotherapy, n (%)				.80
Yes	19 (20)	10 (21)	9 (19)	
No	59 (61)	29 (60)	30 (63)	
Missing	18 (19)	9 (19)	9 (19)	
Surgical location, n (%)				.59
Unilateral	70 (73)	32 (67)	38 (79)	
Bilateral	13 (14)	7 (15)	6 (14)	
Missing	13 (14)	9 (19)	4 (8)	
Lymphadenectomy, n (%)				.55
Sentinel lymph node biopsy	68 (71)	33 (69)	35 (73)	
Axillary lymph node dissection	15 (16)	6 (13)	9 (19)	
Missing	13 (14)	9 (19)	4 (8)	
Active therapy 1 month after surgery, n (%)				
Chemotherapy	12 (13)	4 (8)	8 (17)	.25
Radiation therapy	4 (4)	3 (6)	1 (2)	.29
Active therapy 8 months after surgery, n (%)				
Chemotherapy	1 (1)	0 (0)	1 (2)	.34
Radiation therapy	2 (2)	0 (0)	2 (4)	.16
Active therapy 12 months after surgery, n (%)				
Chemotherapy	2 (2)	0 (0)	2 (4)	.16
Radiation therapy	1 (1)	0 (0)	1 (2)	.32
Self-evaluation on body appearance, mean (SD)				
The difference in appearance between reality and ideality (range 0-10)	4.22 (2.80)	4.27 (2.89)	4.17 (2.73)	.86
The importance of appearance in life (range 0-10)	5.53 (2.58)	5.35 (2.61)	5.71 (2.55)	.50

Decision-Making Quality

Table 3 summarizes the findings of decision-making quality outcomes. Overall, this sample reported an average of DCS score that was higher than the cut-off of 37.5. However, decision conflict declined significantly after the 1-week follow-up for both groups. Furthermore, the interaction of time and group did not reach significance (β=–2.79, standard error [SE]=3.72, *P*=.46), indicating that the decreasing of DCS score in the IG was not greater than that in the CG (Table 3).

The amount of medical information related to BR at 1 week after consultation did not differ between the IG and CG (*P*=.13), which indicates that women in both groups perceived a similar understanding level related to medical information, whether using just a pamphlet or combined with app.

Decision regret did not differ between groups at 1 month (*P*=.51), 8 months (*P*=.66), or 12 months (*P*=.61), and the interaction of time and group also did not reach significance (*P*=.44-.55).

Table 3. Between-group differences using the mixed effect model of decision-making quality.

Outcome measure	Intervention group, mean (SD)	Control group, mean (SD)	*P* value	β	Standard error	95% CI	*P* value
T0 (Baseline)[a]							
Decision Conflict Scale (0-100)							
Total	38.28 (17.22)	42.90 (21.52)	.25				
Informed	48.26 (21.54)	51.39 (26.71)	.53				
Values	51.22 (23.19)	56.94 (27.52)	.27				
Support	25.52 (20.80)	31.25 (24.40)	.22				
Uncertainty	37.67 (20.56)	41.67 (28.71)	.44				
Effective decision	31.12 (20.20)	35.68 (25.78)	.34				
T1 (1-week postconsultation)[b]							
Decision Conflict Scale (0-100)							
Total	19.35 (15.68)	20.87 (16.65)	.67	2.79	3.72	–4.61 to 10.19	.46
Informed	17.05 (19.44)	21.43 (21.00)	.32	–1.79	4.86	–11.44 to 7.87	.71
Values	20.45 (18.10)	24.80 (24.45)	.35	0.90	5.45	–9.93 to 11.73	.87
Support	14.39 (18.71)	14.09 (16.20)	.94	5.87	5.00	–4.06 to 15.80	.24
Uncertainty	26.70 (19.07)	23.81 (20.71)	.50	6.66	5.22	–3.72 to 17.04	.21
Effective decision	18.47 (16.34)	20.39 (16.51)	.60	2.59	4.94	–7.22 to 12.40	.60
Amount of medical information	20.25 (4.91)	21.76 (4.19)	.13				
T2 (1 week after consultation)[c]							
Decision Regret Scale (0-100)	19.52 (15.53)	21.95 (18.00)	.51				
T3 (8 months after surgery)[d]							
Decision Regret Scale (0-100)	21.84 (23.38)	19.63 (20.83)	.66	3.38	5.70	–7.84 to 14.60	.55
T4 (12 months after surgery)[e]							
Decision Regret Scale (0-100)	21.63 (23.95)	19.25 (7.45)	.61	3.69	4.73	–5.66 to 13.04	.44

[a]n=48 in the intervention and control group, respectively.

[b]n=44 and 42 in the intervention and control group, respectively.

[c]n=42 and 41 in the intervention and control group, respectively.

[d]n=38 and 40 in the intervention and control group, respectively.

[e]n=40 in the intervention and control group, respectively.

Psychological Indicators

Body Image and Appearance Satisfaction

Table 4 summarizes the findings of psychological outcomes. Body image distress declined significantly over time for both groups. The interaction of time and group also reached significance, indicating a significant decrease in body image distress from the baseline in the IG compared with the CG after the 12-month (T4) follow-up (β=–2.25, SE=1.01, *P*=.027).

There was also a tendency toward an improvement in appearance satisfaction over time in both groups. The interaction of time and group reached significance from T3 (β=1.15, SE=0.57,

P=.045) to T4 (β=1.17, SE=0.54, *P*=.031), which indicated that the improvement in appearance satisfaction from baseline in the IG compared with the CG was significant after the 8-month follow-up (*P*=.031-.045).

Table 4. Between-group differences using the mixed effect model of psychological distress.

Outcome measure	Intervention group, mean (SD)	Control group, mean (SD)	*P* value	β	Standard error	95% CI	*P* value
T0 (Baseline)[a]							
HADS[b] (0-21)							
Anxiety	8.38 (4.55)	9.98 (4.66)	.09				
Depression	6.12 (3.76)	7.48 (3.85)	.09				
Body satisfaction (0-10)	5.46 (2.16)	6.08 (2.08)	.15				
T2 (1 month after surgery)[c]							
HADS (0-21)							
Anxiety	5.64 (4.33)	5.85 (3.84)	.82	1.23	1.09	–0.90 to 3.37	.256
Depression	5.64 (3.52)	5.93 (4.08)	.74	0.93	1.04	–1.11 to 2.98	.370
Body satisfaction (0-10)	6.64 (1.76)	6.71 (2.33)	.89	0.44	0.58	–0.70 to 1.57	.447
Body Image Distress (0-30)	6.38 (5.55)	6.49 (6.09)	.93				
T3 (8 months after surgery)[d]							
HADS (0-21)							
Anxiety	4.66 (3.68)	4.88 (3.72)	.80	0.96	0.99	–0.99 to 2.90	.334
Depression	5.29 (4.13)	5.03 (3.82)	.77	1.23	0.96	–0.66 to 3.11	.202
Body satisfaction (0-10)	7.08 (1.94)	6.78 (2.38)	.54	1.15	0.57	0.25 to 2.27	*.045* [e]
Body Image Distress (0-30)	6.08 (6.02)	6.93 (6.15)	.54	–0.91	1.32	–3.52 to 1.69	.490
T4 (12 months after surgery)[f]							
HADS (0-21)							
Anxiety	4.25 (3.70)	4.88 (3.52)	.44	0.86	0.78	–0.68 to 2.39	.273
Depression	4.03 (3.93)	5.03 (3.71)	.25	0.24	0.77	–1.27 to 1.76	.752
Body satisfaction (0-10)	7.98 (1.46)	7.03 (2.20)	.41	1.17	0.54	0.11 to 2.24	*.031* [e]
Body Image Distress (0-30)	4.35 (0.69)	7.11 (1.12)	.05	–2.25	1.01	–4.24 to –0.26	.027

[a]n=48 in each group.

[b]HADS: Hospital Anxiety and Depression Scale.

[c]n=42 and 41 in the intervention and control group, respectively.

[d]n=38 and 40 in the intervention and control group, respectively.

[e]Statistical significance.

[f]n=40 and 40 in the intervention and control group, respectively.

Anxiety and Depression

The HADS anxiety scores 1 (*P*=.82), 8 (*P*=.80), and 12 months (*P*=.44) after surgery did not differ between groups. The HADS depression scores 1 (*P*=.74), 8 (*P*=.77), and 12 months (*P*=.25) after surgery also did not differ between groups. There was a tendency toward a decrease in depression and anxiety over time in both groups. However, the interaction of time and group did not reach significance (Table 4).

Choice of Surgery

Choice of surgery differed between the IG and CG. Overall, 56% (27/48) and 46% (22/48) opted for mastectomy plus immediate reconstruction in the IG and CG, respectively (*P*=.05). Moreover, a majority selected implanted-based BR, which did not differ between groups (Table 5).

Table 5. Comparison of surgical decision between groups.

Decision	Total (n=96)	Intervention group (n=48)	Control group (n=48)	*P* value
Breast reconstruction				*.048* [a]
Yes, n (%)	49 (51)	27 (56)	22 (46)	
No, n (%)	36 (38)	14 (29)	22 (46)	
Missing, n (%)	11 (11)	7 (15)	4 (8)	
Breast reconstruction type				.160
Implant based, n/N (%)	46/49 (94)	26/27 (96)	20/22 (91)	
Autologous, n/N (%)	3/49 (6)	1/27 (4)	2/22 (9)	

[a]Statistical significance.

Discussion

Decision-Making Quality

This study evaluated the effects of app-based DA on women's decision quality and postoperative psychological morbidity regarding BR. Women who received the app-based DA reported a similar decline in decision conflict 1 week after consultation compared with women only receiving standard care with a pamphlet. This result is consistent with a study using an interactive web-based training program [30], but is in contrast to previous studies that revealed that breast cancer treatment DA reduced decision conflict to higher levels compared with a standard booklet after consultation [31-33]. In Luan et al's study [33], only postconsultation score was compared, and it was not clear whether the decreasing level of decision conflict between baseline and postconsultation was significant. In Sherman et al's study [32], decision conflict was measured 1 month after baseline, so women who completed surgery or not may bias this outcome. In Lam et al's study [31], women in the CG only received standard information without a take-home booklet, which is different from our study, as we provided a take-home pamphlet also to the CG. Low statistical power may exist due to small differences in the treatments designed in our study compared with those in Lam et al's study [31]. Manne et al's study [30] was similar to our study by providing a pamphlet to the CG, and it revealed no significant change in decision conflict over a short period. The amount of medical information women received was not significant between groups in our study, which suggests that improving knowledge about BR may decrease women's decision conflict over a short period, but its long-term effects should be examined.

There was no difference in decision regret 1, 8, and 12 months after surgery between the 2 groups. Previous studies have revealed lower regret in the IG; however, these studies only measured 1 period and without a follow-up for over 6 months [32,33]. Lam et al' study [31] revealed decreased decision regret 4 and 10 months after surgery. This may be because most participants in their study also had the option to undergo BCS [31]. By contrast, in our study, mastectomy was necessary and there was no option to choose BCS, and this may have contributed to the nonsignificant result in this study.

Psychological Distress

There was no significant difference in appearance satisfaction and body image distress 1 month after surgery between the IG and CG. However, at both 8 and 12 months after surgery, women in the CG reported significantly better appearance and lower body image distress. A limited number of studies examined the effect of DA on body image distress. Using BREAST-Q assessment, Politi et al [34] found that satisfaction with breasts score in the DA group was slightly higher than that in the CG. Luan et al [33] reported that sum scores of sexual well-being satisfaction, satisfaction with breasts, outcome, and care in the DA group were more likely to be higher than those in the CG; however, no statistical significance was revealed in the aforesaid studies. These studies evaluated patients' feedback only over a short period, but our study highlights the importance of accessing both short- and long-term impacts of BR surgery on body image distress. Our study is the first to demonstrate significantly lower body image distress among women in the IG compared with the CG 12 months after surgery. Supporting our hypothesis, adding a values clarification exercise in DA may have helped women to create more realistic expectations about outcomes after BR, decreasing the sense of loss and reducing their body image distress.

Our analysis demonstrated that providing information using a paper or digital format in combination with values clarification did not increase anxiety for either group. This result is consistent with studies with a similar design that provided a pamphlet to the CG with short [30] and long follow-ups [33]. No significant difference between groups regarding depression was also consistent with a recent study [33]. Body image distress is associated with depression. In our study, body image distress significantly improved in the IG compared with the CG at the 12-month follow-up, suggesting that continuous follow-up to clarify the effect of DA is necessary.

Given the characteristics of universal national health insurance and the convenient geographical environment in Taiwan, losing women to follow-ups is usually because they tend to search for a second opinion in other hospitals. Although the attrition rate is not higher (17%), we do acknowledge that this was a pilot randomized controlled trial and a relatively small sample size may underestimate the effects of this study. Second, because breast surgeons are generally familiar with only 1 surgical technique, inclusion of a single breast surgeon in 1 medical

center limits this study's generalization. Lastly, our study used an amount of medical information to evaluate women's perception of their understanding of BR knowledge and only assessed it 1 week after intervention without a preintervention assessment. This limited us from comparing the changes in knowledge between groups.

Conclusions

This was the first trial to examine the long-term effects of an app-based DA on both decision-making quality and psychological morbidity for women only having the option for mastectomy. It demonstrates that DA designed with values clarification exercises can reduce similar decision conflict and depression without increasing anxiety over time compared with only receiving a pamphlet. It also further supports that using values clarification exercises can help women reduce their body image distress and increase body appearance satisfaction. Because low monitors who were highly anxious about detailed information had a greater likelihood of experiencing regret [14], future DA trials should also consider monitoring coping style and DA design that could be tailored to each women's needs. In addition, DA with a value clarification exercise that considers personal value and shows an effect on body image distress supports the utilization of personalized treatments, such as nanomedicine and immune therapy. These therapies specify women's tumor characteristics to increase therapeutic effect but with fewer side effects, empathizing that precision care for women with breast cancer will become a trend in the future [35-37].

Acknowledgments

The authors thank the Ministry of Science and Technology for funding support (grant number 106-2314-B-006-010-MY2). We also thank Dr Chen for assisting with *Pink Journey* prototyping and all women who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 388 KB - mhealth_v9i12e31092_app1.pdf]

References

1. Global Cancer Observatory: Cancer Today. URL: http://gco.iarc.fr/ [accessed 2021-11-14]
2. Taiwan Cancer Registry. Five-Year Survival Rate of Breast Cancer. Taipei, Taiwan: Taiwan Cancer Registry; 2014. URL: https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=269&pid=10227 [accessed 2018-01-29]
3. Kummerow KL, Du L, Penson DF, Shyr Y, Hooks MA. Nationwide trends in mastectomy for early-stage breast cancer. JAMA Surg 2015 Jan 01;150(1):9-16. [doi: 10.1001/jamasurg.2014.2895] [Medline: 25408966]
4. Steiner CA, Weiss AJ, Barrett ML, Fingar KR, Davis PH. Trends in Bilateral and Unilateral Mastectomies in Hospital Inpatient and Ambulatory Settings, 2005–2013: Statistical Brief #201. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville, MD: Agency for Healthcare Research and Quality; 2006 Feb. URL: https://www.ncbi.nlm.nih.gov/books/NBK367629/ [accessed 2021-11-14]
5. Fang S, Chang H, Shu B. Objectified Body Consciousness, Body Image Discomfort, and Depressive Symptoms Among Breast Cancer Survivors in Taiwan. Psychology of Women Quarterly 2014 Oct 02;38(4):563-574. [doi: 10.1177/0361684314552652]
6. Sun L, Ang E, Ang WHD, Lopez V. Losing the breast: A meta-synthesis of the impact in women breast cancer survivors. Psychooncology 2018 Feb 16;27(2):376-385. [doi: 10.1002/pon.4460] [Medline: 28544436]
7. Brandão T, Schulz MS, Matos PM. Psychological adjustment after breast cancer: a systematic review of longitudinal studies. Psychooncology 2017 Jul 12;26(7):917-926. [doi: 10.1002/pon.4230] [Medline: 27440317]
8. Seidel W, Bins-Ely J, Ongaratto Barazzetti D, Della Giustina R, Walter GP, Ferri TA, et al. Breast reconstruction after mastectomy for breast cancer: comparative analysis of early and delayed reconstruction. Minerva Surg 2017 Jun;72(3):188-189. [doi: 10.23736/s0026-4733.17.07314-x]
9. Lai H, Chen S, Chen D, Chen S, Chang T, Kuo S, et al. Current Trends in and Indications for Endoscopy-Assisted Breast Surgery for Breast Cancer: Results from a Six-Year Study Conducted by the Taiwan Endoscopic Breast Surgery Cooperative Group. PLoS One 2016 Mar 7;11(3):e0150310 [FREE Full text] [doi: 10.1371/journal.pone.0150310] [Medline: 26950469]
10. Flitcroft K, Brennan M, Spillane A. Making decisions about breast reconstruction: A systematic review of patient-reported factors influencing choice. Qual Life Res 2017 Sep 10;26(9):2287-2319. [doi: 10.1007/s11136-017-1555-z] [Medline: 28397191]
11. Flitcroft K, Brennan M, Spillane A. Decisional regret and choice of breast reconstruction following mastectomy for breast cancer: A systematic review. Psychooncology 2018 Apr 07;27(4):1110-1120. [doi: 10.1002/pon.4585] [Medline: 29143481]

12. Spector D, Mayer DK, Knafl K, Pusic A. Not what I expected: informational needs of women undergoing breast surgery. Plast Surg Nurs 2010;30(2):70-4; quiz 75. [doi: 10.1097/PSN.0b013e3181dee9a4] [Medline: 20543636]
13. Fang S, Balneaves LG, Shu B. "A struggle between vanity and life": the experience of receiving breast reconstruction in women of Taiwan. Cancer Nurs 2010;33(5):E1-11. [doi: 10.1097/NCC.0b013e3181d1c853] [Medline: 20467306]
14. Sheehan J, Sherman KA, Lam T, Boyages J. Association of information satisfaction, psychological distress and monitoring coping style with post-decision regret following breast reconstruction. Psychooncology 2007 Apr;16(4):342-351. [doi: 10.1002/pon.1067] [Medline: 16874745]
15. Lee CN, Deal AM, Huh R, Ubel PA, Liu Y, Blizard L, et al. Quality of Patient Decisions About Breast Reconstruction After Mastectomy. JAMA Surg 2017 Aug 01;152(8):741-748 [FREE Full text] [doi: 10.1001/jamasurg.2017.0977] [Medline: 28467530]
16. Lam WWT, Bonanno GA, Mancini AD, Ho S, Chan M, Hung WK, et al. Trajectories of psychological distress among Chinese women diagnosed with breast cancer. Psychooncology 2010 Oct;19(10):1044-1051. [doi: 10.1002/pon.1658] [Medline: 20014074]
17. Lam WWT, Li WWY, Bonanno GA, Mancini AD, Chan M, Or A, et al. Trajectories of body image and sexuality during the first year following diagnosis of breast cancer and their relationship to 6 years psychosocial outcomes. Breast Cancer Res Treat 2012 Feb;131(3):957-967. [doi: 10.1007/s10549-011-1798-2] [Medline: 21971728]
18. Kuo N, Kuo Y, Lai H, Ko N, Fang S. The influence of partner involvement in the decision-making process on body image and decision regret among women receiving breast reconstruction. Support Care Cancer 2019 May 21;27(5):1721-1728. [doi: 10.1007/s00520-018-4416-6] [Medline: 30132239]
19. O'Connor AM, Tugwell P, Wells GA, Elmslie T, Jolly E, Hollingworth G, et al. A decision aid for women considering hormone therapy after menopause: decision support framework and evaluation. Patient Educ Couns 1998 Mar;33(3):267-279. [doi: 10.1016/s0738-3991(98)00026-3] [Medline: 9731164]
20. Stacey D, Légaré F, Lewis K, Barry M, Bennett C, Eden K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
21. Paraskeva N, Guest E, Lewis-Smith H, Harcourt D. Assessing the effectiveness of interventions to support patient decision making about breast reconstruction: A systematic review. Breast 2018 Aug;40:97-105. [doi: 10.1016/j.breast.2018.04.020] [Medline: 29730304]
22. Syrowatka A, Krömker D, Meguerditchian AN, Tamblyn R. Features of Computer-Based Decision Aids: Systematic Review, Thematic Synthesis, and Meta-Analyses. J Med Internet Res 2016 Jan 26;18(1):e20 [FREE Full text] [doi: 10.2196/jmir.4982] [Medline: 26813512]
23. Mathijssen EGE, van den Bemt BJF, van den Hoogen FHJ, Popa CD, Vriezekolk JE. Interventions to support shared decision making for medication therapy in long term conditions: A systematic review. Patient Educ Couns 2020 Feb;103(2):254-265. [doi: 10.1016/j.pec.2019.08.034] [Medline: 31493959]
24. Hoffman AS, Volk RJ, Saarimaki A, Stirling C, Li LC, Härter M, et al. Delivering patient decision aids on the Internet: definitions, theories, current evidence, and emerging research areas. BMC Med Inform Decis Mak 2013 Nov 29;13(S2):S13. [doi: 10.1186/1472-6947-13-s2-s13]
25. Lin P, Fang S, Kuo Y. Development and Usability Testing of a Decision Support App for Women Considering Breast Reconstruction Surgery. J Cancer Educ 2021 Feb 28;36(1):160-167. [doi: 10.1007/s13187-019-01612-4] [Medline: 31463810]
26. Create a Randomization List. URL: https://www.sealedenvelope.com/simple-randomiser/v1/lists [accessed 2021-11-21]
27. O'Connor AM. Validation of a Decisional Conflict Scale. Med Decis Making 2016 Jul 02;15(1):25-30. [doi: 10.1177/0272989x9501500105]
28. Lam WWT, Kwok M, Liao Q, Chan M, Or A, Kwong A, et al. Psychometric assessment of the Chinese version of the decisional conflict scale in Chinese women making decision for breast cancer surgery. Health Expect 2015 Apr 21;18(2):210-220 [FREE Full text] [doi: 10.1111/hex.12021] [Medline: 23167846]
29. Hopwood P, Fletcher I, Lee A, Al Ghazal S. A body image scale for use with cancer patients. Eur J Cancer 2001 Jan;37(2):189-197. [doi: 10.1016/s0959-8049(00)00353-1] [Medline: 11166145]
30. Manne SL, Topham N, Kirstein L, Virtue SM, Brill K, Devine KA, et al. Attitudes and Decisional Conflict Regarding Breast Reconstruction Among Breast Cancer Patients. Cancer Nurs 2016;39(6):427-436 [FREE Full text] [doi: 10.1097/NCC.0000000000000320] [Medline: 26780376]
31. Lam WWT, Chan M, Or A, Kwong A, Suen D, Fielding R. Reducing treatment decision conflict difficulties in breast cancer surgery: a randomized controlled trial. J Clin Oncol 2013 Aug 10;31(23):2879-2885. [doi: 10.1200/JCO.2012.45.1856] [Medline: 23835709]
32. Sherman KA, Shaw LE, Winch CJ, Harcourt D, Boyages J, Cameron LD, et al. Reducing Decisional Conflict and Enhancing Satisfaction with Information among Women Considering Breast Reconstruction following Mastectomy. Plastic and Reconstructive Surgery 2016;138(4):592e-602e. [doi: 10.1097/prs.0000000000002538]
33. Luan A, Hui KJ, Remington AC, Liu X, Lee GK. Effects of A Novel Decision Aid for Breast Reconstruction: A Randomized Prospective Trial. Ann Plast Surg 2016 May;76 Suppl 3(Supplement 3):S249-S254. [doi: 10.1097/SAP.0000000000000722] [Medline: 27070681]

34. Politi M, Lee C, Philpott-Streiff S, Foraker R, Olsen M, Merrill C, et al. A Randomized Controlled Trial Evaluating the BREASTChoice Tool for Personalized Decision Support About Breast Reconstruction After Mastectomy. Ann Surg 2020 Feb;271(2):230-237. [doi: 10.1097/SLA.0000000000003444] [Medline: 31305282]
35. Emens LA. Breast Cancer Immunotherapy: Facts and Hopes. Clin Cancer Res 2017 Aug 11;24(3):511-520. [doi: 10.1158/1078-0432.ccr-16-3001]
36. Abdullah CS, Ray P, Alam S, Kale N, Aishwarya R, Morshed M, et al. Chemical Architecture of Block Copolymers Differentially Abrogate Cardiotoxicity and Maintain the Anticancer Efficacy of Doxorubicin. Mol Pharm 2020 Dec 07;17(12):4676-4690 [FREE Full text] [doi: 10.1021/acs.molpharmaceut.0c00963] [Medline: 33151075]
37. Das A, Haque I, Ray P, Ghosh A, Dutta D, Quadir M, et al. CCN5 activation by free or encapsulated EGCG is required to render triple-negative breast cancer cell viability and tumor progression. Pharmacol Res Perspect 2021 Apr 21;9(2):e00753 [FREE Full text] [doi: 10.1002/prp2.753] [Medline: 33745223]

Abbreviations

BCS: breast conservative surgery
BIS: Body Image Scale
BR: breast reconstruction
CG: control group
DA: decision aid
DCS: Decision Conflict Scale
DRS: Decision Regret Scale
HADS: Hospital Anxiety and Depression Scale.
IG: intervention group
ODSF: Ottawa Decision Support Framework
SE: standard error

Edited by L Buis; submitted 06.07.21; peer-reviewed by C Jacob, P Ray; comments to author 29.07.21; revised version received 10.08.21; accepted 27.10.21; published 10.12.21.

Please cite as:
Fang SY, Lin PJ, Kuo YL
Long-Term Effectiveness of a Decision Support App (Pink Journey) for Women Considering Breast Reconstruction Surgery: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(12):e31092
URL: https://mhealth.jmir.org/2021/12/e31092
doi:10.2196/31092
PMID:34890354

Original Paper

Evaluation of a Language Translation App in an Undergraduate Medical Communication Course: Proof-of-Concept and Usability Study

Anne Herrmann-Werner[1,2], MD, Prof Dr; Teresa Loda[1], Dr rer nat; Stephan Zipfel[1,3], MD, Prof Dr; Martin Holderried[4,5], Prof Dr med Dr oec; Friederike Holderried[6], MD, Dr med; Rebecca Erschens[1], Dr rer nat

[1]Department of Internal Medicine VI/Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Tuebingen, Germany

[2]Competence Centre for University Teaching in Medicine, Baden-Wuerttemberg, Faculty of Medicine, Eberhard-Karls University, Tuebingen, Germany

[3]Deanery of Students' Affairs, Faculty of Medicine, Eberhard-Karls University, Tuebingen, Germany

[4]Department of Strategic Medical Development and Quality Management, University of Tuebingen, Tuebingen, Germany

[5]Institute for Healthcare and Public Management, University of Hohenheim, Stuttgart, Germany

[6]eHealth Research-Group, Department of General, Visceral, and Transplant Surgery, University of Tuebingen, Tuebingen, Germany

Corresponding Author:
Anne Herrmann-Werner, MD, Prof Dr
Department of Internal Medicine VI/Psychosomatic Medicine and Psychotherapy
University Hospital Tuebingen
Osianderstr 5
Tuebingen, 72076
Germany
Phone: 49 7071 29 73719
Email: anne.herrmann-werner@med.uni-tuebingen.de

Abstract

Background: Language barriers in medical encounters pose risks for interactions with patients, their care, and their outcomes. Because human translators, the gold standard for mitigating language barriers, can be cost- and time-intensive, mechanical alternatives such as language translation apps (LTA) have gained in popularity. However, adequate training for physicians in using LTAs remains elusive.

Objective: A proof-of-concept pilot study was designed to evaluate the use of a speech-to-speech LTA in a specific simulated physician-patient situation, particularly its perceived usability, helpfulness, and meaningfulness, and to assess the teaching unit overall.

Methods: Students engaged in a 90-min simulation with a standardized patient (SP) and the LTA iTranslate Converse. Thereafter, they rated the LTA with six items—*helpful*, *intuitive*, *informative*, *accurate*, *recommendable*, and *applicable*—on a 7-point Likert scale ranging from 1 (*don't agree at all*) to 7 (*completely agree*) and could provide free-text responses for four items: general impression of the LTA, the LTA's benefits, the LTA's risks, and suggestions for improvement. Students also assessed the teaching unit on a 6-point scale from 1 (*excellent*) to 6 (*insufficient*). Data were evaluated quantitatively with mean (SD) values and qualitatively in thematic content analysis.

Results: Of 111 students in the course, 76 (68.5%) participated (59.2% women, age 20.7 years, SD 3.3 years). Values for the LTA's being *helpful* (mean 3.45, SD 1.79), *recommendable* (mean 3.33, SD 1.65) and *applicable* (mean 3.57, SD 1.85) were centered around the average of 3.5. The items *intuitive* (mean 4.57, SD 1.74) and *informative* (mean 4.53, SD 1.95) were above average. The only below-average item concerned its *accuracy* (mean 2.38, SD 1.36). Students rated the teaching unit as being excellent (mean 1.2, SD 0.54) but wanted practical training with an SP plus a simulated human translator first. Free-text responses revealed several concerns about translation errors that could jeopardize diagnostic decisions. Students feared that patient-physician communication mediated by the LTA could decrease empathy and raised concerns regarding data protection and technical reliability. Nevertheless, they appreciated the LTA's cost-effectiveness and usefulness as the best option when the gold standard is unavailable. They also reported wanting more medical-specific vocabulary and images to convey all information necessary for medical communication.

Conclusions: This study revealed the feasibility of using a speech-to-speech LTA in an undergraduate medical course. Although human translators remain the gold standard, LTAs could be valuable alternatives. Students appreciated the simulated teaching and recognized the LTA's potential benefits and risks for use in real-world clinical settings. To optimize patients' and health care professionals' experiences with LTAs, future investigations should examine specific design options for training interventions and consider the legal aspects of human-machine interaction in health care settings.

(JMIR Mhealth Uhealth 2021;9(12):e31559) doi:10.2196/31559

KEYWORDS

undergraduate medical students; translation app; simulation; physician-patient communication; mHealth; mobile applications; digital health; app development; language translation; translation apps

Introduction

Communication with patients ranks among the most important tasks for physicians and is thus an integral aspect of medical training [1-3]. Many institutional bodies and national catalogues of learning objectives have even designated communication with patients as a core competency [4-8]. However, several circumstances (eg, reduced consciousness and high emotionality) can impair communication with patients [9,10]; of them, language barriers can especially put timely, sufficient medical care at risk [11,12]. In nonmedical contexts, internet- and app-based digital translation services have become widely used to overcome such language barriers.

Albeit to a lesser extent than in the general public, the use of such translation services, particularly language translation apps (LTA), has gained traction in medical settings. Owing to increased globalization, migration, and refugee resettlement during the 21st century, patients often cannot speak the language spoken where they receive medical treatment and may thus be at risk of receiving less effective health care [13-15]. In response, human translators have been shown to benefit health care delivery in numerous ways; hence, various attempts have been made to train medical students or physicians to act as translators [16-20]. Although human translators are currently the gold standard for obtaining information from patients, obtaining their informed consent, and delivering negative news to them [10,21,22], such services may not always be available owing to timing and financial limitations. In such cases, digital technology such as LTAs seem to offer the second-best option [10,23,24]. LTAs generally function in one of three ways: text to text (ie, translation of a word or sentence from text into new text), text to speech (ie, translation of text from a tappable dictionary into voice output), and speech to speech (ie, translation of spoken sentences into voice output) [25].

In any case, LTAs are doubtlessly preferable to ad hoc alternatives such as relying on relatives, who may be too emotionally involved and thus prone to potentially fatal translation errors, or staff members who speak the same language as the patient, which would violate patient confidentiality and data security and could precipitate misunderstandings due to a lack of clinical and medical knowledge [21,26-29]. Nonetheless, guidelines applicable to communication via human translators may also be relevant when using LTAs, including ensuring direct communication, maintaining eye contact, talking to the person instead of the device, and using simple, clear, and sufficiently audible language [30,31].

In emergency medicine, studies have shown that using LTAs can overcome language barriers [32,33]. In particular, participants in those studies reported greater satisfaction with the more domain-specific app QuickSpeak than the generic Google Translate, although in both cases, they were worried about the inaccuracy of the translations [10,34-40]. Other studies have involved investigating the use of LTAs in clinical settings and shown their usefulness in simple communicational situations [34,41,42].

LTAs designed for clinical application often require the use of simple sentences. As a case in point, by using a text-to-speech app that simplified open-ended questions into closed-ended ones, Narang et al [33] found good user satisfaction and improvements in communication with patients with limited English proficiency. At the same time, inaccuracy in machine translation has been documented in various LTAs [40,42-44] and could precipitate misdiagnoses, incorrect prescriptions, and general mistreatment [45,46]. In sum, using LTAs in physician-patient communication demands caution, and physicians need to be trained in the adequate professional use of such apps [34,47].

Against that background, we conducted a proof-of-concept pilot study to examine the use of a speech-to-speech LTA in an undergraduate medical course, particularly its perceived usability, helpfulness, and meaningfulness in a simulated specific physician-patient situation, and to assess the teaching unit overall. Because the setting was simulated, we did not account for legal aspects (eg, data security) that would apply in clinical settings.

Methods

Setting and Participants

The proof-of-concept study was conducted in the Medical Faculty at the University of Tuebingen, Germany, between April and June 2019 in a medical communication course designed for the second preclinical year of medical school. Whereas participation in the course was mandatory, participation in the study was voluntarily, and 111 medical students were invited to participate. All participating students gave their written informed consent to participate, and data were collected anonymously.

Teaching Unit and Study Procedure

The teaching unit in this study was a 90-min seminar within the medical communication course taught by experienced instructors at the Department of Psychosomatic Medicine and Psychotherapy. All students had previously attended a lecture on the general aspects of physician-patient communication and how to inquire into and document a patient's medical history. They had also completed a repeat session on specifically assessing the history of present illness (HPI) in which they had practiced with a standardized patient (SP) presenting with chronic back pain.

The seminar on the LTA began with an introductory lecture on handling language-nonconcordant patients. Students also received information on how to work with human translators, including common pitfalls to avoid, and organizational background information specific to Tuebingen University Hospital (eg, how to request and pay human translators). Afterward, during the encounter with the SP, one medical student per 10-student group acted as the attending physician. The interaction was followed by a feedback session on general communication strategies and the overall management of the situation. The session ended with an interactive discussion on the usage of an LTA in physician-patient communication and appropriate medical strategies in the management of this patient case.

Standardized Patient

The SP was a 20-year-old male from Syria who spoke Arabic, a foreign language chosen owing to its relative frequency among patients in German hospitals and the low probability that participating students would understand or speak it. The SP was a young traveler who had experienced acute-onset nausea and vertigo hours before, which had worsened when he presented at the emergency department at 3 AM. Although the patient could not speak German or English, the attending physician's task was nevertheless to obtain some basic information about the patient and his HPI using the LTA. Full instructions are provided in Textboxes 1 and 2.

Textbox 1. Instruction for students acting as the attending physician.

Setting:

Emergency department, 3 AM.

You are the attending physician on your 4th night shift this week.

Case:

A 20-year-old male presents with acute-onset vertigo and nausea. The highly experienced on-duty nurse tells you, with slight exasperation, that the patient can communicate in Arabic only and that all attempts to gather basic information thus far have been futile. More important, the patient is in obvious distress (e.g. restlessly turning on the stretcher and clutching a kidney dish), and time seems of the essence. Knowing that no other staff on the ward can translate Arabic, you consider the option to request a professional translator. However, you are also aware that procuring a translator won't be easy at 3 a.m. As an alternative, you remember that one of your colleagues had introduced you to a language translation app, and you decide that now is the time to try it. After all, what do you have to lose?

Task:

Take the patient's HPI using the app on the iPhone. You have 10 min.

Textbox 2. Instruction for the standardized patients.

Setting:

Emergency department, 3 AM.

You are _____ (insert name), a 20-year-old from Syria who has been travelling across Europe with a friend for several weeks. Although you do not speak any German or English, you have managed quite well thus far.

This evening, you experienced a sudden onset of nausea and vertigo. You haven't been drinking alcohol or taken any drugs. The vertigo is rotational, similar to being on a merry-go-round, not a sailboat, and you feel the constant urge to vomit, even though you have not vomited thus far. Although lying on your back initially helped, your posture no longer affects your symptoms, and turning your head rapidly especially worsens your vertigo. You have never experienced a comparable condition, and you are unaware of any family history of vertigo.

You are usually an open-minded, easy-going person who loves to travel. You are in Europe for the first time, and so far, you have had lots of fun and appreciated all of the impressions made and opportunities encountered on your journey. Currently, however, you feel rather unwell and slightly scared because you can't judge the seriousness of your situation, and it doesn't help that you don't understand what people are saying. On the plus side, you very much like the young doctor taking care of you. You appreciate their effort to communicate with you on an app and thus try your best to communicate given the circumstances.

Remember that you speak Arabic exclusively. Only respond to whatever the app translates for you, even if you know that the original question in German was somewhat different. Please use simple sentences and only respond to what you've been asked (e.g. don't add information).

If you're asked any question not listed in these instructions, then please improvise. Remember, the session is part of a medical communication course in the second year of medical school. The simulation does not focus on the medical content as much as the general communication techniques and the specific situation of communicating via the app.

The encounter will last approximately 10 min.

LTA: iTranslate Converse

The app used, iTranslate Converse available for Android, Windows, and Mac, was chosen for its benefits identified by Khander et al [48] that we considered important for our simulation—that is, a wide range of available languages, ease of navigation and a high score (2.5/2.7) for "application comprehensiveness." It has also been shown to produce translations of similar quality to that of human translators, at least with simple sentences [35]. Preliminary tests for usability were also conducted by 2 authors (AHW and SZ).

The LTA was downloaded to an iPhone 7 device from the faculty's IT Department; the phone was not connected to the hospital's Wi-Fi, had no SIM card but had its languages preset to German and Arabic. The app was downloaded using Wi-Fi, accessed with the Apple ID of one author (AHW), and the connection was terminated immediately afterward because the LTA can be run offline.

Before students commenced the SP encounter, they were allotted time to become familiar with how the LTA worked. To translate speech, the student, either as the attending physician or patient, had to tap and hold a button while speaking, and releasing the button generated an audio translation. The system recognized the language spoken and automatically switched between the 2 preset languages.

Questionnaire

We developed a questionnaire with reference to the literature, models (eg, Unified Theory of Acceptance and Use of Technology) and ratings by expert panels [10,49-51]. Before the first seminar, the questionnaire had undergone cognitive pretesting by using the so-called "think aloud" method, in which the respondent concurrently verbalizes thoughts when responding to questionnaire items [52,53]. Consequently, minor adaptions to the questionnaire were made, and it was administered after the teaching unit but before the interactive discussion. The questionnaire collected demographic information (ie, age and gender) and ratings of the use of the LTA, the latter with 6 items on a 7-point Likert scale ranging from 1 (*don't agree at all*) to 7 (*completely agree*). The 6 items were (1) *helpful* (ie, able to support the task), (2) *intuitive* (ie, easy to use), (3) *informative* (ie, able to gather all necessary information), (4) *accurate* (ie, able to provide correct translations), (5) *recommendable* (ie, advisable for use by patients and clinical staff), and (6) *applicable* (ie, likely to be employed for personal use). Following those items, complementary free-text responses were requested for four additional items: (7) general impression of the LTA, (8) the LTA's benefits, (9) the LTA's risks, and (10) suggestions for improvement. These questions were added to obtain a deeper insight into students' considerations.

Teaching Unit Evaluation

Students anonymously evaluated the teaching unit on a secure platform for teaching assessment used by the faculty members for all courses at the university's medical school. The grading system used in German schools (1=*excellent*, 6=*insufficient*) was employed.

Respondents and Nonrespondents

At the beginning of the study, a questionnaire was placed on each medical student's desk. Students who answered and submitted the questionnaire were considered respondents, whereas those who left the questionnaire unanswered or did not submit it were considered nonrespondents.

Statistical Analysis

A sample size calculation was conducted with a 95% CI, population proportion of 50%, and a population size of 120, which resulted in a sample size of 92 respondents. The data were evaluated both quantitatively and qualitatively. Quantitative statistical analyses were performed with SPSS for Windows (version 25.0) under the assumption that the variables followed a normal distribution. First, for reliability analysis, the Cronbach α for internal consistency was computed to assess the 6 items in the quantitative part of the questionnaire (ie, Items 1-6). The internal consistency was satisfactory (α=.86), and reliability could not be improved by deleting items [54]. Corrected item–total correlations for all 6 items ranged between .45 and .81, and mean (SD) values were calculated. The final 4 items addressed in free-text responses (ie, Items 7-10) were evaluated in thematic content analysis using Microsoft Excel as coding software [55]. Themes in the data set were identified, analyzed, and documented. During content analysis, the reviewers familiarized themselves with the data and developed codes. After themes were sought, examined, and specified, results of the analysis were interpreted.

Ethics

Ethics approval for the study was obtained by the local ethics committee (No. 443/2018BO2).

Data Availability Statement

Full data are available on reasonable request by the corresponding author.

Results

Demographic Information

Of the 111 students in the course, 76 (68.5%) participated in the study. Most were women (n=45, 59.2%), and all were from 17 to 40 years of age (mean 20.66 years, SD 3.26 years).

Rating of the LTA

The mean rating across the first 6 items (ie, items 1-6) was only slightly above average (mean 3.64, SD 1.36). For the individual items, ratings for *helpful*, *recommendable*, and *applicable* were average. Students rated the LTA's being *intuitive* and *informative* as slightly above average (mean 4.52, SD 1.95) but its *accuracy* as rather below average (mean 2.38, SD 1.36). Table 1 reports the individual ratings of the 6 dimensions.

Table 1. Items regarding features of the language translation app.

Dimension	Rating (1=*don't agree at all*, 7=*completely agree*), mean (SD)
(1) Helpful	3.45 (1.79)
(2) Intuitive	4.57 (1.74)
(3) Informative	4.52 (1.95)
(4) Accurate	2.38 (1.36)
(5) Recommendable	3.33 (1.65)
(6) Applicable	3.57 (1.85)

Analysis of Free-Text Responses

Regarding items 6-10, most general impressions regarding the LTA contained largely critical comments about its accuracy. The students noticed, especially following comments from the SP and bilingual classmates, the possibility of severe translation errors, especially in translations from Arabic to German. Students also reported that the LTA largely failed to compute long, complex, or open-ended questions, and students instead suggested using close-ended questions to "get to the point." Many students reported worrying that planning and administering misguided follow-up or unnecessary interventions owing to linguistic misunderstandings could harm patients. Students additionally raised concerns about the technical challenges that LTAs can present (eg, poor connectivity or updates).

Regarding the LTA's benefits, students considered the app very useful for emergency situations and other brief conversations. Beyond that, they envisioned using the LTA more in hospital contexts than in ambulatory ones. A particularly positive aspect mentioned was that the LTA allows creating transcripts of dialogues, albeit only in its paid upgraded version. Another advantage was the LTA's cost-effectiveness relative to human translators and its potential use in translating uncommon languages and dialects not always known by hospitals' human translators.

Concerning the LTA's risks, students emphasized not only concerns about inaccuracy and its consequences but also the risk of fragmented, ineffective physician-patient communication. By using the LTA as an intermediary, many students experienced increased distance between themselves and the SP and added that the LTA needlessly prolonged the task of taking the SP's medical history. Students also reported worrying about losing empathy for patients and their symptoms by using the LTA. In particular, to assess mental distress or psychological comorbidities, they expressed doubts that the LTA would transmit the interpersonal information correctly. Furthermore, students were concerned that they would accidentally make offensive or politically incorrect statements to patients owing to the LTA's mistranslation. Other feedback focused on the extent to which the LTA guaranteed data confidentiality and whether machines such as LTAs would soon replace human interpreters.

Finally, regarding suggestions for improving the LTA, students generally aligned with their risk assessments by expressing a desire for more accurate translations. Considering the context of application, however, they contemplated the usefulness of predefined questions as a means to simplify the taking of medical history. Along similar lines, students wished for specialized terms adapted to the medical context and a "greater and more diverse vocabulary" both to prevent misunderstandings and to plan more precise interventions. Other students proposed adding pictures or predefined snapshots of difficult situations to improve the LTA's translation accuracy and ease of use. A final suggestion was for the LTA to reproduce the voice of the respective speaker to make taking the medical history more realistic.

Teaching Unit Evaluation

Analysis of the free-text responses in the evaluation of the teaching unit revealed that students were interested in the topic and generally liked the idea of including an app in the course's instruction. They also appreciated the possibility of practicing with the LTA with an SP in a controlled environment and receiving feedback from multiple sources afterward. At the same time, they underscored the topic's lack of connection to other learning content and demanded a better introduction to the topic, including practice with an SP along with a simulated human translator first. On the whole, students quantitatively rated the teaching unit in the official teaching evaluation system as excellent (mean 1.2, SD 0.54).

Discussion

Our proof-of-concept pilot study was designed to gain insight into the use of an LTA in a simulated setting in undergraduate medical education.

Principal Findings

Tested as part of an undergraduate medical curriculum, the LTA was perceived by medical students as being generally useful for the task of taking a HPI during acute care. Students appreciated the teaching unit taught in the seminar, even if they had only general interest in the topic and favored using the gold standard of human translators instead, which corroborates with other published findings [56].

When comparing human variants in translation, the role of the translator demands consideration. Ideally, a translator should act as a "conduit" transferring information neutrally from one party to the other [57,58]. However, depending on the circumstances, additional roles—managers, advocates, cultural mediators, or even co-therapists, to name a few—may equally need to be filled [59,60], none of which LTA can. Despite this

limitation, it does guarantee the basic function demanded of a translator—pure information exchange—and students should be made aware of its possibilities.

Although generally appreciative of the teaching unit, students complained that the challenge of using LTAs can be better confronted with more training, especially simulated training in communication with the aid of a human translator. Such training could easily be accommodated by the educational approach of spiral curriculum design [61].

Students' overall satisfaction with the LTA was high, however, as previous findings have also shown [47]. In particular, medical students considered the LTA easy to handle, possibly owing to its user-friendly interface and the fact that the students' age group is highly familiar with using mobile apps in their day-to-day lives. Nevertheless, the results suggest that students need to be trained in the professional application of LTAs, as recently stressed [47]. Students also acknowledged the potential of acquiring the necessary information with the LTA, information that they could not have obtained without the app, or at least not as rapidly, which confirms a known effect of using LTAs [62]. Even so, the students could readily specify the potential difficulties and pitfalls of using an LTA in real-world practice. In general, students feared that using an LTA to communicate with patients would threaten the physician's empathy, which is another known phenomenon of the replacement of human translators [63]. They were also concerned that translation errors could result in maltreatment or misdiagnosis, among other dangerous mistakes, that would jeopardize the patient's health and life. Their concern echoes findings from other research groups [10,27,35,40].

At the same time, our intended meaning of *errors* needs clarification. So-called "noncatastrophic errors" such as incorrect grammar or awkward translations may be tolerable, whereas critical mistranslations may not only cause confusion but also create the potential for serious harm [40,64]. Students need to be aware of such problems and need to be equipped with strategies to minimize them. After all, professional human translators are as liable to commit translation errors that become medical errors [29]. Similarly, an LTA's disadvantage may be its inflexibility compared to the flexibility that human interaction offers. With a human translator, at least one person can understand both languages and may be able to detect mismatches between speech and reactions and can adapt to cultural differences and communication-related concerns, whereas machines can accomplish neither task. Nevertheless, as Freyne et al [62] have shown, with repeated use of an LTA, health care professionals cultivate confidence in its translation abilities, possibly because they adapt their way of speaking to accommodate the possibilities and limitations of the app's functionality. To aid that process, some students wished for predefined sentences or images as a means to minimize misunderstandings. On that topic, the choice of Arabic as the SP's language might have aggravated the problem in our study because especially rare or non-European languages are prone to translation errors [38,65]. Indeed, more specific apps such as Quick Talk have been shown to be more helpful in emergency medicine settings than Google Translate [10]. Additionally, when used with native speakers on both ends, LTAs can usually produce the correct meaning, even if the translation is not completely accurate [10]. An ideal solution might be a mix of preset questions as options supplemented with images and the additional function of free-text entry.

The reluctance to trust the LTA's accuracy was also reflected by the fact that whereas all ratings for the 6 items correlated with each other, no intercorrelation emerged between the students' rating of the LTA's helpfulness and their assessment of its accuracy, which indicates that the students appreciated using an LTA for collecting the medical history of language-nonconcordant patients but were partly deterred by its technical restrictions.

Limitations

Our study had several limitations. First, it was conducted during only one semester with medical students from only one faculty in Germany. Those constraints upon the sample and the study limit the generalizability of the findings. Second, only one LTA was used in the study, meaning that the findings might not be applicable for other LTAs. Third, we tested only one language, Arabic, chosen as a compromise between a language encountered often enough amongst patients in hospitals in Germany and a language with little risk of being known by many students, which would have jeopardized their learning experience. Because students participated in the study on a voluntary basis, we cannot exclude selection bias; however, given the number of respondents and their age and gender distribution, the sample can be considered to represent the student population at Tuebingen Medical Faculty in general. Finally, the study was designed as a self-report paper-and-pencil survey with quantitative and open-ended questions. Self-report surveys are generally open to bias, and responses to the items were analyzed in accordance with the level of data available. Consequently, there was no need to compute moderator or between-group analyses.

Despite those limitations, we strongly believe that the pilot study offers valuable insight into the use of a speech-to-speech LTA that offers the possibility of speaking freely, in an undergraduate medical curriculum. Those initial data show that such an LTA can be helpful in obtaining the HPIs of patients in simulated acute care settings. It remains unclear whether this app could be reliably integrated into actual patient care where other additional aspects (eg, data protection and legal liability) would have to be considered.

Comparison With Prior Work

To the best of our knowledge, this study was the first to examine an LTA in undergraduate medical education, which offers the possibility of speaking freely and thus approximates a normal conversation without language barriers. Findings concerning the evaluation of the LTA used were primarily in line with published results. However, they additionally showed that students need training in the use of LTAs, which confirms the recently identified need among physicians to be properly prepared for using LTAs.

Conclusions

Our proof-of-concept study revealed that using a speech-to-speech LTA in an undergraduate medical class is feasible. Students primarily benefitted from the feedback from multiple sources as part of the simulation, as well as from becoming familiar with the general possibilities and potential drawbacks of using LTAs.

Although human translators remain the gold standard and are preferred by patients and health care professionals, LTAs might pose a valuable alternative to less favorable options (eg, relying on bystanders and family members) or a valuable addition to the off-the-cuff approach because they do not present the obstacles that human translators often do (eg, timing, cost, and inflexibility) [10,21,26,33,56,66]. Students liked the idea of studying the topic as part of their simulated teaching. However, they also recognized the risks of using such an LTA in clinical settings with real patients.

The COVID-19 pandemic has altered health care in diverse ways, including by increasing the acceptability of telemedical health care solutions. Further investigations should examine changes in the usage and acceptability of LTAs and how training interventions can be designed to optimize patients' and health care professionals' experiences with LTAs. At the same time, legal concerns (eg, data security) need to be addressed in future LTA training courses because they are essential to consider when LTAs are intended for use in clinical practice. As a next step, we propose the development of a full-scale training course for undergraduate medical students that addresses communication with language-nonconcordant patients, including algorithms and strategies for using LTAs and the gold standard: face-to-face or video-based human translation.

Acknowledgments

We thank our information technology department for technical support. We acknowledge support with financing publication fees by "Deutsche Forschungsgemeinschaft" and Open Access Publishing Fund of the University of Tuebingen.

Conflicts of Interest

None declared.

References

1. Ha J, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
2. Stewart M. Effective physician-patient communication and health outcomes: a review. CMAJ 1995 May 01;152(9):1423-1433 [FREE Full text] [Medline: 7728691]
3. Keifenheim KE, Teufel M, Ip J, Speiser N, Leehr EJ, Zipfel S, et al. Teaching history taking to medical students: a systematic review. BMC Med Educ 2015 Sep 28;15:159 [FREE Full text] [doi: 10.1186/s12909-015-0443-x] [Medline: 26415941]
4. Edgar L, Roberts S, Holmboe E. Milestones 2.0: A Step Forward. J Grad Med Educ 2018 Jun;10(3):367-369 [FREE Full text] [doi: 10.4300/JGME-D-18-00372.1] [Medline: 29946411]
5. Fischer M, Bauer D, Mohn K, NKLM-Projektgruppe. Finally finished! National Competence Based Catalogues of Learning Objectives for Undergraduate Medical Education (NKLM) and Dental Education (NKLZ) ready for trial. GMS Z Med Ausbild 2015;32(3):Doc35 [FREE Full text] [doi: 10.3205/zma000977] [Medline: 26677513]
6. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach 2007 Sep;29(7):642-647. [doi: 10.1080/01421590701746983] [Medline: 18236250]
7. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Acad Med 2001 Apr;76(4):390-393. [doi: 10.1097/00001888-200104000-00021] [Medline: 11299158]
8. Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, et al. Doctor-patient communication: the Toronto consensus statement. BMJ 1991 Nov 30;303(6814):1385-1387 [FREE Full text] [doi: 10.1136/bmj.303.6814.1385] [Medline: 1760608]
9. Grow RW, Sztajnkrycer MD, Moore BR. Language barriers as a reported cause of prehospital care delay in Minnesota. Prehosp Emerg Care 2008;12(1):76-79. [doi: 10.1080/10903120701709878] [Medline: 18189182]
10. Turner AM, Choi YK, Dew K, Tsai M, Bosold AL, Wu S, et al. Evaluating the usefulness of translation technologies for emergency response communication: a scenario-based study. JMIR Public Health Surveill 2019 Jan 28;5(1):e11171 [FREE Full text] [doi: 10.2196/11171] [Medline: 30688652]
11. Flores G. Language barriers to health care in the United States. N Engl J Med 2006 Jul 20;355(3):229-231. [doi: 10.1056/NEJMp058316] [Medline: 16855260]
12. Meischke HW, Calhoun RE, Yip M, Tu S, Painter IS. The effect of language barriers on dispatching EMS response. Prehosp Emerg Care 2013;17(4):475-480. [doi: 10.3109/10903127.2013.811565] [Medline: 23952940]
13. Kale E, Syed HR. Language barriers and the use of interpreters in the public health services. A questionnaire-based survey. Patient Educ Couns 2010 Nov;81(2):187-191. [doi: 10.1016/j.pec.2010.05.002] [Medline: 20542656]

14. Ngo-Metzger Q, Sorkin DH, Phillips RS, Greenfield S, Massagli MP, Clarridge B, et al. Providing high-quality care for limited English proficient patients: the importance of language concordance and interpreter use. J Gen Intern Med 2007 Nov;22 Suppl 2:324-330 [FREE Full text] [doi: 10.1007/s11606-007-0340-z] [Medline: 17957419]
15. Waxman MA, Levitt MA. Are diagnostic testing and admission rates higher in non-English-speaking versus English-speaking patients in the emergency department? Ann Emerg Med 2000 Nov;36(5):456-461. [doi: 10.1067/mem.2000.108315] [Medline: 11054199]
16. Jacobs EA, Shepard DS, Suaya JA, Stone E. Overcoming language barriers in health care: costs and benefits of interpreter services. Am J Public Health 2004 May;94(5):866-869. [doi: 10.2105/ajph.94.5.866] [Medline: 15117713]
17. Jacobs EA, Diamond LC, Stevak L. The importance of teaching clinicians when and how to work with interpreters. Patient Educ Couns 2010 Feb;78(2):149-153. [doi: 10.1016/j.pec.2009.12.001] [Medline: 20036480]
18. McEvoy M, Santos MT, Marzan M, Green EH, Milan FB. Teaching medical students how to use interpreters: a three year experience. Med Educ Online 2009 Sep 02;14:12 [FREE Full text] [doi: 10.3885/meo.2009.Res00309] [Medline: 20165526]
19. Bischoff A, Hudelson P. Access to healthcare interpreter services: where are we and where do we need to go? Int J Environ Res Public Health 2010 Jul;7(7):2838-2844 [FREE Full text] [doi: 10.3390/ijerph7072838] [Medline: 20717543]
20. Silva MD, Genoff M, Zaballa A, Jewell S, Stabler S, Gany FM, et al. Interpreting at the end of life: a systematic review of the impact of interpreters on the delivery of palliative care services to cancer patients with limited English proficiency. J Pain Symptom Manage 2016 Mar;51(3):569-580 [FREE Full text] [doi: 10.1016/j.jpainsymman.2015.10.011] [Medline: 26549596]
21. Flores G. The impact of medical interpreter services on the quality of health care: a systematic review. Med Care Res Rev 2005 Jun;62(3):255-299. [doi: 10.1177/1077558705275416] [Medline: 15894705]
22. Panayiotou A, Gardner A, Williams S, Zucchi E, Mascitti-Meuter M, Goh AM, et al. Language translation apps in health care settings: expert opinion. JMIR Mhealth Uhealth 2019 Apr 09;7(4):e11316 [FREE Full text] [doi: 10.2196/11316] [Medline: 30964446]
23. Locatis C, Williamson D, Sterrett J, Detzler I, Ackerman M. Video medical interpretation over 3G cellular networks: a feasibility study. Telemed J E Health 2011 Dec;17(10):809-813 [FREE Full text] [doi: 10.1089/tmj.2011.0084] [Medline: 22011055]
24. Sciarra A, Batigália FF, Oliveira MAB. Technological devices improving system of translating languages: what about their usefulness on the applicability in medicine and health sciences? Braz J Cardiovasc Surg 2015;30(6):664-667 [FREE Full text] [doi: 10.5935/1678-9741.20150087] [Medline: 26934409]
25. Chen K. Top translator apps can help you communicate with patients who have limited English proficiency. OBG Management 2017;29(10):36-37 [FREE Full text]
26. Flores G, Abreu M, Barone CP, Bachur R, Lin H. Errors of medical interpretation and their potential clinical consequences: a comparison of professional versus ad hoc versus no interpreters. Ann Emerg Med 2012 Nov;60(5):545-553. [doi: 10.1016/j.annemergmed.2012.01.025] [Medline: 22424655]
27. Chang D, Thyer I, Hayne D, Katz D. Using mobile technology to overcome language barriers in medicine. Ann R Coll Surg Engl 2014 Sep;96(6):e23-e25 [FREE Full text] [doi: 10.1308/003588414X13946184903685] [Medline: 25198966]
28. Diamond L, Wilson-Stronks A, Jacobs EA. Do hospitals measure up to the national culturally and linguistically appropriate services standards? Med Care 2010 Dec;48(12):1080-1087. [doi: 10.1097/MLR.0b013e3181f380bc] [Medline: 21063229]
29. Flores G, Laws MB, Mayo SJ, Zuckerman B, Abreu M, Medina L, et al. Errors in medical interpretation and their potential clinical consequences in pediatric encounters. Pediatrics 2003 Jan;111(1):6-14. [doi: 10.1542/peds.111.1.6] [Medline: 12509547]
30. Lake VE, Beisly AH. Translation apps: increasing communication with dual language learners. Early Childhood Educ J 2019 Mar 16;47(4):489-496. [doi: 10.1007/s10643-019-00935-7]
31. Robertson K. Tips for successful parent-teacher conferences with bilingual families. Colorín Colorado. 2007. URL: https://www.colorincolorado.org/article/tips-successful-parent-teacher-conferences-bilingual-families [accessed 2021-11-01]
32. Tate RC, Hodkinson PW, Meehan-Coussee K, Cooperstein N. Strategies used by prehospital providers to overcome language barriers. Prehosp Emerg Care 2016;20(3):404-414. [doi: 10.3109/10903127.2015.1102994] [Medline: 26807873]
33. Narang B, Park SY, Norrmén-Smith IO, Lange M, Ocampo AJ, Gany FM, et al. The use of a mobile application to increase access to interpreters for cancer patients with limited English proficiency: a pilot study. Med Care 2019 Jun;57 Suppl 6 Suppl 2:S184-S189 [FREE Full text] [doi: 10.1097/MLR.0000000000001035] [Medline: 31095059]
34. Albrecht U, Behrends M, Schmeer R, Matthies HK, von Jan U. Usage of multilingual mobile translation applications in clinical settings. JMIR Mhealth Uhealth 2013 Apr 23;1(1):e4 [FREE Full text] [doi: 10.2196/mhealth.2268] [Medline: 25100677]
35. Chen X, Acosta S, Barry AE. Machine or human? Evaluating the quality of a language translation mobile app for diabetes education material. JMIR Diabetes 2017 Jun 29;2(1):e13 [FREE Full text] [doi: 10.2196/diabetes.7446] [Medline: 30291084]
36. Randhawa G, Ferreyra M, Ahmed R, Ezzat O, Pottie K. Using machine translation in clinical practice. Can Fam Physician 2013 Apr;59(4):382-383 [FREE Full text] [Medline: 23585608]

37. Taylor RM, Crichton N, Moult B, Gibson F. A prospective observational study of machine translation software to overcome the challenge of including ethnic diversity in healthcare research. Nurs Open 2015 Apr;2(1):14-23 [FREE Full text] [doi: 10.1002/nop2.13] [Medline: 27708797]
38. Turner AM, Dew KN, Desai L, Martin N, Kirchhoff K. Machine translation of public health materials from English to Chinese: a feasibility study. JMIR Public Health Surveill 2015;1(2):e17 [FREE Full text] [doi: 10.2196/publichealth.4779] [Medline: 27227135]
39. Zeng-Treitler Q, Kim H, Rosemblat G, Keselman A. Can multilingual machine translation help make medical record content more comprehensible to patients? Stud Health Technol Inform 2010;160(Pt 1):73-77. [Medline: 20841653]
40. Patil S, Davies P. Use of Google Translate in medical communication: evaluation of accuracy. BMJ 2014 Dec 15;349:g7392 [FREE Full text] [doi: 10.1136/bmj.g7392] [Medline: 25512386]
41. Abujarad F, Vaca FE. Health tool for alcohol use disorders among Latinos in emergency department. Proc Int Symp Hum Factors Ergon Healthc 2015 Jun;4(1):12-19 [FREE Full text] [doi: 10.1177/2327857915041005] [Medline: 26844234]
42. Nguyen-Lu N, Reide P, Yentis SM. 'Do you have a stick in your mouth?'– use of Google Translate as an aid to anaesthetic pre-assessment. Anaesthesia 2010;65(1):96-97. [doi: 10.1111/j.1365-2044.2009.06184_4.x]
43. Beh THK, Canty DJ. English and Mandarin translation using Google Translate software for pre-anaesthetic consultation. Anaesth Intensive Care 2015 Nov;43(6):792-793 [FREE Full text] [Medline: 26603812]
44. Kaliyadan F, Gopinathan Pillai S. The use of Google language tools as an interpretation aid in cross-cultural doctor-patient interaction: a pilot study. Inform Prim Care 2010;18(2):141-143 [FREE Full text] [doi: 10.14236/jhi.v18i2.764] [Medline: 21078237]
45. Butow PN, Goldstein D, Bell ML, Sze M, Aldridge LJ, Abdo S, et al. Interpretation in consultations with immigrant patients with cancer: how accurate is it? J Clin Oncol 2011 Jul 10;29(20):2801-2807. [doi: 10.1200/JCO.2010.34.3335] [Medline: 21670462]
46. Nápoles AM, Santoyo-Olsson J, Karliner LS, Gregorich SE, Pérez-Stable EJ. Inaccurate language interpretation and its clinical significance in the medical encounters of Spanish-speaking Latinos. Med Care 2015 Nov;53(11):940-947 [FREE Full text] [doi: 10.1097/MLR.0000000000000422] [Medline: 26465121]
47. Ji X, Chow E, Abdelhamid K, Naumova D, Mate KK, Bergeron A, et al. Utility of mobile technology in medical interpretation: A literature review of current practices. Patient Educ Couns 2021 Sep;104(9):2137-2145. [doi: 10.1016/j.pec.2021.02.019] [Medline: 33653659]
48. Khander A, Farag S, Chen KT. Identification and evaluation of medical translator mobile applications using an adapted APPLICATIONS scoring system. Telemed J E Health 2018 Aug;24(8):594-603. [doi: 10.1089/tmj.2017.0150] [Medline: 29271702]
49. Brooke J. SUS - A quick and dirty usability scale. 1996. URL: https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf [accessed 2021-11-01]
50. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
51. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
52. van Someren MW, Barnard YF, Sandberg J. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes (volume 12). Cambridge, MA: Academic Press; 1994.
53. Ericsson KA, Simon HA. Verbal reports as data. Psychol Rev 1980;87(3):215-251. [doi: 10.1037/0033-295x.87.3.215]
54. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/bf02310555]
55. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
56. Diamond LC, Schenker Y, Curry L, Bradley EH, Fernandez A. Getting by: underuse of interpreters by resident physicians. J Gen Intern Med 2009 Feb;24(2):256-262 [FREE Full text] [doi: 10.1007/s11606-008-0875-7] [Medline: 19089503]
57. Matthews C, Johnson M, Noble C, Klinken A. Bilingual health communicators: role delineation issues. Aust Health Rev 2000;23(3):104-112. [doi: 10.1071/ah000104] [Medline: 11186042]
58. Sleptsova M, Hofer G, Morina N, Langewitz W. The role of the health care interpreter in a clinical setting--a narrative review. J Community Health Nurs 2014;31(3):167-184. [doi: 10.1080/07370016.2014.926682] [Medline: 25051322]
59. Dysart-Gale D. Clinicians and medical interpreters: negotiating culturally appropriate care for patients with limited English ability. Fam Community Health 2007;30(3):237-246. [doi: 10.1097/01.FCH.0000277766.62408.96] [Medline: 17563485]
60. Hsieh E. Interpreters as co-diagnosticians: overlapping roles and services between providers and interpreters. Soc Sci Med 2007 Feb;64(4):924-937. [doi: 10.1016/j.socscimed.2006.10.015] [Medline: 17126465]
61. Harden RM. What is a spiral curriculum? Med Teach 1999;21(2):141-143. [doi: 10.1080/01421599979752] [Medline: 21275727]

62. Freyne J, Bradford D, Pocock C, Silvera-Tawil D, Harrap K, Brinkmann S. Developing digital facilitation of assessments in the absence of an interpreter: participatory design and feasibility evaluation with allied health groups. JMIR Form Res 2018 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/formative.8032] [Medline: 30684405]
63. Pugh M, Vetere A. Lost in translation: an interpretative phenomenological analysis of mental health professionals' experiences of empathy in clinical work with an interpreter. Psychol Psychother 2009 Sep;82(Pt 3):305-321. [doi: 10.1348/147608308X397059] [Medline: 19208292]
64. Ettelaie E, Gandhe S, Georgiou P, Knight K, Marcu D, Narayanan S, et al. Transonics: A practical speech-to-speech translator for English-Farsi medical dialogues. In: ACLdemo '05: Proceedings of the ACL 2005 on Interactive poster and demonstration sessions. 2005 Presented at: ACL 2005 Interactive Poster and Demonstration session; June 25-30, 2005; Ann Arbor, MI p. 89-92. [doi: 10.3115/1225753.1225776]
65. Kirchhoff K, Turner AM, Axelrod A, Saavedra F. Application of statistical machine translation to public health information: a feasibility study. J Am Med Inform Assoc 2011;18(4):473-478 [FREE Full text] [doi: 10.1136/amiajnl-2011-000176] [Medline: 21498805]
66. Schenker Y, Pérez-Stable EJ, Nickleach D, Karliner LS. Patterns of interpreter use for hospitalized patients with limited English proficiency. J Gen Intern Med 2011 Jul;26(7):712-717 [FREE Full text] [doi: 10.1007/s11606-010-1619-z] [Medline: 21336672]

Abbreviations

HPI: history of present illness
LTA: language translation app
SP: standardized patient

Edited by G Eysenbach; submitted 25.06.21; peer-reviewed by M Lotto; comments to author 19.07.21; revised version received 11.08.21; accepted 25.08.21; published 02.12.21.

Please cite as:
Herrmann-Werner A, Loda T, Zipfel S, Holderried M, Holderried F, Erschens R
Evaluation of a Language Translation App in an Undergraduate Medical Communication Course: Proof-of-Concept and Usability Study
JMIR Mhealth Uhealth 2021;9(12):e31559
URL: https://mhealth.jmir.org/2021/12/e31559
doi:10.2196/31559
PMID:34860678

Original Paper

Selecting and Evaluating Mobile Health Apps for the Healthy Life Trajectories Initiative: Development of the eHealth Resource Checklist

Leigh M Vanderloo[1], MSc, PhD; Sarah Carsley[2], PhD; Payal Agarwal[3], MD; Flavia Marini[4], PhD; Cindy-Lee Dennis[5], PhD; Catherine Birken[1], MD

[1]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada

[2]Department of Health Promotion, Chronic Disease and Injury Prevention, Public Health Ontario, Toronto, ON, Canada

[3]Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[4]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada

[5]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Leigh M Vanderloo, MSc, PhD
Child Health Evaluative Sciences
The Hospital for Sick Children
686 Bay St
Toronto, ON, M5G 0A4
Canada
Phone: 1 5194956306
Email: lvande32@uwo.ca

Abstract

Background: The ubiquity of smartphones and mobile devices in the general population presents an unprecedented opportunity for preventative health. Not surprisingly, the use of electronic health (eHealth) resources accessed through mobile devices in clinical trials is becoming more prevalent; the selection, screening, and collation of quality eHealth resources is necessary to clinical trials using these technologies. However, the constant creation and turnover of new eHealth resources can make this task difficult. Although syntheses of eHealth resources are becoming more common, their methodological and reporting quality require improvement so as to be more accessible to nonexperts. Further, there continues to be significant variation in quality criteria employed for assessment, with no clear method for developing the included criteria. There is currently no single existing framework that addresses all six dimensions of mobile health app quality identified in Agarwal et al's recent scoping review (ie, basic descriptions of the design and usage of the resource; technical features and accessibility; health information quality; usability; evidence of impact; and user engagement and behavior change). In instances where highly systematic tactics are not possible (due to time constraints, cost, or lack of expertise), there may be value in adopting practical and pragmatic approaches to helping researchers and clinicians identify and disseminate e-resources.

Objective: The study aimed to create a set of guidelines (ie, a checklist) to aid the members of the Healthy Life Trajectories Initiative (HeLTI) Canada trial—a preconception randomized controlled clinical trial to prevent child obesity—to assist their efforts in searching, identifying, screening, and including selected eHealth resources for participant use in the study intervention.

Methods: A framework for searching, screening, and selecting eHealth resources was adapted from the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) checklist for systematic and scoping reviews to optimize the rigor, clarity, and transparency of the process. Details regarding searching, selecting, extracting, and assessing quality of eHealth resources are described.

Results: This study resulted in the systematic development of a checklist consisting of 12 guiding principles, organized in a chronological versus priority sequence to aid researchers in searching, screening, and assessing the quality of various eHealth resources.

Conclusions: The eHealth Resource Checklist will assist researchers in navigating the eHealth resource space by providing a mechanism to detail their process of developing inclusion criteria, identifying search location, selecting and reviewing evidence, extracting information, evaluating the quality of the evidence, and synthesizing the extracted evidence. The overarching goal of

this checklist is to provide researchers or generalists new to the eHealth field with a tool that balances pragmatism with rigor and that helps standardize the process of searching and critiquing digital material—a particularly important aspect given the recent explosion of and reliance on eHealth resources. Moreover, this checklist may be useful to other researchers and practitioners developing similar health interventions.

(JMIR Mhealth Uhealth 2021;9(12):e27533) doi:10.2196/27533

KEYWORDS

eHealth resources; applications; quality assessment; preconception health

Introduction

The ubiquity of smartphones or mobile devices in the general population represents an unprecedented opportunity to reach diverse individuals for preventative health. The use of smartphone apps for the provision of health information, promotion, and intervention has become widespread [1], even sparking a new label, "digitized health promotion" [1]. Electronic health (eHealth) interventions or programs use diverse information and communication technologies (web- or mobile-based) to improve or facilitate health behaviors. Recent systematic reviews of trials evaluating eHealth resources in adolescents and adults observed significant reductions in BMI and improvements in dietary behaviors, physical activity, and self-monitoring [2-4]. In addition, eHealth resources (ie, online web resources or apps) designed to enhance healthy behaviors are appealing: they are highly accessible and sustainable [5], can be tailored to specific populations [6,7], and provide low-cost scalable opportunities for population-wide promotion of health behaviors [8]. Although most interventions using eHealth resources are brief and relatively simple, the content varies greatly and there is a lack of standardized methodology to rigorously evaluate the quality and effectiveness of eHealth resources. However, the proliferation of apps for chronic disease management and prevention poses challenges for clinicians, policy makers, and patients in understanding which apps are most likely to provide benefit.

The prevalence of noncommunicable diseases—including cardiovascular diseases, respiratory diseases, diabetes, and mental health issues—is on the rise worldwide and preventive strategies are urgently needed [9]. To address this issue, the Healthy Life Trajectories Initiative (HeLTI) Canada study was designed. This randomized controlled trial aims to evaluate a preconception to early childhood telephone-based public health intervention with tailored eHealth resources for women and their partners to optimize growth and development among children in Canada [10,11]. This clinical trial uses a Developmental Origins of Health and Disease approach, which is based on the notion that environmental factors interact with genes from preconception to early childhood and that this programming affects a child's health into adulthood [12]. HeLTI builds upon the diverse clinical trial research capacity in Canada, while harmonizing the intervention and outcome measures with three other international HeLTI trials (in China, India, and South Africa) to generate evidence that will inform national policy and decision-making for the improvement of health and reduction of noncommunicable diseases starting in preconception [13]. The primary objective of HeLTI Canada is to determine whether a 4-phase intervention, from preconception into pregnancy through to infancy and early childhood, can reduce the rates of child overweight and obesity. Secondary objectives aim to reduce child Z-score of BMI (zBMI) and improve zBMI trajectories, cardiometabolic risk factors, health behaviors (nutrition, physical activity, sedentary behavior, and sleep), and development and school readiness at the age of 5 years. Maternal and paternal health outcomes and parenting behaviors are further examined to provide a family-level evaluation.

In the HeLTI Canada trial, participants in the intervention group are assigned to an experienced public health nurse who provides telephone-based collaborative care to support women and their partners to improve their health, modify their health behaviors, or improve their parenting skills. Nurses perform a detailed telephone assessment to identify preconception risk factors or parenting concerns, develop a structured management plan based on family preference, and conduct scheduled follow-up calls to assist the participants in meeting their outlined health goals. Each woman and their partner will be provided with their own secure login to a website that includes personalized web-based eHealth resources based on their specific goals. This selection of eHealth resources will be curated and customized for the participant and will be used by the nurses to provide individual-based care with resources that are convenient and readily accessible to help them achieve their goals. Given the growing popularity of smartphones, tablets, and apps [14], coupled with the noted shift in how individuals consume health information [15], the inclusion of eHealth resources in the HeLTI Canada trial allows the public health nurse and participant to use evidence-based tools to work collaboratively to address identified health needs.

Working groups were created to identify, evaluate, and recommend eHealth resources specific to health goals (or behaviors of focus) that could be used in the HeLTI Canada intervention. These eHealth resources were meant to be easily accessible on a smartphone, tablet, or computer, and provide personalized, innovative, and engaging support to participants with diverse preventive health needs. The selection, screening, and collation of quality eHealth resources was a necessary component to develop and enhance the HeLTI Canada trial intervention. However, the constant creation and turnover of new apps can make this task difficult and time-consuming. Although syntheses of eHealth resources are becoming more common, their methodological and reporting quality require improvement so as to be more accessible to nonexperts [16]. Further, there continues to be significant variation in quality criteria employed for assessment, with no clear method for

developing the included criteria. Per the recent scoping review by Agarwal and colleagues [16], there is currently no single existing framework that addresses all six identified dimensions of mHealth app quality (ie, basic descriptions of the design and usage of the resource; technical features and accessibility; health information quality; usability; evidence of impact; and user engagement and behavior change). In instances where highly systematic tactics are not possible (due to time constraints, cost, or lack of expertise), there is still value in adopting practical and pragmatic approaches to helping researchers and clinicians navigate this space. Specifically, guiding principles that researchers and clinicians could use to select quality eHealth resources are an identified need [17,18]. As there were no available guidelines to assist the HeLTI Canada app working groups in this task, we aimed to address this gap. As such, we sought to create a set of guidelines (ie, a checklist) to aid researchers and clinicians in searching, identifying, screening, and selecting eHealth resources for use in research or clinical practice. These guidelines were developed through the experience of HeLTI Canada researchers as they selected eHealth resources for the trial intervention.

Methods

Overview

To optimize the rigor, clarity, and transparency of the current guidelines, the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) checklists for systematic [19] and scoping [20] reviews were adapted to provide a framework for screening eHealth resources. Reporting guidelines outline a minimum set of items to include in research reports and have been shown to increase methodological transparency, uptake of research findings, and intervention fidelity [21].

Selection Criteria of eHealth Resources

In this study, eHealth resources were considered for inclusion if the following conditions were met: (1) targeted children or parents influencing behavior change in children; (2) was either a website or app that provided content on health behaviors (physical activity, sedentary behaviors, screen use, nutrition, wellness, healthy weights, active play, healthy habits); (3) had a minimum quality indicator such as a rating of ≥4 stars if the resource was an iOS app or ≥10,000 installs if the resources was an Android app; and (4) was available in English and/or French.

In addition, eHealth resources were excluded for the following reasons: (1) they relied *solely* on data from a paired external device (ie, wearable technology like a Fitbit, with no option of manually inputting data; this was to ensure all apps would be used by all participants without the need to purchase additional hardware), (2) they were not oriented toward individual users (ie, if they were directed toward school or gym programs; this was to ensure open and wide access to the resource), or (3) they had content focused primarily on the management of specific health conditions (ie, heart disease; this exclusion criterion was included to accommodate a universal and more general population approach).

Information Sources: Locating eHealth Resources

When searching for eHealth resources, multiple information sources were considered, including the following: (1) the Apple Store (iOS) and the Google Play Store (Android); (2) literature reviews of eHealth articles; (3) consultations with eHealth experts (author PA and Practical Apps [working group], Women's Health College, Toronto, Canada), reputable public organizations and authorities, and government via email; and (4) recommendations from other experts, including family doctors, pediatricians in primary care, and child caregivers.

Search Strategy

No date restrictions were placed on the search, which was completed in August 2018. Using the previously identified information sources, the following keywords were used to retrieve e-resources: sleep, physical activity, sedentary behaviors, screens, screen time, nutrition, children, smartphone app, online resource, e-resource, eHealth resource, wellness, weight management, healthy weight, play, activity, fitness, development, healthy habits, healthy behavior, behavior change, monitoring, tracking, and health advice. Once retrieved, all resources were exported and saved in an editable Microsoft Excel (Microsoft Corp) spreadsheet via Google Docs (Google LLC) and duplicates were removed manually. Each resource was assigned a unique identification number.

Process for Selecting Resources

Based on the eligibility criteria, a standardized screening form (Multimedia Appendix 1) was developed; initially, 10 resources were selected to pilot test and refine the form (91.3% agreement across 6 researchers). Next, all selected resources that met the eligibility were reviewed collectively as a team and a final suite of eHealth resources was identified for inclusion. All disagreements in selection were discussed and resolved by consensus and mediated where necessary.

Methods for Charting and Extracting Data

All eHealth resources were assessed by the 6 reviewers to determine whether they reported on one or multiple health behaviors of interest, and whether the eHealth resource was child- or parent-focused. All data of interest from the eHealth resources (ie, behavior and population of focus, details about the resources) were entered into an Excel sheet stored in Google Docs.

Quality Assessment of eHealth Resources

With the multitude of health apps identified, it was essential to evaluate the quality of each resource. All eHealth resources were evaluated by a minimum of two team members and—depending on the type of eHealth resource selected—different quality assessment tools were used. When selecting appropriate tools, it is important to consider the needs and preferences of the resource user.

Driven by consultations with eHealth experts, Stoyanov and colleagues' [22] Mobile App Rating Scale (MARS) was used to evaluate the quality of the apps. The scale contains 23 items, each rated on a 5-point scale (where 1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent) and categorized into three sections: classification, app quality, and satisfaction. The

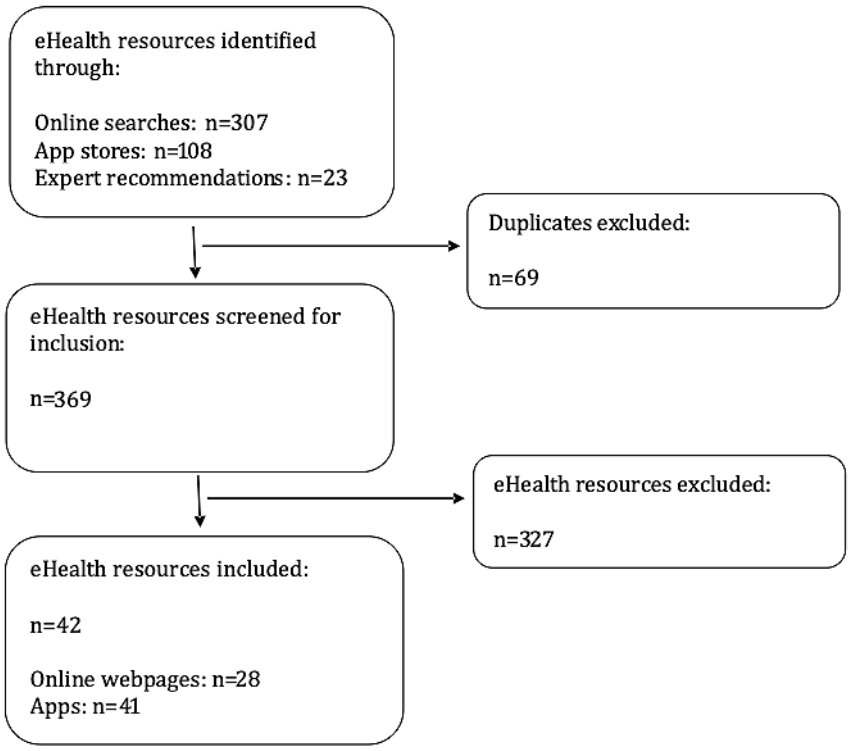
classification section is only used for descriptive purposes. The 19-item app quality section rates apps on 4 subscales: engagement, functionality, aesthetics, and information quality. The subjective quality section contains 4 items evaluating the user's overall satisfaction. The MARS is scored by calculating the mean scores of the app quality subscales and the total mean score.

For online web resources, the DISCERN tool was used [23]. DISCERN is a brief questionnaire that provides users with a valid and reliable way of assessing the quality of written information on treatment choices for a health problem. The tool consists of 15 key questions plus an overall quality rating. Each question represents a separate quality criterion and is rated on a 5-point scale where 1=no, 2-4=partial, and 5=yes. The rating scale has been designed to help researchers decide whether the quality criterion in question is present or has been "fulfilled" by the eHealth resource.

Reporting the Individual eHealth Resources and Key Content

The total number of resources identified, selected, screened, and assessed for inclusion was recorded (Figure 1). Next, the selected eHealth resources were included in a standardized extraction form and grouped based on health behavior of focus (eg, sleep, physical activity, nutrition, weight management, screen time). Key information about the eHealth resource (type of resource, health behavior, target audience [child or parent], type of content or activities offered, science-backed, etc) was charted.

Figure 1. The eHealth resource selection flow diagram.



Process Refinement and Adoption

Once the initial guidelines were drafted by two members of the working group (LMV and SC), an example online resource and an app were pilot tested by 6 reviewers and 1 mediator to ensure each item or "step" progressed logically and was comprehensive. Following this refinement process, the checklist was modified accordingly. Face validity for the guidelines was achieved by sharing the itemized list with all members of the working group and modifying it further. The final approach was shared and adopted by all members of the HeLTI team to assist with their eHealth resource searching and screening efforts.

Results

Resulting from the previously described steps, a checklist consisting of 12 guiding principles was systematically developed—organized in a chronological versus priority sequence to aid researchers in searching, screening, and assessing the quality of various eHealth resources in a pragmatic manner (Table 1).

Table 1. The eHealth resource checklist.

Section	Item	Checklist item	Present? (✓)
Objective			
Purpose	1	Provides an explicit statement of the objectives being addressed concerning the population and behavior/condition of interest.	
Methods			
Eligibility criteria	2	Specific characteristic of the sources of evidence used as eligibility criteria.	
Information sources	3	Describes information sources (App Store, online searches, expert consultation). Provides the date the most recent search was conducted.	
Search	4	Describes the search strategy with enough information so that it is reproducible.	
Selection of evidence sources	5	States the process for selecting resources.	
Data mapping and/or charting	6	Describes the methods of charting data.	
Evaluation and quality assurance	7	Describes the evaluation tools to be used to assess the quality of mHealth (eg, MARS) and eHealth (eg, DISCERN) resources. Note: The needs and preferences of the patient population, as well as the clinical conditions, should be considered when selecting an appropriate evaluation tool.	
Results			
Selection of sources of evidence	8	Provides the number of resources identified, selected, screened, and assessed for inclusion/exclusion.	
Results of the individual e-resources	9	Presents the relevant data that was charted to help address the study's objectives, including evidence of effectiveness.	
Evaluation and quality assurance	10	Uses the MARS (mHealth) or DISCERN (eHealth) tool to assess the quality of the resource.	
Discussion			
Summary of Evidence	11	Summarizes the main findings.	
Limitations	12	Discusses the limitations of the mHealth/eHealth resource review process.	

Discussion

Principal Findings

This paper describes the development of a set of guidelines for pragmatically selecting online resources and apps designed to support a variety of health behaviors as part of the HeLTI Canada trial. Using smartphones for health interventions has the potential to reach many populations, harness the internet's access to information, and use the latest behavioral science to incorporate nudges and reminders to make positive health decisions the default choice [15,24,25]. However, these novel opportunities for eHealth resources, coupled with their exponential proliferation, are not without their challenges [16]. As expertise in the field of eHealth is not always available to researchers, clinicians, and patients, an evidence-informed checklist to assist with navigating the identification, selection, and assessment of such online web resources and apps is needed. We believe the proposed checklist helps address the gaps outlined in the recent scoping review by Agarwal and colleagues [16], providing a pragmatic approach to evaluating apps by striking a balance between the utilization of standardized quality criteria and the need to conduct expeditious and cost-effective reviews.

Limitations

During the process of selecting the eHealth resources for the HeLTI Canada trial, it was clear a more rigorous method for searching and selecting mHealth apps was needed. Not surprisingly, practical challenges and limitations were encountered. First, the sheer volume of apps and resources available related to health behaviors (eg, the Apple Store has just over 300,000 apps available [1] and the Google Play Store has approximately 325,000 apps [14]), coupled with the constantly changing content and quick turnover of apps, was a major challenge. Second, because a full download was required to assess the app, evaluators required the necessary hardware on their mobile devices (ie, space and memory) to store the apps. Third, because some of the apps cost money or required in-app purchases, it was at times difficult to fully assess the quality of the app's contents and features based on the selected quality assessment tools (unless evaluators already had the devices downloaded on their personal devices). Fourth, it was important to ensure that the apps did not endorse private companies and that the recommended apps would not create issues for users' privacy. Lastly, despite the MARS and DISCERN tools being two of the most widely used eHealth assessment tools, certain dimensions of quality are not captured, such as privacy and security, which may be important to users. Given these limitations, the guidelines and checklist developed to search, screen, and assess eHealth resources should be repeated to confirm their rigor and reliability. With this validation work, we anticipate our checklist and outlined principles will address an important need highlighted by experts to effectively classify and evaluate apps suitable for the most

common health conditions through a reliable and valid measurement tool [26].

Future Considerations and Next Steps

The use of health apps is led by consumers and the self-tracking movement (ie, "the quantified self"). However, the quality of these apps is variable and the evidence to support the effectiveness of these interventions on public and population health is limited or unknown. Other quality assessments focus on understanding the features of apps that may be the catalyst for behavior change [27,28]. Additionally, each health specialty or specific health behavior has developed its own methods to critically appraise eHealth resources [29,30]. Our process provides a more general method to mitigate some of the limitations previously identified in the literature, particularly the large volume of potentially useful apps.

It is contended that many digitized health promotion strategies focus on individual responsibility for health and fail to recognize the social, cultural, and political dimensions of digital technology use. What is particularly noticeable about how digitized health promotion is employed in most programs is that most strategies render health even more individualized and draw attention away from the social determinants of health to a greater degree than ever before. This is despite the current emphasis on health promotion policy that seeks to take a broader approach to alleviate socioeconomic disadvantages and inequities rather than focusing on individuals' specific health-related behaviors. In the specific context of the HeLTI Canada trial, a public health nurse develops an individualized goal setting plan with each participant; each participant's context regarding social determinants of health is assessed and the participant is provided with eHealth resources that would work in tandem with their situation, thus helping to alleviate or overcome equity limitations.

Conclusions

Much like academics have come together to define checklists of critical elements for reporting in clinical trials and systematic reviews, researchers and clinicians planning to use eHealth resources in health behavior interventions require a standardized approach to identify, select, and evaluate these resources. General critical appraisal methods could help researchers from multiple disciplines select and use eHealth tools in their research.

Acknowledgments

The authors would like to thank the members of the Child E-health Resource Working Group (Stephanie Carson, Dr Shelley Vanderhout, Dr Evelyn Constantin, Dr Patricia Li, and Dr Robyn Stremler) for their assistance in reviewing the apps and e-resources for this portion of the larger Healthy Life Trajectories Initiative project. Special thanks are also extended to the Child Health Behaviour Advisory Committee: Drs Mark Tremblay, Sheila Costford, Allyson Shorkey, Cynthia Colapinto, Hasan Hutchison, Heather Manson, Jonathon Maguire, Patricia Li, Patricia Parkin, Peter Szatmari, Robyn Stremler, Rulan Parekh, and Evelyn Constantin.

LMV was supported by a Canadian Institutes of Health Research Fellowship Award.

Authors' Contributions

LMV and SC conceptualized this paper. LMV, SC, PA, and CSB contributed to the development of the screen and assessment tools. All authors have reviewed and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The eHealth resources screening tool.
[DOCX File , 14 KB - mhealth_v9i12e27533_app1.docx]

References

1. Statista. Number of mHealth apps available in the Apple App Store from 1st quarter to 3rd quarter 2021. 2015. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2020-08-20]
2. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
3. Turner T, Spruijt-Metz D, Wen CKF, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409. [doi: 10.1111/ijpo.12002] [Medline: 25641770]
4. 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]
5. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Computers in Human Behavior 2009 May;25(3):749-760. [doi: 10.1016/j.chb.2009.02.003]

6. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
7. Poorman E, Gazmararian J, Parker RM, Yang B, Elon L. Use of text messaging for maternal and infant health: a systematic review of the literature. Matern Child Health J 2015 May;19(5):969-989. [doi: 10.1007/s10995-014-1595-8] [Medline: 25081242]
8. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
9. NCD Countdown 2030 collaborators. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. Lancet 2018 Sep 22;392(10152):1072-1088. [doi: 10.1016/S0140-6736(18)31992-5] [Medline: 30264707]
10. The Healthy Life Trajectories Initiative (HeLTI). 2019. URL: https://helti-net.org/about-helti/ [accessed 2020-09-23]
11. Dennis C, Marini F, Dick JA, Atkinson S, Barrett J, Bell R, et al. Protocol for a randomised trial evaluating a preconception-early childhood telephone-based intervention with tailored e-health resources for women and their partners to optimise growth and development among children in Canada: a Healthy Life Trajectory Initiative (HeLTI Canada). BMJ Open 2021 Feb 10;11(2):e046311 [FREE Full text] [doi: 10.1136/bmjopen-2020-046311] [Medline: 33568380]
12. Heindel JJ, Vandenberg LN. Developmental origins of health and disease: a paradigm for understanding disease cause and prevention. Curr Opin Pediatr 2015 Apr;27(2):248-253 [FREE Full text] [doi: 10.1097/MOP.0000000000000191] [Medline: 25635586]
13. Mobile Fact Sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-08-21]
14. 325,000 mobile health apps available in 2017 - Android now the leading mHealth platform. Research 2 Guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2020-08-12]
15. Kratzke CK, Cox C. Smartphone technology and apps: rapidly changing health promotion. International Electronic Journal of Health Education 2012;15(1):72-82 [FREE Full text]
16. Agarwal P, Gordon D, Griffith J, Kithulegoda N, Witteman HO, Sacha Bhatia R, et al. Assessing the quality of mobile applications in chronic disease management: a scoping review. NPJ Digit Med 2021 Mar 10;4(1):46 [FREE Full text] [doi: 10.1038/s41746-021-00410-x] [Medline: 33692488]
17. Zawati MH, Lang M. Mind the App: Considerations for the Future of Mobile Health in Canada. JMIR mHealth uHealth 2019 Nov 04;7(11):e15301 [FREE Full text] [doi: 10.2196/15301] [Medline: 31682580]
18. Guiding Principles for Physicians Recommending Mobile Health Applications to Patients. Canadian Medical Association. 2015. URL: https://www.cma.ca/guiding-principles-physicians-recommending-mobile-health-applications-patients [accessed 2020-09-07]
19. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339:b2700. [doi: 10.1136/bmj.b2700]
20. 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]
21. Altman D, Simera I. Using reporting guidelines effectively to ensure good reporting of health research. In: Moher WE, editor. Guidelines for Reporting Health Research: A User's Manual. Hoboken: Wiley; 2014:32-40.
22. 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]
23. Shepperd S, Charnock D, Needham G, Gann B, Arnold I. The DISCERN Instrument. 1997. URL: http://www.discern.org.uk/discern_instrument.php [accessed 2020-09-03]
24. What is persuasive design? Interaction Design Foundation. 2015. URL: https://www.interaction-design.org/literature/topics/persuasive-design [accessed 2020-08-04]
25. Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014 Feb;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
26. 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]
27. Reyes A, Qin P, Brown CA. A standardized review of smartphone applications to promote balance for older adults. Disabil Rehabil 2018 Mar 20;40(6):690-696. [doi: 10.1080/09638288.2016.1250124] [Medline: 27868438]
28. Weekly T, Walker N, Beck J, Akers S, Weaver M. A Review of Apps for Calming, Relaxation, and Mindfulness Interventions for Pediatric Palliative Care Patients. Children (Basel) 2018 Jan 26;5(2):16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]

29. Shang J, Wei S, Jin J, Zhang P. Mental Health Apps in China: Analysis and Quality Assessment. JMIR mHealth uHealth 2019 Nov 07;7(11):e13236 [FREE Full text] [doi: 10.2196/13236] [Medline: 31697245]
30. Thornton L, Quinn C, Birrell L, Guillaumier A, Shaw B, Forbes E, et al. Free smoking cessation mobile apps available in Australia: a quality review and content analysis. Aust N Z J Public Health 2017 Dec;41(6):625-630. [doi: 10.1111/1753-6405.12688] [Medline: 28749591]

Abbreviations

HeLTI: Healthy Life Trajectories Initiative
MARS: Mobile App Rating Scale
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
zBMI: Z-score of BMI

Edited by L Buis; submitted 27.01.21; peer-reviewed by C Jacob, N Khalili-Mahani, C Reis, B Eapen; comments to author 06.05.21; revised version received 30.05.21; accepted 28.10.21; published 02.12.21.

Please cite as:
Vanderloo LM, Carsley S, Agarwal P, Marini F, Dennis CL, Birken C
Selecting and Evaluating Mobile Health Apps for the Healthy Life Trajectories Initiative: Development of the eHealth Resource Checklist
JMIR Mhealth Uhealth 2021;9(12):e27533
URL: https://mhealth.jmir.org/2021/12/e27533
doi:10.2196/27533
PMID:34860681

Original Paper

Reliability and Confirmatory Factor Analysis (CFA) of a Paper-Versus App-Administered Resilience Scale in Scottish Youths: Comparative Study

Julienne Mcgeough[1*], PhD.; Thomas Gallagher-Mitchell[1*], PhD; Dan Philip Andrew Clark[1*], PhD; Neil Harrison[1*], PhD

Department of Psychology, Liverpool Hope University, Liverpool, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Julienne Mcgeough, PhD.
Department of Psychology
Liverpool Hope University
Hope Park
Liverpool, L16 9JD
United Kingdom
Phone: 44 1512913073
Email: mcgeouj@hope.ac.uk

Abstract

Background: Adequately measuring resilience is important to support young people and children who may need to access resources through social work or educational settings. A widely accepted measure of youth resilience has been developed previously and has been shown to be suitable for vulnerable youth. While the measure is completed by the young person on paper, it has been designed to be worked through with a teacher or social worker in case further clarification is required. However, this method is time consuming and, when faced with large groups of pupils who need assessment, can be overwhelming for schools and practitioners. This study assesses app software with a built-in avatar that can guide young persons through the assessment and its interpretation.

Objective: Our primary objective is to compare the reliability and psychometric properties of a mobile software app to a paper version of the Child and Youth Resilience measure (CYRM-28). Second, this study assesses the use of the CYRM-28 in a Scottish youth population (aged 11-18 years).

Methods: Following focus groups and discussion with teachers, social workers, and young people, an avatar was developed by a software company and integrated into an android smartphone app designed to ask questions via the device's inbuilt text-to-voice engine. In total, 714 students from 2 schools in North East Scotland completed either a paper version or app version of the CYRM-28. A cross-sectional design was used, and students completed their allocated version twice, with a 2-week period in between each testing. All participants could request clarification either from a guidance teacher (paper version) or from the in-built software glossary (app version).

Results: Test and retest correlations showed that the app version performed better than the paper version of the questionnaire (paper version: r_{303}=0.81; *P*<.001; 95% CI 0.77-0.85; app version: r_{413}=0.84; *P*<.001; 95% CI 0.79-0.89). Fisher r to z transformation revealed a significant difference in the correlations (Z=–2.97, *P*<.01). Similarly, Cronbach α in both conditions was very high (app version: α=.92; paper version: α=.87), suggesting item redundancy. Ordinarily, this would lead to a possible removal of highly correlated items; however, our primary objective was to compare app delivery methods over a pen-and-paper mode and was hence beyond the scope of the study. Fisher r to z transformation revealed a significant difference in the correlations (Z=–3.69, *P*<.01). A confirmatory factor analysis supported the 3-factor solution (individual, relational, and contextual) and reported a good model fit (χ^2_{15}=27.6 [n=541], *P*=.24).

Conclusions: ALEX, an avatar with an integrated voice guide, had higher reliability when measuring resilience than a paper version with teacher assistance. The CFA reports similar structure using the avatar when compared against the original validation.

(JMIR Mhealth Uhealth 2021;9(12):e11055) doi:10.2196/11055

KEYWORDS

resilience; psychometrics; app administration; cyberpsychology

Introduction

Resilience

Resilience has traditionally been conceptualized as an individual difference. For example, early research in the field showed that some children, even when exposed to a chaotic family life or early life stressors (eg, bereavement) displayed surprisingly healthy behaviors; for example, coping ability [1-3]. Indeed, a child with high levels of resilience will be able to overcome stressors to achieve a sense of well-being [4]. Furthermore, in a review, Panter-Brick and Leckman [5] established a pathway between childhood resilience and adult well-being. However, as work on resilience has progressed, it has become increasingly recognized that factors external to the child may also influence later personal and academic success [1-3]. Luthar, Lyman, and Crossman [6] categorized subfactors of resilience into three themes, namely "Attributes of the individual," "Family influences," and "Wider social environments." Ungar [7,8] further expanded on these categories to develop a dynamic concept of resilience that places society at the center of a child's ability to develop resilience and coping strategies. Ungar's *ecological model of resilience* is culturally sensitive, and while it does accept that there are individual differences in coping, it argues that the environment surrounding the individual is crucial in providing appropriate resources. For example, while Ungar's definition and subsequent measurement includes differential aspects of the ability to maintain friendships, it also measures whether the young persons have been provided with the tools to do so. Ungar [7-9] further suggests that resilience definitions should reflect both ontological and ecological variability and states the following:

> *In the context of exposure to significant adversity, resilience is both the capacity of individuals to navigate their way to the psychological, social, cultural, and physical resources that sustain their well-being, and their capacity individually and collectively to negotiate for these resources to be provided and experienced in culturally meaningful ways.*

In Scotland (the setting for this study), pupils are currently supported via guidance teachers within the Getting It Right For Every Child (GIRFEC) framework set by the Government, and well-being is conceptualized within SHANARRI. SHANARRI has 8 indicators of well-being: Safe, Healthy, Active, Nurtured, Achieving, Respected, Responsible, and Included [10,11]. Guidance teachers lead the pastoral support for pupils of all ages, generally with approximately 200-250 pupils within their care, and with whom they will have Personal and Social learning classes each week, along with additional support if required [12]. It is within this setting that well-being, resilience, and SHANARRI are measured. While there is a positive perception among pupils and parents regarding the support offered by guidance teachers, this is not consistent with a large minority of parents who argue that the system does not support their child [13]. The challenge for schools across Scotland is the government-led initiative in which they are expected to assess the risks and vulnerability of each child [14]. Clearly, this should easier to accomplish with an app that can measure resilience and well-being easily while engaging each pupil. Furthermore, the system is under strain as funding decreases, with the education system reducing the number of guidance teachers [15,16].

Psychometric Measurement Using Apps

Ungar and Liebenberg [17,18] developed a scale of resilience, which reflected this definition of resilience and was expressed in 3 factors (individual, relational, and contextual). Sample items are "I cooperate with people around me" (individual), and "my caregivers watch me closely" (relational). The questionnaire is designed to be used as a verbally administered questionnaire, conducted by a professional within the setting, with responses measured on a Likert scale from 1 to 7. However, this is time-consuming and difficult to administer on an individual basis to large groups of pupils requiring assessment. Further studies have changed verbal administration of the questionnaire to a more traditional paper-based version to widen participation [19]. However, this obviously loses the verbal aspect of the questionnaire, which, according to Ungar [9], increases participants' understanding. Therefore, an alternative to personal administration with each child is to use software that allows questions to be read if the participant requires it.

This study seeks to address the issue of scalability while retaining the verbal aspect and reducing the need for competent reading skills. A further advantage is the benefit of software-based data collection, which, according to current research, reduces the chances of incorrect or missing input and therefore increases validity and reliability [20]. Furthermore, there is evidence that internal consistency and concurrent validity are retained when transitioning to an app-based questionnaire. Importantly, app-based scales have consistently been shown to have higher completion rates among studies included in a large-scale meta-analysis [21]. However, it cannot be assumed that transitioning from a paper version to an app version will automatically carry over psychometric properties, though there is growing evidence that the transfer to computer-based measures does not result in a loss of psychometric properties [22]. However, this is transference of psychometric properties is by no means universal; for example, when transferring pen-and-paper psychometric questionnaires, Booth-Kewley et al [23] found that a level of disinhibition crept in to measures regarding such topics as alcohol consumption and risky sexual behaviors. Therefore, it is still necessary to validate the development of a software-based app. It is of crucial importance that this is undertaken when the design of the app differs from the original scale administration format, as in this study where an avatar is used to deliver the items. Traditionally, data collection on the internet was designed to closely resemble that with paper questionnaires; however, recent studies have explored nonhuman interaction (Bot) with humans and their tendency to disclose, with provide evidence that self-disclosure increases with the use of nonhuman interviewees [24].

This Study

Our affinity for smartphones has been explained by various theories ranging from Bowlby's attachment theory, addiction-based models, and emotional needs theories [25-27]. Indeed, it has been suggested that even larger portable technology, such as laptops, can be seen to be an extension of our identity and selves, given that we store memories through photographs and access social media on them [28]. For this study, these identity processes and dynamics are identified as being drivers in the adolescent relationship with their technological companions, which may be seen as an extension of "self" [29]. Furthermore, adolescents have been described as a population that is difficult to reach for research purposes; therefore, a smartphone app such as the one tested in this study should increase usability [30]. It has been proposed that the interaction of the aforementioned dynamics will encourage honesty in this population and therefore increase the reliability of the questionnaire, as reported in other studies exploring issues of well-being in hard-to-reach populations [31]. "Avatar as a researcher" is an emerging concept, and previous studies have shown increased trust and openness, thus increasing the reliability and confidence in data when discussing sensitive topics [32]. Identification with avatars and robots occurs with both humanoid and nonhumanoid avatars. For example, even computer-driven triangle shapes are perceived to have intentionality [33,34]. Therefore, it is expected that this study will see improved reliability, increased completion rates, and similar psychometric properties retained following validity analysis, in the app-based delivery. Additionally, this study aims to validate the use of the CYRM-28 among a Scottish population.

Methods

App Development

Feedback on a number of avatar designs was gathered from 30 professionals, including social workers, educational psychologists, and teachers, at the 2015 *Pathways to Resilience* Conference. The outcome of the discussions was to avoid humanoid-like avatars of similar ages to the participants, and to opt for one that would be considered gender neutral. ALEX has facial elements that move (eyes and mouth), and uses the speech-to-text engines of the device that is running the app. ALEX moves and bounces in response to screen touches. Further focus groups with young people confirmed that ALEX was user-friendly, approachable, and liked by a wide range of ages of both sexes. Participants in the app group were asked to complete a usability questionnaire following the resilience questionnaire.

Design

Recruitment was carried out in schools that agreed to take part in trials. Information sheets were sent to parents electronically and parents could access a website about the research and agree to participate via web-based surveys. A cross-sectional design was used, which aimed at comparing the performances of pen-and-paper to that of an app-based CYRM-28 scale [17]. Two schools included all of their pupils, and classes were randomly designated as either app versus paper with age groups represented in each group. All groups were presented with the scale twice, with a 2-week retest design. Data collection was completed in Personal and Social Education (PSE) classes, and took approximately 10 minutes for the majority of the students. This was preceded by a short explanation regarding the administration of the scale and a reminder of their ethical rights. A guidance teacher and a member of the data collection team were present during the session. As with the original CYRM-28, participants could request further information and clarification from the researcher regarding the item statements (paper version) or an in-built glossary that could be accessed when the pupil highlighted a word or phrase. All research took place during the second term of the academic year (January to March 2017). A third school took part in 1 app-based data collection during the Summer term (July 2017) under the same conditions as described above, but further participation was prevented owing to end-term examination. These data are included only in the CFA.

Participants

The participants were 714 students from 2 North-East Scotland coeducational schools, aged 11-17 years (males: n=354, mean age 14.3 years, SD 2.42 years; females: n=360, mean age 14.6 years, SD 2.37 years). Areas in Scotland are divided into 5 broad groupings of deprivation (1=most deprived to 5=least deprived) and are reported with the Scottish Index of Multiple Deprivation [35]. School 1 (n=403) includes a high-income area, and the majority of pupils fall into bands 4 and 5 (relatively high socioeconomic status [SES] in accordance with the Government's deprivation bands). School 2 (n=311) is in an urban setting classified as a high deprivation area (all pupils are classed as being in the top 2 levels of deprivation). The final school draws from a wide range of SES bands. All 3 schools are comprehensives and therefore mixed-ability schools with intakes of pupils aged 16-18 years. The schools used mixed-ability groups, and each of the schools have approximately similar numbers on the roll.

Materials

The app version ran on Kindle Fires (HD), which were disconnected from the internet, and other software could not be accessed. The app presents the questions via the ALEX avatar. ALEX is gender-neutral and is displayed in diagram 1 below, along with a typical question. As with the paper version, the students were required to respond on a 1-7–point Likert scale (strongly disagree to strongly agree), yielding a possible data range of 28-196, with a higher score indicating stronger resilience. The app version has a computerized voice, which is able to read the question to the participant, and a glossary of available terms. These had been tested by adolescents who had trialed the software and had indicated where they thought help would be required. In the pen-and-paper version, help was given if requested by the participant at the time, and adults provided the same answers as given by the predetermined glossary. There were no reports of pupils asking questions outside of this set. The scale has previously been found to have good reliability scores (individual: α=.803; relational: α=.833; contextual: α=.794), and adequate validity after exploratory and confirmatory analyses [17]. The project received ethical

approval from the Liverpool Hope University Ethics board (S040417 SFREC 001), and students were required to read a short participation information sheet or screen after a short verbal reminder of their right to withdraw from the research. Parents had provided informed consent to their children's participation. Demographic information and data regarding the usability of the app were collected.

Statistical Analysis

For demographic descriptive statistics, only results from time 1 were included. All data met parametric assumptions. Items in the app condition were grouped and calculated to form 3 factors in accordance with an a priori theory developed by Liebenberg and Ungar [18]. The first factor (individual) was composed of 11 items which were further conceptualized as personal skills, peer support, and social skills. The second factor of relationships with caregivers included 7 items divided into physical and psychological care. The final factor was labeled as contextual and had 3 subfactors (educational, spiritual, and cultural).

Data from 12 respondents were removed prior to a CFA, following identification as multivariate outliers using the Mahalanobis Distance (MD) method. AMOS 24 was used to complete the CFA using a Maximum Likelihood Model. Files have been archived on the Open Science Forum [36].

Results

Usability Results

In total, 262 of the pupils took part in the usability questionnaire. The majority of the participants rated the app as easy to very easy to use (87.4%), compared to those who rated it hard or very hard (4.4%). Additionally, users were positive about their experience regarding interaction with ALEX. However, participants were moderately negative with the voice that read the instructions, with 31% stating that it needed to be changed. They were also encouraged to leave comments regarding improvements; in this field, the most common suggestion was to include a game.

Assessment Results

Descriptive statistics for resilience are reported in Table 1. These data show that males and females reported similar scores and suggest minor differences in resilience across schools. Resilience scores decreased with age, with the youngest pupils aged 11 years reporting higher levels (mean 113.05, SD 11.85) than those aged >16 years (mean 103.50, SD 15.10). Pearson correlation analysis indicated a significant relationship between age and resilience (r=0.81; P=.006; 95% CI 0.02-2.73).

Table 1. Summary of the scores for each sample.

	Sample size, n		Score, mean (SD)		95% CI	
	S1[a]	S2[b]	S1	S2	S1	S2
Paper version						
Total	82	126	107.85 (13.66)	104.22 (11.64)	104.08-110.09	102.17-106.27
Males	36	50	108.06 (12.93)	103.06 (10.52)	103.68-112.24	100.04-106.08
Females	45	76	106.53 (14.41)	105.01 (12.38)	102.20-110.86	102.19-107.84
App version						
Total	234	183	107.45 (13.71)	105.95 (13.33)	105.69-109.21	104.01-107.90
Males	135	97	107.65 (13.69)	106.38 (14.00)	105.30-110.02	10354-109.22
Females	99	84	107.57 (13.65)	105.90 (12.57)	104.84-110.29	103.17-108.63

[a]S1: school 1 (deprivation groups 4 and 5).

[b]S2: school 2 (deprivation groups 1 and 2).

There was no difference between the schools in terms of resilience (school 1: mean 107.24, SD 12.87; school 2: mean 105.79, SD 13.15; t_{720}=1.38; P=.18). In the paper version, scores on the CYRM-28 ranged from 63 to 131 (mean 106.98, SD 13.51); however, in the app version, the equivalent results were 56-135 (mean 106.79, SD 13.62). An independent samples t test was conducted between the 2 conditions and reported no significant difference (t_{720}=–0.632; P=.53; 95% CI –2.55 to 1.31).

Psychometric Properties

Cronbach α in both conditions was very high (app: α=.92; paper: α=.87). Fisher r to z transformation revealed a significant difference in the correlations (Z=–3.69, P<.01). Test-retest results (Pearson correlation coefficients) were significant in both conditions, although the app version had higher reliability (paper version: r_{303}=0.81; P<.001; 95% CI 0.77-0.85; app version: r_{413}=0.84; P<.001; 95% CI 0.79-0.89). As SPSS was used to calculate the 95% CIs with a linear regression model, z scores were used to calculate 95% CIs. Fisher r to z transformation revealed a significant difference in the correlations (Z=–2.97, P<.01). Additionally, intraclass correlation (2,1) estimates and their 95% CIs were calculated using SPSS (SPSS Inc), the absolute-agreement, single rater model indicates that the reliability of the app version of the questionnaire was similar to the paper version (Table 2).

Table 2. Intraclass correlations in SPSS using an absolute-agreement, single rater model.

	Intraclass correlation	95% CI	*F* test with a true value of 0			
			Value	df_1	df_2	Sig
App	0.842	0.812-0.868	11.689	416	416	0.000
Paper	0.810	0.783-0.834	9.526	721	721	0.000

The 3-factor structure of the 28-item CYRM-28, based on the model confirmed by Liebenberg and Unger [18], was estimated using a CFA with the Time 1 data set in AMOS 24. A maximum likelihood estimation CFA model was found to be parsimonious; however, the significant results on chi-square analysis indicate that the model did not adequately fit the data (χ^2_{15}=27.6 [n=541], *P*=.24). As large sample sizes can increase the likelihood of significant chi-square results, other indices of model fit are of particular interest. Table 3 includes a range of fit indices, all of which are within acceptable parameters.

Table 3. Model fit summary for the app version of CYRM-28[a] confirmatory factor analysis.

	Chi-square (*df*)	*P* value	Goodness of fit	Comparative fit index	Root mean square error of approximation
Original model	43.8 (17)	>.01	0.94	0.98	0.54
Second model	27.59 (15)	.24	0.98	0.99	0.39

[a]CYRM-28: child and youth resilience measure.

Modification indices were examined, and several items were found to have significant shared error variance, including the following: relational (physical) and contextual (spiritual); individual (personal) and individual (peer). An exploration of the items included in each of these factors for multicollinearity between the items suggested that no item was so redundant with another item that it could be dropped (e1-e2, tolerance=1.00, variance inflation factor=1.00; e4-e8, tolerance=–1.00, variance inflation factor=1.00). As the shared error variance between all of these pairs of items was conceptually consistent with the domain assessed, a final model was respecified to free these correlated errors. This model was found to fit the data moderately well, and increased goodness of fit (χ^2_{15}=27.6 [n=541], *P*=.24); further details of fit can be seen in Table 2. The final confirmatory factor analytic model of the CYRM-28 indicated that the items were strongly correlated within factors rather than across factors, this replicates the findings from the original validity study [18]. Diagram 2 shows the error-covariances added to improve the model goodness of fit; each of these were low (*r*=0.12 and *r*=–0.15).

Discussion

Principal Findings

The aim of the study was to establish the adequacy of an app version of a previously validated paper version of a scale to measure resilience. The app and the paper versions of the scale presented the text of the items using Likert scales. The paper version allowed pupils to ask staff for support while in the app version, this was built into the device. The results indicate that the app had significantly better reliability in a test-retest analysis and had significantly higher internal consistency, as measured with the Cronbach α score. Scores across the demographic groups between the paper and app versions did not differ, indicating that the app version matches the paper version on the CYRM-28 when measuring resilience. Finally, the study supports the use of the CYRM-28 in a Scottish youth population [9-16].

Comparison With Prior Work

Ungar [7] previously reported that resilience was not only a function of the individual, but also that environmental influences are important. The CFA reflected this understanding of resilience and further confirmed by Liebenberg and Ungar [17] earlier reported a 3-factor solution (individual, family relationships, and contextual). Furthermore, the CYRM-28 was designed to be used with the support of an adult professional (teacher or social worker) [18], and while this ensures that young people have understood the statements, it is not cost-effective and therefore is of use only to small groups of children who have been identified as vulnerable. Additionally, the pastoral system within Scottish schools is increasingly under strain. This study provides evidence that a sizable percentage of children would not seek support from their guidance teachers. The purpose of this study was to develop a low-cost scalable version of the questionnaire, which depends on an avatar to support understanding and encourages openness in adolescents. As discussed by Palmier-Claus [37], the app's increased reliability, as evident from its high internal consistency, and in addition, participants were more likely to provide similar responses across time periods when using the app version. Previous studies indicated that the use of the avatar in the app would be a positive experience, and this has been replicated in this study. The students who completed the supplementary usability questions were generally positive about the avatar. It can be assumed that while app usage was time-limited, the participants were able to develop a relationship of trust with ALEX and were therefore open in their responses.

Limitations

This study sought to explore how effective an avatar was in connecting with young people and collecting data about their home-lives and feelings. Our findings show that the app performed well at this level of data collection and a proof of concept has been met. However, for ethical reasons, it was decided to test this in a general population of young people, rather than adolescents who have been identified as vulnerable.

Furthermore, while it can be argued that resilience is more observable among people who are facing trauma or difficult situations, the CYRM-28 has previously been used in general populations [17,19]. Nonetheless, further research that includes vulnerable participants would be warranted.

The final version of the app was designed to allow the participant as well as the professional to access information about the pupil. While it is important to develop highly reliable but easy-to-administer assessments, it is important that the results are of use to the teacher or social worker in helping support pupils. In this study, the reports were only available to guidance teachers and were for research purposes only. It is possible that knowledge of this had an impact on the participants' answers. However, both groups (app and paper) were exposed to this variable. Furthermore, among the usability questions, pupils were asked about whether they had thought this knowledge had affected their answer, with the majority stating that it had not. Additionally, the app will be used in a setting in which reports will be available to experts such as teachers, educational psychologists, and social workers. It was important that this was incorporated in the trial. Parents had consented to reports being used in future studies about the usability of reports, and both groups of pupils were informed of this prior to the study as part of the assent process.

Current studies are exploring how professionals utilize feedback from an app, but another question not answered here is how the young people themselves react to instant feedback on an aspect of their psychological life. Additionally, a discussion on the use of the app within a broader health and social education setting should be developed. The authors strongly suggest that the app would be well-suited in ongoing curricula designed around assessing and developing aspects of well-being. Education practitioners and social workers should be involved in developing good practice in relation to the use of such apps. It is recommended that this forms part of a conversation between guidance teachers and young people, rather than the end result of an assessment. To that end, future research should consider how assessment apps can enable participants to communicate with their guidance teachers; this feature is of particular interest, given the findings of our study on the reluctance of pupils to approach their teachers.

Conclusions

The app technology utilized in this study has shown strong reliability and validity in measuring resilience in young adult populations. Our findings demonstrate the efficacy of moving the CYRM-28 "gold-standard" measure of resilience to a web-based app-based platform. The benefits of avatar-led questioning in relation to young people's understanding of resilience are evident; however, future studies should address how technology can be effectively integrated into existing practitioner-led support services within schools.

Acknowledgments

We would like to acknowledge APPA-Wellbeing.com for having conceptualized and designed the smartphone app.

Conflicts of Interest

None declared.

References

1. Garmezy N, Streitman S. Children at risk: the search for the antecedents of schizophrenia. Part I. Conceptual models and research methods. Schizophr Bull 1974(8):14-90. [doi: 10.1093/schbul/1.8.14] [Medline: 4619494]
2. Ong AD, Bergeman CS, Bisconti TL, Wallace KA. Psychological resilience, positive emotions, and successful adaptation to stress in later life. J Pers Soc Psychol 2006 Oct;91(4):730-749. [doi: 10.1037/0022-3514.91.4.730] [Medline: 17014296]
3. Masten AS. Global perspectives on resilience in children and youth. Child Dev 2014;85(1):6-20. [doi: 10.1111/cdev.12205] [Medline: 24341286]
4. Panter-Brick C, Leckman JF. Editorial Commentary: Resilience in child development--interconnected pathways to wellbeing. J Child Psychol Psychiatry 2013 Apr;54(4):333-336. [doi: 10.1111/jcpp.12057] [Medline: 23517424]
5. Luthar SS, Lyman EL, Crossman EJ. Resilience and Positive Psychology. In: Handbook of Developmental Psychopathology. Boston, MA: Springer; 2014:125-140.
6. Ungar M. A Constructionist Discourse on Resilience. Youth & Society 2016 Aug 19;35(3):341-365 [FREE Full text] [doi: 10.1177/0044118X03257030] [Medline: 22973420]
7. Ungar M. Resilience across Cultures. Br J Soc Work 2006 Nov 08;38(2):218-235. [doi: 10.1093/bjsw/bcl343]
8. Ungar M. Community resilience for youth and families: Facilitative physical and social capital in contexts of adversity. Child Youth Serv Rev 2011 Sep;33(9):1742-1748. [doi: 10.1016/j.childyouth.2011.04.027]
9. The Scottish Government, Children and Young People Act (2014). URL: https://www.gov.scot/publications/children-young-people-scotland-act-2014-national-guidance-part-12/pages/3/ [accessed 2021-11-26]
10. Daniel B, Scott J, Burgess C, Mulley K. Noticing and Helping Neglected Children in Scotland: 'sometimes it's just too hard to talk'. Res Pol Plan 2016;32(1):11-23.

11. Supporting pupils, A Study of guidance and pupil support in Scottish schools. The Scottish Government (2005). URL: https://www.webarchive.org.uk/wayback/archive/20150220045536/http://www.gov.scot/Publications/2005/02/20693/52522 [accessed 2019-11-21]
12. Black C, Eunson J, Murray L, Zubairi S, Bowen L. Behaviour in Scottish Schools Research. 2016. URL: https://www.gov.scot/publications/behaviour-scottish-schools-research-2016/ [accessed 2019-11-22]
13. Stoddart E. The 'named person': surveillance and the wellbeing of children and young people in Scotland. SS 2015 Feb 09;13(1):102-116. [doi: 10.24908/ss.v13i1.5231]
14. Provision of Teachers and Pupil Support Assistants to support children and young people with additional supports needs within secondary schools. Scottish Secondary Teachers' Association (SSTA). 2017. URL: https://archive2021.parliament.scot/S5_Education/General%20Documents/20170317SSTAAsnSubmission.pdf [accessed 2019-11-15]
15. Number of additional support teachers in Scotland sinks to lowest level since 2007. 2017. URL: https://inews.co.uk/news/education/number-additional-support-teachers-scotland-sinks-lowest-level-since-2007-39762 [accessed 2019-11-21]
16. Abualkibash S, Lera M. Psychological resilience among Palestinian school students: an exploratory study in the West Bank. Int Humanit Stud 2015;2(3):1-20 [FREE Full text]
17. Ungar M, Liebenberg L. Assessing Resilience Across Cultures Using Mixed Methods: Construction of the Child and Youth Resilience Measure. J Mix Methods Res 2011 Mar 07;5(2):126-149. [doi: 10.1177/1558689811400607]
18. Liebenberg L, Ungar M, Vijver FVD. Validation of the Child and Youth Resilience Measure-28 (CYRM-28) Among Canadian Youth. Res Soc Work Pract 2011 Nov 24;22(2):219-226. [doi: 10.1177/1049731511428619]
19. Erbe D, Eichert HC, Rietz C, Ebert D. Interformat reliability of the patient health questionnaire: Validation of the computerized version of the PHQ-9. Internet Interv 2016 Sep;5:1-4 [FREE Full text] [doi: 10.1016/j.invent.2016.06.006] [Medline: 30135800]
20. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res 2014 Jun 30;217(1-2):124-127. [doi: 10.1016/j.psychres.2014.03.009] [Medline: 24679993]
21. Bush NE, Skopp N, Smolenski D, Crumpton R, Fairall J. Behavioral screening measures delivered with a smartphone app: psychometric properties and user preference. J Nerv Ment Dis 2013 Nov;201(11):991-995. [doi: 10.1097/NMD.0000000000000039] [Medline: 24177488]
22. Marcano-Belisario JS, Gupta AK, O'Donoghue J, Morrison C, Car J. Tablet computers for implementing NICE antenatal mental health guidelines: protocol of a feasibility study. BMJ Open 2016 Jan 22;6(1):e009930 [FREE Full text] [doi: 10.1136/bmjopen-2015-009930] [Medline: 26801468]
23. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. J Med Internet Res 2014 Dec 03;16(12):e268 [FREE Full text] [doi: 10.2196/jmir.3395] [Medline: 25472463]
24. Booth-Kewley S, Larson GE, Miyoshi DK. Social desirability effects on computerized and paper-and-pencil questionnaires. Comput Hum Behav 2007 Jan;23(1):463-477. [doi: 10.1016/j.chb.2004.10.020]
25. Hasler B, Tuchman P, Friedman D. Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. Comput Hum Behav 2013 Jul;29(4):1608-1616. [doi: 10.1016/j.chb.2013.01.004]
26. Thorsteinsson G, Page T. User attachment to smartphones and design guidelines. IJMLO 2014;8(3/4):201. [doi: 10.1504/IJMLO.2014.067020]
27. Balakrishnan V, Raj RG. Exploring the relationship between urbanized Malaysian youth and their mobile phones: A quantitative approach. Telemat Inform 2012 Aug;29(3):263-272. [doi: 10.1016/j.tele.2011.11.001]
28. Hassenzahl M, Diefenbach S, Göritz A. Needs, affect, and interactive products – Facets of user experience. Interacting with Computers 2010 Sep;22(5):353-362. [doi: 10.1016/j.intcom.2010.04.002]
29. Waugh M. 'My laptop is an extension of my memory and self': Post-Internet identity, virtual intimacy and digital queering in online popular music. Pop Mus 2017 Jun 19;36(2):233-251. [doi: 10.1017/S0261143017000083]
30. Park BK, Calamaro C. A systematic review of social networking sites: innovative platforms for health research targeting adolescents and young adults. J Nurs Scholarsh 2013 Sep;45(3):256-264. [doi: 10.1111/jnu.12032] [Medline: 23676115]
31. Jones L, Saksvig BI, Grieser M, Young DR. Recruiting adolescent girls into a follow-up study: benefits of using a social networking website. Contemp Clin Trials 2012 Mar;33(2):268-272 [FREE Full text] [doi: 10.1016/j.cct.2011.10.011] [Medline: 22101207]
32. Schnall E, Schnall D. Positive Psychology in Jewish Education: Gratitude in the School and Synagogue Classroom. Religious Education 2016 Dec 07;112(2):160-171. [doi: 10.1080/00344087.2016.1224002]
33. Pickard MD, Roster CA, Chen Y. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? Computers in Human Behavior 2016 Dec;65:23-30. [doi: 10.1016/j.chb.2016.08.004]
34. Müller BCN, Oostendorp AK, Kühn S, Brass M, Dijksterhuis A, van Baaren RB. When triangles become human. IS 2015 Aug 17;16(1):54-67. [doi: 10.1075/is.16.1.03mul]
35. Scottish Index of Multiple Deprivation 2020. Scottish Government. URL: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/ [accessed 2021-11-16]

36. McGeough J, Gallagher-Mitchell T, Clark D, Harrison N. APPA phase 1. OSF. 2019. URL: https://osf.io/kr6g9/ [accessed 2019-11-16]
37. Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. BMC Psychiatry 2012 Oct 17;12:172 [FREE Full text] [doi: 10.1186/1471-244X-12-172] [Medline: 23075387]

Abbreviations

CYRM-28: child and youth resilience measure
GIRFEC: getting it right for every child
SHANARRI: Safe, Healthy, Active, Nurtured, Achieving, Respected, Responsible, Independent.

Edited by G Eysenbach; submitted 15.05.18; peer-reviewed by W Symes, R Froud; comments to author 02.09.18; revised version received 14.06.19; accepted 15.11.19; published 07.12.21.

Please cite as:
Mcgeough J, Gallagher-Mitchell T, Clark DPA, Harrison N
Reliability and Confirmatory Factor Analysis (CFA) of a Paper- Versus App-Administered Resilience Scale in Scottish Youths: Comparative Study
JMIR Mhealth Uhealth 2021;9(12):e11055
URL: https://mhealth.jmir.org/2021/12/e11055
doi:10.2196/11055
PMID:34878995

Original Paper

Willingness to Share Data From Wearable Health and Activity Trackers: Analysis of the 2019 Health Information National Trends Survey Data

Camella J Rising[1], MS, PhD, RDN; Anna Gaysynsky[1], MPH; Kelly D Blake[1], ScD; Roxanne E Jensen[2], PhD; April Oh[3], MPH, PhD

[1]Behavioral Research Program, Division of Cancer Control and Population Sciences, US National Cancer Institute, Rockville, MD, United States

[2]Health Care Delivery Research Program, Division of Cancer Control and Population Sciences, US National Cancer Institute, Rockville, MD, United States

[3]Implementation Science, Office of the Director, Division of Cancer Control and Population Sciences, US National Cancer Institute, Rockville, MD, United States

Corresponding Author:
Camella J Rising, MS, PhD, RDN
Behavioral Research Program
Division of Cancer Control and Population Sciences
US National Cancer Institute
9609 Medical Center Drive
Rockville, MD, 20850
United States
Phone: 1 240 276 5262
Email: camella.rising@nih.gov

Abstract

Background: Sharing data from wearable health and activity trackers (wearables) with others may improve the health and behavioral outcomes of wearable users by generating social support and improving their ability to manage their health. Investigating individual factors that influence US adults' willingness to share wearable data with different types of individuals may provide insights about the population subgroups that are most or least likely to benefit from wearable interventions. Specifically, it is necessary to identify digital health behaviors potentially associated with willingness to share wearable data given that the use of and engagement with various technologies may broadly influence web-based health information–sharing behaviors.

Objective: This study aims to identify sociodemographic, health, and digital health behavior correlates of US adults' willingness to share wearable data with health care providers and family or friends.

Methods: Data for the analytic sample (N=1300) were obtained from the 2019 Health Information National Trends Survey of the National Cancer Institute. Digital health behavior measures included frequency of wearable device use, use of smartphones or tablets to help communicate with providers, use of social networking sites to share health information, and participation in a web-based health community. Multivariable logistic regression analysis of weighted data examined the associations between digital health behaviors and willingness to share wearable device data, controlling for sociodemographics and health-related characteristics.

Results: Most US adults reported willingness to share wearable data with providers (81.86%) and with family or friends (69.51%). Those who reported higher health self-efficacy (odds ratio [OR] 1.97, 95% CI 1.11-3.51), higher level of trust in providers as a source of health information (OR 1.98, 95% CI 1.12-3.49), and higher level of physical activity (OR 2.00, 95% CI 1.21-3.31) had greater odds of willingness to share data with providers. In addition, those with a higher frequency of wearable use (OR 2.15, 95% CI 1.35-3.43) and those who reported use of smartphones or tablets to help communicate with providers (OR 1.99, 95% CI 1.09-3.63) had greater odds of willingness to share data with providers. Only higher level of physical activity was associated with greater odds of willingness to share wearable data with family or friends (OR 1.70, 95% CI 1.02-2.84). Sociodemographic factors were not significantly associated with willingness to share wearable data.

Conclusions: The findings of this study suggest that, among US adult wearable users, behavior-related factors, rather than sociodemographic characteristics, are key drivers of willingness to share health information obtained from wearables with others. Moreover, behavioral correlates of willingness to share wearable data are unique to the type of recipient (ie, providers vs family

or friends). Future studies could use these findings to inform the development of interventions that aim to improve the use of patient-generated data from wearable devices in health care settings.

(JMIR Mhealth Uhealth 2021;9(12):e29190) doi:10.2196/29190

KEYWORDS

mobile health; population health; health communication; survey methodology; mobile apps; devices; online social networking; mobile phone

Introduction

Background

In 2019, nearly one-quarter of US adults reported using wearable health and activity trackers (wearables) [1-3], which is approximately twice the reported use in 2015 [4]. Wearables are mobile health (mHealth) technologies worn on the body that can detect, record, and report information about behaviors (eg, step count and dietary intake) and health indicators (eg, heart rate and calories burned) [5]. Wearables may improve health and behavioral outcomes, such as physical activity participation [1,6] and weight status [7], by prompting users to set health goals, providing automated personalized feedback about health and activity data, motivating healthy habit formation, and encouraging social data sharing and competition [6-9].

Connected devices, such as mHealth apps on smartphones or computer tablets, enable users to share health information from wearables with others, such as health care providers and family or friends. Sharing wearable data may improve the health and behavioral outcomes of users by generating social support and improving their ability to manage their health [9-12], increasing patient–provider engagement, and facilitating individualized counseling and clinical decision making [12-15]. Studies also suggest that wearable data sharing among behavioral intervention participants may increase intervention effectiveness [16].

When used as a health communication tool, the potential of wearable technologies to improve health may not be fully realized without an understanding of the willingness to share wearable data, particularly with providers. Reasons for not sharing wearable data with providers include lack of awareness of the social sharing features of wearables, uncertainty about the relevance or usefulness of the data to providers, low expectation of supportive feedback, and concerns about privacy or control over data shared from the device [12-14]. Past studies of patient data sharing from various mHealth technologies also suggest that willingness to share wearable data may vary by individual characteristics, such as sociodemographics (eg, sex, age, race, or ethnicity), health-related factors (eg, weight status or having a chronic condition), and trust [17-19]. However, the correlates of willingness to share data from wearable health and activity trackers, specifically, need further examination, as the characteristics of mHealth users vary by type of device used for health and behavioral tracking [20].

Following evidence that the use of and engagement with digital health can influence health and communication behavior [21-23], digital health behaviors may be additional factors associated with the willingness to share wearable data. For example, individuals who use mHealth technologies such as smartphones or tablets to help communicate with providers may have greater technology self-efficacy [21], which may influence their willingness to share health information from wearable devices. Frequency of wearable use, one aspect of engagement with mHealth technologies for health and behavioral tracking [22], could also be a factor associated with willingness to share data from these devices. Moreover, other digital health behaviors, such as sharing health information on social networking sites (SNSs; eg, Facebook) or within web-based health communities (eg, online cancer support groups), may be associated with willingness to share wearable data if social sharing of health information across digital media is broadly perceived as useful or beneficial [9,11,23]. However, to the best of our knowledge, such relationships between digital health behaviors and willingness to share data from wearables have not been examined in a nationally representative sample of US adults.

A better understanding of the factors that influence willingness to share data from wearables could have implications for the use of patient-generated data in clinical practice [24], particularly given the growing number of interventions that use wearable devices to track health and activity [25] and recent calls for integration and use of these data in interventions [26,27]. Identifying the sociodemographic, health, and behavioral correlates of willingness to share data from wearables could provide insights on the population subgroups that are most or least likely to engage with, and benefit from, wearable interventions or multicomponent behavioral interventions that involve wearable use. In addition, exploring the correlates of willingness to share wearable data with providers, as well as with family or friends, may identify different drivers of willingness to share data with different types of recipients.

Objectives

This study has 2 primary aims to address gaps in the literature. The first aim is to describe the sociodemographic and health-related correlates of the reported willingness of wearable users to share data with health care providers and with family or friends. The second aim is to investigate the relationship between different digital health behaviors (ie, use of smartphones or tablets to help communicate with providers, frequency of wearable use, sharing health information on SNSs, and participation in a web-based health community) and the willingness of users to share wearable data with health care providers and with family or friends.

Methods

Sample Population

Data from the 2019 Health Information National Trends Survey (HINTS) of the National Cancer Institute were analyzed. HINTS is a nationally representative, probability-based cross-sectional survey. Self-administered questionnaires were completed by adult, civilian, noninstitutionalized individuals (N=5438) between January and April 2019 (Multimedia Appendix 1). Respondents completed mailed paper questionnaires (paper-only group) or completed the questionnaire on the web as part of a push-to-web pilot study. Individuals participating in the web pilot were randomly assigned to a web-option group (choice of responding by paper or web) or a web-bonus group (choice of responding by paper or web, with a US $10 bonus incentive for responding via web). There were no significant differences in response rates for the paper-only group (30.2%), the web-option group (29.6%), and the web-bonus group (31.5%). Additional information about HINTS data, resources, and methodology has been described elsewhere [28], and information specific to HINTS 5 Cycle 3 (2019) can be found in publicly available methods reports [29].

To be included in the analytic sample, respondents had to report the use of a wearable device to track their health or activity. Thus, respondents were included if they selected "yes" (vs "no") in response to the item "In the past 12 months, have you used an electronic wearable device to monitor or track your health or activity? For example, a Fitbit, Apple Watch, or Garmin Vivofit." In addition, respondents must have reported data for the outcome variables of interest, as described in the following sections.

Measures

Willingness to Share Wearable Data

Willingness to share wearable data with providers was measured with the item "Would you be willing to share health data from your wearable device with your health care provider?" ("yes" or "no"). Willingness to share wearable data with family or friends was measured by asking respondents, "Would you be willing to share health data from your wearable device with your family or friends?" ("yes" or "no").

Digital Health Behaviors

To evaluate web-based health information sharing, the reported use of SNSs to share health information and participate in a web-based health community were examined. The use of SNSs was measured with the item "In the past 12 months, have you used the Internet for any of the following reasons? To share health information on social networking sites, such as Facebook or Twitter" ("yes" or "no"). To measure participation in a web-based health community, respondents were asked: "In the past 12 months, have you used the Internet for any of the following reasons? To participate in an online forum or support group for people with a similar health or medical issue" ("yes" or "no").

The use of mHealth technologies to help communicate with providers was measured with the item "Has your tablet or smartphone helped you in discussions with your health care provider?" ("yes" or "no"). Frequency of wearable use was evaluated by asking respondents: "In the past month, how often did you use a wearable device to track your health?" Responses were dichotomized into higher frequency use ("almost every day" or "every day") and lower frequency use ("1-2 times per week," "less than once per week," or "did not use a wearable device in the past month").

Health-Related Characteristics

Health-related correlates included perceived health status, health self-efficacy, BMI, multimorbidity, and level of physical activity. Perceived health status was measured with the item "In general, would you say your health is...?" Responses were dichotomized into good health ("excellent," "very good," or "good") and "fair" or "poor" health. Health self-efficacy was measured with the item "Overall, how confident are you about your ability to take good care of your health?" Responses were dichotomized into higher health self-efficacy ("very confident" or "completely confident") and lower health self-efficacy ("somewhat confident," "a little confident," or "not confident at all"). Self-reported height and weight were used to calculate and classify BMI [30]; underweight respondents were excluded from analysis due to low frequency of BMI indicative of underweight (BMI<18.5) among wearable users (n=13). A composite multimorbidity variable (0 conditions, 1 condition, or ≥2 conditions) combined data from items that assessed history of chronic conditions ("yes" or "no"), including diabetes, heart disease, lung disease, depression or anxiety, and any cancer except nonmelanoma skin cancer. On the basis of the Physical Activity Guidelines for Americans for minutes per week of moderate-intensity physical activity [31], level of physical activity was assessed with a discrete numerical response to the item "On the days that you do any physical activity or exercise of at least moderate intensity, how long do you typically do these activities?" Responses were dichotomized as higher level of physical activity (≥150 minutes per week) versus lower level of physical activity (<150 minutes per week).

Additional health-related measures included having a regular health care provider, trust in health information from a physician, and trust in health information from family or friends. Having a regular health care provider ("yes" or "no") was measured with the item "Not including psychiatrists and other mental health professionals, is there a particular doctor, nurse, or other health professional that you see most often?" Trust in health information from a physician was evaluated with the item "In general, how much would you trust information about health or medical topics from each of the following? A doctor." Trust in health information from family or friends was evaluated with the item "In general, how much would you trust information about health or medical topics from each of the following? Family or friends." Response options for both trust items were dichotomized as higher trust ("a lot") versus lower trust ("some," "a little," or "not at all").

Sociodemographic Characteristics

Sociodemographic variables included sex (women and men), age (18-34, 35-49, 50-64, and ≥65 years), race (White, Black, and other races, which combined low-frequency responses for

American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, and other Pacific Islander), ethnicity (Hispanic and non-Hispanic), education (high school graduate or less; technical, vocational, or some college; and college graduate or postgraduate), annual household income in US dollars (<US $35,000, US $35,000-$49,999, US $50,000-$74,999, US $75,000-$99,999, and ≥US $100,000), and geographic area (urban vs rural [32]). Due to the relatively high proportion of missing data in the annual household income measure, an imputed variable provided in the data set was used to avoid losing respondents in the analytic sample.

Statistical Analysis

Frequencies, weighted percentages, and chi-square statistics were calculated to describe the distribution of US adults who reported using a wearable device to track health or activity. Binomial logistic regression analysis was conducted to examine correlations between individual characteristics (sociodemographic, health-related, and digital health behavior variables) and willingness to share wearable data with health care providers, and correlations between individual characteristics and willingness to share wearable data with family or friends. In total, 2 regression models were constructed.

For the model predicting willingness to share wearable data with providers, all sociodemographic and health-related variables (excluding trust in health information from family or friends) were entered first to address the first aim of this study. Because digital health behaviors were factors of particular interest, the second aim was addressed by adding digital health behavior variables stepwise in the following order: frequency of wearable use, use of mHealth technologies to help communicate with providers, use of SNSs to share health information, and participation in a web-based health community. The order in which variables were added to the model was based on the extent of supporting literature [21-23] that suggests a potential association between the respective digital health behavior and willingness to share wearable data, such that variables with a greater evidence base were added to the model first. Pseudo R^2 was examined after adding each digital health behavior variable to the model to determine how much variability could be explained by each of these key predictor variables. For the model predicting willingness to share wearable data with family or friends, covariates were entered using a similar stepwise approach; however, variables pertaining to interactions with providers were excluded (ie, use of mHealth technologies to help communicate with providers, having a regular health care provider, and trust in health information from a physician), whereas trust in health information from family or friends was added.

Statistical analysis was conducted using SAS (version 9.4; SAS Institute). Complete-case analysis with listwise deletion was used for the regression models. Group differences by survey modality (paper-only, web-option, and web-bonus) for the outcome variables of interest were assessed using the jackknife replication variance estimation method, applying a final sample weight and replicate weights created using the Rizzo method [33]. Because group differences were not significant in our analysis of outcomes by modality, the full-sample weight was applied to calculate population estimates for the combined sample without controlling for group differences by survey modality. Replicate weights were also used to compute SEs of estimates using the jackknife replication method for the combined sample without controlling for group differences by survey modality.

Results

Sample Population

The analytic sample comprised 1300 wearable users. Analysis of weighted data showed that women (55.03%) and men (44.97%) each constituted approximately half of the sample. The majority were under 50 years of age (64.1%), urban residents (88.07%), non-Hispanic White (64.69%), and reported having an education beyond high school (85%). Individuals with annual household income under US $75,000 comprised 44.02% of the sample, with the remaining 55.98% having an annual household income of US $75,000 or more (Table 1). The characteristics of the HINTS analytic sample can be referenced alongside the characteristics of the analytic sample of wearable users in Multimedia Appendix 2. Similar to other health-based surveys, HINTS respondents tend to be female, older, non-Hispanic White, urban-dwelling, and more educated, and have a higher annual household income than the general population [28].

Most individuals reported having a good health status (89.65%), higher health self-efficacy (77.15%), and a regular health care provider (64.73%). More individuals reported higher (vs lower) trust in health information from a physician (75.55%) and lower (vs higher) trust in health information from family or friends (90.89%). A majority of individuals had a BMI≥24.9 (69.44%), and just over half reported having one or more chronic conditions (53.71%). Approximately half of the individuals reported a higher (48.54%) versus lower (51.46%) level of physical activity.

Most individuals included in the analytic sample reported using their wearables “every day” or “almost every day” (72%). They were relatively evenly divided on the use of other mHealth technologies (eg, smartphones and tablets) to help communicate with providers (47.93% “yes” vs 52.07% “no”). A minority of individuals reported sharing health information on SNSs (19.54%) or participating in a web-based health community (11.95%).

Table 1. Weighted, unadjusted population estimates for sociodemographic and health-related characteristics of wearable users willing to share data with providers and with family or friends, HINTS 2019[a] (N=1300).

Characteristics	Users, n (weighted %, SE)		
	Wearable users (N=1300)	Willing to share data with providers (n=1033[b])	Willing to share data with family or friends (n=853[c])
Sex			
Men	486 (44.97, 2.08)	387 (44.44, 2.24)	310 (44.68, 2.62)
Women	787 (55.03, 2.08)	630 (55.56, 2.24)	527 (55.32, 2.62)
Age (years)			
18-34	272 (33.62, 2.08)	234 (36.42, 2.46)	211 (36.87, 3.03)
35-49	339 (30.48, 1.85)	265 (29.82, 2.32)	237 (30.62, 2.49)
50-64	411 (26.37, 1.88)	315 (24.81, 2.11)	245 (23.91, 2.3)
≥65	257 (9.53, 0.81)	205 (8.95, 0.87)	152 (8.6, 0.97)
Race and ethnicity			
White, non-Hispanic	788 (64.69, 1.83)	645 (66.21, 2.03)	533 (66.04, 2.6)
Black, non-Hispanic	141 (8.86, 1.19)	114 (9.26, 1.54)	96 (9.82, 1.92)
Hispanic	169 (17.66, 1.64)	129 (16.29, 1.81)	103 (15.93, 1.86)
Other race or ethnicity	106 (8.79, 1.18)	78 (8.24, 1.32)	70 (8.21, 1.47)
Education			
College graduate or postgraduate	796 (42.55, 1.97)	647 (43.98, 2.41)	541 (43.69, 2.63)
Technical, vocational, or some college	344 (42.45, 2.38)	276 (43.38, 2.84)	220 (41.83, 3.1)
High school graduate or less	130 (15, 2.07)	90 (12.64, 2.05)	76 (14.48, 2.57)
Annual household income (US $)			
≥100,000	526 (38.26, 2.25)	424 (38.41, 2.17)	357 (39.39, 2.76)
75,000-99,999	211 (17.72, 1.56)	175 (19.18, 1.94)	138 (17.5, 2.1)
50,000-74,999	219 (16.18, 1.58)	176 (16.02, 1.74)	142 (16.33, 2.03)
35,000-49,999	142 (13.8, 1.81)	105 (12.6, 2.1)	96 (13.38, 2.22)
<35,000	192 (14.04, 1.77)	145 (13.79, 1.82)	112 (13.4, 1.85)
Geographic area			
Urban	1196 (88.07, 1.91)	951 (88.25, 2.26)	780 (86.07, 2.65)
Rural	104 (11.93, 1.91)	82 (11.75, 2.26)	73 (13.93, 2.65)
Perceived health status			
Poor or fair	113 (10.35, 1.65)	85 (9.67, 1.85)	51 (7.42, 2.1)
Good	1174 (89.65, 1.65)	938 (90.33, 1.85)	793 (92.58[d], 2.1)
Health self-efficacy			
Lower	278 (22.85, 1.97)	209 (20.32, 2.17)	161 (18.54, 2.33)
Higher	1006 (77.15, 1.97)	813 (79.68, 2.17)	683 (81.46[e], 2.33)
Regular health care provider			
No	379 (35.27, 2.28)	288 (33.6, 2.74)	Not examined[f]
Yes	904 (64.73, 2.28)	735 (66.4, 2.74)	Not examined[f]
Trust health information from physician			
Lower	321 (24.45, 1.9)	226 (20.96, 2.24)	Not examined[f]
Higher	955 (75.55, 1.9)	794 (79.04[g], 2.24)	Not examined[f]

Characteristics	Users, n (weighted %, SE)		
	Wearable users (N=1300)	Willing to share data with providers (n=1033[b])	Willing to share data with family or friends (n=853[c])
Trust health information from family or friends			
Lower	1167 (90.89, 1.64)	Not examined[h]	772 (89.55, 2.42)
Higher	89 (9.11, 1.64)	Not examined[h]	66 (10.45, 2.42)
BMI			
18.5-24.9 (normal)	389 (30.56, 1.92)	313 (30.28, 2.34)	263 (30.09, 2.39)
25-29.9 (overweight)	472 (38.87, 2.44)	374 (39.45, 2.93)	321 (40.46, 3.1)
≥30 (obese)	395 (30.57, 2.15)	318 (30.27, 2.41)	244 (29.45, 2.78)
Multimorbidity			
0 conditions	545 (46.29, 2.3)	429 (46.04, 2.83)	368 (47.62, 3.02)
1 condition	398 (31.56, 2.31)	320 (31.46, 2.6)	271 (33.35, 3.02)
≥2 conditions	326 (22.15, 2.02)	263 (22.5, 2.55)	193 (19.03, 2.38)
Level of physical activity			
Lower	674 (51.46, 2.18)	521 (47.97, 2.53)	430 (47.06, 2.81)
Higher	595 (48.54, 2.18)	490 (52.03[i], 2.53)	408 (52.94[j], 2.81)

[a]HINTS 2019: Health Information National Trends Survey 5, Cycle 3.

[b]A total of 18 wearable users had missing data for willingness to share wearable data with providers, therefore the denominator for weighted percentages in this column is 1282.

[c]A total of 22 wearable users had missing data for willingness to share wearable data with family or friends, therefore the denominator for weighted percentages in this column is 1278.

[d]χ^2_1=5.9; *P*=.02.

[e]χ^2_1=7.2; *P*=.01.

[f]Provider-specific variables were not examined in the model predicting willingness to share wearable data with family or friends (see section *Statistical Analysis*).

[g]χ^2_1=8.0; *P*=.007.

[h]Family or friend-specific variables were not examined in the model predicting willingness to share wearable data with providers (see section *Statistical Analysis*).

[i]χ^2_1=12.5; *P*<.001.

[j]χ^2_1=7.6; *P*=.008.

Willingness to Share Wearable Data

A small number of wearable users had missing data regarding willingness to share wearable data with providers (n=18) and willingness to share wearable data with family or friends (n=22); therefore, the analytic sample comprised 1282 respondents for willingness to share wearable data with providers and 1278 respondents for willingness to share wearable data with family or friends. A majority of individuals reported that they would be willing to share wearable data with health care providers (81.86%) and with family or friends (69.1%). In the bivariate analyses, willingness to share wearable data with providers was significantly associated with trust in health information from a physician and level of physical activity. Willingness to share wearable data with family or friends was significantly associated with perceived health status, health self-efficacy, and level of physical activity (Table 1).

Willingness to share wearable data with providers was also significantly associated with each of the 4 measured digital health behaviors: frequency of wearable use, use of mHealth technologies to help communicate with providers, use of SNSs to share health information, and participation in a web-based health community. Only the use of SNSs to share health information was significantly correlated with reported willingness to share wearable data with family or friends (Table 2).

Regression analysis showed that individuals who reported higher (vs lower) health self-efficacy (odds ratio [OR] 1.97, 95% CI 1.11-3.51), higher (vs lower) trust in health information from a physician (OR 1.98, 95% CI 1.12-3.49), and higher (vs lower) levels of physical activity (OR 2.00, 95% CI 1.21-3.31) had significantly greater odds of reported willingness to share wearable data with providers. Among the digital health behaviors, higher (vs lower) frequency of wearable use (OR 2.15, 95% CI 1.35-3.43) and use of mHealth technologies to help communicate with providers (OR 1.99, 95% CI 1.09-3.63) were significantly associated with willingness to share wearable data with providers (Table 3). On the basis of pseudo R^2 values,

the model fit improved with the addition of each digital health behavior variable.

In the regression analysis, only individuals who reported higher (vs lower) levels of physical activity had higher odds of reported willingness to share wearable data with family or friends (OR 1.70, 95% CI 1.02-2.84; *P*=.04). Of the 3 digital health behaviors included in the model, none were significantly associated with willingness to share wearable data with family or friends (data not shown). As in the first model, the model fit improved with the addition of each digital health behavior variable.

Table 2. Weighted, unadjusted population estimates for digital health behaviors of wearable users willing to share data with providers and with family or friends, HINTS 2019[a] (N=1300).

Characteristic	Users, n (weighted %, SE)		
	Wearable users (N=1300)	Willing to share data with providers (n=1033[b])	Willing to share data with family or friends (n=853[c])
Frequency of wearable use			
Lower	396 (28, 1.69)	295 (25.84, 2.02)	237 (26.52, 2.42)
Higher	888 (72, 1.69)	735 (74.16[d], 2.02)	615 (73.48, 2.42)
Use of mHealth[e] technologies to help communicate with providers			
No	635 (52.07, 2.18)	482 (47.98, 2.68)	Not examined[f]
Yes	617 (47.93, 2.18)	521 (52.02[g], 2.68)	Not examined[f]
Use social networking sites to share health information			
No	1056 (80.46, 1.75)	830 (78.51, 2.24)	670 (77.57, 2.35)
Yes	229 (19.54, 1.75)	192 (21.49[h], 2.24)	174 (22.43[i], 2.35)
Participating in an online health community			
No	1143 (88.05, 1.58)	901 (86.52, 2.08)	737 (86.7, 2.21)
Yes	145 (11.95, 1.58)	122 (13.48[j], 2.08)	108 (13.3, 2.21)

[a]HINTS 2019: Health Information National Trends Survey 5, Cycle 3.

[b]A total of 18 wearable users had missing data for willingness to share wearable data with providers, therefore the denominator for weighted percentages in this column is 1282.

[c]A total of 22 wearable users had missing data for willingness to share wearable data with family or friends, therefore the denominator for weighted percentages in this column is 1278.

[d]χ^2_1=4.27; *P*=.04.

[e]mHealth: mobile health.

[f]Provider-specific variables were not examined in the model predicting willingness to share wearable data with family or friends.

[g]χ^2_1=11.13; *P*=.002.

[h]χ^2_1=5.55; *P*=.02.

[i]χ^2_1=5.56; *P*=.02.

[j]χ^2_1=5.67; *P*=.02.

Table 3. Correlates of willingness to share wearable data with providers, weighted, fully adjusted binomial logistic regression model, HINTS 2019[a] (n=1070).

Characteristic	OR[b] (95% CI)	*P* value
Sex		
Men	Reference	Reference
Women	1.13 (0.68-1.89)	.63
Age (years)		
18-34	Reference	Reference
35-49	0.49 (0.25-0.98)	.05
50-64	0.51 (0.24-1.09)	.08
≥65	0.45 (0.17-1.22)	.11
Race and ethnicity		
White, non-Hispanic	Reference	Reference
Black, non-Hispanic	1.26 (0.50-3.18)	.62
Hispanic	0.53 (0.25-1.11)	.09
Other race or ethnicity	0.54 (0.20-1.49)	.23
Education		
College graduate or postgraduate	Reference	Reference
Technical, vocational, or similar college	1.05 (0.58-1.91)	.87
High school graduate or less	0.73 (0.34-1.58)	.42
Annual household income (US $)		
≥100,000	Reference	Reference
75,000-99,999	1.69 (0.76-3.76)	.20
50,000-74,999	1.38 (0.72-2.64)	.32
35,000-49,999	0.86 (0.37-2.01)	.72
<35,000	1.61 (0.64-4.06)	.30
Geographic area		
Urban	Reference	Reference
Rural	0.56 (0.25-1.22)	.14
Perceived health status		
Poor or fair	Reference	Reference
Good	0.96 (0.33-2.76)	.94
Health self-efficacy		
Lower	Reference	Reference
Higher	1.97 (1.11-3.51)	.02
Regular health care provider		
No	Reference	Reference
Yes	1.40 (0.75-2.61)	.28
Trust in health information from a physician		
Lower	Reference	Reference
Higher	1.98 (1.12-3.49)	.02
BMI		
18.5-24.9 (normal)	Reference	Reference
25-29.9 (overweight)	1.00 (0.48-2.08)	.99

Characteristic	OR[b] (95% CI)	*P* value
≥30 (obese)	0.97 (0.46-2.05)	.93
Multimorbidity		
0 conditions	Reference	Reference
1 condition	0.98 (0.53-1.83)	.94
≥2 conditions	1.16 (0.49-2.76)	.74
Level of physical activity		
Lower	Reference	Reference
Higher	2.00 (1.21-3.31)	.008
Frequency of wearable use		
Lower	Reference	Reference
Higher	2.15 (1.35-3.43)	.002
Use of mHealth[c] technologies to help communicate with providers		
No	Reference	Reference
Yes	1.99 (1.09-3.63)	.03
Use of SNSs[d] to share health information		
No	Reference	Reference
Yes	1.50 (0.72-3.12)	.27
Participation in a web-based health community		
No	Reference	Reference
Yes	1.64 (0.65-4.15)	.29

[a]HINTS 2019: Health Information National Trends Survey 5, Cycle 3.

[b]OR: odds ratio.

[c]mHealth: mobile health.

[d]SNS: social networking site.

Discussion

Principal Findings

The purpose of this study was to describe the willingness to share health information collected on wearable health and activity trackers with health care providers and family or friends in a nationally representative sample of US adult wearable users. The findings of this study suggest that most individuals who used wearables were willing to share data generated from these devices with providers (approximately 80%), as well as with family or friends (approximately 70%); however, willingness to share this information varied with behavior-related factors. Health self-efficacy, trust in providers as an information source, frequency of wearable use, use of other mHealth technologies to help communicate with providers, and being physically active appeared to be key factors that influenced willingness to share wearable data with providers. Being physically active also appeared to play an important role in willingness to share data from wearables with family or friends, whereas other factors such as sociodemographics, health-related characteristics, and digital health behaviors played a less prominent role.

These findings contribute to the literature by identifying individual characteristics associated with willingness to share data from wearable health and activity trackers in the adult population and distinguishing the correlates of willingness to share on the basis of the recipient of the data. Interestingly, our study revealed no differential willingness to share according to the sociodemographic characteristics of wearable users. Although the HINTS response rate was relatively low for all survey modalities (approximately 30%), differences by survey modality group (paper, web, and web-bonus) were not significant for the response rate and for the outcome variables of interest.

Willingness to Share Wearable Data With Health Care Providers

The findings of this study suggest that willingness to exchange health- and activity-related information with providers via mHealth technologies may be increasing. For example, in 2013, approximately 50% of US adults who used smartphones, tablets, or other mobile devices reported that they would be “somewhat” or “very” willing to use these technologies to exchange health information about lifestyle behaviors with a provider [18]. Similar to the results of this study, Hyde et al [2] found that approximately 76% of adults reported willingness to share data from wearable health and activity monitors or fitness trackers with providers.

In contrast to prior studies examining willingness or sharing of data from mHealth technologies, factors such as sex, age, weight status [17], race and ethnicity [17,34], income [18,19], and education [18] were not significantly associated with willingness to share wearable data with providers. In addition, having a regular health care provider and having a chronic condition, such as diabetes or hypertension, were not key predictors of willingness to share wearable data. Past studies of US adults have found these to be significant correlates of reported data sharing from electronic medical devices (eg, glucometers and blood pressure monitors) [19]. Therefore, along with prior research [20], the findings of this study demonstrate the importance of examining behavioral predictors and outcomes of patient-generated data sharing for different types of technology.

In this study, level of physical activity and health self-efficacy were significant health-related correlates of willingness to share wearable data with providers. Previous research has shown that wearable users tend to be more physically active than the general population [1]; however, the association between physical activity and willingness of wearable users to share data with providers has been unclear to date [2]. Therefore, our study results contribute to the literature that, among wearable users, those with higher versus lower levels of physical activity may be more willing to share their data with providers. Moreover, although research suggests that using a wearable device may increase health self-efficacy [21,35], our findings suggest that users with relatively high versus low health self-efficacy may be more willing to share their wearable data. Because wearable users with lower levels of physical activity or lower health self-efficacy may benefit the most from sharing wearable data with receptive providers (eg, individualized counseling), future mHealth intervention studies could include these factors in intervention design and explore how to overcome these potential barriers to data sharing. An increasingly acknowledged digital divide that has arisen from disparate health or behavioral outcomes among technology users [36] makes such research particularly important.

The results of this study also showed that trust in health information from providers is a strong predictor of willingness to share wearable data with them. Previous studies of US adults also found an association between trust in providers and willingness to exchange lifestyle behavior information via mHealth technologies, such as smartphones or tablets [18]. To increase willingness to share wearable data among those with lower levels of trust in providers as information sources, future mHealth intervention studies could explore ways to build trust in health information exchange within the patient–provider–technology relationship.

The second aim of our study was to investigate the relationship between digital health behaviors and willingness of users to share wearable data. The study findings showed that those who reported using their wearables every day or almost every day were more likely to report willingness to share data with providers than those who used them less often. Those who reported using (vs not using) smartphones or tablets to help communicate with providers were also more likely to report willingness to share. Consistent with prior research [21,22,25], these results suggest that greater use and technology self-efficacy, specifically in the context of health care and the patient–provider relationship, may increase the intention to share wearable health information with providers. These may be important targets for future intervention research focused on increasing health information exchange with providers via wearables.

By contrast, using SNSs to share health information and participating in a web-based health community were not significantly correlated with willingness to share health information from wearables with providers. These findings suggest that health information–sharing behaviors may vary based on the context (eg, health care setting or online support group), the audience or recipient (eg, health care providers or peers), and the technology through which the information is shared. Because SNSs and web-based health communities may be helpful to individuals through the visibility, availability, control, and reach they offer [37,38], future mHealth intervention studies that aim to improve wearable data sharing with providers could consider how to incorporate these factors into the intervention design.

Willingness to Share Wearable Data With Family or Friends

This study also aimed to explore the correlates of willingness to share data with family and friends, as there may be different drivers of willingness to share data based on the recipient of the information [2]. Controlling for other factors, including sociodemographics, health-related characteristics, and digital health behaviors (frequency of wearable use, use of SNSs to share health information, and participation in a web-based health community), only higher (vs lower) levels of physical activity were significantly associated with willingness to share wearable data with family and friends. As shown by Hyde et al [2], our findings suggest that there are distinctive drivers of intention to share health information via wearable health and activity trackers. However, we contribute to the literature the finding that physical activity of US adult wearable users appears to be a particularly important individual factor associated with willingness to share wearable health information given that physical activity was a strong predictor of willingness to share with both providers and family or friends.

One explanation for these findings is that individuals already engaged in health-promoting behaviors have higher health self-efficacy and are more willing to share their data because these data improve their ability to manage their health. Because social support and health self-efficacy are beneficial outcomes of sharing wearable health data with family or friends [9-11], individuals who may need support the most (those with low levels of physical activity) may be missing these benefits. To overcome barriers to sharing wearable data, such as lack of confidence in level of physical activity, mHealth interventions could be designed to work with participants in web-based health communities to focus on progress and on generating esteem support rather than focusing predominantly on social competition.

Limitations

One of the limitations of this study is the reliance on self-reported and cross-sectional data. In addition, our study was limited by the inability to distinguish between various types of wearable health and activity trackers, which can vary considerably in their functionality. Due to limitations of the data set, we also could not assess other factors that potentially affect willingness to share wearable data with others, such as technology self-efficacy or concerns about privacy or data security.

Conclusions

This study contributes to understanding the willingness of US adults to share data from wearable health and activity trackers with health care providers and family or friends. Several behavior-related factors were independently associated with willingness to share wearable data with providers, including level of physical activity, health self-efficacy, information-related trust in providers, frequency of wearable use, and use of mHealth technologies to help communicate with providers. Only level of physical activity was significantly associated with willingness to share wearable data with family or friends, controlling for other factors. Future behavioral surveillance research could assess attitudes associated with willingness to share wearable data, as well as factors that may influence these attitudes (eg, concerns about privacy), given the strong relationship between attitudes and behavioral intention [39]. In addition, given that attitudes about mHealth technologies and use of patient-generated data from wearables involve both patients and providers, researchers could use participatory action approaches that include these stakeholders in intervention design and implementation. When used as a communication tool, the potential of wearables to improve population health may not be fully realized without attention to these individual and relational factors.

Acknowledgments

No external funds were received to conduct this study, and the authors have no financial disclosures to report. No copyrighted materials were used in this study. The opinions expressed by the authors are their own, and this material should not be interpreted as representing the official viewpoint of the US Department of Health and Human Services, the National Institutes of Health, or the National Cancer Institute.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Health Information National Trends Survey 5, Cycle 3 instrument.
[PDF File (Adobe PDF File), 1425 KB - mhealth_v9i12e29190_app1.pdf]

Multimedia Appendix 2
Weighted, unadjusted population estimates for characteristics of the analytic sample of the Health Information National Trends Survey and that of wearable users.
[DOCX File , 20 KB - mhealth_v9i12e29190_app2.docx]

References

1. Xie Z, Jo A, Hong Y. Electronic wearable device and physical activity among US adults: an analysis of 2019 HINTS data. Int J Med Inform 2020 Dec;144:104297. [doi: 10.1016/j.ijmedinf.2020.104297] [Medline: 33059242]
2. Hyde ET, Omura JD, Fulton JE, Weldy A, Carlson SA. Physical activity surveillance using wearable activity monitors: are US adults willing to share their data? Am J Health Promot 2020 Jul;34(6):672-676. [doi: 10.1177/0890117119900587] [Medline: 31957481]
3. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2021-03-13]
4. Omura JD, Carlson SA, Paul P, Watson KB, Fulton JE. National physical activity surveillance: users of wearable activity monitors as a potential data source. Prev Med Rep 2017 Mar;5:124-126 [FREE Full text] [doi: 10.1016/j.pmedr.2016.10.014] [Medline: 28101443]
5. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018 Sep;15(5):429-448 [FREE Full text] [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
6. 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]

7. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
8. Jin D, Halvari H, Maehle N, Olafsen AH. Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. Behav Inform Technol 2020 Aug 05:1-20. [doi: 10.1080/0144929x.2020.1801840]
9. Zhu Y, Dailey SL, Kreitzberg D, Bernhardt J. "Social Networkout": connecting social features of wearable fitness trackers with physical exercise. J Health Commun 2017 Dec;22(12):974-980. [doi: 10.1080/10810730.2017.1382617] [Medline: 29173072]
10. Kreitzberg DS, Dailey SL, Vogt TM, Robinson D, Zhu Y. What is your fitness tracker communicating?: exploring messages and effects of wearable fitness devices. Qual Res Rep Commun 2016 Sep 03;17(1):93-101. [doi: 10.1080/17459435.2016.1220418]
11. Oeldorf-Hirsch A, High AC, Christensen JL. Count your calories and share them: health benefits of sharing mhealth information on social networking sites. Health Commun 2019 Sep;34(10):1130-1140. [doi: 10.1080/10410236.2018.1465791] [Medline: 29683721]
12. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: an interview study. AMIA Annu Symp Proc 2016;2016:1303-1312 [FREE Full text] [Medline: 28269928]
13. Shin GD, Feng Y, Gafinowitz N, Jarrahi MH. Improving patient engagement by fostering the sharing of activity tracker data with providers: a qualitative study. Health Info Libr J 2020 Sep;37(3):204-215. [doi: 10.1111/hir.12300] [Medline: 32144876]
14. Luo Y, Oh CY, Jean BS, Choe EK. Interrelationships between patients' data tracking practices, data sharing practices, and health literacy: onsite survey study. J Med Internet Res 2020 Dec 22;22(12):e18937 [FREE Full text] [doi: 10.2196/18937] [Medline: 33350960]
15. Li LC, Feehan LM, Xie H, Lu N, Shaw CD, Gromala D, et al. Effects of a 12-week multifaceted wearable-based program for people with knee osteoarthritis: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19116 [FREE Full text] [doi: 10.2196/19116] [Medline: 32618578]
16. Arigo D. Promoting physical activity among women using wearable technology and online social connectivity: a feasibility study. Health Psychol Behav Med 2015 Dec 31;3(1):391-409. [doi: 10.1080/21642850.2015.1118350]
17. Pevnick JM, Fuller G, Duncan R, Spiegel BM. A large-scale initiative inviting patients to share personal fitness tracker data with their providers: initial results. PLoS One 2016;11(11):e0165908 [FREE Full text] [doi: 10.1371/journal.pone.0165908] [Medline: 27846287]
18. Serrano KJ, Yu M, Riley WT, Patel V, Hughes P, Marchesini K, et al. Willingness to exchange health information via mobile devices: findings from a population-based survey. Ann Fam Med 2016;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
19. Turner K, Jo A, Wei G, Tabriz AA, Clary A, Jim HS. Sharing patient-generated data with healthcare providers: findings from a 2019 national survey. J Am Med Inform Assoc 2021 Mar 15;28(2):371-376. [doi: 10.1093/jamia/ocaa272] [Medline: 33180896]
20. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the US population by patterns of mobile health use for health and behavioral tracking: analysis of the National Cancer Institute's Health Information National Trends Survey Data. J Med Internet Res 2020 May 14;22(5):e16299 [FREE Full text] [doi: 10.2196/16299] [Medline: 32406865]
21. Rahman MS, Ko M, Warren J, Carpenter D. Healthcare Technology Self-Efficacy (HTSE) and its influence on individual attitude: an empirical study. Comput Hum Behav 2016 May;58:12-24. [doi: 10.1016/j.chb.2015.12.016]
22. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13:254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
23. Jackson DN, Chou WS, Coa KI, Oh A, Hesse B. Implications of social media use on health information technology engagement: data from HINTS 4 cycle 3. Transl Behav Med 2016 Dec 14;6(4):566-576 [FREE Full text] [doi: 10.1007/s13142-016-0437-1] [Medline: 27628436]
24. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39:115-121 [FREE Full text] [doi: 10.1200/EDBK_238919] [Medline: 31099626]
25. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
26. Riley WT, Oh A, Aklin WM, Wolff-Hughes DL. National Institutes of Health support of digital health behavior research. Health Educ Behav 2019 Dec;46(2_suppl):12-19. [doi: 10.1177/1090198119866644] [Medline: 31742453]
27. Riley WT, Oh A, Aklin WM, Sherrill JT, Wolff-Hughes DL, Diana A, et al. Commentary: pediatric digital health supported by the National Institutes of Health. J Pediatr Psychol 2019 Apr 01;44(3):263-268 [FREE Full text] [doi: 10.1093/jpepsy/jsy108] [Medline: 30597095]

28. Rutten LJ, Blake KD, Skolnick VG, Davis T, Moser RP, Hesse BW. Data resource profile: the National Cancer Institute's Health Information National Trends Survey (HINTS). Int J Epidemiol 2020 Feb 01;49(1):17 [FREE Full text] [doi: 10.1093/ije/dyz083] [Medline: 31038687]
29. Health Information National Trends Survey, methodology reports. US National Cancer Institute. URL: https://hints.cancer.gov/data/methodology-reports.aspx [accessed 2021-03-13]
30. Defining adult overweight and obesity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/obesity/adult/defining.html#:~:text=Adult%20Body% 20Mass%20Index%20(BMI)&text=If%20your%20BMI%20is%20less,falls%20within%20the%20obese%20range [accessed 2021-03-13]
31. President's council on sports, fitness and nutrition. US Department of Health and Human Services - Office of Disease Prevention and Health Promotion. URL: https://www.hhs.gov/fitness/be-active/physical-activity-guidelines-for-americans/index.html#:~:text=For%20substantial%20health%20benefits%2C%20adults,or%20an%20equivalent%20combination%20of [accessed 2021-11-05]
32. Rural-urban continuum codes. US Department of Agriculture - Economic Research Service. 2013. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx [accessed 2021-03-13]
33. Rizzo L, Moser R, Waldron W, Wang Z, Davis W. Analytic methods to examine changes across years using HINTS 2003 and 2005 data. National Institutes of Health. 2008. URL: https://hints.cancer.gov/docs/HINTS_Data_Users_Handbook-2008.pdf [accessed 2021-03-13]
34. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
35. Gualtieri L, Rosenbluth S, Phillips J. Can a free wearable activity tracker change behavior? The impact of trackers on adults in a physician-led wellness group. JMIR Res Protoc 2016 Nov 30;5(4):e237 [FREE Full text] [doi: 10.2196/resprot.6534] [Medline: 27903490]
36. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
37. Rains SA. Coping With Illness Digitally. Cambridge, MA: MIT Press; 2018:1-240.
38. Wright KB. New technologies and health communication. In: O'Hair HD, O'Hair MJ, editors. The Handbook of Applied Communication Research. Hoboken, New Jersey, United States: John Wiley & Sons, Inc; 2020:863-878.
39. Montaño DE, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer BK, Viswanath KV, editors. Health Behavior: Theory, Research, and Practice. Hoboken, New Jersey, United States: Wiley; 2018:95-124.

Abbreviations

HINTS: Health Information National Trends Survey
mHealth: mobile health
OR: odds ratio
SNS: social networking site

Edited by L Buis; submitted 29.03.21; peer-reviewed by A Jo, J Pevnick; comments to author 27.04.21; revised version received 21.06.21; accepted 08.10.21; published 13.12.21.

Please cite as:
Rising CJ, Gaysynsky A, Blake KD, Jensen RE, Oh A
Willingness to Share Data From Wearable Health and Activity Trackers: Analysis of the 2019 Health Information National Trends Survey Data
JMIR Mhealth Uhealth 2021;9(12):e29190
URL: https://mhealth.jmir.org/2021/12/e29190
doi:10.2196/29190
PMID:34898448

Original Paper

A Wearable Activity Tracker Intervention With and Without Weekly Behavioral Support Emails to Promote Physical Activity Among Women Who Are Overweight or Obese: Randomized Controlled Trial

Melissa Black[1], MA; Jennifer Brunet[1,2,3], PhD

[1]School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada

[2]Institut du savoir Monfort, Hôpital Monfort, Ottawa, ON, Canada

[3]Cancer Therapeutic Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

Corresponding Author:
Jennifer Brunet, PhD
School of Human Kinetics
Faculty of Health Sciences
University of Ottawa
125 University Private
Ottawa, ON, K1N6N5
Canada
Phone: 1 6135625800 ext 3068
Email: jennifer.brunet@uottawa.ca

Abstract

Background: Physical activity (PA) plays a fundamental role in combating the current obesity epidemic; however, most women who are overweight or obese are generally physically inactive. Wearable activity tracker interventions can help increase the PA levels in this population. Supplementing such interventions with behavioral support emails may further improve their effectiveness, but this remains to be confirmed.

Objective: This study aims to determine if adding behavioral support emails to a wearable activity tracker intervention can further increase PA levels among women who are overweight or obese in comparison to a wearable activity tracker–only intervention and a control condition.

Methods: Women with a BMI ≥25 kg/m^2 who were not meeting the Canadian PA guidelines for aerobic and strength training were randomized into 1 of 3 groups. Group 1 received 6 weekly behavioral support emails, a wearable activity tracker, and a copy of the Canadian PA guidelines. Group 2 received a wearable activity tracker and a copy of the Canadian PA guidelines, and group 3 (control condition) received a copy of the Canadian PA guidelines. Self-reported data for walking and moderate to vigorous intensity PA were collected preintervention (week 0; prerandomization), postintervention (7 weeks postrandomization), and at follow-up (21 weeks postrandomization) and analyzed as metabolic equivalent of task minutes per week. In addition, potential mechanisms of behavior change (ie, basic psychological needs satisfaction and motivational regulations) were assessed for within- and between-group differences at all 3 time points. Data were analyzed using nonparametric statistical tests.

Results: A total of 49 women were recruited; data from 47 women (mean age 37.57 years, SD 11.78 years; mean BMI 31.69 kg/m^2, SD 5.97 kg/m^2) were available for analysis. Group 1 reported a significant increase in walking from preintervention to postintervention (χ^2_2=7.5; *P*=.02) but not in moderate to vigorous intensity PA (*P*=.24). Group 1 also reported significant increases in perceptions of competence from preintervention to follow-up (χ^2_2=7.6; *P*=.02) and relatedness from preintervention to follow-up (χ^2_2=8.7; *P*=.005). Increases in perceived autonomy were observed for group 2 (χ^2_2=7.0) and group 3 (χ^2_2=10.6). There were no significant changes in the motivational regulations within the groups. The difference between the groups was not significant for any outcome variable.

Conclusions: The results suggest that adding behavioral support emails to a wearable activity tracker intervention may help to increase time spent walking and perceptions of competence and relatedness for PA among women who are overweight or obese.

Trial Registration: ClinicalTrials.gov NCT03601663; http://clinicaltrials.gov/ct2/show/NCT03601663

(JMIR Mhealth Uhealth 2021;9(12):e28128) doi:10.2196/28128

KEYWORDS

behavior change; motivation; obesity; physical activity; women; mobile phone

Introduction

Background

According to the World Health Organization, obesity is a major risk factor for serious conditions, such as diabetes, certain cancers, and heart diseases [1]. In North America, the prevalence of people who are overweight or obese is a public health concern reported to affect approximately 64% and 28% of the population, respectively [2]. Regular engagement in physical activity (PA) helps with weight management and can reduce the risk of developing certain health conditions associated with obesity [3]. However, large community-based surveys show that only 16% of adults living in Canada meet the current PA guidelines of 150 minutes of moderate to vigorous intensity aerobic PA per week [4], with lower rates being observed among women who are overweight or obese [5]. Therefore, there is a critical need to promote PA among women living in Canada, especially those who are overweight or obese, because it could help to reduce their risk of developing several health conditions, lower their risk of all-cause mortality, and offer them a better quality of life [6].

Previous interventions aimed at increasing PA levels among women who are overweight or obese have shown promising results [7-11], with many of these interventions being delivered face-to-face. Advancements in technology provide the opportunity to build on existing knowledge and develop PA behavior change interventions that are less time- and resource-intensive and more accessible for those who face barriers to attending face-to-face programming (eg, limited transportation options, rural communities, or anxiety about attending in person), which may allow more women to increase their PA levels. Technologies including email and messaging platforms allow specialists to deliver interventions and share well-established behavior change techniques with participants to promote PA, whereas activity tracking devices enable users to self-monitor their PA behavior and make changes accordingly. Several studies have shown that providing participants with a wearable activity tracker to self-monitor their PA behavior is associated with increases in PA [12,13]. Cadmus-Bertram et al [11] observed a 62-minute per week increase in moderate to vigorous intensity PA immediately following a 4-week intervention that provided women who were overweight or obese with a wearable activity tracker and an instructional session. Accordingly, several interventions for which efficacy has been demonstrated to increase PA levels, now provide participants with a wearable activity tracker [12,13].

Although wearable activity tracker interventions may help to increase PA levels initially, researchers have noted a decrease in PA levels following initial exposure to the device [14-16] and a lack of evidence regarding the effectiveness of wearable activity tracker use beyond the initial intervention phase [12]. Moreover, some studies have found that using a wearable activity tracker may undermine autonomous motivation for PA and associated processes [17-19]. Specifically, Kerner and Goodyear [18] found that providing participants with a wearable activity tracker can decrease basic psychological needs satisfaction and autonomous motivation for PA. Mendoza et al [19] found that providing participants with a wearable activity tracker can increase introjected motivation—a controlled form of motivation in which people behave to avoid feelings of guilt or enhance feelings of pride [20]. This is critically important because autonomous motivation is a significant and robust predictor of PA behavior and is associated with PA adherence, whereas controlled motivation is a neutral or negative predictor of long-term PA engagement [21]. Teaching other effective strategies that are based on relevant literature may help augment the short-term benefits of self-monitoring to increase PA by developing autonomous motivation to ensure that the changes are sustained over time [22,23]. Accordingly, it is possible that participants combining a wearable activity tracker with a theory-based behavioral intervention that targets core predictors of PA (eg, psychological needs satisfaction, motivation) may optimize increases in PA levels [24,25].

Harnessing technology to provide self-directed materials explaining effective behavior change techniques that align with contemporary theories of health behaviours could help to further increase and sustain PA levels. Email is a common tool for delivering self-directed materials within interventions. Not only is emailing free and familiar to most adult women, it also provides the opportunity to access the materials when it is suitable to them. Self-determination theory is a suitable theory to guide the development of such emails because it provides a powerful framework for explaining women's PA behavior [10,26,27] and has previously been used to develop effective interventions among women who are overweight or obese [7-10]. Self-determination theory is a macrotheory of human motivation, in which motivation exists along a continuum from amotivation (ie, complete lack of motivation) through controlled motivation (ie, engagement in behavior for external reasons including rewards, pride, or guilt) to autonomous motivation (ie, engagement in behavior for its own sake) [20,27].More autonomous forms of motivation are more positively associated with PA behavior, with the most autonomous form (ie, intrinsic motivation) being the most predictive of long-term PA adherence [21,26]. The use of motivational and behavior change techniques within interventions can be used to enhance autonomous motivation [23] and elicit behavior changes [20] by fostering perceptions of autonomy, competence, and relatedness. Interventionists can deliver many of these techniques to participants over email; for example, they can provide participants with choices, encourage experimentation, and teach strategies for goal setting, self-monitoring, and addressing obstacles.

In addition, much of the literature examining interventions to promote PA has focused on increasing moderate to vigorous intensity PA. Few studies have focused on increasing low-intensity PA (eg, low-intensity walking) despite the growing evidence suggesting that it may have significant health benefits [28,29] and that it is often rated as more enjoyable and more pleasant in low-active women who are obese [30,31]. Walking, regardless of intensity, may also be easier to integrate into one's everyday routines than other types of PA (eg, by replacing some driving with walking; going on social walks with a friend, spouse, or child; taking a lunchtime walk; parking further away from stores). Thus, interventions aimed at promoting walking may help increase PA levels in low-active women who are overweight or obese.

Drawing on self-determination theory [20,27] and previous research [18,19,32], the aim of this randomized controlled trial is to determine whether women who received a multicomponent intervention would increase their PA levels more than women who received fewer components. The multicomponent intervention consisted of 6 weekly autonomy-supportive emails designed to increase perceptions of competence, autonomy, and relatedness as well as autonomous motivation for walking and moderate to vigorous intensity PA; a wearable activity tracker to facilitate self-monitoring; and a paper copy and verbal explanation of the Canadian PA guidelines to establish a target for behavior change (group 1) as compared with receiving a wearable activity tracker to facilitate self-monitoring and a paper copy and verbal explanation of the Canadian PA guidelines (group 2) or a paper copy and verbal explanation of the Canadian PA guidelines only (group 3).

Objective

The aim of this study is to assess changes in PA levels over time within each group and to determine if there were significant differences in changes in PA levels between the groups. A secondary objective is to explore changes in PA-related basic psychological needs satisfaction and motivational regulations within and between groups to gain more insight into any observed changes in PA.

Methods

Study Design

This study was a 3-arm parallel group randomized controlled trial featuring a 6-week intervention designed to increase PA levels among low-active women who were overweight or obese. The study was conducted in Ontario, Canada. The primary outcome of the trial was PA, and the secondary outcomes were PA-related basic psychological needs satisfaction and motivational regulations. Data were collected at preintervention (prerandomization; week 0), postintervention (week 7), and at follow-up (week 21) using a combination of self-report questionnaires and direct measurements. The reporting of this study is in accordance with the CONSORT 2010 statement [33] and the CONSORT guidelines for eHealth interventions [34]. This trial was registered at ClinicalTrials.gov (NCT03601663) on July 26, 2018, and was approved by the institutional review board at the University of Ottawa (H-06-18-437). All participants provided informed consent digitally through a web-based form.

Recruitment and Study Sample

A convenience sample of women was recruited between September 2018 and March 2019 by advertising through social media (ie, Facebook), web-based boards (ie, Kijiji, Craigslist, or local classifieds), and posters in publicly accessible areas (ie, community centers or physician's offices). Advertisements encouraged women to contact the research team for further information and eligibility screening.

Women were eligible if they met the following inclusion criteria: (1) identified as female; (2) aged 18-65 years; (3) BMI ≥25 kg/m^2; (4) could read and write in English; (5) answered *no* to the question: *Do you have any health concerns that could prevent you from safely engaging in PA?*; (6) were not pregnant or lactating; (7) reported engaging in <150 minutes of moderate to vigorous intensity PA and strength or resistance training (eg, free weights, weight machines, resistance bands, and exercises using body weight) <2 times per week; (8) had access to internet and an active email account; (9) had not used a wearable activity tracking device in the past 12 months; and (10) lived <50 km from the University of Ottawa.

Data Collection

Eligible participants were informed about all relevant aspects of the study before enrolling and then the digital consent was obtained. After providing consent, they were directed to a web-based platform (ie, SurveyMonkey) to complete the baseline questionnaires. The questionnaires were designed to collect sociodemographic information, health status, self-reported PA, basic psychological needs satisfaction for PA, and motivational regulations for PA. Once participants completed the questionnaires, they were invited to meet with the first author at the location of their preference, either their home or the University of Ottawa, to measure their height, weight, body mass, and body composition. After the measurements were taken and recorded, the first author opened an opaque envelope revealing the participants' group allocation. Subsequent questionnaires were completed on the web by participants at postintervention (week 7) and at follow-up (week 21).

Randomization

The randomization sequence was generated by an independent researcher using permuted blocks of 3 and 6 using a web-based randomization software program (Sealed Envelope Ltd, 2017). It was not possible to blind participants or the researchers because of the nature of the intervention and their role in delivering the intervention, respectively.

Intervention Groups

Group 1

Participants randomized to group 1 received a paper copy and brief verbal explanation of the Canadian PA guidelines for adults aged 18-64 years. They also received a Polar A300 activity monitor with a charging cable and access to the Polar Flow web and smartphone apps for the duration of the 6-week intervention.

All materials were provided when they met with the first author for the baseline assessment. They were instructed to wear the device on their wrist daily during waking hours for the 6-week intervention period, except when swimming or bathing, beginning the day following the baseline assessment. The first author provided instructions on how to navigate the device and assisted participants in syncing the device with their smartphone and/or computer so that they could review their PA data in greater detail.

In addition, during the 6-week intervention, participants received standardized emails from the first author on a weekly basis. The emails featured established motivational and behavior change techniques that align with self-determination theory and were written in a noncontrolling language [23,35] to enhance perceptions of autonomy (ie, perceived control over one's actions), competence (ie, perceived mastery of tasks and skills), and relatedness (ie, perceived belonging and connection to others) and in turn enhance autonomous motivation for PA [36,37]. Key techniques included goal setting, action planning, contingency planning, and self-monitoring. Other recurring themes throughout the emails included learning from trial and error, focusing on making small changes, choosing enjoyable activities, and aligning plans with personal beliefs and values. A detailed overview of the contents and techniques included in the emails is provided in Multimedia Appendix 1.

Group 2

Participants randomized to group 2 received a paper copy and brief verbal explanation of the Canadian PA guidelines for adults aged 18-64 years and a Polar A300 activity monitor when they met with the first author at week 0. Group 2 was a comparison arm for testing if the combined intervention group 1 received was more effective than providing people with a wearable activity tracker alone, as few studies have isolated the effect of this component [38,39]. After completing the questionnaires at follow-up, group 2 participants were provided with a copy of the weekly emails to thank them for their participation in the study.

Group 3

Participants randomized to group 3 received a paper copy and brief verbal explanation of the Canadian PA guidelines for adults aged 18-64 years when they met with the first author at week 0 but no further treatment as group 3 represented the control condition. After completing the questionnaires at follow-up, group 3 participants were provided with a copy of the weekly emails to thank them for their participation in the study.

Sample Size Determination

A priori, the target sample size was estimated using G*Power [40] to ensure sufficient power for the primary outcome of the total metabolic equivalent of task (MET)-minutes per week of PA. Using a mixed repeated-measures analysis of variance with 3 groups and 3 repeated measures, the target sample size was 36 participants, assuming an effect size of 0.25, α of .05, power of 0.80, and a correlation coefficient of among repeated measures of 0.50. These assumptions were made based on findings from a meta-analysis of pedometer-based PA interventions [41] and other interventions with overlapping features developed to promote PA in similar populations [11,42].

Measures

Sociodemographic and Health Characteristics

Sociodemographic and health information were collected from participants before the intervention. Sociodemographic measures included age, marital status, race, highest level of education attained, number of children and their age, annual household income, and employment status. Health measures included self-reported history of chronic diseases, smoking history, and self-rated health. Self-rated health was measured using the first question of the 36-item Short Form Health Survey [43], which asks, "In general, how would you say your health is?" and provides 5 response categories: (1) *excellent*, (2) *very good*, (3) *good*, (4) *fair*, and (5) *poor*. Self-reported health was reassessed postintervention.

Anthropometrics

The height (m), body mass (kg), body composition, and waist circumference (cm) of the participants were measured preintervention and postintervention without shoes and with light clothes. Body mass and composition were measured using a hospital-grade body weight scale (TBF 300A, Tanita Corporation of America Inc). Height was measured using a portable wall-mounted height rod (HR-200, Tanita Corporation of America Inc). Waist circumference was measured over participants' clothing with a measuring tape midway between the 10th rib and the top of the iliac crest. Before the measurements, participants were asked to refrain from the following: (1) drinking alcohol or engaging in moderate to vigorous intensity PA for 12 hours before meeting with the first author, (2) eating or drinking for 3 hours before the meeting, and (3) eating *excessively* or *restrictively* within 24 hours of the meeting [44].

PA Behavior

PA behavior (primary outcome) was assessed at all 3 time points using the International Physical Activity Questionnaire (IPAQ) Short Form. Participants were asked to report the number of days and average duration over the past week that they engaged in sedentary behaviors, walking, and moderate to vigorous intensity PA. The number of days was multiplied by the average duration to estimate the number of minutes per week for each category. Scores for vigorous, moderate, and walking activities were multiplied by 8.0, 4.0, and 3.3, respectively, to calculate the total number of MET minutes per week, which reflects the amount of energy expended in each category throughout the week. The scores for moderate and vigorous intensity PA were then summed to calculate moderate to vigorous intensity PA. Both moderate to vigorous intensity PA and walking MET minutes per week were analyzed as outcome variables. Scores on the IPAQ Short Form have demonstrated good reliability and validity for use in adult populations [45].

PA-Related Basic Psychological Needs Satisfaction

Basic psychological needs satisfaction in relation to PA (secondary outcome) was measured at all 3 time points using the Psychological Need Satisfaction in Exercise (PNSE) scale

[46]. The PNSE scale consists of 18 statements that were used to calculate 3 subscale scores, which measure perceived autonomy, competence, and relatedness for exercise. All items were rated using a 6-point Likert scale ranging from (1) *false* to (6) *true*, wherein lower scores represent less needs satisfaction. For this study, the scale was modified by replacing the word *exercise* with *physical activity*. Internal reliability coefficients for each subscale of the PNSE in this study are presented in Multimedia Appendix 2.

Motivational Regulations for PA

Motivational regulations for PA (secondary outcome) were assessed at all 3 time points using the Behavioral Regulation in Exercise Questionnaire, version 3 (BREQ-3) [47,48]. The BREQ-3 includes 24 items divided into 6 subscales assessing all 6 motivational regulations; each motivational regulation was assessed as a separate outcome in this study given the limitations associated with using a combined score [49]. Participants were asked to respond to each item using a 5-point Likert scale ranging from (0) *not true for me* to (4) *very true for me*, wherein lower scores represent less of that motivational regulation. For this study, the scale was modified by replacing the word *exercise* with *physical activity*. The internal reliability coefficients of the BREQ-3 in this study are presented in Multimedia Appendix 2.

Statistical Analysis

Data were analyzed using SPSS Statistics (version 26; IBM Corporation) following intent-to-treat principles in which data from all participants who were randomized were analyzed. Initially, all data were screened for missingness, outliers, and normality. Item-level missing data were imputed by calculating the mean score of the subscale to which the missing item belonged; person-level missing data (ie, data missing because of participant attrition) was imputed by replacing the missing subscale scores with the last observation. Descriptive statistics were calculated for all variables. Pairwise correlations were estimated between moderate to vigorous intensity PA, walking, BMI, age, education, and depressive symptoms at baseline to identify any potential covariates. No correlations were statistically significant, and thus were not included in the analyses.

The Shapiro-Wilk test was used to check for normality, as recommended for sample sizes <50. As the data were found to be nonnormally distributed, the Friedman test (ie, the nonparametric equivalent to the 2-way analysis of variance) was used to test for significant differences in median values between time points within groups. Pairwise comparisons were performed using the Wilcoxon signed-rank test with a Bonferroni correction for multiple comparisons to locate differences. The Kruskal-Wallis H test was used to assess for differences in change scores between groups. For these tests, change scores were calculated between the time points by subtracting the later values (eg, postintervention) from the former values (eg, preintervention). Statistically significant differences between the groups were further analyzed with post hoc analyses, namely pairwise comparisons using the Dunn procedure with a Bonferroni correction for multiple comparisons. Bonferroni adjusted *P* values <.05 were considered statistically significant.

Results

Participants

In total, 88 women contacted the first author to express interest in this study. A total of 63 patients were screened for eligibility, of which 49 (78%) provided consent. Of these 49 women, 2 (4%) dropped out after completing the web-based questionnaires but before meeting with the first author in person for anthropometric measurements because of unforeseen time constraints (n=1) and unwillingness to comply with the wearable activity tracker protocol (n=1). The remaining 96% (47/49) of participants were randomized and included in the analyses. Among the participants, 9% (4/47) did not complete the postintervention questionnaires for unspecified reasons (2/4, 50%) or were lost to follow-up (2/4, 40%); an additional 11% (5/47) were lost to follow-up (ie, at the follow-up assessment). A CONSORT flow diagram showing the flow of participants through the trial is shown in Figure 1.

The characteristics of the analytical sample preintervention are presented in Table 1. There were no statistically significant differences found between the groups for the main study variables (ie, PA, basic psychological needs satisfaction, and motivational regulations), age, BMI, or waist circumference. For the 47 participants randomized, the average age was 37.57 years (SD 11.78 years), average BMI was 31.69 kg/m^2 (SD 5.97 kg/m^2; with 26/47, 55% classified as obese; BMI ≥30 kg/m^2), and average waist circumference was 98.54 cm (SD 15.13 cm). Participants reported a history of chronic diseases, including stroke (1/47, 2%), diabetes (3/47, 6%), high blood pressure (7/47, 15%), high cholesterol (5/47, 11%), arthritis (7/47, 15%), asthma (7/47, 15%), and moderate to severe depression (13/47, 28%). The medians and IQRs for all outcome variables are shown in Table 2. Mean MET minutes per week by group for moderate to vigorous intensity PA and walking are presented in Figures 2 and 3, respectively. The effect sizes and 95% CI for the post hoc analyses are presented in Multimedia Appendix 3. Bivariate correlations between change in primary (ie, moderate to vigorous intensity PA and walking) and secondary outcomes (ie, basic psychological needs and motivational regulations) are presented in Multimedia Appendix 4.

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

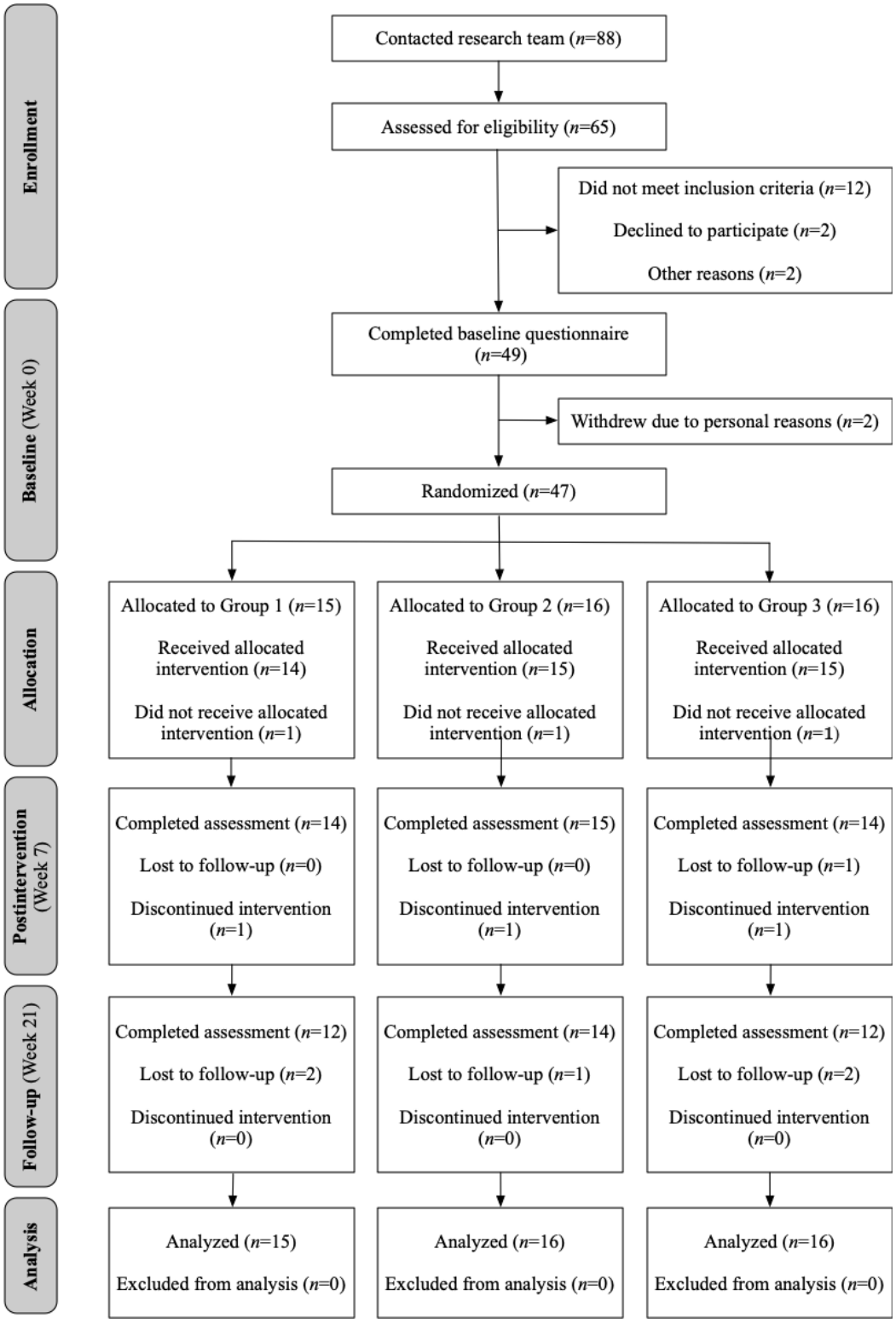


Table 1. Baseline characteristics of participants randomized (N=47).

Variable	Group 1 (n=15)	Group 2 (n=16)	Group 3 (n=16)	Value, range
Age (years; n=47), mean (SD)	32.6 (7.8)	39.2 (11.6)	40.6 (14.0)	18-63
BMI (kg/m^2; n=47), mean (SD)	31.3 (5.8)	32.9 (5.7)	30.8 (6.5)	23.6-45.7
Body composition (% fat; n=46), mean (SD)	41.1 (5.8)	42.5 (6.0)	40.1 (7.1)	24.2-55
Waist circumference (cm; n=47), mean (SD)	97.4 (13.1)	102.8 (16.2)	95.3 (15.7)	71-141
Self-rated health (n=47), n (%)				N/A[a]
Poor	1 (7)	1 (6)	1 (6)	
Fair	5 (33)	4 (25)	5 (31)	
Good	9 (60)	8 (50)	8 (50)	
Very good	0 (0)	3 (19)	2 (13)	
Excellent	0 (0)	0 (0)	0 (0)	
Smoking status (n=47), n (%)				N/A
Never smoked	13 (87)	6 (38)	14 (88)	
Previously smoked	1 (7)	5 (31)	1 (6)	
Currently smokes	1 (7)	5 (31)	1 (6)	
Education (n=47), n (%)				N/A
High school	0 (0)	1 (6)	1 (6)	
Some college or university	1 (7)	4 (25)	4 (25)	
College or university	13 (87)	7 (44)	10 (63)	
Graduate degree	1 (7)	4 (25)	1 (6)	
Employment status (n=46), n (%)				
Unemployed	2 (13)	3 (19)	4 (25)	
Student	3 (20)	1 (6)	2 (13)	
Part-time worker	3 (20)	1 (6)	3 (19)	
Full-time worker	7 (47)	10 (63)	7 (44)	
Annual household income (CAD $; US $), n (%)				N/A
≤49,999 (34,999)	6 (40)	6 (38)	4 (25)	
50,000-99,999 (35,000-69,999)	3 (20)	5 (31)	2 (12)	
>100,000 (70,000)	4 (27)	5 (31)	4 (25)	
Race (n=47), n (%)				N/A
White	11 (73)	14 (88)	14 (88)	
Other	4 (27)	2 (13)	2 (13)	

[a]N/A indicates the value is not applicable, as the data are presented as number and frequency.

Table 2. Median and IQR for all outcome variables.

Variables	Group 1				Group 2				Group 3			
	Preintervention, median (IQR)	Postintervention, median (IQR)	Follow-up, median (IQR)	*P* value	Preintervention, median (IQR)	Postintervention, median (IQR)	Follow-up, median (IQR)	*P* value	Preintervention, median (IQR)	Postintervention, median (IQR)	Follow-up, median (IQR)	*P* value
Physical activity[a]												
MVPA[a,b]	40.00 (720.00)	0.00 (832.00)	360.00 (800.00)	.24	300.00 (480.00)	280.00 (1140.00)	820.00 (2040.00)	.24	60.00 (440.00)	250.00 (920.00)	600.00 (1890.00)	.34
Walking	676.50 (792.00)	1386.00 (1798.00)	1386.00 (2178.00)	.02[c]	198.00(903.38)	594.00 (2557.50)	693.00 (1311.75)	.162	429.00 (1641.75)	643.00 (2165.63)	643.5.00 (928.13)	.49
Basic psychological needs satisfaction												
Autonomy	5.50 (1.83)	5.17 (1.00)	5.67 (2.00)	.27	4.83 (1.83)	5.33 (1.46)	5.00 (2.08)	.03[c]	4.83 (1.71)	5.83 (1.13)	5.67 (1.83)	.005[c]
Competence	3.67 (2.83)	4.00 (0.83)	4.00 (1.17)	.02[c]	3.67 (1.71)	3.75 (1.58)	4.00 (3.08)	.34	3.75 (1.25)	3.67 (3.13)	3.42 (2.33)	.55
Relatedness	3.00 (2.50)	3.83 (2.50)	3.89 (2.67)	.01[c]	3.29 (2.35)	3.75 (2.58)	3.33 (2.21)	.50	3.67 (1.25)	2.83 (3.83)	3.83 (3.00)	.92
Motivational regulations												
Amotivation	0.00 (1.00)	0.00 (0.75)	0.00 (0.75)	.67	0.00 (0.50)	0.00 (0.44)	0.12 (0.94)	.70	0.38 (0.94)	0.38 (1.19)	0.12 (0.90)	.80
External	1.00 (2.00)	0.50 (2.25)	0.25 (1.50)	.20	0.75 (1.25)	0.50 (1.25)	0.62 (1.19)	.87	1.00 (0.90)	1.00 (1.19)	0.75 (2.00)	.77
Introjected	2.50 (0.75)	2.00 (1.75)	2.00 (1.00)	.12	2.62 (1.00)	2.50 (0.88)	2.25 (0.94)	.42	2.12 (2.38)	2.00 (2.56)	1.88 (3.19)	.56
Identified	2.50 (1.00)	2.75 (1.25)	2.50 (1.50)	.59	2.62 (1.00)	2.62 (1.13)	2.25 (1.38)	.09	2.25 (1.38)	2.25 (1.50)	2.12 (2.00)	.86
Integrated	2.00 (0.75)	2.00 (1.50)	2.00 (1.25)	.63	1.50 (1.31)	2.00 (1.69)	1.75 (1.44)	.77	1.75 (1.88)	1.88 (2.06)	1.88 (1.81)	.92
Intrinsic	2.25 (1.00)	2.75 (1.50)	2.75 (1.50)	.28	2.25 (1.44)	2.62 (1.44)	2.50 (1.38)	.32	2.12 (1.69)	2.12 (2.25)	2.25 (1.69)	.56

[a]Current physical activity guidelines recommend at least 150 minutes of moderate intensity aerobic PA or at least 75 minutes of vigorous intensity aerobic PA, which is equivalent to at least 450 metabolic equivalent of task (MET) minutes per week to meet PA. PA has been shown to have a dose-response relationship with subsequent health-benefits, therefore higher MET minute scores are considered better.

[b]Moderate to vigorous intensity physical activity.

[c]Indicates a significant within-group difference ($P<.05$).

Figure 2. Mean metabolic equivalent of task minutes of moderate to vigorous intensity physical activity per week by group. MET: metabolic equivalent of task; MVPA: moderate to vigorous intensity physical activity.

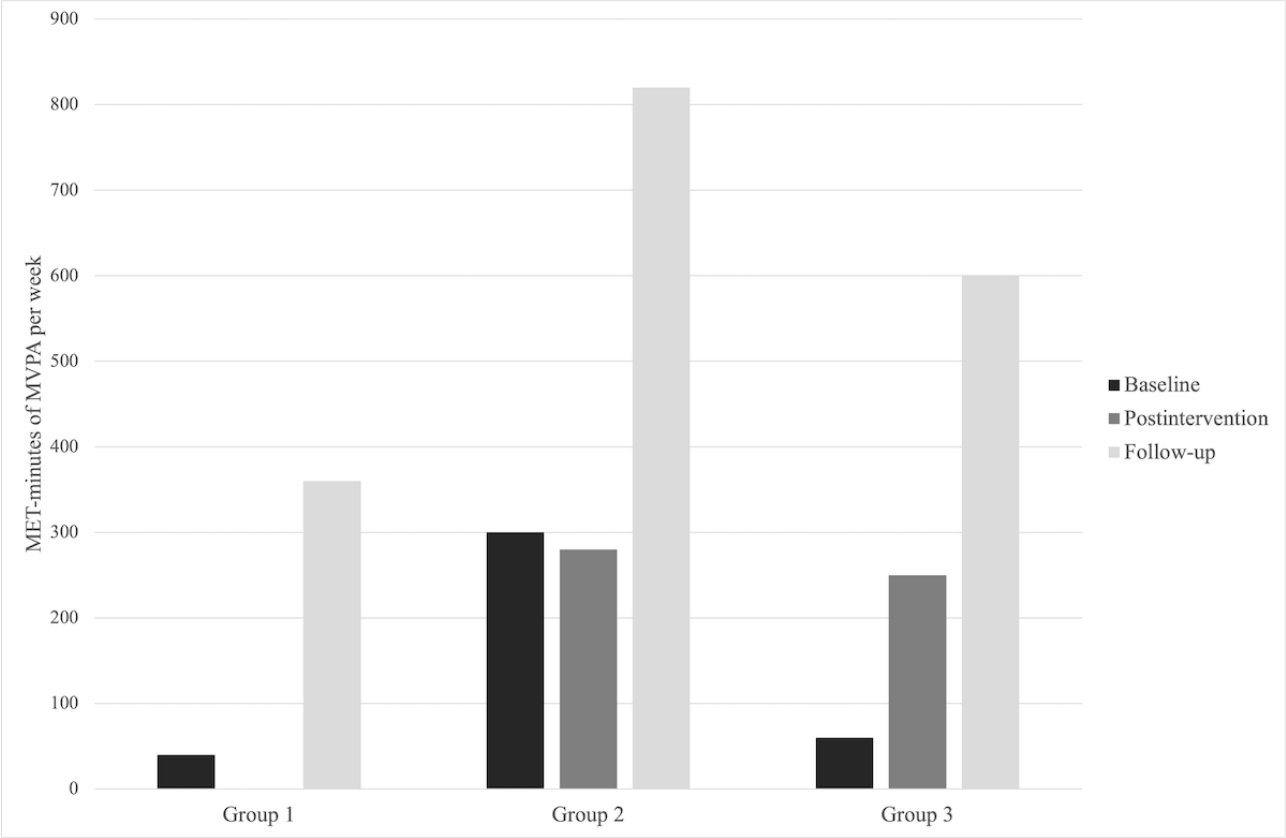
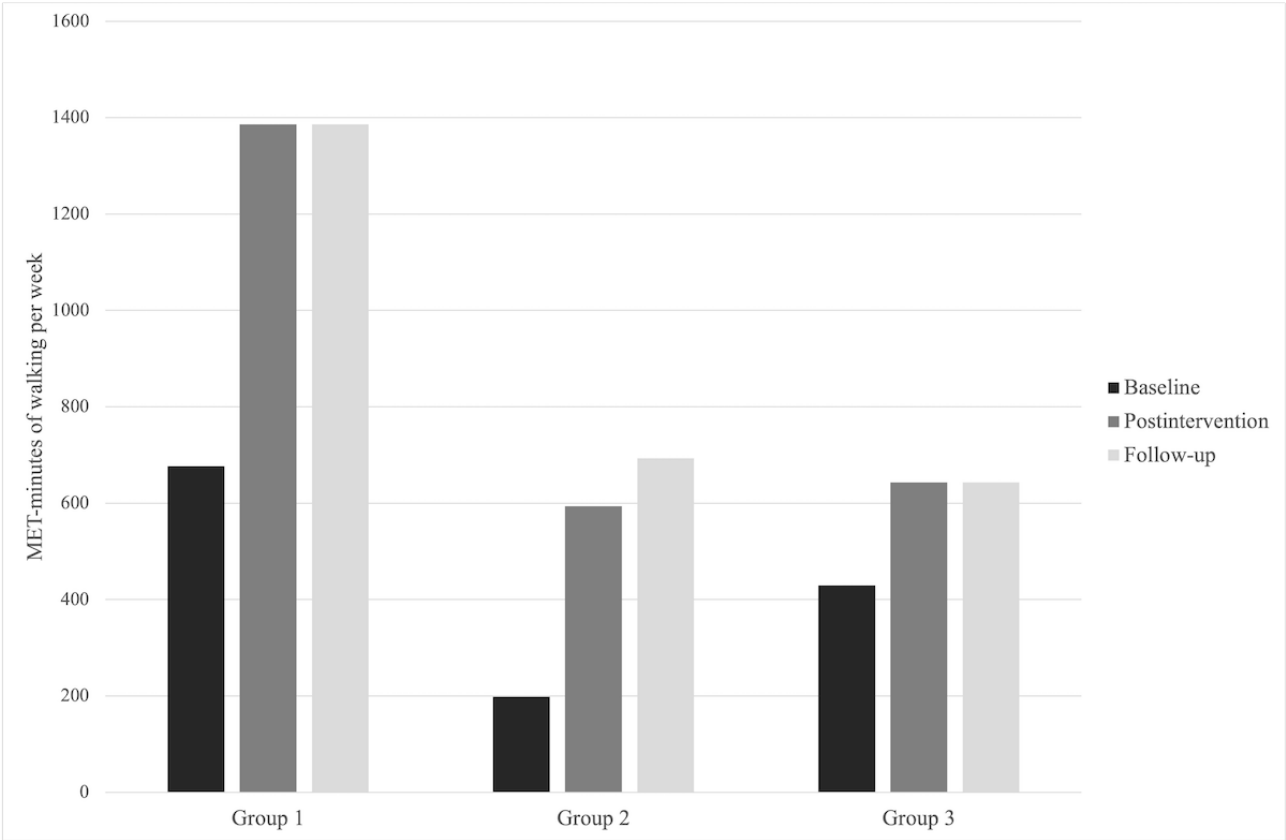


Figure 3. Mean metabolic equivalent of task minutes of walking per week by group. MET: metabolic equivalent of task.



PA Behavior

There was a significant increase in MET minutes per week of walking within group 1 (χ^2_2=7.5; *P*=.02) but not in group 2 (*P*=.16) or group 3 (*P*=.49). Pairwise comparisons using the Wilcoxon signed-rank test revealed that walking increased for group 1 from preintervention to postintervention (t_1=−0.87; *r*=0.16; *P*=.05) although not statistically significant with the Bonferroni correction, and there were no significant changes in walking from preintervention to follow-up (*P*=.25) or from postintervention to follow-up (*P*=.99). In addition, there were no significant changes in moderate to vigorous intensity PA within group 1 (*P*=.24), group 2 (*P*=.24), or group 3 (*P*=.34). There were no significant differences in change scores between the groups for either outcome from preintervention to postintervention (walking, *P*=.26; moderate to vigorous intensity PA, *P*=.40), preintervention to follow-up (walking, *P*=.43; moderate to vigorous intensity PA, *P*=.40), or postintervention to follow-up (walking, *P*=.98; moderate to vigorous intensity PA, *P*=.97).

PA-Related Basic Psychological Need Satisfaction

A significant change in perceived autonomy was observed within group 2 (χ^2_2=7.0; *P*=.03) and group 3 (χ^2_2=10.6; *P*=.005). Pairwise comparisons indicated that perceived autonomy decreased from postintervention to follow-up in group 2 (t_1=0.78; *r*=−0.39; *P*=.08) but not from preintervention to postintervention (*P*=.65) or from preintervention to follow-up (*P*=.99). In addition, perceived autonomy increased significantly from preintervention to postintervention in group 3 (t_1=−1.00; *r*=0.50; *P*=.02) but not from preintervention to follow-up (*P*=.28) or from postintervention to follow-up (*P*=.75).

There was also a significant difference in change between the groups from preintervention to postintervention (H_2=5.99; *P*=.05) and from postintervention to follow-up (H_2=6.70; *P*=.04) but not from preintervention to follow-up (*P*=.09). Pairwise comparisons with adjusted *P* values showed that group 3 had a greater increase in perceived autonomy from preintervention to postintervention than group 1 (*P*=.04), but there were no significant differences between groups 1 and 2 (*P*=.67) or groups 2 and 3 (*P*=.63). In addition, from postintervention to follow-up, there was a smaller decrease in perceived autonomy in group 1 compared with group 2 (*P*=.03) but no significant differences between group 1 and group 3 (*P*=.56) or group 2 and group 3 (*P*=.59).

A significant change in perceived competence was observed within group 1 (χ^2_2=7.6; *P*=.02). Pairwise comparisons revealed a significant increase from preintervention to follow-up (T=−0.90; *r*=.45; *P*=.04) but not from preintervention to postintervention (*P*=.30) or from postintervention to follow-up (*P*=.99). No significant changes in perceived competence were observed in group 2 (*P*=.34) or group 3 (*P*=.55), and there were no significant differences between the groups for change in perceived competence from preintervention to postintervention (*P*=.76), preintervention to follow-up (*P*=.10), or postintervention to follow-up (*P*=.34).

A significant change in perceived relatedness was observed in group 1 (χ^2_2=8.7; *P*=.01). Pairwise comparisons revealed a significant increase from preintervention to follow-up (T=−0.93; *r*=0.47; *P*=.03) but not from preintervention to postintervention (*P*=.20) or from postintervention to follow-up (*P*=.99). No significant changes in perceived relatedness were found within group 2 (*P*=.50) or group 3 (*P*=.92), and there were no significant differences between the groups for change in perceived relatedness preintervention to postintervention (*P*=.28), preintervention to follow-up (*P*=.07), or postintervention to follow-up (*P*=.15).

Motivational Regulations for PA

There were no significant changes in amotivation, external, introjected, identified, integrated, or intrinsic motivational regulations for PA within groups across time points (Table 2). There were no significant differences between the groups for changes in any of the motivational regulations for PA. Specifically, there was no significant difference between groups for amotivation (preintervention to postintervention, *P*=.72; preintervention to follow-up, *P*=.75; or postintervention to follow-up, *P*=.67); external (preintervention to postintervention, *P*=.73; preintervention to follow-up, *P*=.42; or postintervention to follow-up, *P*=.67); introjected (preintervention to postintervention, *P*=.60; preintervention to follow-up, *P*=.48; or postintervention to follow-up, *P*=.82); identified (preintervention to postintervention, *P*=.80; preintervention to follow-up, *P*=.21; or postintervention to follow-up, *P*=.30); integrated (preintervention to postintervention, *P*=.95; preintervention to follow-up, *P*=.74; or postintervention to follow-up, *P*=.99); or introjected regulations (preintervention to postintervention, *P*=.60; preintervention to follow-up, *P*=.35; or postintervention to follow-up, *P*=.22).

Discussion

Principal Findings

Low PA among women who are overweight and obese is a cause for concern, considering the numerous health benefits associated with regular engagement in PA [6]. The principal finding of this study indicates that adding email behavioral support to a wearable activity tracker intervention yielded an increase in walking of 709.50 MET minutes per week over the course of the 6-week intervention, though such increases were not statistically significantly more than providing women with a wearable activity tracker without emails (M_{change}=600.19 MET minutes) or only providing them with a copy of PA guidelines (M_{change}=109.31 MET minutes). Furthermore, no significant changes in moderate to vigorous intensity PA were observed in any of the groups. The findings also show that increases in autonomy preintervention to postintervention were greater for group 2 and 3 participants than for group 1 participants, that perceptions of autonomy decreased toward preintervention values from postintervention to follow-up for group 2, and that perceptions of competence and relatedness increased from preintervention to follow-up for group 1, although increases were not significantly greater than for groups 2 or 3. Collectively, the findings from this study suggest that the

autonomy-supportive email intervention received by group 1 could help to promote walking among women who are overweight or obese, but further research is necessary to confirm the current results and perhaps optimize the intervention.

Results in the Context of Other Literature

Previous studies suggest that, in general, behavior change interventions (including those delivered by email) may be effective in increasing walking among people who are sedentary [50]. The increase in weekly walking observed from preintervention to postintervention for group 1 participants, though not statistically significantly different from the increases observed in groups 2 and 3, supports this assertion. Walking was incorporated into the examples provided in the emails to the participants given evidence suggesting engaging in low-intensity PA (including low-intensity walking) in lieu of sedentary activity may confer meaningful health benefits [29], and may be more accessible and enjoyable for women who are sedentary and overweight or obese. For example, the following excerpt was included in week 2 email:

> *Walking is an excellent and accessible way to be physically active and improve your health. Running is a higher intensity activity that is also a good choice. The important thing is that you are moving, doing something you enjoy, and that your physical activity choices fit with your lifestyle. Walking is a great way to do this.*

In addition, the increase in weekly walking observed in group 1 and group 2 may be explained by the provision of a wearable activity tracker. Consistent with previous studies [12,13], providing participants in both groups with a tracker may have fostered self-monitoring, encouraging participants to walk more each week. Finally, providing participants in all 3 groups with a copy and verbal explanation of the PA guidelines may have contributed to the observed increases. Despite no additional counseling component (for groups 2 and 3), it is possible that receiving the PA guidelines prompted participants to reflect on their current PA behavior and make changes accordingly, and since walking can be easily incorporated into many activities of daily living, it may have been an option for them as they worked toward meeting PA guidelines.

In contrast to research showing that autonomy-supportive interventions and wearable activity tracker interventions can increase moderate to vigorous intensity PA levels [7-10,13], no significant changes in moderate to vigorous intensity PA were observed in this study, regardless of group allocation. Across groups, participants' ability to increase their moderate to vigorous intensity PA levels, which can be more difficult to incorporate into activities of daily living than walking, may have been hindered by commonly reported barriers (eg, inflexible work schedules, long working hours, household responsibilities and chores, weather conditions, and other commitments) [51,52]. Furthermore, although emails attempted to help group 1 participants overcome such barriers, there remains a need for further improvement. Possibly, 1-way message delivery from the first author to the participants may be insufficient. If so, incorporating 2-way interactive counseling by adding a counseling component that could be delivered over the phone or via videoconferencing to help participants identify and overcome barriers may help to achieve the desired increases in moderate to vigorous intensity PA levels. Another possible explanation for the nonsignificant finding across groups is related to the measure used to assess PA behavior. The IPAQ Short Form is a self-report measure, and people tend to overestimate their PA behavior [53]. In addition, when completing the IPAQ Short Form, participants were asked to provide the amount of time spent per week in vigorous intensity PA, moderate intensity PA, and walking. To ensure consistency of interpretation, explanations of the types of activities included in each category were provided within the questionnaire. However, walking may incur a higher energy expenditure for women who are overweight or obese than for those who are normal weight [54]; therefore, increases in moderate to vigorous intensity walking may have consequently been captured in the walking variables rather than in the moderate to vigorous intensity PA variable. In the future, it may be beneficial to add objective measures of PA (eg, accelerometers) to assess PA intensity (ie, light, moderate, or vigorous) and energy expenditure alongside self-report to capture PA context (ie, work, leisure, transportation, exercise, or walking) to delineate the effects of interventions on different PA outcomes.

A significant increase in perceptions of competence and relatedness for PA was observed in group 1, though not significantly more than in groups 2 and 3, and this may be related to the increases observed in walking. Although PA enjoyment was not assessed in this study, other studies have shown that low-intensity PA such as low-intensity walking can be enjoyable for women who are overweight or obese [30,31], and it can be performed by most women. As such, participants who engaged in walking may have had higher perceptions of competence to engage in PA because they enjoyed it and did not find it hard to do as a result. It is also possible that participants who increased their walking did so in the company of others (eg, friends, spouse, or coworkers) providing opportunities for them to connect and bond with others, and thus increase their perceptions of relatedness. Indeed, some of the content within the emails may have prompted them to do so, as exemplified by the recurring suggestion of *going for a walk with a friend or family member*. Future research should consider how PA context (eg, type of activity, presence of PA companions, or location) is related to basic psychological needs satisfaction, motivational regulations, and PA behavior among women who are overweight or obese.

Moreover, perceived autonomy increased from preintervention to postintervention in groups 2 and 3, and the observed increases were greater than those observed in group 1. Despite no additional behavioral support for groups 2 and 3, simply participating in the study and receiving a copy and verbal explanation of the PA guidelines may have prompted participants to reflect on their PA behavior and consider options for making changes on their own. In addition, consistent with previous studies [12,13], providing group 2 with a wearable activity tracker may have enabled self-monitoring, empowering them to increase their PA levels independently, however they wished to. For group 1, despite recruitment materials providing a description of the intervention, it is possible that participants

in group 1 anticipated receiving more support from the facilitator than was provided. Although the emails featured various established motivational and behavior change techniques [23,55], the absence of reciprocal interaction with a facilitator may have limited the amount of autonomy support that could be derived from the emails. Techniques such as encouraging participants to ask questions, using demonstrations, using empathetic listening, providing opportunities for ongoing support, and offering clear, constructive, and relevant feedback were absent from this intervention because of its asynchronous nature. These techniques may provide participants with the support they need to feel confident about their own choices, which in turn could help to increase their perceptions of autonomy. An improvement to the current intervention may be a mix of emails and synchronous sessions through a web-based platform with a specialist who is present in real time and can emphasize certain autonomy-supportive techniques in a personally relevant manner [56,57]. Because of the positive association observed between perceptions of autonomy and PA in previous studies [21], it is critical to determine if such revisions can lead to gains in perceived autonomy.

Finally, unlike Silva et al [9], who found significantly higher introjected, identified, and intrinsic regulations in the intervention group following a 12-month behavior change intervention for PA, no changes were observed in this study for motivational regulations for PA. Levels of amotivation and external regulation were relatively low preintervention, which is consistent with the possibility that the convenience sample of women recruited to this study was, in general, motivated to make changes in their PA behavior. Although participants' stage of change related to PA was not assessed in this study, it is possible that most were, at minimum, in the contemplation (ie, thinking about change, not yet engaging in change) or preparation stage of change (ie, intending to change in the next 6 months) at the start of the study, given the study eligibility criteria. In this sense, those in the precontemplation stage (ie, unwilling to change or not aware of a problem), which shares features of amotivation, may not have signed up for this study that was focused on promoting PA. Regardless, low levels of, and limited variance in, amotivation and external regulation preintervention likely made it impossible to observe hypothesized decreases in these variables within the current sample. Although the lack of increase in the remaining motivational regulations is unexpected, the maintenance of preintervention moderate levels, particularly among identified regulation and intrinsic motivation, is very important because endorsement of these motivational regulations is associated with higher PA adherence [21]. It is recommended to use qualitative research in future trials to identify reasons associated with the lack of change in motivational regulations and seek out strategies to improve autonomous regulations.

Limitations

The limitations of this study include the sample size for examining secondary outcomes, possible selection bias toward women who volunteered to participate in this study who may have been more motivated to increase their PA than the general population, and the use of *physical activity* in the stems of the questionnaires used to assess basic psychological needs satisfaction and motivational regulations (rather than *physical activity* and *walking*). The use of self-report to assess PA, although necessary for reasons of feasibility and to avoid the risk of a Hawthorne effect, is also a limitation. Finally, data on whether participants read the emails, and if so, how many times and for how long, were not collected; these use data may have provided valuable information regarding fidelity to guide improvements to the intervention. Relatedly, participants allocated to groups 1 and 2 were asked to self-report how often they wore the wearable activity tracker and how often they looked at their PA data (on their wrist and through the web application); however, the accuracy of these self-report data could not be verified. Thus, although seemingly high as of the 31 participants who received a wearable activity tracker, 79% (n=23) reported wearing their tracker >4 days per week for >12 hours per day, underuse may have led to an underestimation of the effects.

Conclusions

This study represents an important step toward developing interventions that promote PA among women who are overweight and obese. It sought to evaluate the additional benefit of adding email counseling to a wearable activity tracker intervention (with the provision of PA guidelines). The findings suggest that, among the 3 interventions tested, providing women with behavior support weekly emails in addition to a wearable activity tracker and PA guidelines may help to increase walking behavior over a 6-week period. However, this study has also revealed possible areas for improvement as there were no significant increases in moderate to vigorous intensity PA observed and providing participants with weekly behavior support emails did not foster basic psychological needs satisfaction and motivational regulations for PA over time. These issues need to be addressed in future trials, possibly by adding synchronous sessions with a PA specialist to the weekly behavior support emails and the wearable activity tracker intervention.

Acknowledgments

JB holds a Tier II Canada Research Chair in Physical Activity Promotion for Cancer Prevention and Survivorship. Both authors would like to thank the women who participated in this study.

Authors' Contributions

MB coconceptualized the study, designed, and delivered the interventions, collected, analyzed, and interpreted the data, drafted the manuscript, and approved the submitted version of this manuscript. JB coconceptualized the study, contributed to the formulation

of the intervention, oversaw all aspects of the study conduct (ie, recruitment, intervention delivery, data collection, and management), contributed to the analytic plan and interpretation of the results, critically revised and edited the manuscript, and approved the submitted version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of weekly emails.
[DOCX File , 17 KB - mhealth_v9i12e28128_app1.docx]

Multimedia Appendix 2
Properties of all outcome variables.
[DOCX File , 20 KB - mhealth_v9i12e28128_app2.docx]

Multimedia Appendix 3
Effect sizes and 95% CI for all outcome variables for each group.
[DOCX File , 23 KB - mhealth_v9i12e28128_app3.docx]

Multimedia Appendix 4
Pairwise correlations between change scores in moderate-to-vigorous intensity physical activity, walking, basic psychological needs satisfaction, and motivational regulations.
[DOCX File , 28 KB - mhealth_v9i12e28128_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 408 KB - mhealth_v9i12e28128_app5.pdf]

References

1. Obesity and overweight: key facts. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-07-02]
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. King NA, Hopkins M, Caudwell P, Stubbs RJ, Blundell JE. Beneficial effects of exercise: shifting the focus from body weight to other markers of health. Br J Sports Med 2009 Dec;43(12):924-927. [doi: 10.1136/bjsm.2009.065557] [Medline: 19793728]
4. Physical Activity, Sedentary Behaviour and Sleep (PASS) indicators data tool, 2020 edition. Center for Surveillance and Applied Research, Public Health Agency of Canada. 2020. URL: https://health-infobase.canada.ca/pass/ [accessed 2021-07-05]
5. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep 2011 Mar;22(1):7-14 [FREE Full text] [Medline: 21510585]
6. Warburton DE, Bredin SS. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
7. Edmunds J, Ntoumanis N, Duda JL. Testing a self-determination theory-based teaching style intervention in the exercise domain. Eur J Soc Psychol 2008 Mar;38(2):375-388. [doi: 10.1002/ejsp.463]
8. Moustaka F, Vlachopoulos S, Kabitsis C, Theodorakis Y. Effects of an autonomy-supportive exercise instructing style on exercise motivation, psychological well-being, and exercise attendance in middle-age women. J Phys Act Health 2012 Jan;9(1):138-150. [doi: 10.1123/jpah.9.1.138] [Medline: 22232500]
9. Silva MN, Vieira PN, Coutinho SR, Minderico CS, Matos MG, Sardinha LB, et al. Using self-determination theory to promote physical activity and weight control: a randomized controlled trial in women. J Behav Med 2010 Apr 11;33(2):110-122. [doi: 10.1007/s10865-009-9239-y] [Medline: 20012179]
10. Silva MN, Markland D, Vieira PN, Coutinho SR, Carraça EV, Palmeira AL, et al. Helping overweight women become more active: need support and motivational regulations for different forms of physical activity. Psychol Sport Exerc 2010 Nov;11(6):591-601. [doi: 10.1016/j.psychsport.2010.06.011]
11. 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]

12. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
13. 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]
14. Gaudet J, Gallant F, Bélanger M. A Bit of Fit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
15. Hermsen S, Moons J, Kerkhof P, Wiekens C, De GM. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
16. Karapanos E, Gouveia R, Hassenzahl M, Forlizzi J. Wellbeing in the making: peoples' experiences with wearable activity trackers. Psych Well-Being 2016 Jun 14;6(1):4. [doi: 10.1186/s13612-016-0042-6] [Medline: 27376017]
17. Attig C, Franke T. I track, therefore I walk – Exploring the motivational costs of wearing activity trackers in actual users. Int J Hum-Comput Stud 2019 Jul;127:211-224. [doi: 10.1016/j.ijhcs.2018.04.007]
18. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on fitbits with adolescents. Am J Health Edu 2017 Jul 28;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
19. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. Pediatr Blood Cancer 2017 Dec 15;64(12):e26660. [doi: 10.1002/pbc.26660] [Medline: 28618158]
20. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066X.55.1.68]
21. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
22. Robinson SA, Bisson AN, Hughes ML, Ebert J, Lachman ME. Time for change: using implementation intentions to promote physical activity in a randomised pilot trial. Psychol Health 2019 Feb;34(2):232-254 [FREE Full text] [doi: 10.1080/08870446.2018.1539487] [Medline: 30596272]
23. Teixeira PJ, Marques MM, Silva MN, Brunet J, Duda JL, Haerens L, et al. A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. Motiv Sci 2020 Dec 05;6(4):438-455. [doi: 10.1037/mot0000172]
24. 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]
25. Ellis DA, Piwek L. Failing to encourage physical activity with wearable technology: what next? J R Soc Med 2018 Sep 23;111(9):310-313 [FREE Full text] [doi: 10.1177/0141076818788856] [Medline: 30032696]
26. Silva MN, Markland D, Carraça EV, Vieira PN, Coutinho SR, Minderico CS, et al. Exercise autonomous motivation predicts 3-yr weight loss in women. Med Sci Sports Exerc 2011 Apr;43(4):728-737. [doi: 10.1249/MSS.0b013e3181f3818f] [Medline: 20689448]
27. Ryan R, Deci E. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York: Guilford Press; 2018:1-756.
28. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020 Jun 20;17(1):78 [FREE Full text] [doi: 10.1186/s12966-020-00978-9] [Medline: 32563261]
29. Colley RC, Michaud I, Garriguet D. Reallocating time between sleep, sedentary and active behaviours: associations with obesity and health in Canadian adults. Health Rep 2018 Apr 18;29(4):3-13 [FREE Full text] [Medline: 29668028]
30. Krinski K, Machado DG, Lirani LS, DaSilva SG, Costa EC, Hardcastle SJ, et al. Let's walk outdoors! Self-paced walking outdoors improves future intention to exercise in women with obesity. J Sport Exerc Psychol 2017 Apr;39(2):145-157. [doi: 10.1123/jsep.2016-0220] [Medline: 28787251]
31. Decker ES, Ekkekakis P. More efficient, perhaps, but at what price? Pleasure and enjoyment responses to high-intensity interval exercise in low-active women with obesity. Psychol Sport Exerc 2017 Jan;28:1-10. [doi: 10.1016/j.psychsport.2016.09.005]
32. Toner J. Exploring the dark-side of fitness trackers: normalization, objectification and the anaesthetisation of human experience. Perform Enhanc Health 2018 Aug;6(2):75-81. [doi: 10.1016/j.peh.2018.06.001]
33. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
34. Eysenbach G, CONSORT- E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

35. 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]
36. 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(1):18 [FREE Full text] [doi: 10.1186/1479-5868-9-18] [Medline: 22385676]
37. Kayser JW, Cossette S, Alderson M. Autonomy-supportive intervention: an evolutionary concept analysis. J Adv Nurs 2014 Jun 27;70(6):1254-1266. [doi: 10.1111/jan.12292] [Medline: 24279721]
38. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126. [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
39. 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]
40. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
41. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
42. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
43. Ware JE, Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36): conceptual framework and item selection. Med Care 1992;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
44. Tanita technial bulletin: body composition measurement repeatability. Tanita. 2013. URL: https://www.tanita.com/data/File/AdditionalResearch/BodyCompositionMeasurementRepeatabilityp2.pdf?rev=EC5D [accessed 2018-05-28]
45. 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]
46. Wilson PM, Rogers WT, Rodgers WM, Wild TC. The Psychological Need Satisfaction in Exercise Scale. J Sport Exerc Psychol 2006 Sep;28(3):231-251. [doi: 10.1123/JSEP.28.3.231]
47. Wilson P, Rodgers W, Loitz C, Scime G. "It's Who I Am … Really!' The importance of integrated regulation in exercise contexts. J Appl Biobehav Res 2006;11(2):79-104. [doi: 10.1111/j.1751-9861.2006.tb00021.x]
48. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exerc Psychol 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
49. Wilson PM, Sabiston CM, Mack DE, Blanchard CM. On the nature and function of scoring protocols used in exercise motivation research: an empirical study of the behavioral regulation in exercise questionnaire. Psychol Sport Exerc 2012 Sep;13(5):614-622. [doi: 10.1016/j.psychsport.2012.03.009]
50. Ogilvie D, Foster CE, Rothnie H, Cavill N, Hamilton V, Fitzsimons CF, Scottish Physical Activity Research Collaboration. Interventions to promote walking: systematic review. Br Med J 2007 Jun 09;334(7605):1204 [FREE Full text] [doi: 10.1136/bmj.39198.722720.BE] [Medline: 17540909]
51. Eyler AE, Wilcox S, Matson-Koffman D, Evenson KR, Sanderson B, Thompson J, et al. Correlates of physical activity among women from diverse racial/ethnic groups. J Womens Health Gend Based Med 2002 Apr;11(3):239-253. [doi: 10.1089/152460902753668448] [Medline: 11988134]
52. Dugan S, Karavolos K, Lynch E, Hollings CS, Fullam F, Lange-Maia BS, et al. A multimethod investigation into physical activity in midlife women. J Phys Act Health 2016 Dec;13(12):1351-1359. [doi: 10.1123/jpah.2015-0565] [Medline: 27633615]
53. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000 Jun 13;71 Suppl 2(sup2):1-14. [doi: 10.1080/02701367.2000.11082780] [Medline: 25680007]
54. Wilms B, Ernst B, Thurnheer M, Weisser B, Schultes B. Correction factors for the calculation of metabolic equivalents (MET) in overweight to extremely obese subjects. Int J Obes (Lond) 2014 Nov 7;38(11):1383-1387. [doi: 10.1038/ijo.2014.22] [Medline: 24503879]
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. Samdal GB, 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]

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

Abbreviations

BREQ-3: Behavioral Regulation in Exercise Questionnaire, version 3
IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalent of task
PA: physical activity
PNSE: Psychological Need Satisfaction in Exercise

Edited by L Buis; submitted 22.02.21; peer-reviewed by G Watson, A Sullivan, K Knittle; comments to author 07.05.21; revised version received 31.07.21; accepted 28.10.21; published 16.12.21.

Please cite as:
Black M, Brunet J
A Wearable Activity Tracker Intervention With and Without Weekly Behavioral Support Emails to Promote Physical Activity Among Women Who Are Overweight or Obese: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(12):e28128
URL: https://mhealth.jmir.org/2021/12/e28128
doi:10.2196/28128
PMID:34927590

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org